Development and Applications of Cyanobacterial Metabolic Models

By

Trang T. Vu

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy
(Chemical Engineering)

at the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 5/17/13

The dissertation is approved by the following members of the Final Oral Committee:
Jennifer L. Reed, Assistant Professor, Chemical and Biological Engineering
John Yin, Professor, Chemical and Biological Engineering
Christos T. Maravelias, Associate Professor, Chemical and Biological Engineering
Brian F. Pfleger, Assistant Professor, Chemical and Biological Engineering
Daniel R. Noguera, Professor, Civil and Environmental Engineering

To my family and friends

for their endless love and support

Acknowledgements

As I am completing this dissertation, I feel extremely thankful for having this chance to express my gratitude for a great number of people for their help and support during my time as a graduate student at UW-Madison.

To my advisor, Dr. Jennifer Reed, thank you for giving me the opportunity to work in your lab. I still remember how excited I was when I first learned about systems biology through your presentation during the faculty presentations in my first year. Thank you for giving me the freedom to solve the problems on my own yet always providing me guidance when I am off track. Thank you for being so supportive over the years and for giving me many opportunities to become a better researcher.

To my dissertation committee members: Prof. Yin, Prof. Maravelias, Prof. Pfleger, and Prof. Noguera, thank you for giving me critical comments and sharing your insights which are extremely valuable to my work.

To all the members of Reed's group, thank you for your constructive criticism and helpful suggestions on every single presentation and practice talk. I would like to especially thank Dr. David Baumler and Klaus Lovendahl for helping me with experiments. My small-group-meeting members, Josh Hamilton, and Wai Kit Ong, thanks for always listening and sharing your thoughts whenever I run into a roadblock. To Joonhoon Kim, who has been my friend and officemate since our first year of graduate school, thank you for patiently answering my technical questions and troubleshooting my codes. I have learned so much from you.

I would like to thank the Pacific Northwest National Laboratory (PNNL) for funding my research and all the collaborators at PNNL especially Dr. Sergey Stolyar, Dr. Alex Beliaev, and Dr. Grigoriy Pinchuk for their insightful discussion and experimental support.

Without the love and support from my family and friends, I would not be what I am today. To my parents, thank you for your unconditional love and for encouraging me every step of the way. You both are the reason that keeps me going. To Nhi and Tam, thank you for always caring for and supporting me through all my ups and downs. To my ‘TrueCore’ buddies Truc, An, and Long, you have made my last few years in Madison so much fun and memorable. I will never forget our summer days biking around Madison, and I am looking forward to sharing more memories with you in the upcoming years. Vu, I am forever grateful for your love and care through all these years.

Table of Contents

List of Tables

List of Figures

Chapter 1

Cyanobacteria – a potential platform for production of biochemicals

Energy crisis is one of the biggest problems the world is facing nowadays and in the future. Over many years, energy production has heavily relied on the availability of fossil fuels via chemical-based processes that are becoming increasingly limited and expensive. In addition, these energy resources are not distributed uniformly throughout the world, resulting in political and economic imbalances among nations [1]. Carbon dioxide emission from fossil fuel usage also raises health and environmental concerns regarding pollution and global warming [2]. The disadvantages of fossil fuels directly make renewable fuels stand out as a promising future energy source for they are renewable and environmental friendly [2]. The limitations of the traditional methods motivate researchers to develop new and "clean" processes to generate renewable fuels and chemicals, among which biological processes receive a great deal of research interests and investments. In this chapter, we reviewed the use of cyanobacteria as a potential platform for production of bio-based chemicals. We discussed different bio-production processes, as well as the issues and challenges associated with them. Recent advances in metabolic engineering cyanobacteria for bio-chemicals together with some future perspectives were also highlighted to provide some useful insights about the potential for this new and exciting technology.

1.1 Advantages of cyanobacteria

Cyanobacteria, also known as blue-green algae, are oxygenic photosynthetic bacteria found in most ecological niches from fresh water to marine, terrestrial and extreme environments [3]. Cyanobacteria can possess 3 basic growth modes: photoautotrophy, photoheterotrophy and heterotrophy. During photoautotrophic growth, cyanobacteria use energy from sunlight and electrons donated from water to convert CO_2 to essential precursors via photosynthesis. While all cyanobacteria can grow photoautotrophically, not all of them can grow heterotrophically (in the absence of light and presence of some organic carbon sources other than CO_2 such as glycerol or glucose), or photoheterotrophically (in the presence of light and presence of organic carbon sources) [4].

Cyanobacteria are known to have similar photosynthesis mechanism to that of higher plants but possess much simpler physiological structure and genetics, and therefore naturally become the most suitable model to study oxygenic photosynthesis. The structure and physiology of cyanobacterial cell resemble the chloroplast in plant cells and therefore cyanobacteria and chloroplast are believed to be evolutionarily related [5]. In addition, cyanobacteria are the only known prokaryotes that exhibit circadian rhythm and this behavior is believed to control metabolic processes inside the cyanobacterial cells [6].

Cyanobacteria are considered important microorganisms on Earth for they play key roles in the carbon and nitrogen cycles of ecosystems [7]. Compared to other terrestrial photosynthetic systems such as plants and algae, cyanobacteria have faster growth rates and higher photosynthetic conversion efficiency (10% compared to 1 - 5% in higher plants or algae [8]). Due to their ability to growth in minimal environments (only require mineral, CO_2, sunlight, water) and to inhabit various environments, biofuel production in cyanobacteria does not

compete for arable lands like other processes that require plant-derived (e.g., sugar cane, starch, switchgrass) biomass [9]. In addition, cyanobacteria are known to synthesize high-energy storage components such as lipids, proteins, and carbohydrates, which can be converted to energy [9]. Genome sequences of 41 cyanobacteria strains have been published and genetic modification tools have been developed for a few strains, forming the basis to apply synthetic and systems biology tools for engineering cyanobacteria for biochemical production. The properties mentioned above make cyanobacteria an attractive microbial system for biofuel research.

1.2 Biochemical production in cyanobacteria

There are two common processes that are often employed in engineering cyanobacteria to produce various native or non-native chemicals. These processes can be categorized based on the types of carbon sources (inorganic carbon or organic carbon sources) and the environmental conditions (light or dark) in which the cyanobacteria grow (Table 1.1).

Table 1.1: Biochemical production processes in cyanobacteria

Processes	Carbon source	Environment	Example of engineered products
Photofermentation (photanol)	CO_2	Light	1-butanol, isobutanol, isoprene, ethylene, 2-methyl-1-butanol
Dark-fermentation (photofermentation)	Organic carbon	Dark	hydrogen, lactate, succinate, formate, acetate, ethanol

Photo-fermentation (or photanol) refers to the direct conversion of CO_2 to end products via photosynthesis and the addition of exogenous fermentative pathways to cyanobacteria (Figure 1.1) [10]. In contrast, in a dark-fermentation process, fermentative products are generated from degrading either exogenous organic carbon sources such as glucose, glycerol, or fructose or endogenous carbohydrate such as glycogen accumulated during photoautotrophic growth (Figure 1.1) [11, 12] . Cyanobacteria hosts that are commonly used to study the biochemical production are *Synechocystis* sp. PCC 6803 (*Synechocystis* 6803) and *Synechococcus* elongatus PCC 7942

(*Synechococcus* 7942) since the genetic manipulation tools have been developed for these strains [13].

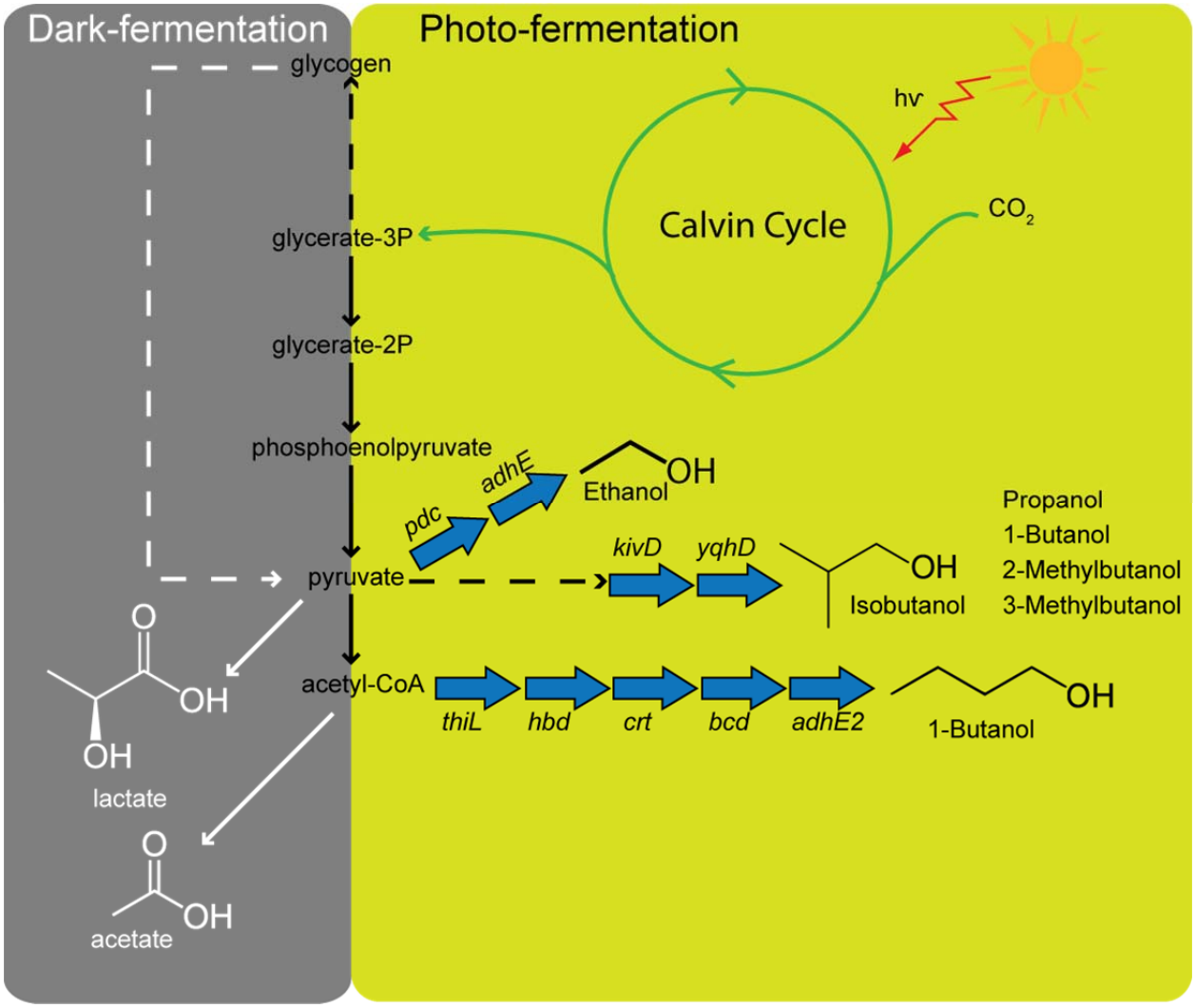

Figure 1.1: Schematic illustration of biochemical production processes in cyanobacteria
Pyruvate and acetyl-CoA derived from fixed CO_2 are major precursors of the biosynthesis of various native chemicals (e.g., lactate, acetate, succinate) and non-native chemicals (e.g., ethanol, 1-butanol, isobutanol). Blue arrows represent the foreign genes of exogenous pathways introduced to the cyanobacteria to convert the precursors to non-native chemicals. Dash-arrows represent multi-step reactions. *kivD* – ketoacid decarboxylase from *Lactobacillus lactis*, *yqhD* – alcohol dehydrogenases from *Escherichia coli*; Genes from *Clostridium acetobutylicum*: *thiL* –acetyl-coA acetyltransferase, *hbd* – β-hydroxybutyryl-CoA dehydrogenase, *crt* – crotonase, bcd – butyryl-CoA dehydrogenase, *adhE2* – bifunctional aldehyde/alcohol dehydrogenase; Genes from *Zymomonas mobilis*, *pdc* – pyruvate decarboxylase, *adhE* – acohol dehydrogenase [13].

1.2.1 Biochemical production via photo-fermentation in cyanobacteria

Cyanobacteria have the ability to fix CO_2 using light energy, however they do not normally produce biofuels and chemicals under this condition. In contrast, many fermentative bacteria have developed biosynthesis machinery for a variety of compounds such as ethanol, 1-butanol, and lactic acid. Fortunately, advances in genetic engineering and synthetic biology allow the integration of exogenous pathways from other bacteria into cyanobacteria to realize the full potential of the combined system (photofermenting bacteria). Photosynthesis provides energy (in the form of ATP) and reductant (in the form of NADPH) for carbon fixation process. As illustrated in Figure 1.1, CO_2 is transformed into glycerate-3-phosphate, which is eventually converted to two major precursors (pyruvate and acetyl-CoA) for biosynthesis of various chemicals. It should be noted that the term 'photo-fermentation' in cyanobacteria described here is different from that used in the context of biological production of hydrogen in other photosynthetic bacteria such as green sulfur or purple non-sulfur bacteria. In these other photosynthetic bacteria, organic substrate is fermented to hydrogen and carbon dioxide, using energy from sun light and electrons extracted from non-water sources such as sulfur or hydrogen sulfide [14]. In addition, biochemical production via photo-fermentation is not necessary a true fermentation process, i.e. process occurs in an anaerobic environment, because the production occurs in the presence of light and oxygen constantly evolves from photosynthesis.

Cyanobacteria have been engineered to produce a number of high-value products via photo-fermentation such as isobutyraldehyde, isobutanol, 1-butanol, 2-methyl-1-butanol, 1, 2-propanediol, 2, 3-butanediol, isoprene, ethylene, lactic acid, ethanol, fatty acids, and fatty alcohols [15-25]. Similar to engineering other microorganisms, the genetic modifications involved in creating 'photo-fermenting' cyanobacteria require the homologous combination of genes from one or several microorganisms that encode highly-efficient enzymes into the

cyanobacterial genome. For example, *Synechococcus* 7942 was engineered to produce 1-butanol via a modified fermentative pathway derived from *Clostridium acetobutylicum* [16]. Instead of introducing all 5 genes from *C. acetobutylicum* (Figure 1.1), genes *thl* and *bcd* were replaced with *atoB* from *E. coli* and *ter* from *Treponema denticola*, respectively because the corresponding enzymes of these genes have been shown to have higher specific activities [16]. In many cases, the biochemical production can be increased by improving CO_2 fixation activity in the cyanobacteria as this is often a limiting step in generating biofuel precursors [15]. Overexpression of Rubisco, the enzyme responsible for carbon fixation, has led to an improvement in isobutanol production by two fold in an engineered strain of *Synechococcus* 7942 [15]. Oxygen evolved during photosynthesis poses a potential problem for introducing fermentative pathways into cyanobacteria because most often, the fermentative enzymes are oxygen-sensitive and thus would not be functional under aerobic condition. Solutions to these issues have been addressed in a number of engineered systems including inhibiting oxygen evolution by treating the engineered cyanobacteria cell culture with a photosystem II activity inhibitor [26], or searching for oxygen-insensitive enzymes [19]. Other engineering strategies involve the replacement of enzymes such that the modified enzymes can use cofactors that are more available to cyanobacteria. For example, the NADH-dependent enzymes involved in the synthetic pathways of 1, 2-propanediol and 1-butanol have been replaced by the NADPH-dependent enzymes to make use of the NADPH pool generated during photosynthesis [16, 18].

1.2.2 Biochemical production via dark-fermentation in cyanobacteria

Although cyanobacteria are oxygenic phototrophic bacteria, many of them live in habitats in which absence of light creates an anoxic (no oxygen evolved from photosynthesis) dark environment. Therefore, some cyanobacteria also develop the ability to generate energy from

organic substrates to survive in the dark [27]. Biochemical production via dark-fermentation therefore refers to the break-down of organic carbon sources in the absence of light to precursors of fermentative products [27] (Figure 1.1). The organic carbon sources can come from the uptake of exogenous carbon such as glucose, fructose, glycerol, or from accumulated endogenous carbon, such as glycogen (autofermentation). The fermentative products often are lactate, formate, acetate, hydrogen, and ethanol [27]. Compared to photo-fermentation process, there are fewer metabolic engineering studies on dark-fermentation in cyanobacteria. This is mainly due to the limited numbers of dark-fermenting cyanobacteria strains, which often are not well-studied and thus genetic modification systems have not yet been developed. In addition, the uptake rates of exogenous carbon sources are relatively slow which results in a slow growth and product formation [27]. Genetic engineering approaches reported thus far for biochemical production in dark fermentation conditions involved redirecting reductant toward desired products by deleting competing pathways [12, 28]. Recently, an engineered cyanobacteria strain *Synechococcus* 7002 lacking a lactate-dehydrogenase enzyme has been shown to produce hydrogen at a level 5 times higher than that of the wildtype via autofermentation of glycogen [12]. In another study, a *Synechococcus* 7002 strain lacking pyruvate:ferredoxin oxidoreductase secreted a ~ 2 fold increase in lactate and succinate production compared to wildtype strain [28].

1.3 Challenges and prospects of biochemical production in cyanobacteria

As a comparison, highest reported yields of selected chemicals produced in engineered cyanobacteria are compared with those in the well-studied bacteria *E. coli* (Table 1.2). It was also estimated that in most engineered cyanobacteria systems, only 5-6% of the captured carbon dioxide were directed to the desired products [29], while the conversion of glucose to desired

products were at least 25% theoretical maximum efficiency in other microorganisms [29]. Clearly, efficiency of the chemical production in cyanobacteria is much less than that in other engineered microorganisms. It should be noted that, however, the use of *E. coli* or yeast in metabolic engineering have been studied for many years and the genetic modification systems for these organisms are much more diverse and developed than for cyanobacteria. With the advances in synthetic biology and molecular biology, more genetic tools will be developed for cyanobacteria. In addition, systems biology approaches can be employed to improve our understanding of cyanobacteria metabolism and regulation, which would be beneficial in metabolic engineering cyanobacteria strains with enhanced photosynthesis and carbon fixation efficiency, and optimized pathways for biochemical production of value products.

Table 1.2: Comparison of reported production yields for various chemicals produced in cyanobacteria and in *E. coli*

Organisms	**Cyanobacteria**		***E. coli***	
Products	**Titer**	**Host/ References**	**Titer**	**References**
Ethanol	550 mg/L (6 days)	Synechocystis 6803 [23]	36.33 g/L (3 days)	[30]
Isobutanol	450 mg/L (6 days)	*Synechococcus* 7942 [15]	22 g/L (5 days)	[31]
Isobutyraldehyde	1100 mg/L (8 days)	*Synechococcus* 7942 [15]	35 g/L (5 days)	[32]
1-Butanol	30 mg/L (18 days)	*Synechococcus* 7942 [16]	375 mg/L (1 day)	[33]
2-Methyl-1-butanol	200 mg/L (12 days)	*Synechococcus* 7942 [17]	1.25 g/L (1 day)	[34]
Fatty acids	197 mg/L (17 days)	*Synechocystis* 6803 [24]	4.8 g/L (2 days)	[24]
Hydrogen	3.26 g/L/h	*Cyanothece* 51142 [11]	23.6 g/L/h	[35]

Chapter 2

Metabolic network reconstruction and constraint-based modeling

The importance of model development in different research areas has been realized for many years as models are useful in testing hypotheses, making predictions, reducing significant numbers of experiments, and contributing to knowledge discovery. Genome-scale metabolic models are important in systems biology and metabolic engineering for many practical applications such as generating testable hypotheses, identifying strain-design strategies, or integrating high-throughput data. As more genome sequences become available, the need for accurate metabolic models becomes more important. In this chapter, we report the reconstruction of three genome-scale metabolic models that will be analyzed in later chapters, *Cyanothece* 51142 (*i*Cce806), *Synechococcus* 7002 (*i*Syp708) and *Shewanella* W3181 (*i*W31818_794). The analyses of the *Cyanothece* 51142 metabolic model are presented in chapter 3. Chapter 4 describes the computational evaluation of the *Synechococcus* 7002 model for chemical production of various compounds. Chapter 5 describes the development of a co-culture model using the metabolic models of *Synechococcus* 7002 and *Shewanella* W3181.

2.1 Metabolic network reconstruction

2.1.1 Databases and draft metabolic network reconstruction

A metabolic network is reconstructed on the basis of a genome annotation of a particular organism combined with knowledge of enzymes and biochemical reactions. A reconstructed

metabolic network is composed of a list of three entities, the genes whose annotations suggest metabolic functions, the enzymatic proteins encoded by those genes, the biochemical reaction equations that are catalyzed by those proteins, and the gene-protein-reaction (GPR) association that connects those entities (Figure 2.1). Genome annotation databases for specific organisms include EcoCyc (specific for *Escherichia coli* K12 MG1655) [36], SGD (*Sacharomyces* Genome Database, specific for yeast) [37], CyanoBase (specific for cyanobacteria) [38]. Other databases such as CMR (Comprehensive Microbial Resources) [39], SEED [40] or NCBI (National Center for Biotechnology Information) [41] contain a collection of genome annotations for a variety of organisms.

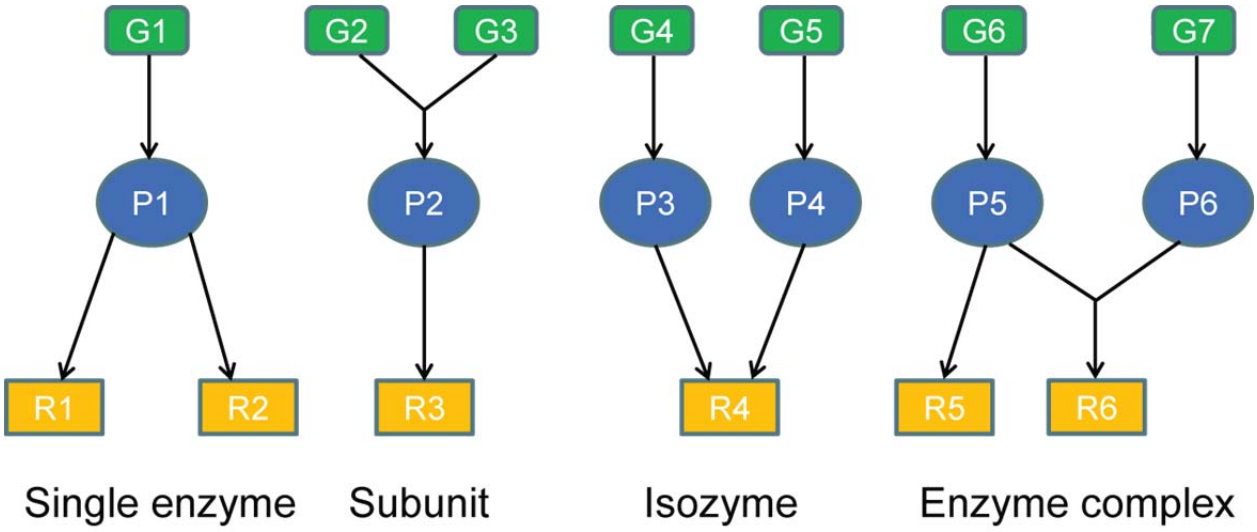

Figure 2.1: Different presentations of Gene-Protein-Reaction (GPR) association
G1, G2, …, G7 represent genes. P1, P2, …, P6 represent proteins. R1, R2, …, R6 represent reactions. Arrows connecting genes to proteins to reactions describe the GPR association. P1 is a single enzyme which encoded by gene G1. P1 can catalyze 2 reactions R1 and R2. P2 is an enzyme that has two subunits encoded by gene G2 and G3. P3 and P4 are isozymes for they can both catalyze the same reaction R4. P5 and P6 are enzyme complexes because they are both needed to create a functional enzyme that catalyzes R6.

Reaction and enzyme databases such as KEGG (Kyoto Encyclopedia of Genes and Genomes) [42], BRENDA (Braunschweig Enzyme Database) [43], MetaCyc [44], TransportDB (Transport Database) [45] provide information on reaction properties such as stoichiometry,

products, substrates, and reaction directionality. Manual metabolic network reconstruction is a laborious and time-consuming process. Fortunately, a number of semi-automatic approaches have been developed to speed up the reconstruction process. These approaches include SimPheny (Genomatica, San Diego, CA), Model SEED [46], and Pathway Tools [47]. Pathway Tools creates model-organism databases for different organisms and allows users to graphically visualize and data-mine the contents of the databases [47]. SimPheny and Model SEED provide users draft metabolic models for which the GPR associations have already automatically constructed [46].

2.1.2 Manual curation of draft reconstruction

The draft metabolic models need to be further curated by modelers to ensure accurate presentations of the gene-protein-reactions, and proper functions of the network. Metabolic network curation normally involves filling network gaps, correctly assigning GPR associations, and constructing biomass equation.

Metabolic gaps are metabolites that are either only produced or consumed in the network, which create holes in the network. The presence of these gaps is likely due to errors generated during the reconstruction of draft model or incomplete knowledge of the organism' metabolism [48, 49]. An example of network gap and gap-filling solution in the reconstruction of the *Cyanothece* 51142 metabolic model is shown in Figure 2.2. In this example, the draft model of *Cyanothece* 51142 could not produce isoleucine due to a gap in producing the isoleucine precursor (2-oxobutanoate, 2obut). Normally, 2obut is produced from threonine via a threonine deaminase enzyme (EC-4.3.1.19), which is missing in the genome annotation of *Cyanothece* 51142. By genome comparison with *Geobacter sulfureducens* [50], a new pathway converting pyruvate to 2obut was introduced to the draft model, and consequently the model can produce

isoleucine. This pathway was later confirmed experimentally in *Cyanothece* 51142 [51]. A number of automatic gap-filling methods have been developed such as GapFind/ GapFill [48], Model SEED [46], MetaFlux [52], and SMILEY [49]. These methods often suggest one of the following gap-filling strategies: changing the directionalities of existing reactions, and adding transport or metabolic reactions from reaction databases specific to the organism or from other organisms [48, 52]. The suggested strategies need to be validated with experimental data before adding to the metabolic models as it was reported that 50% of the gap-filling reaction candidates suggested by MetaFlux have no experimental support and hence were disregarded [52]. Because our knowledge of an organism's metabolism is incomplete and limited, it is generally accepted that genome-scale metabolic models can contain a number of metabolic gaps and these gaps may not affect model predictions in most cases. Gap-filling is an iterative process for which when new knowledge or experimental data become available, the number of gaps can be narrowed.

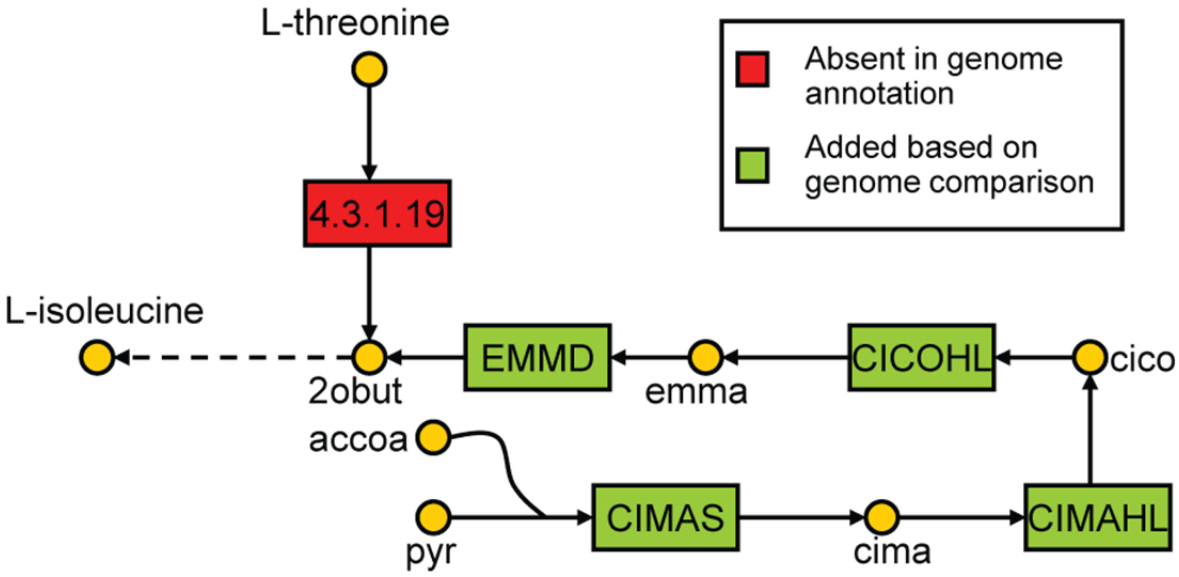

Figure 2.2: Example of network gap and gap-filling solution in the *Cyanothece* 51142 model
Draft model of *Cyanothece* 51142 could not produce isoleucine due to the gap in producing its precursor 2-oxobutanoate (2obut). Threonine deaminase (reaction in red), which can convert threonine to 2obut is missing in the genome annotation. A new pathway (reactions in green) converting pyruvate to 2obut was introduced to the draft model based on genome comparison with *Geobacter sulfureducens*. Metabolites and reactions abbreviations can be found online [53].

Correct GPR associations in a metabolic network reconstruction allow for accurate network modifications when simulating gene deletions. Only one gene can be associated with one mRNA and one protein, based on central dogma. Any GPR associations that violate this rule need to be fixed. An example of incorrect assigning GPR in the early draft model of *Cyanothece* 51142 produced by SimPheny and its corrected version in the published model *i*Cce806 [53] are shown in Figure 2.3 below. In this example, two genes (cce_2605 and cce_2693) were associated with AspB which is an aspartate aminotransferase, an enzyme that involves in the transfer of amine group from one amino acid to another. This enzyme catalyzes two reactions ASPTA1 (which interconverts aspartate and α-ketoglutarate to oxaloacetate and glutamate) and TYRTA (which interconverts 3-(4-hydroxyphenyl)-pyruvate and glutamate to tyrosine and α-ketoglutarate). This GPR association is incorrect because the protein AspB is encoded by two genes that are not protein subunits (Figure 2.1). According to the genome annotation of *Cyanothece* 51142 [54], gene cce_2605 encodes L,L-diaminopimelate aminotransferase, an enzyme that converts L,L-diaminopimelate to glutamate. Therefore, this GPR association is modified such that gene cce_2605 is dissociated from the draft GPR, and is associated with enzyme L,L-diaminopimelate aminotransferase (Dapat), which catalyzes a new reaction DAPAT (which interconverts glutamate to L,L-diaminopimelate). Consequently, gene cce_2693 is associated with protein AspB1 which catalyzes both ASPTA1 and TYRTA reactions. Further examinations of the genome annotation and the draft model of *Cyanothece* 51142 reveal new GPR associations for reactions ASPTA1 and TYRTA. The enzyme AspB2, which is encoded by gene cce_1121, is an isozyme of AspB1 and therefore can also catalyze ASPTA1 and TYRTA. The enzyme HisC, which is encoded by cce_3291 and was associated with histidinol phosphate transaminase reaction (HISPT) in the draft model, can also catalyze TYRTA. All three enzymes

AspB1, AspB2, and HisC can also interconvert glumate and phenylpyruvate to α-ketoglutarate and phenylalanine via reaction PHETA1.

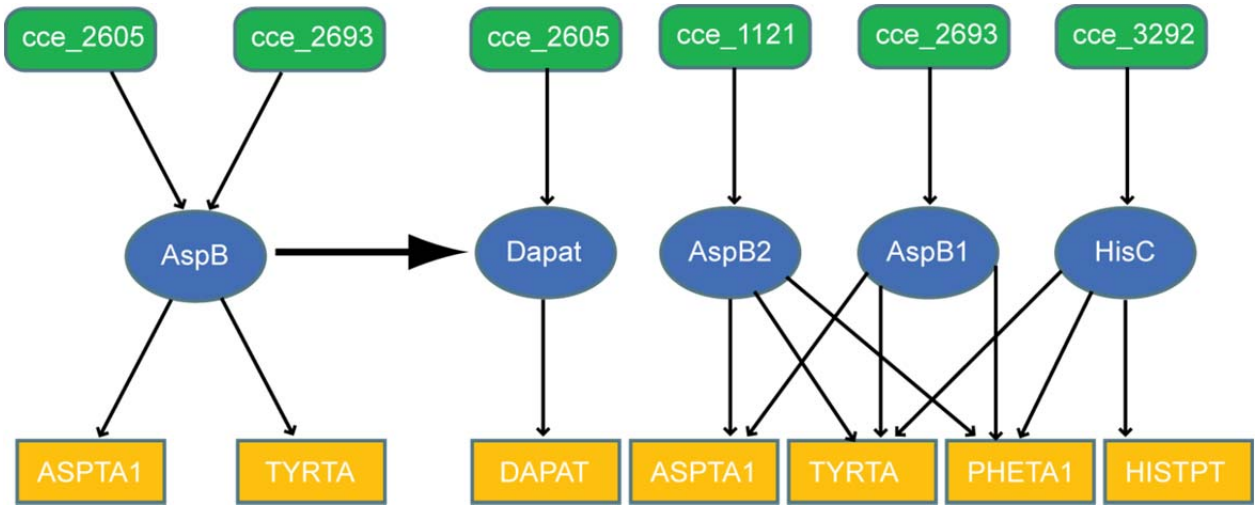

Figure 2.3: Example of incorrect and correct GPR associations
Left panel shows an incorrect GPR association. Protein AspB, which catalyzes reactions ASPTA1 and TYRTA, is associated with two genes cce_2605 and cce_2693 which are not protein subunits. Right panel shows correct and new GPR associations. Gene cce_2605 now encodes protein Dapat (an L,L-diaminopimelate aminotransferase) which catalyzes DAPAT reaction. AspB1 (now encoded by cce_2693) and AspB2 (encoded by cce_1121) are isozymes and both catalyze ASPTA1, TYRTA, and PHETA1 reactions. HisC, which is a histidinol phosphate transaminase, and was originally only catalyzes HISPT reaction, can also catalyze TYRTA and PHETA1 reactions.

The biomass formation equation is crucial for growth simulations of genome-scale models. Biomass compositions measured under different growth conditions and for different organisms are different and will certainly affect the model-predicted flux distribution quantitatively, and qualitatively. Therefore, an accurate representation of biomass formation equation is also an important step in reconstructing a metabolic network. All genome-scale models that have been developed consist of at least one biomass equation that represents the production of 1 unit (in gram dry weight – g DW or ash-free gram dry weight g AFDW) of biomass. In many genome-scale models such as early models of *E. coli* (*i*JE660) [55], or *Synechocystis* 6803 [56], the biomass equation was represented as a linear combination of various central metabolic precursors (such as glucose-6-phosphate, pyruvate, acetyl-coA) and

cofactors (such as nadhp, nadh). This representation is less intuitive since it is difficult to measure concentrations of these precursors. Therefore, the biomass formation equation has recently been replaced by a more straight-forward representation, which is a linear combination of easily quantified macromolecules such as protein, lipids, DNA, RNA, pigments, carbohydrates, and cofactors. The three genome-scale models of *Cyanothece* 51142, *Synechococcus* 7002, and *Shewanella* W3181 developed in this work have the biomass composition constructed in this manner.

2.1.3 Reconstruction of metabolic network for *Cyanothece* 51142

2.1.3.1 Network reconstruction

A draft metabolic network of *Cyanothece* 51142 was reconstructed in SimPheny (Genomatica, San Diego, CA) using a previously described automated model-building process [57]. Metabolic reactions and gene- protein- reaction (GPR) associations from other models were incorporated into the reconstruction if good BLAST hits could be found between genes in *Cyanothece* 51142 and genes in other modeled organisms. Additional reactions were added as necessary to produce known biomass constituents or utilize known nutrients; detailed literature, database, and BLAST searches were then carried out to find genes encoding these reactions in *Cyanothece* 51142 genome. This resulted in several new GPR associations that were incorporated into the reconstruction. The resulting model is referred to as *i*Cce806, and its content can be found online [53]. The resulting *i*Cce806 network contains 806 genes, 587 metabolites, and 667 metabolic and transport reactions (see Table 2.1 for network details). Most of the 42 reactions without genes associated with them were added to complete metabolic pathways needed for biomass production. The final reconstruction encompasses central metabolic pathways such as the Calvin-Benson cycle, the pentose phosphate pathway (PPP),

reactions within the tricarboxylic acid (TCA) cycle, as well as, the complete set of anabolic pathways involved in biosynthesis of glycogen, cyanophycin, amino acids, lipids, nucleotides, vitamins, and cofactors. Pathways for glycolate synthesis (*via* ribulose-1,5-bisphosphate carboxylase/oxygenase, i.e., photorespiration), glycolate conversion to serine, and glycerol catabolism are also included.

Table 2.1: Statistics of the *Cyanothece* 51142 genome-scale metabolic model (*i*Cce806)

Included genes		806
Reactions	With gene association	625
	Without gene association	42
	Exchange reactions	52
Metabolites		587
Network gaps		95

2.1.3.2 Photosynthetic and respiratory pathways

The heart of the metabolic network of *Cyanothece* 51142 lies at the photosynthetic and respiratory pathways. Photosynthetic electron transfer associated with the thylakoid membrane is represented as a set of four separate reactions, including light capture by photosystem II (PS II) and photosystem I (PS I), electron transfer between the two photosystems, and cyclic electron transfer around PS I. Similarly, respiratory electron transfer is represented by reactions catalyzed by terminal cytochrome *c* oxidase (COX), quinol oxidases (QOX, both *bd*- and *bo*-types), NADH dehydrogenases (NDH, type 1 and 2), and succinate dehydrogenase. In addition, two reactions ($NADP^+$- and ferredoxin- requiring) for flavin-dependent reduction of O_2 (i.e., Mehler reactions) were included. A simplified scheme of the photosynthetic and respiratory electron transfer reactions in *i*Cce806 is shown in Figure 2.4.

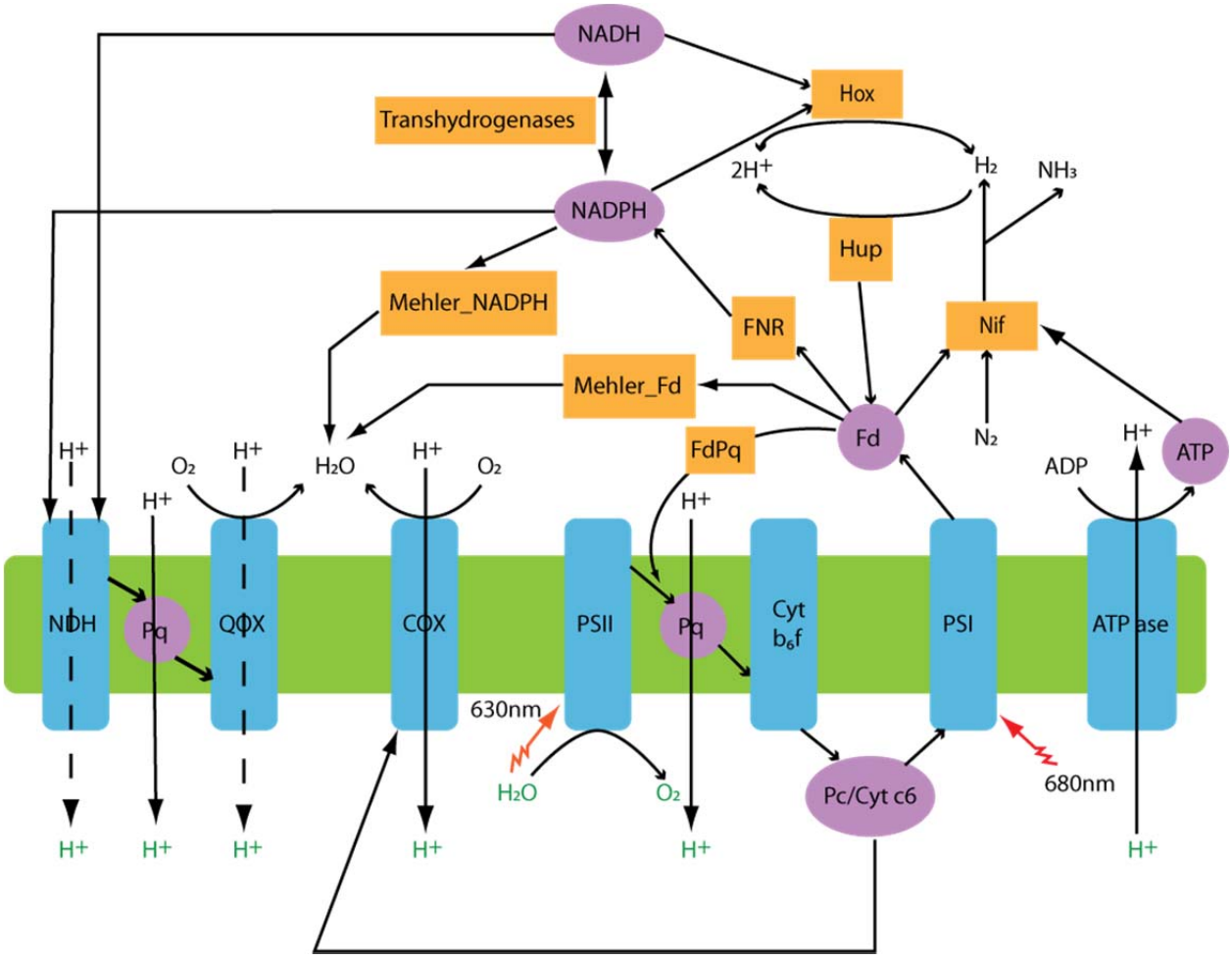

Figure 2.4: Schematic representation of the electron transport and reductant partitioning pathways in the *Cyanothece* 51142 model

Linear photosynthetic electron transfer: electrons from photosystem II (PS II) to photosystem I (PS I) are transferred through plastoquinone (Pq), cytochrome b_6f complex (Cyt b6f), plastocyanin (Pc) and cytochrome c_6 (Cyt c6). From PS I electrons can be transferred to ferredoxin (Fd) *via* ferredoxin:$NADP^+$ reductase (FNR) and subsequently to generate reductant in the form of NADPH. *Cyclic photosynthetic electron transport*: electrons can flow from Fd to Pq (FdPq reaction). *Respiratory electron transfer*: includes two cytochrome oxidases (COX), two cytochrome-quinol oxidases (QOX), and two types of NADH dehydrogenases (NDH-1 and NDH-2). *Alternative sinks for reductant beyond CO_2 fixation*: reduced Fd can be used by the nitrogenase (Nif) and by Mehler reactions to reduce O_2. Bidirectional hydrogenase (Hox) can reversibly produce H_2 using NAD(P)H as an electron donor, while the uptake hydrogenase (Hup) consumes H_2 using Fd as an electron acceptor. Protons transferred across the thylakoid membrane are used by the ATPase to drive ATP synthesis.

2.1.3.3 Constructing biomass equation

As mentioned earlier, the biomass formation equation of *Cyanothece* 51142 was constructed as a linear combination of macromolecules whose concentrations were determined

experimentally from two growth conditions, light-limited and ammonia limited chemostats (Figure 2.5).

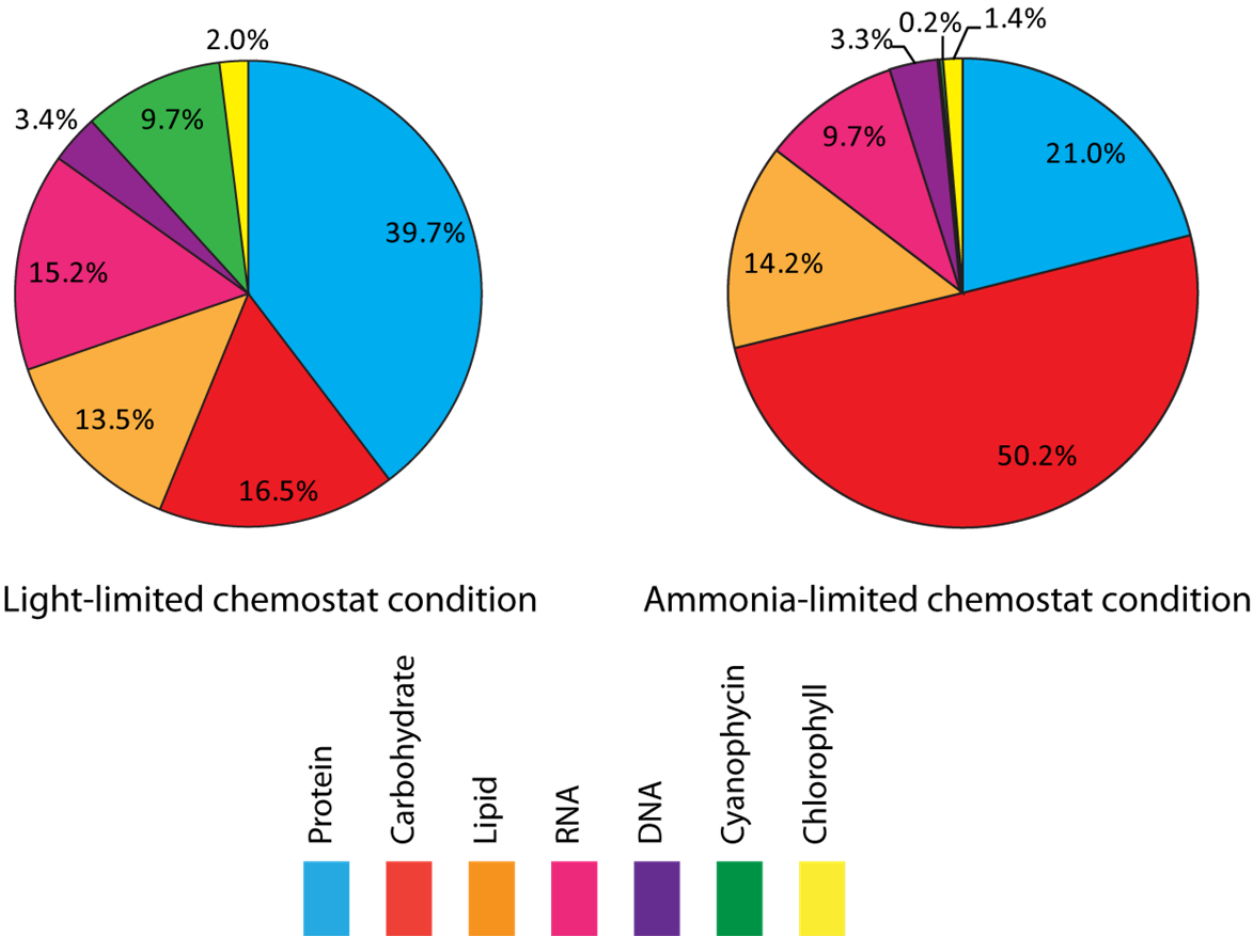

Figure 2.5: Biomass composition (in g/ g AFDW) of *Cyanothece* 51142 measured under light-limited and ammonia-limited chemostat conditions
Cyanophycin concentration was not measured experimentally but estimated using a computational approach.

Since cyanophycin concentration was not determined experimentally, we estimated the concentration using a computational approach. Firstly, the amino acid composition used in the protein synthesis equation (PROTSYN_CN) in the *i*Cce806 model and total number of amino acids (except Cys, Met, and Trp) in protein (P) and cyanophycin (C) were obtained by first solving an optimization problem that minimizes the difference between the relative amino acid

composition in protein $\left(x_P^i\right)$ and the theoretical amino acid composition from genomic data $\left(x_P^{i,theoretical}\right)$ excluding unmeasured amino acids (Cys, Met, and Trp) (Eq. 2.1).

$$\min \quad \sum_{i\in AA\backslash\{Cys,Met,Trp\}}\left(x_P^i - x_P^{i,theoretical}\right)^2 \qquad \text{(Eq. 2.1)}$$

$$s.t \quad P.x_P^i = M^i, i \in AA\backslash\{Cys, Met, Trp, Arg, Asn, Asp, Gln, Glu\} \qquad \text{(Eq. 2.2)}$$

$$P.x_P^{Arg} + C.x_C^{Arg} = M^{Arg} \qquad \text{(Eq. 2.3)}$$

$$P.\left(x_P^{Asn} + x_P^{Asp}\right) + C.x_C^{Asp} = M^{Asn+Asp} \qquad \text{(Eq. 2.4)}$$

$$P.\left(x_P^{Gln} + x_P^{Glu}\right) = M^{Gln+Glu} \qquad \text{(Eq. 2.5)}$$

$$\sum_{i\in AA\backslash\{Cys,Met,Trp\}} x_P^i = 1 \qquad \text{(Eq. 2.6)}$$

$$x_C^{Asp}, x_C^{Arg} = 0.5 \qquad \text{(Eq. 2.7)}$$

$$x_P^i \geq 0 \qquad \text{(Eq. 2.8)}$$

In this method, the experimentally measured amino acid compositions (M^i, μmol/g AFDW) were used from our two chemostat experiments for light-limited (LL) and ammonia-limited (AL) conditions to constrain the values of P, C and x_p^i, such that the estimated amounts of each type of amino acid in protein and cyanophycin equaled their measured values (Eq. 2.2 – 2.5). The theoretical fractions $\left(x_P^{i,theoretical}\right)$ and measured values $\left(M^i\right)$ used in the problem are listed in Table 2.2 below. Cyanophycin contains an equal amount of Asp and Arg, and therefore the fraction of these two amino acids in cyanophycin $\left(x_C^i\right)$ is 0.5.

It should be noted that the compositions for Cys, Met, and Trp are missing from the table as these amino acids were not stable enough to be measured and hence were excluded from the optimization. Consequently, the amount P estimated by the above method represents the number of all amino acids except Cys, Met, and Trp contained in proteins. Additionally, Glu and Gln, as

well as, Asn and Asp are measured as pooled metabolites, so the reported measured values in Table 2.2 for Asp and Glu are actually $M^{Asn+Asp}$ and $M^{Gln+Glu}$, respectively.

Table 2.2: Experimental amino acid compositions of *Cyanothece* 51142 measured under light-limited and ammonia-limited chemostats

Conditions	LL	AL	Theoretical		
Amino acids	M^i **(umol/gDW)**	M^i **(umol/gDW)**	**Fraction from protein sequences**	**Adjusted fraction (exclude Cys, Met, Trp)** $x_P^{i,theoretical}$	**Molecular weight (g/mol)**
Asp	450.507	219.702	0.050	0.053	132.098
Ser	122.154	116.098	0.063	0.066	105.096
Glu	298.283	291.201	0.066	0.069	146.124
Gly	177.042	181.462	0.065	0.068	75.07
His	27.968	31.493	0.019	0.020	155.162
Arg	307.891	108.699	0.045	0.047	175.22
Thr	115.397	120.574	0.057	0.060	119.122
Ala	214.011	201.620	0.066	0.069	89.096
Pro	106.474	104.816	0.046	0.048	115.132
Cys	NA	NA	0.010	Excluded	121.162
Tyr	75.157	66.697	0.033	0.035	181.188
Val	116.127	115.411	0.061	0.064	117.148
Met	NA	NA	0.020	0.000	149.214
Lys	96.985	102.871	0.055	0.058	147.2
Ile	98.619	98.032	0.076	0.079	131.174
Leu	164.729	165.450	0.111	0.116	131.174
Phe	73.953	71.227	0.041	0.043	165.188
Trp	NA	NA	0.014	Excluded	204.226
Asn	NA	NA	0.049	0.051	132.124
Gln	NA	NA	0.052	Excluded	146.15

Solving the above optimization problem, we obtained estimated amounts for P, C and mole fractions for all amino acids in protein, except Cys, Met and Trp. These mole fractions were then readjusted to account for these unmeasured amino acids. The mole fractions $\left(y_{P^*}^i\right)$ for all 20 amino acids in protein (P^*) were calculated using the three equations (Eq. 2.9 – 2.12) listed below, and the resulting values are reported in Table 2.3.

$$y_{P^*}^i = theoretical\ fraction, \forall i \in \{Cys, Met, Trp\} \quad \text{(Eq. 2.9)}$$

$$P^* = \frac{P}{1-\sum_{i\in\{Cys,Met,Trp\}} y_{P^*}^i} \quad \text{(Eq. 2.10)}$$

$$y_{P^*}^i = \frac{P.x_P^i}{P^*}, \forall i \in AA\backslash\{Cys, Met, Trp\} \quad \text{(Eq. 2.11)}$$

Table 2.3: Estimated amino acid compositions in the *Cyanothece* 51142 model

Amino acids	LL Conditon		AL Condition	
	x_P^i	y_{P*}^i	x_P^i	y_{P*}^i
Asp	0.059	0.056	0.054	0.052
Ser	0.061	0.058	0.059	0.056
Glu	0.080	0.076	0.080	0.076
Gly	0.088	0.084	0.092	0.088
His	0.014	0.013	0.016	0.015
Arg	0.046	0.044	0.050	0.048
Thr	0.057	0.054	0.061	0.058
Ala	0.106	0.101	0.102	0.097
Pro	0.053	0.051	0.053	0.051
Cys	Excluded	0.010	Excluded	0.010
Tyr	0.037	0.035	0.034	0.032
Val	0.058	0.055	0.058	0.055
Met	Excluded	0.020	Excluded	0.020
Lys	0.048	0.046	0.052	0.050
Ile	0.049	0.047	0.050	0.048
Leu	0.082	0.078	0.084	0.080
Phe	0.037	0.035	0.036	0.034
Trp	Excluded	0.014	Excluded	0.014
Asn	0.058	0.055	0.053	0.051
Gln	0.067	0.064	0.066	0.063
***P* (umol AA /gDW)**	2014.285		1976.969	
***C* (umol AA /gDW)**	431.012		18.384	
***P** (umol AA /gDW)**	2108.126		2069.071	

The mass ratio of the total protein to cyanophycin (g protein/g cyanophycin) was calculated using P* and C and the weighted average molecular weight for amino acids in each macromolecule. This mass ratio was then used to calculate the protein and cyanophycin concentrations, provided measured total protein concentration (Table 2.4). The concentration of soluble metabolites that are also part of the biomass equations were taken from [58] and [59]. Appendix 1 contains a more detailed description of the complete biomass equations constructed for *Cyanothece* 51142 metabolic model. The biomass compositions of macro molecules were adjusted for each condition so that the total biomass composition added up to 1 g AFDW.

Table 2.4: Biomass compositions of *Cyanothece* 51142 measured under light-limited (LL) and ammonia-limited (AL) chemostats

	LL Condition			AL Condition		
Biomass components	Raw data values (g/L)	Raw data + computed data (g/gAFDW)	Adjusted data (g/gAFDW)	Raw data values (g/L)	Raw data + computed data (g/gAFDW)	Adjusted data (g/gAFDW)
Protein	0.028	0.473	0.397	0.017	0.211	0.210
Carbohydrates	0.012	0.197	0.165	0.040	0.502	0.502
Lipids	0.010	0.161	0.135	0.011	0.142	0.142
RNA	0.011	0.181	0.152	0.008	0.097	0.097
DNA	0.002	0.040	0.034	0.003	0.033	0.033
Cyanophycin [a)]	NA	0.116	0.097	NA	0.002	0.002
Chlorophyll [b)]	0.0014	0.024	0.020	0.001076	0.014	0.014
Ash-free dry weight (g AFDW/L)	0.059			0.079		
Dry weight (gDW)	0.179			0.1814		
Total		1.191	1.000		1.000	1.000

a) Cyanophycin concentration was not measured experimentally, but computed using macromolecular protein concentration measurements, and mass ratio of total protein (P*) to cyanophycin (C) obtained from solving the above optimization problem.

b) Chlorophyll concentration was measured under both chemostat conditions, following methods described by Meeks et al. [60]. Since the total composition of biomass components in LL condition is not equal to 1, we rescaled the data so that the total fraction equals 1.

2.1.4 Updating existing metabolic network for *Synechococcus* 7002

We also updated the previously developed draft metabolic model (*i*Syp611) for *Synechococcus* 7002 [61] using updated genome annotations from a Pathway/Genome Database (provided by Margrethe Serres, personal communication) and new biochemical data [12, 62]. This involved expanding the metabolic network to include genes that were responsible for new metabolic reactions, as well as some existing reactions whose gene-protein-reaction (GPR) associations were missing in *i*Syp611. In addition, we constructed a biomass reaction for the new *Synechococcus* 7002 model based on experimentally-measured biomass composition data (normalized to ash-free dry weight (AFDW) obtained for *Synechococcus* 7002 grown in chemostats under light-limited, nitrogen-limited and carbon-limited conditions. The biomass

composition data was used to generate a biomass reaction, as described previously [53]. The resulting model is referred to as *i*Syp708. This model includes 708 genes and 599 metabolic reactions (Table 2.5). The contents of this model can also be found in [63]

Table 2.5: Statistics of the previous (*i*Syp611) and the current (*i*Syp708) metabolic model of *Synechococcus* 7002

Models		***i*Syp611**	***i*Syp708**
Genes		611	708
Reactions (metabolic and transport)	GPR	517	568
	Non-GPR	35	34
Exchange reactions		37	44
Metabolites		544	581

The addition of 99 new genes (Appendix 2) led to the introduction of 45 new reactions and modifications to previous GPRs (Appendix 3). About half of the newly-included genes were responsible for new metabolic functions, while the rest were added as either isozymes or subunits of existing enzymes in the previous model, *i*Syp611. There were a few cases, where genes were removed from GPRs of one reaction and assigned to another. For example, gene A2770, which was previously included as a subunit of acetolactate synthase (EC-2.2.1.6), has been assigned to a new reaction that converts α-ketoglutarate to succinyl aldehyde. This modification reflected a recent discovery of the enzymatic functions of two genes that are responsible for completing the TCA cycle in cyanobacteria [62]. The model was also updated to include a biomass reaction based on the average biomass composition that was measured under three different photoautotrophic growth conditions. Similar to what we previously observed for *Cyanothece* ATCC 51142 [53], the carbohydrate and protein fractions of biomass showed the greatest variability between light- and nitrogen-limited conditions (Figure 2.6, Appendix 4). The incorporation of this new biomass equation into the model will enable more accurate predictions of cellular phenotypes.

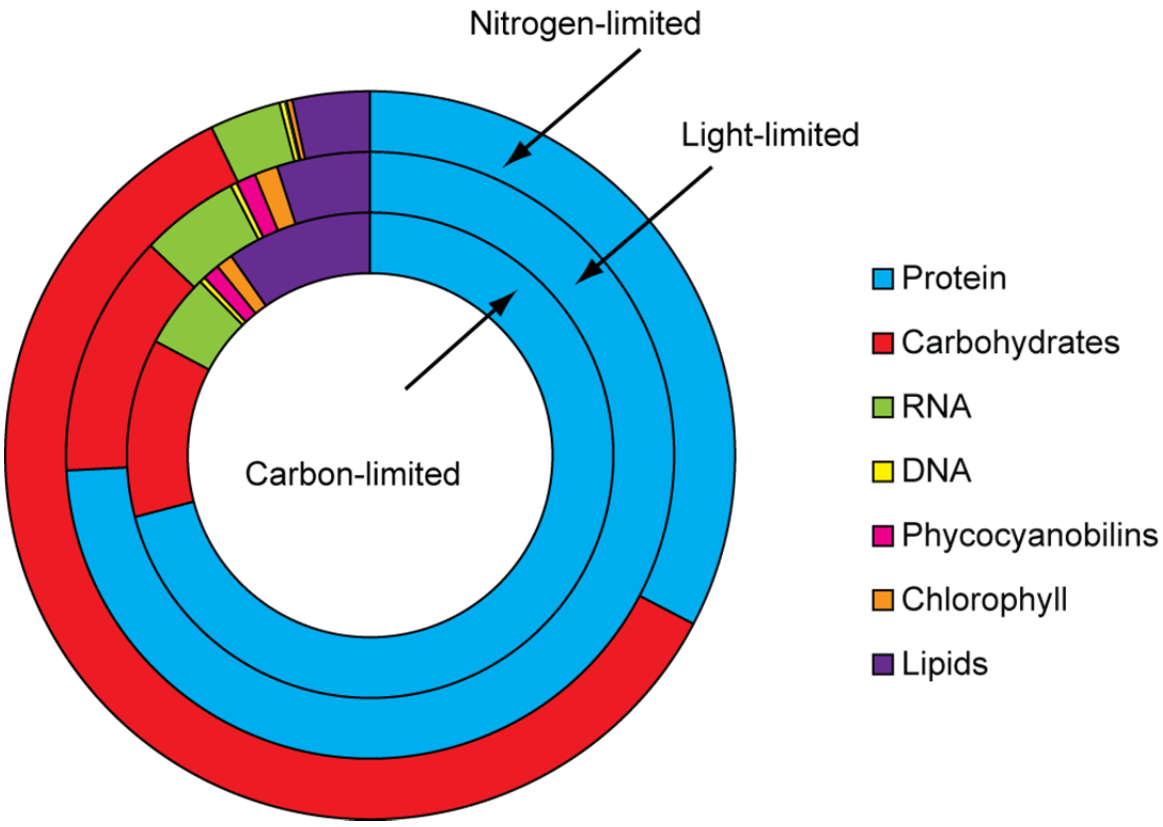

Figure 2.6: Biomass composition of *Synechococcus* 7002 measured under carbon-, nitrogen-, and light-limited chemostat conditions
The biomass compositions of Synechococcus 7002 measured under carbon-, nitrogen-, and light-limited chemostat conditions. The biomass compositions used in the model were scaled so that the total composition equals 1 g AFDW.

Cofactors were also included as part of the biomass requirement and their values were similar to those used in the metabolic model (*i*Cce806) of *Cyanothece* 51142 [53]. While *Synechococcus* 7002 does not produce formate experimentally during photoautotrophic growth, the earlier *i*Syp611 model predicted that formate would be produced under this condition since enzymes degrading formate are missing from the genome. Therefore, in the new *Synechococcus* 7002 (*i*Syp708) model, we removed 10-formyl-tetrahydrofolate (10fthf), flavin adenine dinucleotide (fad) and 5, 10-methylenetetrahydrofolate (mlthf) from the biomass reaction because their biosynthesis reactions produce formate as by-product during photoautotrophic growth. We expect that removing these metabolites would not affect our simulation results

significantly since these metabolites were only present in very small amounts (0.05-0.07 % g/ g AFDW).

2.1.5 Reconstruction of the metabolic network for *Shewanella* W3181

The genome-scale metabolic model of *Shewanella* W3181 was constructed based on genome comparison with the existing metabolic model of *Shewanella* MR1 *i*SO783 [64]. A list of orthologous genes between *Shewanella* MR1 and W3181 (provided by Margrethe Serres, personal communication) was used as the basis of genome comparison. Specifically, if a gene for which orthologous gene can be found in *Shewanella* MR1, corresponding protein and reaction associations would be copied from the *i*SO783 model. The genes for which no orthologous genes can be found in *Shewanella* were evaluated for additional metabolic functions. The reconstruction process involved the removal of some existing reactions in the *i*SO783 model, the modification of existing GPR due to the loss or gain of isozymes or protein subunits, and the addition of several new reactions that were not present in *i*SO783. The resulting model is referred to as *i*W3181_794. This model includes 794 metabolic genes, and 812 metabolic and transport reactions (Table 2.6). Of 794 genes included in the *i*W3181_794 model, there were 708 orthologous genes included in the *i*SO783 model. Since biomass composition was not determined for *Shewanella* W3181, we used the same biomass equation that appeared in *Shewanella* MR1 model *i*SO783. Details of the reconstruction of *Shewanella* W3181 can be found in Appendix 5.

Table 2.6: Statistics of the genome-scale metabolic models of *Shewanella* MR1 (*i*SO783) and *Shewanella* W3181 (*i*W3181_794)

Models	*i*SO783	*i*W3181_794
Genes	783	794
Reactions (metabolic and transport)	774	812
Exchange reactions	85	106
Metabolites	634	645

2.2 Constraint-based modeling

Once the metabolic network is reconstructed, it is necessary to convert it to a mathematical model so that computational approaches can be developed and applied to analyze the model. A metabolic network can be visualized as a network flow composed of links (biochemical reactions) connecting nodes (metabolites) and this network can perform cellular functions such as growth, ATP production, and metabolite production. The links within such networks can be mathematically represented by a stoichiometric matrix (S) with the rows are the metabolites (I) and the columns are the reactions (J) [65]. In this mathematical model, the variables (v) are fluxes through reactions in the network. There are different modeling approaches that have been developed for biological networks including dynamic and constraint-based modeling methods. While dynamic models, which are described by differential equations of metabolite concentrations and involve with reaction kinetics, can efficiently describe the dynamics of a system, it is almost impossible to obtain the kinetic parameters (e.g., rate constants) *in vivo* for every single reaction in the network due to the fact that these parameters often vary with organisms and time through evolution and this makes it difficult to obtain a complete set of kinetic parameters for a specific organism [65]. In addition, solving such large systems of differential equations is computational expensive and challenging task [66]. Therefore, instead of trying to find a single solution, which usually is the collection of flux values of reactions in the network at a given time, constraint-based models seek a solution space that satisfies all the constraints specified by modelers such as thermodynamics, material balances, and enzyme capacity (Eq. 2.13 – 2.15). The mass balance equation (Eq. 2.13) ensures the conservation of metabolites (I) within the model based on steady-state assumption. Thermodynamic constraint (Eq. 2.14) is reflected through the reversibility of each reaction in the

network such that flux value of irreversible reactions (J_{irr}) are non-negative while that of reversible reactions can be negative or positive. Enzyme capacity constraint (Eq. 2.15) is represented by applying lower (α) and upper bounds (β) to the flux values. The solution space is called a feasible region and it contains the flux values (v) that satisfy all of the constraints.

$$\sum_j S_{ij}.v_j = 0, i \in I \quad \text{(Eq. 2.13)}$$

$$v_j \geq 0, j \in J_{irr} \quad \text{(Eq. 2.14)}$$

$$\alpha \leq v_j \leq \beta \quad \text{(Eq. 2.15)}$$

In order to identify single flux distributions from a solution space, a number of constraint-based methods have been developed such as flux balance analysis (FBA) [67], and minimization of metabolic adjustments (MOMA) [68]. Since FBA and MOMA are often used in later chapter, the mathematical formulations of these algorithms are shown in equations below (Eq. 2.16 – 2.18). FBA solves for flux distribution assuming that the metabolic network is at steady state and all of the fluxes through the network are balanced [67]. In addition, the solution is found by optimizing the network for certain objectives such as maximizing growth rate, ATP production, or metabolite production [67]. This optimization problem is often formulated as a linear programming problem because the objective functions as well as the constraints are often linear. MOMA seeks solutions (flux values) that minimize the change in fluxes between the unperturbed wildtype or a parental strain (w) and mutant strain (v) (i.e. knockout strain, for which some fluxes are set to zero. Eq. 2.18) [68]. Due to the nature of the objective functions, MOMA is formulated as a quadratic programming. Another useful constraint-based method for characterizing the metabolic networks is robustness analyses, which measures the change in the optimal value of objective function with respect to change in the values of fluxes through some specific reactions [69]. Similarly, flux variability analysis (FVA) can also be used to determine

the range of values each flux can take that are consistent with the applied constraints by maximizing and minimizing each flux individually [70]. The constraint-based methods mentioned above are widely used to analyze metabolic networks.

$$FBA: \quad min/max \quad c^T.v \qquad \text{(Eq. 2.16)}$$

$$s.t. \quad (Eq.2.13-2.15)$$

$$MOMA: \quad min \quad (v-w)^2 \qquad \text{(Eq. 2.17)}$$

$$s.t. \quad (Eq.2.13-2.15)$$

$$v_j = 0, j \in A \qquad \text{(Eq. 2.18)}$$

Chapter 3

Development and analyses of a genome-scale metabolic model for *Cyanothece* sp. ATCC 51142

Cyanothece spp. are unicellular, diazotrophic cyanobacteria that temporally separate light-dependent oxygenic photosynthesis and glycogen accumulation from N_2 fixation at night [71]. When grown under nutrient excess, *Cyanothece* sp. strain ATCC 51142 (thereafter *Cyanothece* 51142) cells can accumulate significant amounts of storage polymers including glycogen, polyphosphates, and cyanophycin [72]. The inter-thylakoid glycogen granules are significantly larger in size than those found in other cyanobacteria, which points at an unusual branching pattern and packaging of this compound. From a biotechnological perspective, this presents an intriguing theoretical possibility to accumulate substantially higher amounts of polyglucose without any significant increase in the number of granules [73]. *Cyanothece* 51142 is also of interest for bioenergy applications due to its ability to evolve large quantities of H_2. Remarkably, H_2 production in this organism can occur under light conditions in the presence of O_2 and is mediated by nitrogenase [11, 74]

Sequencing of the *Cyanothece* 51142 genome [54] has enabled application of high-throughput genomic approaches to study the unique physiological and morphological features of this organism. Transcriptomic and proteomic studies have been conducted to analyze global gene expression patterns under a variety of environmental conditions and infer regulatory pathways that govern the organism's diurnal growth [75, 76]. The availability of genomic information also provides means to construct genome-scale constraint-based models of metabolism, which are powerful tools for systems-level analysis and prediction of biological systems response to

environmental cues and genetic perturbations [77, 78]. Such models have been developed for a variety of biological systems [77] but only in a few studies has this approach been applied to photosynthetic microorganisms, including *Synechocystis* sp. PCC 6803 including *Synechocystis* sp. PCC 6803 [56, 79-82], *Rhodobacter sphaeroides* [83], and *Chlamydomonas reinhardtii* [84, 85]. However, the modeling of metabolism in oxygenic photoautotrophs is an intriguing problem due to the complexity of photosynthetic and respiratory electron transport chains, and the potential effects of two distinct photosystems upon the generation and fate of reductant and energy that drives the remainder of metabolism.

In chapter 2, we have reconstructed the first genome-scale metabolic network (*i*Cce806) of *Cyanothece* 51142. In this chapter, we used a combination of computation and experimental approaches to investigate how photosynthetic and respiratory fluxes affect metabolism. Discrete representation of PS II and PS I and their integration with multiple respiratory pathways enabled modeling of photon fluxes and electron flux distributions under conditions of variable light quality and intensity. The predicted changes in growth rates of *Cyanothece* 51142 in response to changes in light input were experimentally tested using a photobioreactor with controlled sources of monochromatic 630 and 680 nm light. We also carried out computational and experimental analyses of light- and nitrogen-limited chemostat growth of *Cyanothece* 51142 and used mRNA and protein expression data to constrain model-predicted flux distributions. Both *in silico* and experimental data suggest that respiratory electron transfer plays a significant role in balancing the reductant (NADPH) and ATP pools in the cells during photoautotrophic growth. This study is a first step towards a systems-level analysis of cyanobacterial metabolism, as it integrates information into a genome-scale reconstruction to understand metabolism qualitatively and

quantitatively through a constraint-based analysis [77]. We also discuss strategies for improving internal flux distributions through integration of *in silico* simulations and data.

3.1 Results

3.1.1 Metabolic model prediction validation

For initial testing, we examined the ability of the constraint-based model of *i*Cce806 to predict growth under photoautotrophic (using light and fixing CO_2), heterotrophic (using glycerol in the dark), and photoheterotrophic (using glycerol and light) conditions with different nitrogen sources. *In silico* calculated biomass yields, which simulated carbon or light- limited growth (Figure 3.1), qualitatively agreed with previously reported growth data for *Cyanothece* 51142 [71, 72, 86]. Other non-growth conditions that were simulated with the model included nitrogen fixation as occurs during the dark phase of *Cyanothece*'s circadian cycle [71]. In this case, the oxidation of glycogen provides reductant and ATP for nitrogenase, and we examined the model's ability to quantitatively predict the amount of nitrogen (N_2) that could be fixed and stored in the dark, by maximizing cyanophycin production from glycogen. Although H_2 is an obligate co-product of the nitrogenase reaction, no H_2 was produced in the initial simulations under dark N_2-fixing conditions, contradicting experimental observations. Model examination revealed that all of the nitrogenase-generated H_2 was utilized by hydrogenases to reduce NAD(P) and ferredoxin, which ultimately increased cyanophycin production. When the three hydrogenase reactions (HDH_1, HDH_2, and UPHYDR) were eliminated from the model, the predicted ratio of fixed N_2 to consumed glycogen depended on the non-growth associated ATP requirement (NGAR), and was estimated to be 0.3 (NGAR=2.8) or 0.67 (NGAR=0) mole N_2/mole glycogen, which was in accordance with an experimentally measured value of 0.51 [72]. Under this

condition, the model predicted that H_2 production would have same yields as fixed N_2 (0.3 to 0.67 mole H_2/mole glycogen) due to the stoichiometry of the nitrogenase reaction.

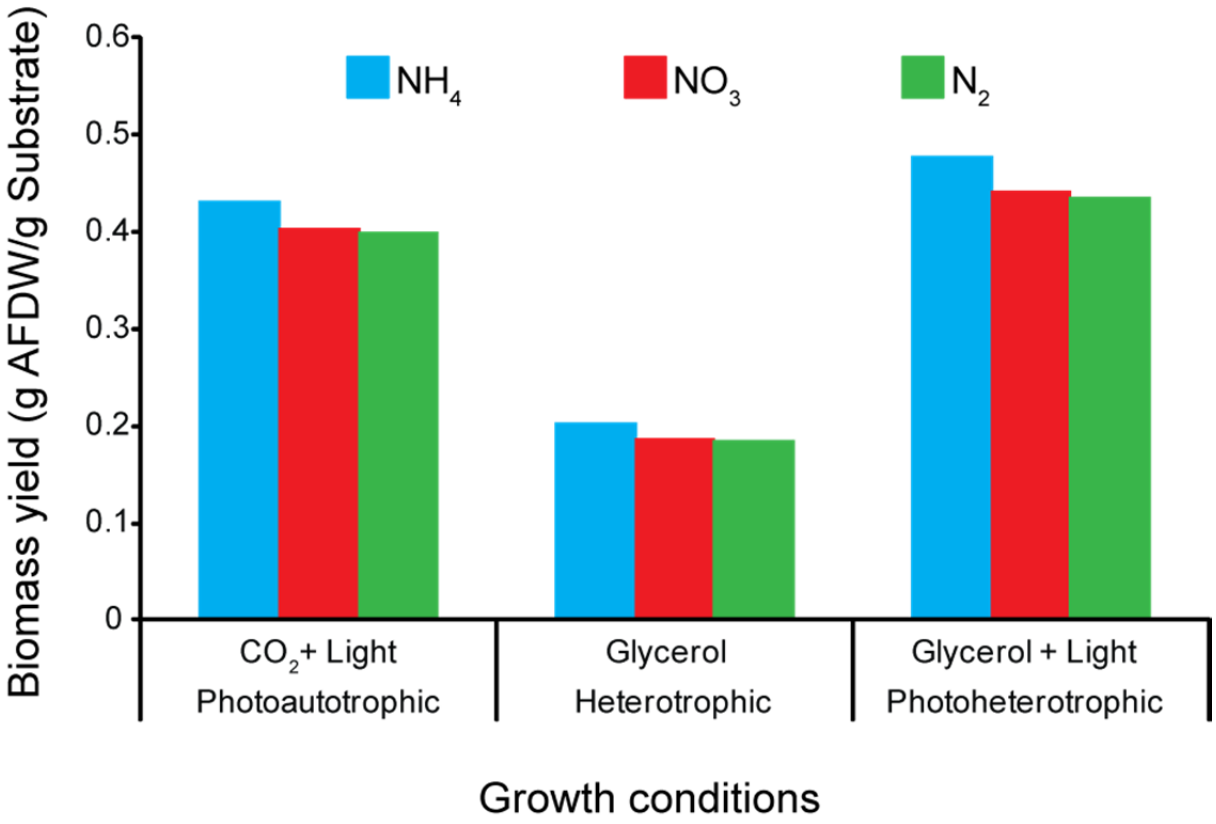

Figure 3.1: *In silico* predictions for *Cyanothece* 51142's biomass yields under photoautotrophic, heterotrophic and photoheterotrophic conditions
Comparison of maximal biomass yields per g of C substrate when different nitrogen sources are used. Under photoautotrophic conditions CO_2 uptake flux was fixed at 1 mmol·g^{-1} AFDW·h^{-1} and photon uptake fluxes at PSI and PSII were fixed at 10 mmol·g^{-1} AFDW·h^{-1}. In the heterotrophic simulations glycerol was the limiting nutrient. Glycerol uptake was fixed at 1 mmol·g^{-1} AFDW·h^{-1} and maximal biomass yields were calculated under dark conditions. In photoheterotrophic simulations both glycerol and light were limiting (so an increase in either would improve growth rates). In this case, glycerol uptake rate was fixed at 1 mmol·g^{-1} AFDW·h^{-1}, while photon uptake fluxes for PSI and PSII were both fixed at 10 mmol·g^{-1} AFDW·h^{-1}. Since light was limiting in the photoheterotrophic condition CO_2 was predicted to be secreted and not used as an additional carbon source.

We also evaluated how fluxes through electron transfer reactions are affected by the nitrogenase flux under N_2-fixing dark conditions. With glycogen being the sole source of reductant for both ATP-generating oxidative phosphorylation and N_2 reduction, a balance between fluxes through respiratory pathways and nitrogenase reaction is needed. In the absence

of the hydrogenase reactions, the model predicted that O_2 reduction via COX, QOX, or Mehler reactions are required to consume NADH resulting from glycogen catabolism (Figure 3.2).

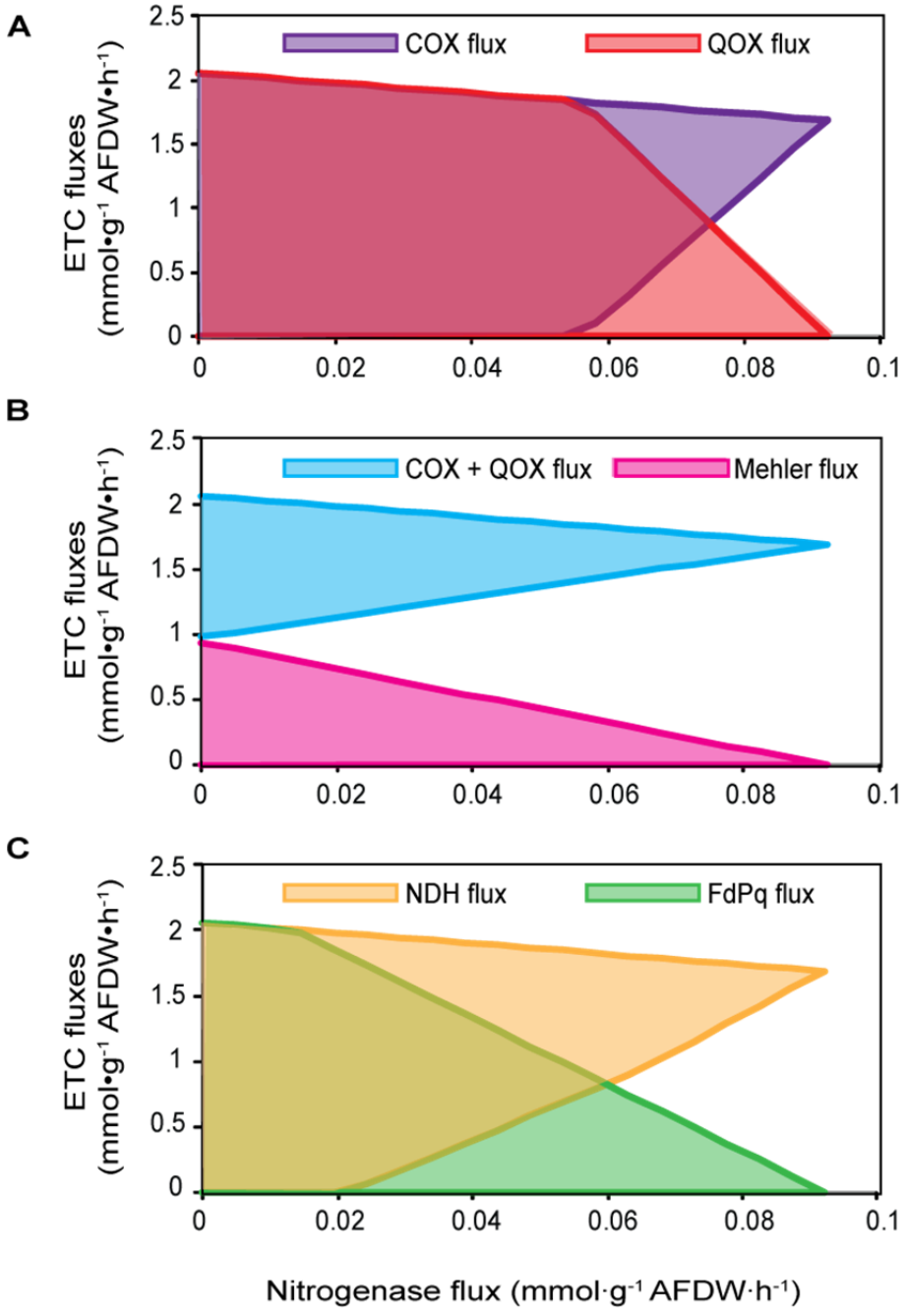

Figure 3.2: Effects of distribution of fluxes through electron transport chains (ETC) on nitrogenase flux in the *Cyanothece* 51142 model

Nitrogen fixation (nitrogenase) flux was varied while fluxes through ETC reactions were maximized and minimized under dark N_2-fixing condition with all hydrogenase reactions eliminated from the model. Under this condition the amount of H_2 produced is equal to the nitrogenase flux. A glycogen demand reaction was added to the model (→ Glycogen; allowing for glycogen consumption) and its flux was

limited to 0.171 mmol·g^{-1} AFDW·h^{-1}. A) Effects of distribution of fluxes through cytochrome c oxidases (COX) and cytochrome-quinol oxidases (QOX) on nitrogenase flux. B) Effects of distribution of total flux through COX and QOX, and flux through Mehler reactions on nitrogenase flux. C) Effects of distribution of fluxes through NADH dehydrogenase reactions (NDH) and Fd-depdent cyclic reaction (FdPq) on nitrogenase flux. Shaded regions indicate ETC reactions can have multiple values for a particular nitrogenase flux.

The model predicts that the COX reaction is required to achieve the maximum N_2 fixation rate since it generates more ATP than the QOX or Mehler pathways (~9 O_2 are needed per N_2 fixed). This is consistent with the results from recent proteomic studies showing the CoxB1 (cce_1977) subunit of COX is more predominant during the dark [87, 88]. These results suggest terminal oxidases are important under dark N_2-fixing conditions not only to generate an intracellular anaerobic environment for nitrogenase, but also to provide ATP for nitrogenase activity.

As photosynthesis and respiratory electron transport chains are interconnected in cyanobacteria [5], these pathways were allowed to interact in the *i*Cce806 model. To perform model robustness analysis, we computationally explored the impact of key photosynthetic and respiratory pathways on growth rate and intracellular flux distributions under varying photon uptake flux for PS I, while the photon uptake flux for PS II was fixed at 20 mmol·g^{-1} AFDW·h^{-1} (Figure 3.3). First, the model was evaluated assuming only linear photosynthetic electron transfer. In this case, all alternative reductant sinks including the proton and O_2 reduction as well as cyclic photosynthetic reactions around PS I were eliminated from the model (Figure 3.3A). Under this condition, growth only occurred at one value of photon uptake flux for PS I and extracellular organic products (ethanol, lactate and/or alanine with trace amounts of formate) would have to be secreted in order to generate enough ATP to support biomass production. Second, when cyclic photosynthetic reactions were added back, the photon uptake flux for PS I

could vary with a fixed photon uptake flux for PS II, but significant amounts

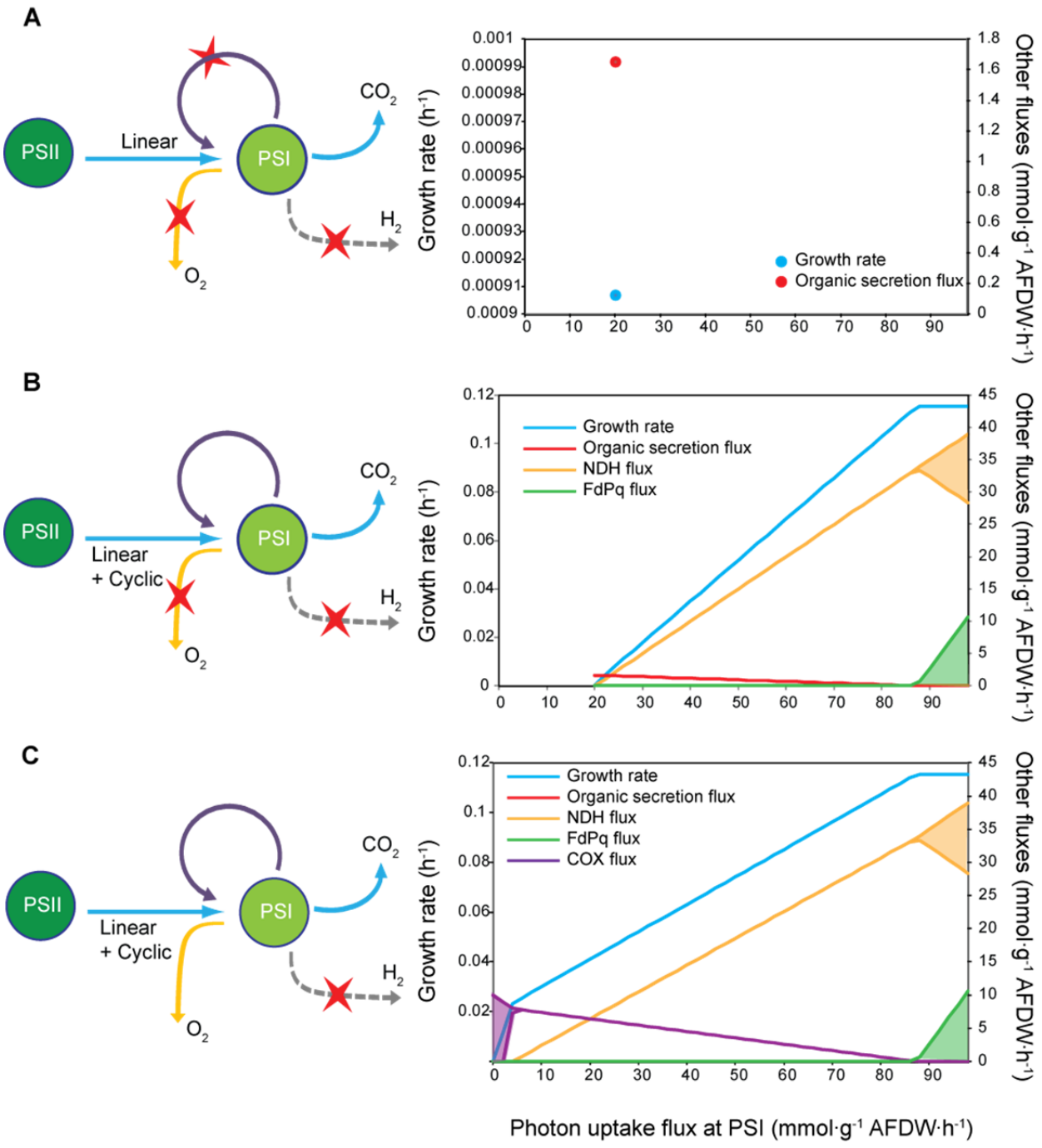

Figure 3.3: Analyses of the operation of electron transport pathways upon growth and metabolism of the *Cyanothece* 51142 model

(A) Effects of removing cyclic photosynthesis (*via* NDH-1, NDH-2, FdPq, G3PD_PQ, and SUCD_PQ) and alternative reductant sinks (H_2 production, COX, QOX, and Mehler reactions). (B) Effect of removing alternative reductant sinks but including all routes for cyclic photosynthesis. Shaded regions indicate that multiple flux values can achieve maximal growth rate. (C) All photosynthetic and respiratory electron flow routes operate, except H_2 production.

of extracellular products were still formed until the photon uptake flux for PS I exceeded ~85 mmol·g^{-1} AFDW·h^{-1} (Figure 3.3B). No growth occurred unless PS I photon uptake flux was greater than or equal to the photon uptake flux for PS II. Only when the model was allowed to use both cyclic photosynthesis and O_2 reduction reactions were no extracellular products predicted and the photon uptake flux for PS I could be less than that for PS II (Figure 3.3C). Since experimental data does not indicate that any by-products including H_2 or organic acids are produced by *Cyanothece* 51142 at a detectable level during photoautotrophic growth with excess ammonium, a plausible mechanism for balancing growth through the generation of additional ATP may involve activity of the cytochrome oxidases.

3.1.2 Effect of light quality on cellular growth and pathway utilization

The discrete representation of PS II- and PS I-mediated reactions and their interactions with multiple respiratory reactions in *i*Cce806 enabled further *in silico* analysis of growth and electron flux distributions under photoautotrophic conditions of variable light quality and intensity. In this case, the complete model was used to explore which reactions would be used to support maximal photoautotrophic growth rates for different levels of PS II and PS I photon uptake fluxes. To predict the corresponding growth rates under light-limited conditions, we constrained the photon uptake fluxes (ranging from 0 to 60 mmol·g^{-1} AFDW·h^{-1}) through each photosystem. The resulting phenotypic phase plane (PhPP) contained three distinct regions (Figure 3.4A): in two regions growth was limited only by fluxes through PS II (region 1) or PS I (region 3), while in region 2 growth was limited by both PS II and PS I photon uptake fluxes (i.e. increases in either flux would improve growth rate). By adding artificial ATP or NADPH generating reactions (ADP + HPO_4 + H → ATP + H_2O and NADP + H → NADPH) to the model and analyzing changes in predicted maximal growth rates, we were able to identify that in

regions 1 and 3 growth was NADPH/reductant-limited, while in region 2 it was limited by energy supply (Figure 3.4A).

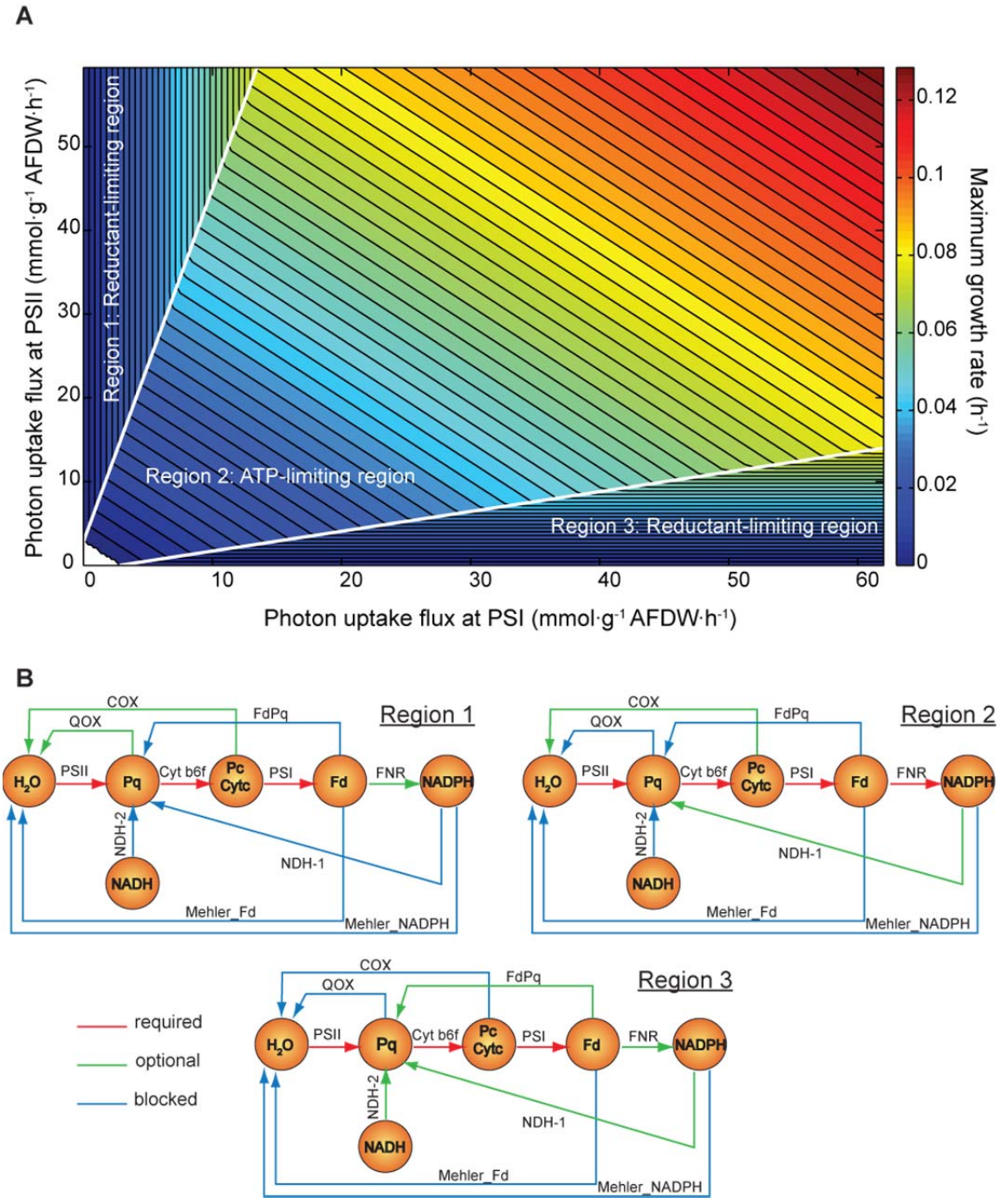

Figure 3.4: Predictions of the effects of varying photon uptake rates on growth and energy metabolism in the *Cyanothece* 51142 model

(A) 2-D phenotypic phase plane (PhPP) displaying maximum growth rates for varying photon uptake rates. The PhPP has 3 distinct regions – in regions 1 and 3, flux through a single photosystem limit

growth rates, whereas in region 2 flux increases through either photosystem will increase growth rate. (B) Pathway maps of electron transfer reactions in different PhPP regions. PhPP flux variability analysis was performed to see which flux is always required (red arrows), optional (green arrows), and blocked (blue arrows) across each of the three PhPP regions.

To analyze the effect of photon uptake rates on electron flux distributions, we calculated the flux ranges using flux variance analysis (FVA) for all photosynthetic and respiration reactions within each PhPP region (Figure 3.4B). In this instance, PhPP FVA was run with constraints that restrict the model to a given region and to the maximum growth for each point in the region (in contrast, standard FVA is used at a single point in a region). Using PhPP FVA, we identified active (both minimum and maximum flux values are positive or negative), inactive/blocked (minimum and maximum fluxes are both zero), and optional (which could have at least one zero and one non-zero flux value somewhere in the region) reactions leading to optimal solutions in each PhPP region. This new analysis technique allowed classification of reaction usage across entire regions of the PhPP and is not restricted to fixed points within a region. While linear photosynthesis was active and Mehler reactions were blocked across the entire PhPP, there were differences in the usage of photosynthetic and respiratory reactions observed within all three regions (Figure 3.4B). Surprisingly, while generation of NADPH from reduced ferredoxin *via* linear photosynthesis is the key source of reductant, ferredoxin-$NADP^+$ oxidoreductase (FNR) was predicted to be active in region 2, but optional in regions 1 and 3. Closer examination of *in silico* calculated electron flux distributions revealed that, in addition to FNR, the model utilized a cycle involving glutamine synthetase, glutamate synthase and transhydrogenase, resulting in ATP-driven NADPH production. In regions 1 and 3, the model predicts there is excess ATP, and so this cycle can be used instead of FNR to move electrons from ferredoxin to NADPH. However, this cycle is unlikely to be of any physiological relevance since there has been no experimental data supporting this route for making NADPH, and FNR is

essential for photoautotrophic growth in unicellular cyanobacteria such as *Synechococcus* 7002 [89]. Differences in the predicted usage of respiratory reactions were also found. In region 1, where growth is limited by the flux through PS I, at least one of the COX and QOX reactions must be active to oxidize excess electron carriers (Pc, Cyt c_6, or Pq) generated from PS II. Similarly, in region 3 under PS II flux limitation, excess electron carriers (Pq, Fd) must be reduced *via* NDH-1 or –2 or ferredoxin-dependent cyclic electron transfer (FdPq). Conversely, due to ATP limitation in region 2, the model favored reactions with higher proton pumping capacities and so both the QOX and FdPq reactions were inactive. The usage of COX was optional in region 2 and depended on photon uptake rates (*e.g.*,, COX reaction was inactive at the boundary between regions 2 and 3).

The model predictions (Figure 3.4A) were compared to batch growth experiments in the LED-photobioreactor which allowed instantaneous measurements of initial growth and photon uptake rates by *Cyanothece* 51142 cells exposed to different intensities and ratios of 630 and 680 nm light (Table 3.1). When *Cyanothece* 51142 cultures were illuminated with both 630 nm and 680 nm light, initial growth rates generally correlated with the total photon flux through PS II and PS I, with higher growth rates observed at 80 $mmol \cdot g^{-1}$ $AFDW \cdot h^{-1}$ total photon flux and 630 nm:680 nm light ratio of 2:1. When cultures were exposed to only a single wavelength of light (batch experiments 6 - 10), i.e., either 630 or 680 nm, *Cyanothece* 51142 cells displayed a similar trend with higher growth rates observed at higher photon flux intensities. The predicted growth rates were within 7% of the experimentally measured values, except for the two cases where single 630 nm wavelength irradiances were used (Table 3.1). The reasons for this are unclear but may be due to other physiological and/or biochemical phenomena such as state transitions that are not contained within the model but are operating *in vivo*.

Table 3.1: **Comparison of predicted growth rates and measured growth rates of *Cyanothece* 51142 grown in batch cultures.**

Batch	Photon uptake rate[c] at 630 nm	Photon uptake rate[c] at 680 nm	Total photon uptake rate[c]	Measured growth rate[d]	Predicted growth rate[d]
1[a]	19.0 ± 1.1	15.5 ± 0.9	34.5	0.035 ± 0.0068	0.035 ± 0.0022
2[a]	15.6 ± 1.1	26.0 ± 1.6	41.6	0.041 ± 0.0076	0.043 ± 0.0032
3[a]	33.4 ± 1.0	13.6 ± 0.4	47.0	0.051 ± 0.0053	0.049± 0.0018
4[a]	34.6 ± 1.4	35.0 ± 1.3	69.6	0.079 ± 0.0062	0.074 ± 0.0033
5[a]	53.6 ± 2.8	26.4 ± 1.0	80.0	0.080 ± 0.0052	0.085 ± 0.0044
6[b]	0	32.1 ± 2.0	32.1	0.032 ± 0.0012	0.032 ± 0.0025
7[b]	0	33.0 ± 2.1	33.0	0.037 ± 0.00014	0.033 ± 0.0026
8[b]	0	37.2 ± 1.5	37.2	0.040 ± 0.00032	0.038 ± 0.0017
9[b]	21.1 ± 1.7	0	21.1	0.016 ± 0.010	0.020 ± 0.0021
10[b]	28.0 ± 1.7	0	28.0	0.036 ± 0.014	0.028 ± 0.0021

[a] Experimental photon uptake and growth rates from batches 1-5 were used to calculate ATP requirement parameters GAR and NGAR

[b] For computational predictions of the growth rate for batches 6 - 10, the total photon uptake flux measurements at 630 nm and 680 nm was used to constrain the total photon uptake flux in the model

[c] Average and standard deviation of the instantaneously measured photon uptake rates (in $mmol \cdot g^{-1}$ $AFDW \cdot h^{-1}$) were calculated over the first 5 hours.

[d] Average and standard deviation of the instantaneously measured growth rate (in h^{-1}) were calculated over the first 5 hours.

Data from these batch experiments (batch experiments 1 – 5, Table 3.1) were also used to estimate the growth (GAR) and non-growth (NGAR) associated ATP requirements. NGAR is the amount of energy spent to maintain the cell (i.e., maintenance energy). GAR is defined as energy expenditures used on protein and mRNA turnover or repair, proton leakage, and maintenance of membrane integrity; it does not include ATP spent on polymerization reactions, which are already accounted for in the macromolecular synthesis pathways of the network. The time-averaged growth and photon uptake rates were used to constrain the model and the maximal amount of ATP hydrolysis was calculated (Figure 3.5) for each batch experiment. A plot of growth rate versus maximum ATP hydrolysis flux was generated and a linear fit used to estimate the GAR and NGAR values [90]. Specifically, the slope of the fitted line is the GAR (544 $mmol \cdot g^{-1}$ $AFDW \cdot h^{-1}$), and the y-intercept is NGAR (2.8 $mmol \cdot g^{-1}$ $AFDW \cdot h^{-1}$).

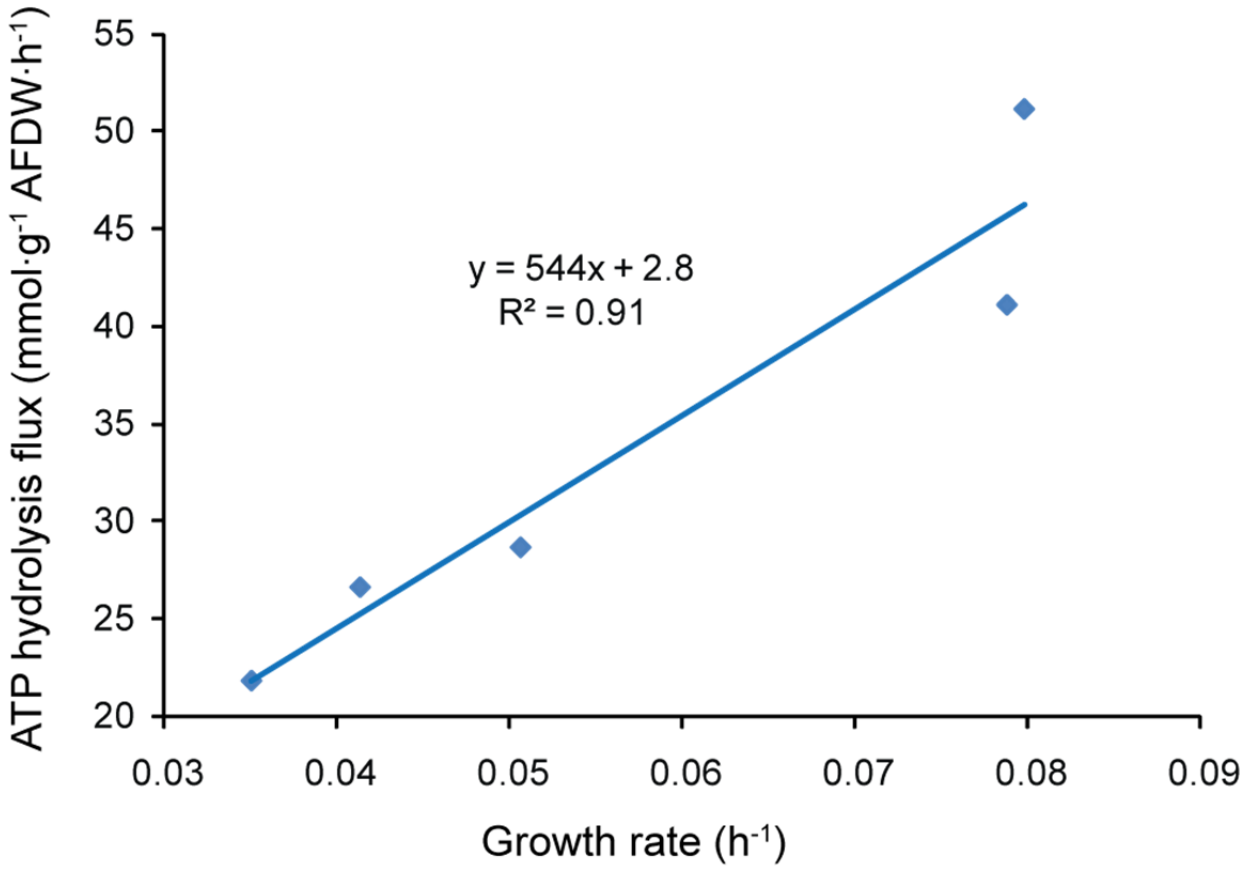

Figure 3.5: Estimating ATP requirements for the *Cyanothece* 51142 model using batch data
Average growth rates and photon uptake fluxes from batch experiments were used to constrain the model. The maximum ATP hydrolysis flux (flux through the ATPM reaction) was calculated using these measurement constraints. The data points represent the calculated maximal ATP hydrolysis values for different batch experiments. The growth-associated ATP requirement (GAR, slope) and non-growth associated ATP requirement (NGAR, y-intercept) were estimated by linear regression of these data.

The estimated GAR value is significantly higher than those reported from other bacteria [64]; however, these model estimates assume that all absorbed photons lead to photosynthetic fluxes (100% quantum efficiency) and that the overall efficiency of ATP production via all electron transfer reactions (photosynthetic and respiratory) are accurate. Depending on the growth condition the quantum yields can change, and for *Cyanothece* 51142 this value was reported to be between ~70-100% for photoautotrophic growth [86]. Upon further analysis, we found the estimated *Cyanothece* ATP requirements were most sensitive to reductions in quantum efficiency and the amount of ATP generated by photosynthesis and respiration (Table 3.2). Since

neither quantum efficiency nor combined photosynthetic and respiratory ATP production were experimentally measured for *Cyanothece* 51142, the original estimates, GAR=544 and NGAR=2.8 were used in all growth simulations.

Table 3.2: Effects of changing simulation conditions on ATP requirement parameters of the *Cyanothece* 51142 model

Simulation conditions	Estimated GAR[a)]	Estimated NGAR[b)]
Complete *i*Cce806 model (no changes to the model are made)	544	2.8
Remove cytochrome c oxidase reactions (COX_PC, COX_CYC)	514	1.1
Reduce quantum efficiency to 70% (only 70% of absorbed photons contributed to PS I and PS II fluxes)	272	2
Reduce ATP efficiency of photosynthesis and respiration by 50% (change H^+/ATP ratio in ATPS4r reaction from $4H^+$/ATP to $8H^+$/ATP)	200	1.4

a) GAR – Growth-associated ATP requirement (mmol·g^{-1} AFDW)
b) NGAR – Non-growth associated ATP requirement (mmol·g^{-1} AFDW·h^{-1})

3.1.3 Using experimental measurements and *in silico* mutagenesis to restrict the range of predicted flux distributions

Since there may be more than one flux distribution that is consistent with the experimentally measured rates of growth, photon uptake, and O_2 production we used FVA to identify required (flux must be non-zero), optional (flux may or may not be zero), or inactive (flux must be zero) reactions under light- and ammonium-limited growth conditions. As our initial simulations (Table 3.3) produced a large number of optional reactions (170 out of 667 for both growth conditions), that represent uncertainty regarding usage, we subsequently used the transcriptome and proteome data (TPD) to further constrain the model. Using a modification to a previously developed approach [91], we obtained a flux distribution that was consistent with measured rates and TPD while reducing the overall flux magnitude.

Table 3.3: **Flux variability analysis of the *Cyanothece* 51142 model in light-limited and ammonium-limited chemostat conditions**

	Without protein and mRNA expression data	With protein and mRNA expression data	
		Light-limited	NH_4-limited
Required reactions	287	364	366
Optional reactions	170	74	76
Inactive reactions	210	229	225

In this analysis, flux was favored through reactions for which proteins were detected and disfavored through reactions associated with undetected proteins and transcriptome data less than a given threshold (e.g.,, $\log_2$ of mRNA expression level is less than 8). The model constrained by TPD predicted that the majority of reactions in central metabolism would be active under both chemostat conditions (Figure 3.6). In addition, we subsequently applied FVA employing additional constraints arising from the TPD. Comparison between FVA results with and without TPD constraints demonstrated a significant decrease in the number of ambiguities (the optional reaction set) when TPD is used (Table 3.3).

While the number of optional reactions was reduced by incorporating TPD into the model, the flux spans (difference between maximum and minimum values) of individual fluxes was still large (>30 $mmol \cdot g^{-1}$ $AFDW \cdot h^{-1}$ for some central metabolic reactions). These large flux spans could arise from cycles or alternative pathways in the model, and deleting these features from the model could subsequently reduce the flux spans.

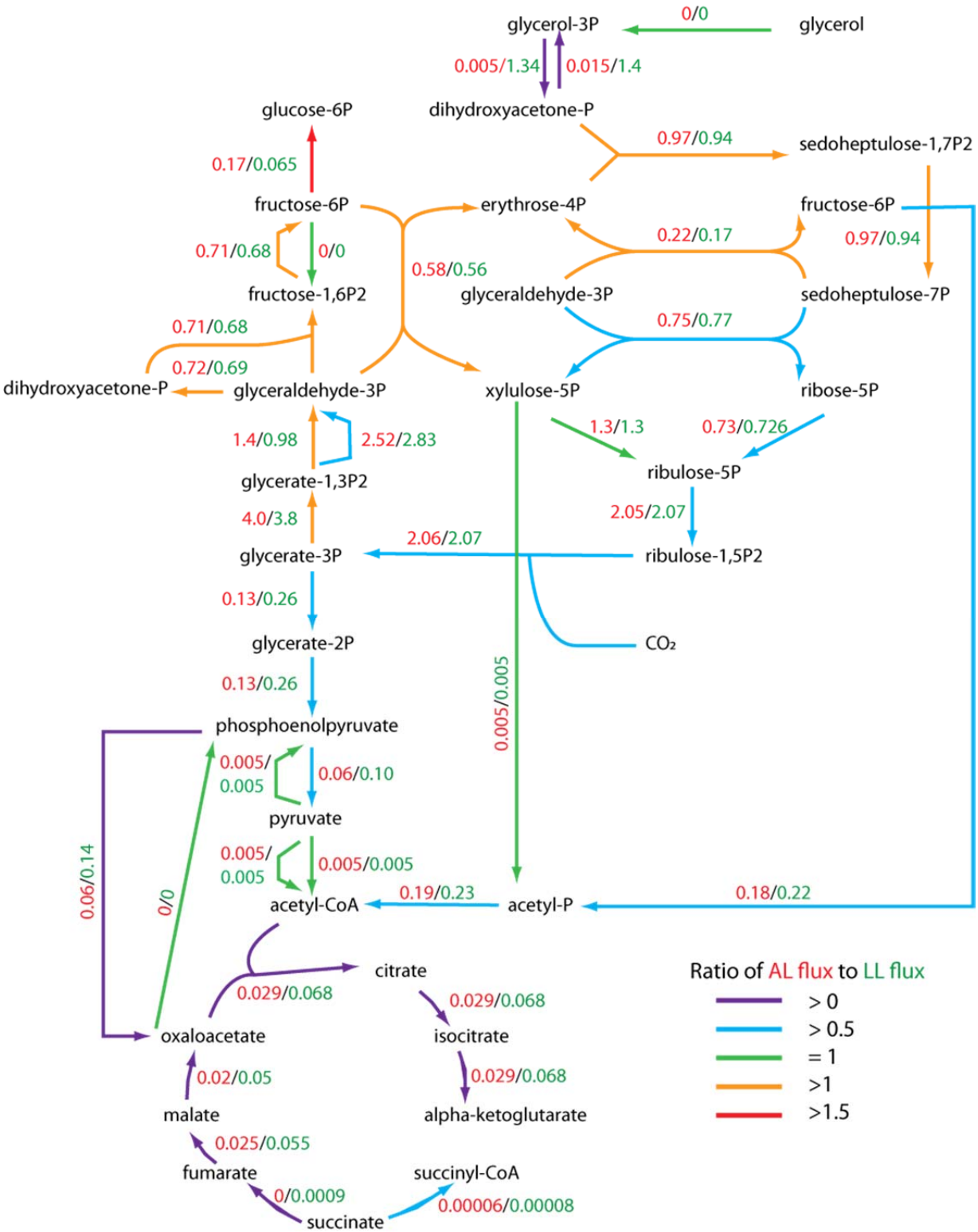

Figure 3.6: Predicted flux distributions in central metabolism of *Cyanothece* 51142 with transcriptome and proteome data (TPD) as constraints

The flux values (mmol·g^{-1} AFDW·h^{-1}) are those where the flux distribution best matches the TPD while also minimizing the magnitude of all fluxes in the network. The flux values in red and green represent

ammonia-limited (AL) and light-limited (LL) conditions, respectively. Arrow colors indicate relative flux ratios between AL and LL conditions.

FVA was repeated using measured growth, photon uptake, and O_2 release rates under light-limited conditions as constraints and with optional reactions were deleted (similar results were found for ammonia limited conditions, data not shown). Flux spans for reactions in central metabolism (Figure 3.6) were then calculated for a series of single or double reaction deletions *in silico*. The purpose of this analysis was to identify those reactions that exert the greatest impact on the flux span in central metabolism (Figure 3.8A).

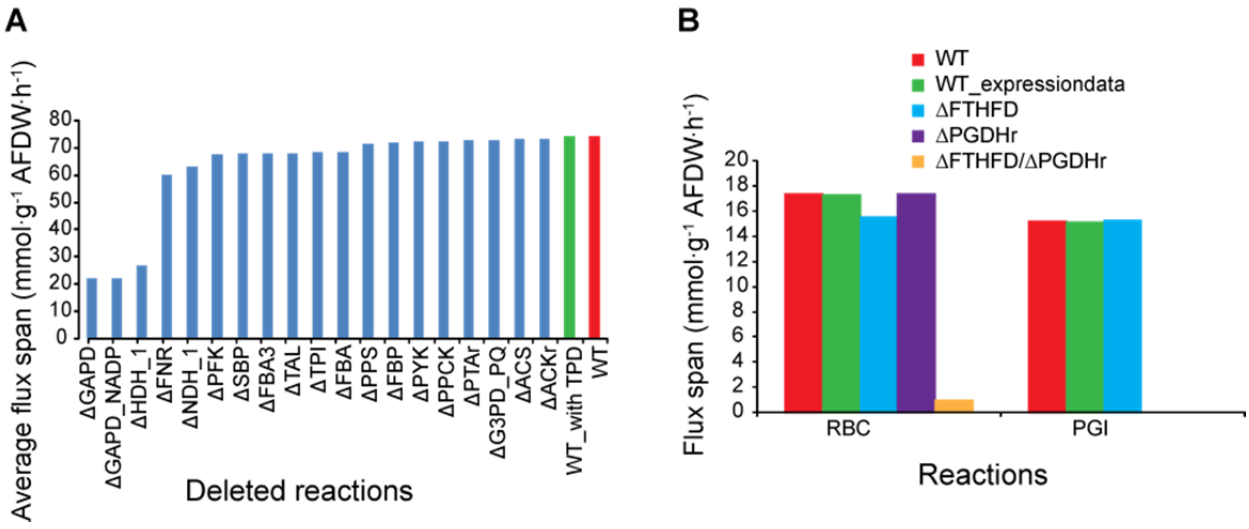

Figure 3.7: Effects of *in silico* reaction deletions on the span of fluxes of the *Cyanothece* 51142 model under light-limited conditions
(A) Effects of deletions are compared to the cases where no reactions were deleted (red bar), or TPD were used as constraints (green bar). The values represent the average flux span across all reactions in central metabolism. Only deletions which lower the flux span by at least > 1 mmol·g^{-1} AFDW·h^{-1} are presented. (B) Changes in flux spans for specific reactions catalyzed by ribulose bisphosphate carboxylase (RBC) and phosphoglucose isomerase (PGI) between simulations that (i) use TPD data as a constraint (green bars), (ii) delete single reactions (blue and purple bars), (iii) delete two reactions (yellow bar) or (iv) impose no additional constraints (red bars). Reaction abbreviations can be found online [53].

Single deletions of glyceraldehyde-3-phosphate dehydrogenase (GAPD or GAPD_NADP) or hydrogenase (HDH_1) reduced the average central metabolic flux span the most (from 74 to 22 mmol·g^{-1} AFDW·h^{-1}). Other single deletions with significant effects included FNR and NDH-1, which are involved in photosynthesis and respiration. The reaction

deletions shown in Figure 3.7A all had a larger impact on reducing average central metabolic flux span than did imposition of constraints based on TPD. There were cases where single deletions had large effects on other specific reactions, but only modest effects on overall central metabolic flux spans. For example, a single deletion in phosphogluconate dehydrogenase (PGDHr) reduced the span for glucose-6-phosphate isomerase flux (PGI) to 0 (Figure 3.7B), but only reduced the average central metabolic flux span by ~ 0.7 $mmol \cdot g^{-1}$ $AFDW \cdot h^{-1}$. The *in silico* analysis of double reaction deletions did not yield any new double deletions that would reduce the average central metabolic flux span significantly. However, some double deletions strategies did reduce flux spans of individual reactions.

3.2 Discussion

Several cyanobacterial metabolic models (all for *Synechocystis* PCC 6803) have been published, which represented photosynthesis as two lumped reactions [56, 79] for linear (PSII, Cyt b_6f, PSI, and FNR) and cyclic (PS I and Cyt b_6f) pathways. In this study, we modeled photosynthesis as a larger set of separate reactions [80] as this structuring allowed analysis of the effects of different illumination on the production and partitioning of reductant through photosynthetic and respiratory reactions, as well as the contribution of different electron transfer pathways to growth. Our PhPP FVA results showed how different photosynthetic and respiratory electron transport chain components are used to maximize biomass production under different lighting regimes. It was not surprising that linear photosynthesis was active in all three regions because the cell needs photons from both PSI and PSII to grow under photoautotrophic conditions. However, the Mehler reactions were inactive in all three regions when we only consider maximal growth rate solutions. In regions 1 and 3, reducing equivalents (e.g.,, NADPH) limit growth and the Mehler reactions would lower the amount of reducing equivalents available

for growth. The Mehler reactions are less energetically efficient than NADH dehydrogenase and cytochrome oxidase so the model would not use them in region 2, where ATP is limiting. So while the Mehler reactions can carry flux in the model, using these reactions lowers the maximum growth rate making them inactive (blocked reactions) in our PhPP analysis. A recent study showed that the Mehler reactions are operational in *Synechocystis* sp. PCC 6803, serving as a sink for excess electrons [92]. These reactions are also likely to be active in *Cyanothece* 51142, since the associated proteins were detected in the proteomic data [53]. As a result the model only predicted non-zero Mehler fluxes when the proteomic data were used to constrain the model [53].

In the absence of cyclic photosynthesis, other products including water (produced by COX, QOX or Mehler reactions), H_2 (*via* hydrogenase), or small organic compounds (alanine, ethanol, lactate and formate) were predicted to be necessary in order to balance the electrons and ATP needed to support growth. In the presence of linear and cyclic photosynthesis reactions, these products must also be produced unless significant amounts of cyclic photosynthesis occurs (>3 times the amount of linear photosynthesis). Since H_2 and small organic compounds are not generally produced under photoautotrophic conditions with excess ammonium, any additional energy is most likely supplied by cytochrome oxidase activities that reduce photosynthetically produced O_2. Interestingly, in the absence of cytochrome oxidase activities in the model, the PS I fluxes must always be greater than or equal to the PS II fluxes. It was shown that the marine cyanobacteriium *Synechococcus* has a PS I/PS II protein ratio >1, which has been explained as a mechanism to protect PS II from photo-damage [93]. Under conditions with high levels of PS II activity, cytochrome oxidase activity may ensure an adequate supply of oxidized plastoquinone (needed for PS II) and reduce O_2 concentrations to limit photorespiration.

Similarly, cyclic electron flow *via* NADH dehydrogenase- or ferredoxin-dependent routes have also been experimentally demonstrated to play important roles in balancing the amount of NADPH and ATP produced via photosynthesis. *Synechocystis* 6803 mutants lacking *ndhD* genes (encoding subunits of NDH-1) had significantly lower cyclic photosynthesis activity [94]. Although the mechanism of electron transfer from ferredoxin to the plastoquinone pool (without using NDH) is still unclear, its activity has been demonstrated in green algae [95] and higher plants [96]. Our computational simulations also showed that, under light-limited photoautotrophic conditions, cyclic electron transfer involving NADH dehydrogenase (NDH-1) is needed for maximal growth if ATP (rather than NADPH) is limiting. In an environment where PS I photon availability is high relative to PS II, cyclic electron transport is needed (Figure 2) to increase availability of PS I substrates (reduced PC or Cyt c_6) and protect against photo-damage. Cyclic electron flow has been experimentally shown to help protect the photosynthetic apparatus from photo-damage [97-99]

In addition to studying the interactions between components of the photosynthetic and respiratory components computationally, we also experimentally evaluated cells grown under continuous light conditions in light- and ammonia-limited chemostats. The measured 630 nm and 680 nm photon uptake and O_2 production rates suggests that reductant was being directed towards O_2 *via* the Mehler, QOX, and/or COX reactions. In both chemostat conditions, the model predicted that steady-state growth rate could have been achieved using lower photon uptake rates by decreasing the amount of reductant generated by PS II that was predicted to reduce O_2.

A limitation to flux balance analysis is that a wide range of flux values may be consistent with the constraints in the computational model. An iterative application of computational and

experimental methods is an important strategy to improve the comprehensive understanding of cyanobacterial metabolism. We have begun to apply this iterative approach, by including mRNA and protein expression datasets as additional constraints beyond biomass composition and physiological rate measurements. Experimentally-measured TPD were successfully used to further constrain the model, and thereby reduce uncertainty and increase the number of required (that is, metabolically active) reactions (Table 3.3). However, there remained discrepancies in that the model did not predict flux through all reactions for which proteins were experimentally detected. Such discrepancies can be used to subsequently improve the model with previously developed approaches [100-102]. For example, an earlier version of the model did not predict flux through proline oxidase, even though proteome data demonstrated that proline oxidase was synthesized. This prediction arose because the model did not contain a reaction in which $FADH_2$ (a product of the proline oxidase reaction) could be reoxidized to FAD. After experimental confirmation that proline can be used as a nitrogen source (implying activity of proline oxidase) by *Cyanothece* 51142, a $FADH_2$ recycling reaction was included in the final *i*Cce806 model.

Even with these additional TPD constraints, a wide range of flux values remained feasible (Figure 3.7). We should note that we did not take real enzymatic activities into account (which can be affected by post-translational modifications), as we did not have this type of data for the two conditions examined. Such data, if available, could be used as additional factors for determining whether to favor or disfavor fluxes through associated reactions (See Material and Methods). Other constraint-based methods for incorporating gene expression data use similar Boolean on/off type of constraints to restrict fluxes [91, 103, 104] and would be expected to yield results similar to those described herein. Thus, novel computational methods which can more quantitatively constrain the metabolic flux values are still needed. The strategy of

evaluating fluxes for reaction deletions *in silico* can be used to identify knockout mutants that can potentially improve the resolution of intracellular flux distributions. A flux that is well resolved would have a small span meaning we can more definitively state its value. If the mutants show no growth defects then the corresponding reactions may not be used under the conditions tested, or alternative pathways not included in the model may occur. Either way, this information could be used to better resolve the intracellular flux distribution or improve the metabolic model. For *Cyanothece* 51142, this would require development of a genetic system (such a system already exists for another *Cyanothece* strain [105]) as experiments with mutants would have the most potential to improve resolution of central metabolic fluxes during photoautotrophic growth. Also, as a complement to the *in silico* reaction knockouts that our simulations predict would reduce the flux spans associated with central metabolic reactions, the photobioreactor employed here provides a system whereby cultivation conditions can be rigorously controlled and some aspects of physiological state monitored continuously. In addition, cells from steady-state or perturbed cultures can be interrogated via physiological or biochemical analyses to experimentally test the predictions of the computational models for wild type or mutants. As the number of available cyanobacterial models continues to grow, cross-species physiological, genomic, and metabolic comparisons will enable the identification of core networks and contribute towards improving our understanding of metabolic processes in cyanobacteria.

3.3. Materials and methods

In this study, all model simulations were performed in GAMS software (General Algebraic Modeling System, GAMS Development Corporation, Washington, D.C.). To further constrain the models based on mRNA or protein expression data, a modified version of the

method developed by Shlomi *et al.* [91] was used. Here, we identified a single flux distribution that best agreed with measured transcriptome and proteome data (TPD) and minimized flux usage. Reactions with experimentally measured fluxes belong to set R_E (which included biomass production and exchange fluxes for oxygen, 630 nm and 680 nm photons) and were constrained to their measured values. Reactions associated with detected proteins were included in the high reaction set(R_H). Reactions associated with undetected proteins and genes with low mRNA expression levels (whose mRNA expression was less than the lowest mRNA expression of detected proteins) were included in the low reaction set(R_L). The method finds a flux distribution that maximizes the number of active reactions ($v \neq 0$) and inactive reactions ($v = 0$) in reaction sets R_H and R_L respectively. For reactions in setR_H, binary variables x and y indicate whether a reaction is active, meaning its flux is greater than a positive threshold($x = 0, y = 1$), or smaller than a negative threshold $-\varepsilon$ $(x = 1, y = 0)$ for reversible reactions. If both x and y are zero then the reaction is inactive and its flux value is zero. Likewise, a binary variable z is used for reactions in set R_L such that if $z = 1$ then the reaction is inactive($v = 0$). The original method [91] has alternate solutions, which can contain unrealistically high flux values due to the presence of cycles (e.g.,, futile cycles and circulations) in the network. To identify a solution that minimizes the use of these cycles, the objective function was modified to also minimize the sum of squared fluxes through the network.

The mixed integer quadratic programming formulation to identify a flux distribution that best matches TPD while minimizing flux magnitude is given below (Eq. 3.1 – 3.8):

$$\max \quad \sum_{j \in R_H} (x_j + y_j) + \sum_{j \in R_L} z_j - \sum_{j \in R} v_j^2$$

$$s.t. \quad S \cdot v = 0 \qquad \text{(Eq. 3.1)}$$

$$\alpha_j \le v_j \le \beta_j \qquad j \in R \qquad \text{(Eq. 3.2)}$$

$$v_j = v_j^{exp} \qquad j \in R_E \qquad \text{(Eq. 3.3)}$$

$$v_j + x_j(\beta_j + \varepsilon) \le \beta_j \qquad j \in R_H \qquad \text{(Eq. 3.4)}$$

$$v_j + y_j(\alpha_j - \varepsilon) \ge \alpha_j \qquad j \in R_H \qquad \text{(Eq. 3.5)}$$

$$(1 - z_j)\alpha_j \le v_j \le (1 - z_j)\beta_j \qquad j \in R_L \qquad \text{(Eq. 3.6)}$$

$$x_j + y_j \le 1 \qquad j \in R_H \qquad \text{(Eq. 3.7)}$$

$$x_j, y_j, z_j \in \{0,1\} \qquad \text{(Eq. 3.8)}$$

Additionally, to find the flux ranges consistent with the TPD, flux variability analysis (FVA) was performed by minimizing and maximizing the flux through each reaction in the network. In these FVA simulations, the same constraints described above were included (Eq. 3.1 – 3.8) and the binary variables (x, y, z) were further constrained by their optimal values $(x^{opt}, y^{opt}, z^{opt})$ found in the original problem (Eq. 3.9 – 3.10).

$$\max \text{ (or min)} \quad v_j$$

$$s.t. \quad \text{Eq. } 3.1 - 3.8$$

$$x_j + y_j = x_j^{opt} + y_j^{opt} \quad j \in R_H \qquad \text{(Eq. 3.9)}$$

$$z_j = z_j^{opt} \quad j \in R_L \qquad \text{(Eq. 3.10)}$$

Chapter 4

Computational evaluation of *Synechococcus* sp. PCC 7002 metabolism for chemical production

To date, almost all metabolic engineering efforts have focused on either *Synechocystis* sp. PCC 6803 or *Synechococcus elongatus* PCC 7942 [9]. These strains have been engineered to produce a variety of chemicals including lactic acid, ethanol, isobutanol, 1-butanol, isoprene, fatty acids, and ethylene [15, 16, 20, 21, 23, 26, 106, 107]. While *Synechococcus* 7002 has not been as extensively used in metabolic engineering, it has a number of attributes that make it a strong candidate for metabolic engineering. Compared to other cyanobacteria strains, it has a fast doubling time (~ 3.5 hours compared to ~12-24 hours) indicating a high inherent metabolic rate. *Synechococcus* 7002 also can grow under high light conditions (up to ~ 4.5 mE m^{-2} s^{-1} compared to its optimal light intensity of 250 μE m^{-2} s^{-1}) [108], can grow in salt water (obviating a requirement for fresh water) [109] and can be genetically manipulated [110-112]. In addition, the strain's genome has been sequenced and a variety of high-throughput experimental and computational tools have been used to evaluate it, including gene expression and genome-scale metabolic modeling [61, 113, 114].

For metabolic engineering purposes, a wide variety of computational approaches have been developed to calculate maximum theoretical yields or to predict the metabolic outcome of genetic perturbations. Example of such methods include flux balance analysis (FBA), minimization of metabolic adjustment (MOMA), regulatory on/off minimization (ROOM) and minimization of relative metabolic change (RELATCH [68, 115, 116] (See chapter 2 for the

formulation of FBA and MOMA). MOMA minimizes the sum of squared differences in flux distributions between mutant and parental strains [68], while ROOM minimizes the number of significant flux changes between mutant and parental strains [115]. On the other hand, RELATCH first limits increases in flux levels and then minimizes relative flux changes and latent pathway activation between mutant and parental strains. RELATCH has been shown to better predict flux distributions in unevolved and adaptively evolved strains of *Escherichia coli* for four different single-gene knockout mutants with higher accuracy than FBA, MOMA, and ROOM [116]. These methods can and have been used to predict how chemical production will be affected by the deletion of different genes in the host organism.

For strain-design purposes, a number of bi-level approaches have been developed which can identify what genetic manipulations are needed to improve chemical production by considering reaction deletions (OptKnock), gene deletions (OptORF, BiMOMA), reaction additions (OptStrain) or flux changes (OptForce) [117-121]. Most of these approaches try to couple biomass and chemical production so that adaptive evolution, where growth is used as a selection pressure, will improve both growth and chemical production rates. OptKnock has been successfully used to develop *E. coli* strains which produce lactate, succinate, and 1, 4 butanediol [122-124]. OptORF is similar to OptKnock but it identifies gene deletions (instead of reaction deletions) and regulatory changes needed to couple growth and chemical production [118].

In this chapter, we used the updated genome-scale metabolic model for *Synechococcus* 7002 (*i*Syp708, Chapter 2) and apply various computational methods including MOMA, RELATCH, and OptORF to predict metabolic engineering strategies that improve production of both native and non-native chemicals that have been studied in other cyanobacteria. This study

provides a thorough assessment of the yields, requirements and genetic strategies for utilizing *Synechococcus* 7002 as a background strain for metabolic engineering.

4.1 Results

4.1.1 Maximum theoretical yields of native and non-native products

The updated metabolic model was used to calculate the maximum theoretical yields for a variety of native and non-native products under photoautotrophic, dark anoxic, and dark oxic conditions. The maximum theoretical yields (mol product/ mol CO_2) for all considered products produced under photoautotrophic conditions were predicted to be greater than or equal to the yields under dark conditions (Table 4.1).

Table 4.1: Predicted theoretical yields (mol product/ mol CO_2 or photon) of different chemicals produced by *Synechococcus* 7002 under different conditions

Products	Photoautotrophic (mol /mol photon)	Photoautotrophic (mol /mol CO_2)	Dark anoxic (mol /mol CO_2)	Dark oxic (mol /mol CO_2)
Acetate[a)]	0.063	0.50	0.50	0.50
Alanine[a)]	0.042	0.33	0.33	0.33
Lactate[a)]	0.042	0.33	0.33	0.33
Succinate[a, b)]	0.031	0.25	0.19	0.24
Hydrogen[a,c)]	--	--	2	2
Fatty acid[d)]	0.005	0.056	0.039	0.039
Isobutanol[d)]	0.021	0.25	0.17	0.17
Isoprene[d)]	0.018	0.20	0.12	0.14
2-Methyl-1-butanol[d)]	0.017	0.20	0.13	0.13
1-Butanol[d)]	0.021	0.25	0.17	0.17
3-Methyl-1-butanol[d)]	0.017	0.20	0.13	0.13
1-Propanol[d)]	0.028	0.33	0.22	0.22
2-Phenylethanol[d)]	0.013	0.13	0.096	0.099
Ethylene[d)]	0.022	0.40	0.00	0.18
Ethanol[d)]	0.042	0.50	0.33	0.33

a) Products that can be produced *via* naturally occurred pathways in *Synechococcus* 7002
b) Products whose production is limited by energy under dark anoxic conditions
c) Products whose production is limited by reductant under dark anoxic conditions
d) Products whose production is limited by both energy and reductant under dark anoxic conditions

In addition, the photoautotrophic maximum theoretical yields of all products, except for ethylene, were equal to the inverse of the number of carbons in the products, indicating that the simulated conditions were carbon-limited. The predicted photoautotrophic maximum theoretical yield of ethylene (0.4 mol/mol CO_2), however, was less than the expected value (0.5 mol/mol CO_2). This was because the ethylene forming reaction that was added to our model produces succinate, guanidine, and 1-pyrroline-5-carboxylate in addition to ethylene. While succinate and 1-pyrroline-5-carboxylate can be re-consumed by reactions in the network, enzymes metabolizing guanidine do not appear to be present in *Synechococcus* 7002, and so guanidine was co-produced with ethylene, lowering the maximum theoretical yield for ethylene to 0.4 mol ethylene/ mol CO_2. Since the ethylene forming enzyme requires oxygen, ethylene could not be produced under dark anoxic conditions. It should be noted that under photoautotrophic and dark oxic conditions, we excluded hydrogen from the maximum theoretical yields calculations because hydrogenases are inhibited by oxygen [125].

Under photoautotrophic conditions, the model-predicted chemical production rates (and thus maximum theoretical yields) were limited by carbon; however, under dark conditions reductant and/or ATP limited most yields. The lower maximum theoretical yields for succinate, fatty acid, and other non-native products predicted under the two dark conditions imply that CO_2 was not the only factor limiting chemical production. While cells can use light energy to generate more energy and reductant (in the form of ATP and NADPH, respectively) under photoautotrophic conditions, they cannot do so under dark conditions. Therefore, besides CO_2, energy and reductant can potentially limit the production of desired products in the absence of light. We determined whether energy and/or reductant further limited production under dark conditions by adding artificial ATP- or NADPH-generating reactions ($ADP + HPO_4 + H \rightarrow ATP$

+ H_2O and NADP → NADPH + H) to the model and re-calculating the maximum theoretical yield for each product under the dark conditions. The resulting model predicted that under dark oxic and dark anoxic conditions, with unlimited amount of NADPH and ATP, the maximum theoretical yields increased to the values calculated under photoautotrophic conditions. Under the dark oxic condition, the presence of O_2 enabled the production of ATP from NADPH using the respiratory enzymes allowing for ATP synthesis (Figure 4.1). Therefore, under the dark oxic condition, adding free NADPH alone was sufficient to increase the maximum theoretical yields of products to values obtained under photoautotrophic conditions. In contrast, under the dark anoxic condition, succinate was predicted to be the only product whose maximum theoretical yield was limited only by ATP. The maximum theoretical yields of other products (except for acetate, alanine and lactate) only increased to the values obtained under photoautotrophic conditions when both free ATP and NADPH sources were present. Moreover, the model predicted that under dark anoxic conditions, acetate would be produced as a co-product when maximizing the production of isoprene and 2-phenylethanol. This co-product was made to generate extra ATP via the acetyl-coA synthetase reaction; however, with free ATP production, acetate was no longer predicted to be a co-product with these two products under dark anoxic conditions.

4.1.2 Predicted phenotypes of gene-deletion mutants under photoautotrophic and dark-anoxic conditions

The flux distributions obtained by fitting subsets of measured fluxes (from ^{13}C MFA or external flux data [12, 126]) and gene expression data ([113]) provided a reference point for methods used to predict the effects of gene deletions (MOMA and RELATCH [68, 116]). These reference flux distributions through central metabolism for photoautotrophic and dark anoxic conditions are shown in Figure 4.2. Using these reference flux distributions, we predicted flux

distributions in single-gene knockout mutants using MOMA and RELATCH and evaluated which mutants were predicted to improve production of the chemicals listed in Table 4.1.

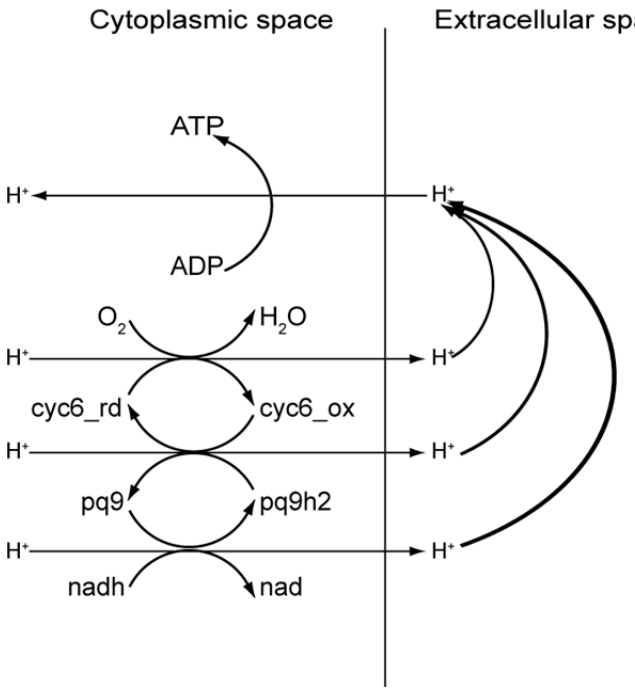

Figure 4.1: Schematic presentation of electron transport in *Synechococcus* 7002
In the presence of oxygen (O_2), there is a coupling between cytochrome oxidase reaction (COX_CYC), and NADH dehydrogenase reaction (NADHPQ9) via plastoquinone-cytochrome c oxidoreductase (PQCYCOR), which leads to a proton gradient across the membrane that can drive ATP synthesis. The reduced and oxidized forms of electron donor and electron acceptors for COX_CYC, NADH_PQ9, and PQCYCOR are cyc6_rd/cyc6_ox, nadh/nad, and pq9h2/pq9, respectively.

We identified mutants that improved product yields by at least 10% (if the products were already secreted by the wildtype strain), or enabled production of new products (if the products were not already secreted by the wildtype strain). Furthermore, for the photoautotrophic condition, we excluded gene deletions that were predicted to be lethal since the *Synechococcus* 7002 mutants need to be able to grow photoautotrophically.

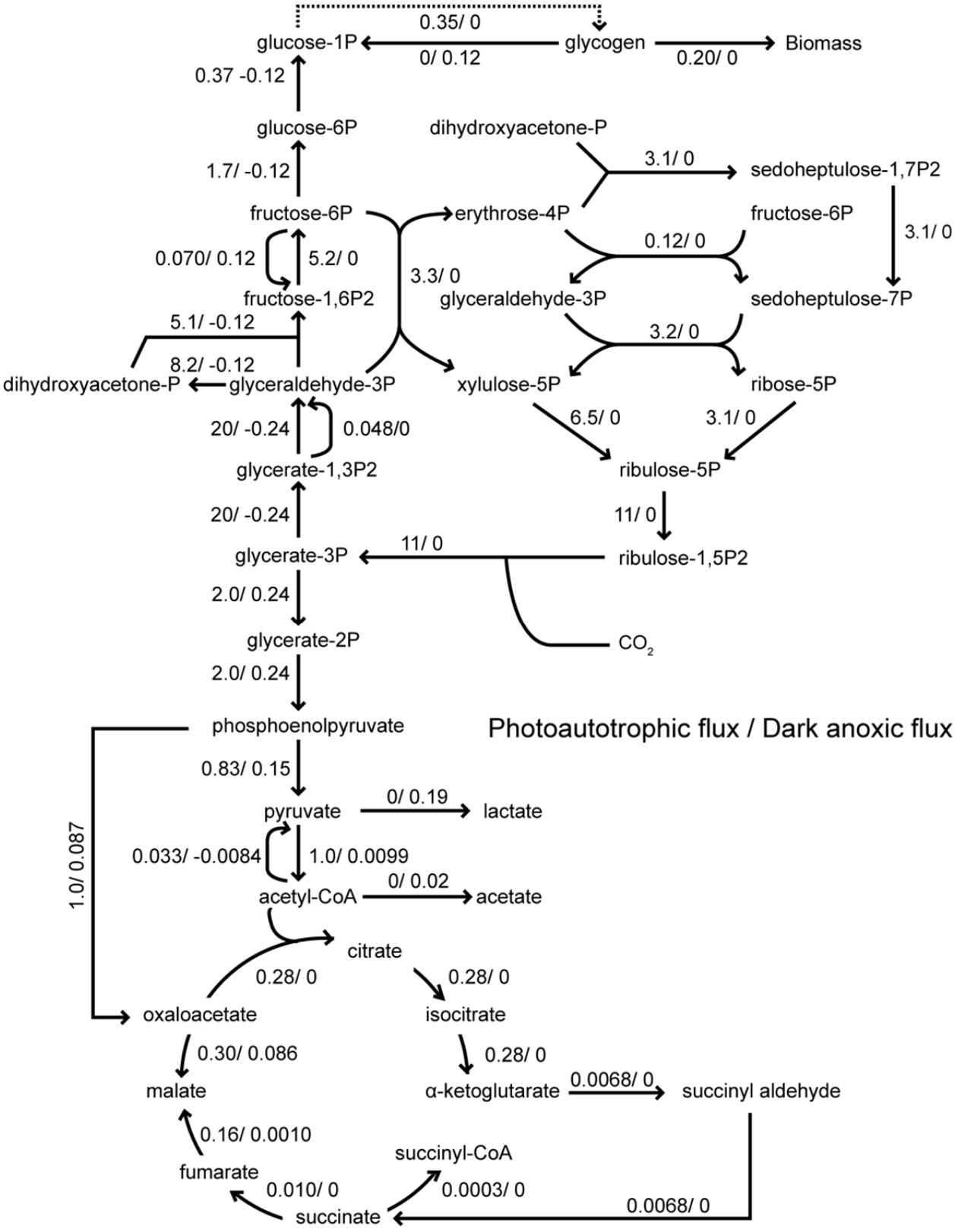
0.35/ 0
glucose-1P
glycogen
Biomass
0/ 0.12
0.20/ 0
0.37 -0.12
glucose-6P
dihydroxyacetone-P
1.7/ -0.12
3.1/ 0
sedoheptulose-1,7P2
fructose-6P
erythrose-4P
fructose-6P
3.1/ 0
0.12/ 0
0.070/ 0.12
5.2/ 0
3.3/ 0
fructose-1,6P2
glyceraldehyde-3P
sedoheptulose-7P
5.1/ -0.12
3.2/ 0
8.2/ -0.12
dihydroxyacetone-P
glyceraldehyde-3P
xylulose-5P
ribose-5P
6.5/ 0
3.1/ 0
20/ -0.24
0.048/0
glycerate-1,3P2
ribulose-5P
20/ -0.24
11/ 0
glycerate-3P
11/ 0
ribulose-1,5P2
2.0/ 0.24
glycerate-2P
CO2
2.0/ 0.24
phosphoenolpyruvate
Photoautotrophic flux / Dark anoxic flux
0.83/ 0.15
pyruvate
0/ 0.19
lactate
1.0/ 0.087
0.033/ -0.0084
1.0/ 0.0099
acetyl-CoA
0/ 0.02
acetate
citrate
0.28/ 0
0.28/ 0
oxaloacetate
isocitrate
0.30/ 0.086
0.28/ 0
malate
α-ketoglutarate
0.0068/ 0
succinyl aldehyde
0.16/ 0.0010
succinyl-CoA
fumarate
0.010/ 0
0.0003/ 0
0.0068/ 0
succinate

Figure 4.2: Predicted flux distributions of *Synechococcus* 7002 under photoautotrophic and dark anoxic conditions
The flux map shows the reference flux distributions, which were estimated using the MFA data for *Synechocystis* 6803 (measured under photoautotrophic condition) or secretion rates for *Synechococcus* 7002 (measured under dark anoxic condition) through central metabolic reactions. Flux values are reported in units of $mmol \cdot g^{-1}$ $AFDW \cdot h^{-1}$ for photoautotrophic and dark anoxic conditions. Reaction directions match those predicted for photoautotrophic condition. Negative flux values for the dark anoxic condition indicate the flux occurs in the opposite direction of the reaction. The dotted line represents the conversion of glucose-1P to glycogen via three sequential reactions that carry equal flux values (glucose-1-phosphate adenylyltransferase, glycogen synthase, and glycogen branching enzyme).

Our results show that under photoautotrophic conditions, MOMA was able to identify gene-knockout mutants with improved phenotypes for all 14 target products considered. On the other hand, RELATCH with tight parameter values, which predicts the behavior of unevolved mutants, only found unevolved mutants that were predicted to produce succinate, ethylene, fatty acids, isoprene, 2-methyl-1-butanol, 1-propanol, 2-phenylethanol, or 1-butanol (Figure 4.3A). When we used RELATCH with relaxed parameter values, which predicts the behavior of mutants that have been adaptively evolved, we were able to find mutants that were predicted to have improved production for almost all products, except for acetate and ethanol. Under dark anoxic conditions, both MOMA and RELATCH (with tight and relaxed parameters) methods were able to identify mutants with increased production for all of the target products (Figure 4.3B). Except for fatty acids, there were surprisingly few mutants (<30) that were predicted by both methods to improve chemical production. For both photoautotrophic and dark anoxic conditions, RELATCH predicted that adaptive evolution of mutants could lead to formation of desired products that might not be observed in unevolved mutants. In addition, we expected that products would be formed by more mutants under dark anoxic conditions than under photoautotrophic conditions since lethal mutants under photoautotrophic conditions were excluded from consideration.

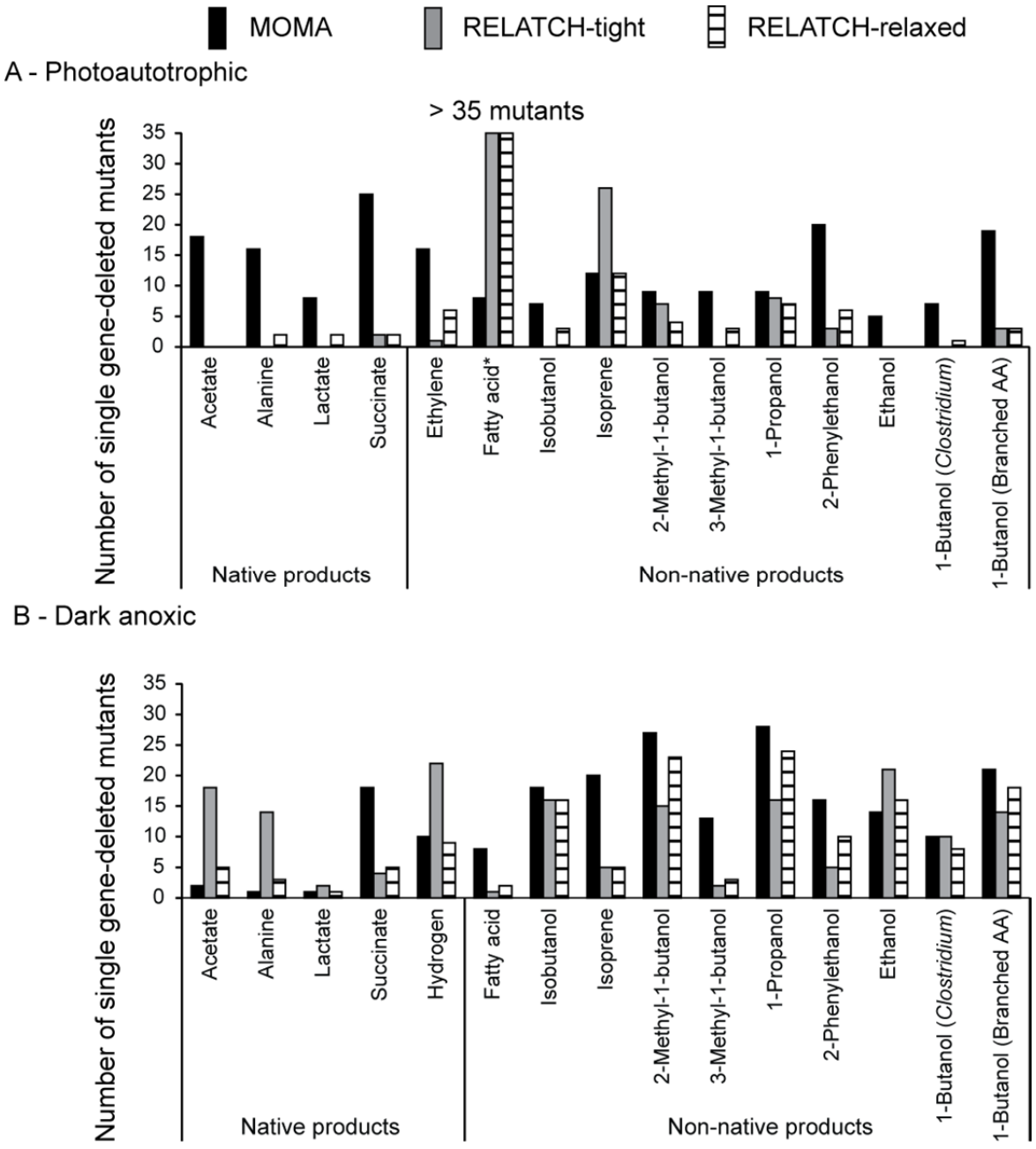

Figure 4.3: Number of *Synechococcus* 7002 mutants predicted by RELATCH and MOMA to produce different products under photoautotrophic and dark anoxic conditions

The bar charts show the numbers of single-gene deletions predicted by MOMA (black bars), RELATCH with tight parameters (grey bars) and RELATCH with relaxed parameters (striped bars) to improve production of a given product. (A) For photoautotrophic conditions, only non-lethal mutants with at least a 10% increase over the wildtype production levels are included. (B) For dark anoxic conditions, only mutants with at least a 10% increase over the wildtype production levels are included. The * indicates the actual number of mutants with improved fatty acid production is 425 and 421 for RELATCH with tight and relaxed parameters respectively.

Both MOMA and RELATCH predicted, however, that most non-native products could be formed by more mutants under dark anoxic conditions than photoautotrophic conditions, while some native products could be formed by more mutants under photoautotrophic conditions than dark anoxic conditions. As shown in Figure 4.4 and Figure 4.5, the number of central metabolic genes that can be deleted to improve chemical production under photoautotrophic condition was higher than under dark anoxic conditions.

4.1.3 Identifying adaptive evolutionary strategies for photoautotrophic chemical production

The mutants predicted by MOMA and RELATCH to have improved chemical production only involved single-gene deletion mutants and they do not guarantee coupling between growth and production of target chemicals. This coupling allows for the selection of improved chemical production by using a growth rate selection pressure in adaptive evolutionary experiments [117, 122]. We ran the OptORF algorithm to identify multiple gene-knockout mutants that would couple target chemical production to cellular growth. OptORF predicted that with a large number of gene deletions, coupling could be achieved, forcing *Synechococcus* 7002 to produce chemicals during optimal growth under photoautotrophic conditions (Figure 4.6). Using a maximum of 10 gene deletions, OptORF was able to identify strategies for 9 out of the 14 target chemicals, including: acetate, alanine, succinate, 1-butanol, ethylene, ethanol, 2-methyl-1-butanol, 3-methyl-1-butanol, and isoprene. Common strategies suggested by OptORF for all products involved blocking reactions or cycles that consume reducing power in the form of NAD(P)H (e.g.,, NADH dehydrogenase and transhydrogenase). Most of the strategies required 9 or 10 deletions and resulted in predicted growth rates between 0.15 to 0.20 h^{-1} and product yields between 15 – 34 % of the maximum theoretical yield. Because 1-butanol can be synthesized via either the branched amino acid pathway or the engineered CoA-dependent pathway [31, 126], we

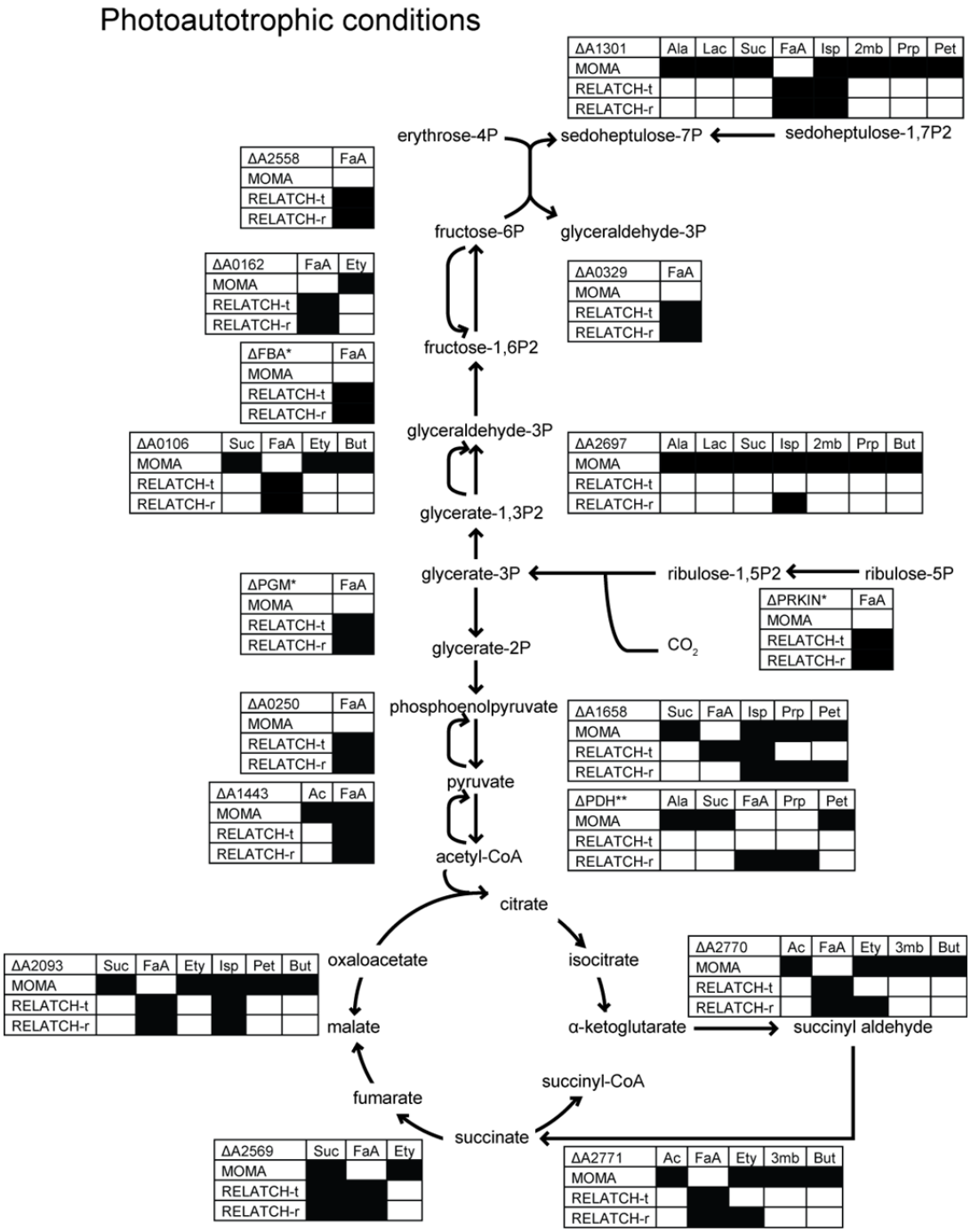

Figure 4.4: Gene deletions in the central metabolism of *Synechococcus* 7002 predicted to improve chemical production under photoautotrophic conditions

Each table represents a mutant which was predicted to improve chemical production of different products under photoautotrophic conditions. The mutants were predicted by MOMA, RELATCH with tight parameters (RELATCH-t), or RELATCH with relaxed parameters (RELATCH-r). The products are abbreviated as Ac – acetate, Ala – alanine, Lac – lactate, Suc – succinate, FaA - fatty acid, Ety – ethylene, Isp – isoprene, 2mb – 2-methyl-1-butanol, 3mb – 3-methyl-1-butanol, Prp – 1-propanol, Pet – 2-phenylethanol, and But – 1-butanol. The following abbreviations were used for reactions with isozymes, FBA – fructose bisphosphate aldolase, PGM – phosphoglucomutase, and PRKIN – phosphoribulokinase. Each entry in the table represents the mutant identified by the corresponding method as either predicted

(black) or not predicted (white) to improve chemical production. The * indicates that the reaction has isozymes. The ** indicates that pyruvate dehydrogenase (PDH) has 4 subunits (A1126, A0353, A0655, and A0110). RELATCH with relaxed parameters predicted that deletion of A1126 did not produce 1-propanol, while deletion of other subunits produced 1-propanol.

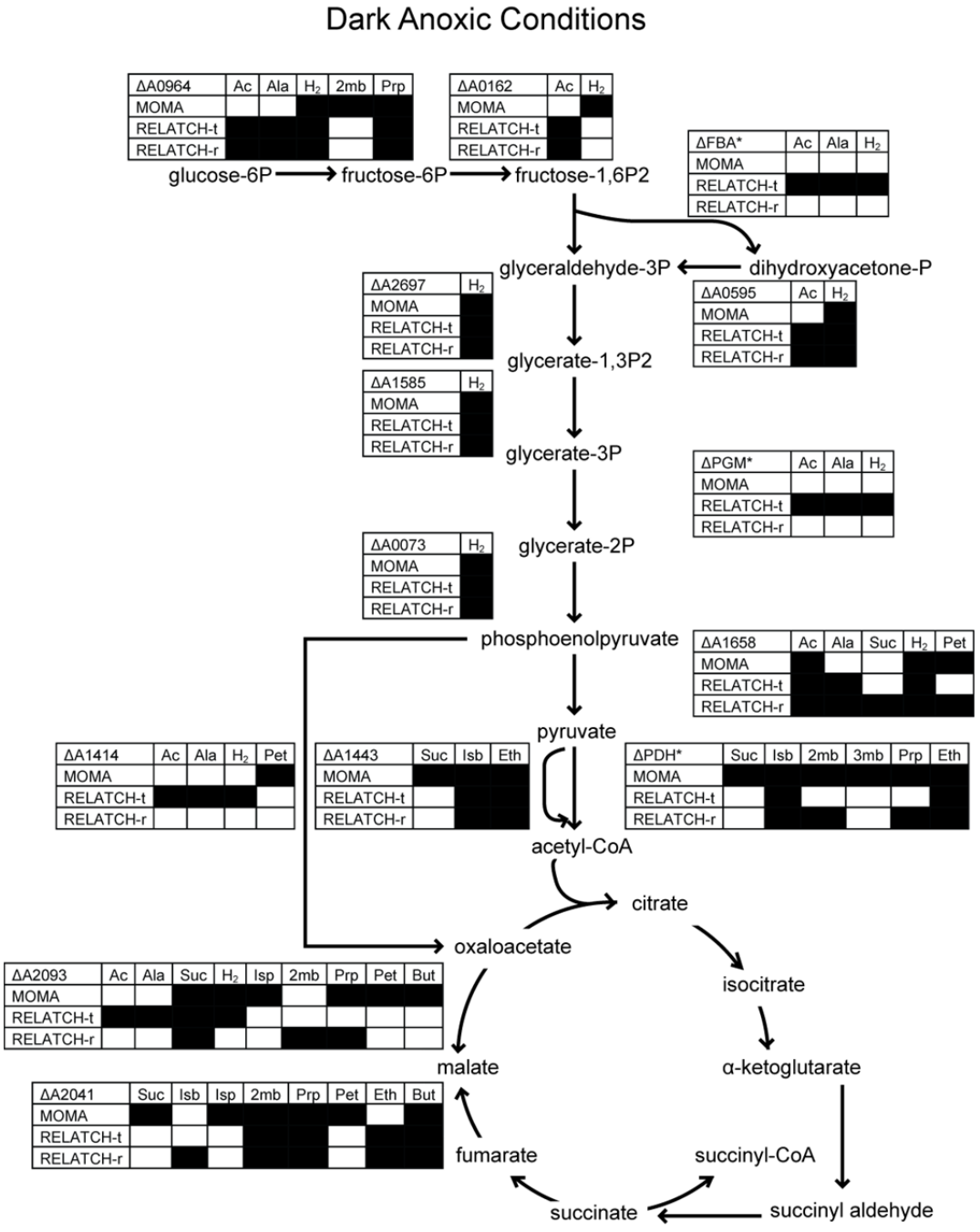

Figure 4.5: Gene deletions in central metabolism of *Synechococcus* 7002 predicted to improve chemical production under dark anoxic conditions

Each table represents a mutant which was predicted to improve chemical production of different products under dark anoxic conditions. The mutants were predicted by MOMA, RELATCH with tight parameters (RELATCH-t), or RELATCH with relaxed parameters (RELATCH-r). The abbreviations match those

used in Figure 3, with the addition of H_2 – hydrogen, and Isb – isobutanol. Each entry in the table represents the mutant identified by the corresponding method as either predicted (black) or not predicted (white) to improve chemical production. The * indicates that the reaction has isozymes.

ran OptORF separately for 1-butanol using the two different biosynthetic pathways. We found that the predicted yield for OptORF mutants using the CoA-dependent pathway (0.045 mol 1-butanol/ mol CO_2) was lower than the predicted yields for OptORF mutants using the branched amino acid pathway (0.085 mol 1-butanol/ mol CO_2), even though the maximum theoretical yields using the two pathways are the same under CO_2-limited photoautotrophic conditions.

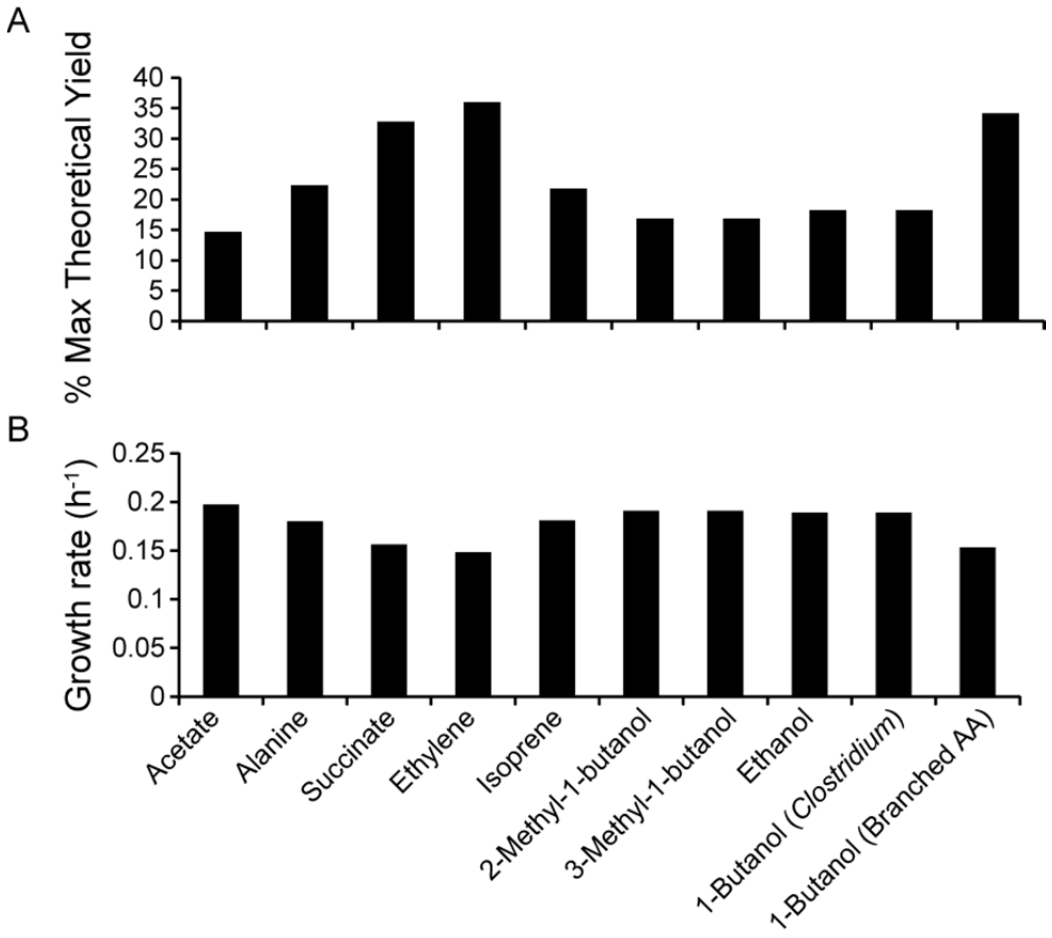

Figure 4.6: Predicted yields and growth rates for OptORF-designed *Synechococcus* 7002 mutants under photoautotrophic conditions
Mutants (using up to 10 gene deletions) with coupling between chemical production and cellular growth rates were found. The predicted production yields are shown in (A) as a percent of the maximum theoretical yields (the latter of which are reported in Table 4.1). The predicted growth rates are shown in (B), where the experimental wildtype growth rate is 0.198 h^{-1}.

4.2 Discussion

In this chapter used the genome-scale metabolic model of *Synechococcus* 7002 (*i*Syp708) to systematically evaluate the organism's potential for producing a range of biofuel precursors and chemicals. Since *Synechococcus* 7002 metabolism changes in response to the day-night cycle, we evaluated the model under both photoautotrophic, dark anoxic, and dark oxic conditions. We also estimated the maximum theoretical yields for a variety of products and identified single and multiple gene deletion strategies for improving chemical production rates and/or yields.

The maximum theoretical yields calculated under photoautotrophic and dark conditions assume no biomass or co-product formation (except for isoprene, 2-phenylethanol, and ethylene). Therefore, it is expected that the actual yield of each target chemical obtained *in vivo* would always be lower under photoautotrophic growth conditions. The maximum theoretical yields calculated under photoautotrophic conditions assumed that carbon was limiting chemical production. Since light and CO_2 uptake fluxes can affect chemical production rate, we also examined the maximum theoretical yield of several products including isobutanol, and lactate when varying both light and CO_2 uptake fluxes (Figure 4.7). We found that for a given CO_2 uptake flux, there was an optimal photon uptake flux that would result in the maximum theoretical yields of products. If the photon uptake fluxes were lower than the optimal level for a fixed amount of CO_2, the yield would be lower and byproducts (such as formate) were also secreted. In order to determine which products have the lowest light requirements, we also calculated the minimum amount of photons (in moles) required to produce one mole of target product and the corresponding water requirements (Table 4.2). Our results showed that the photon requirements generally increased as the number of carbons in the products increased.

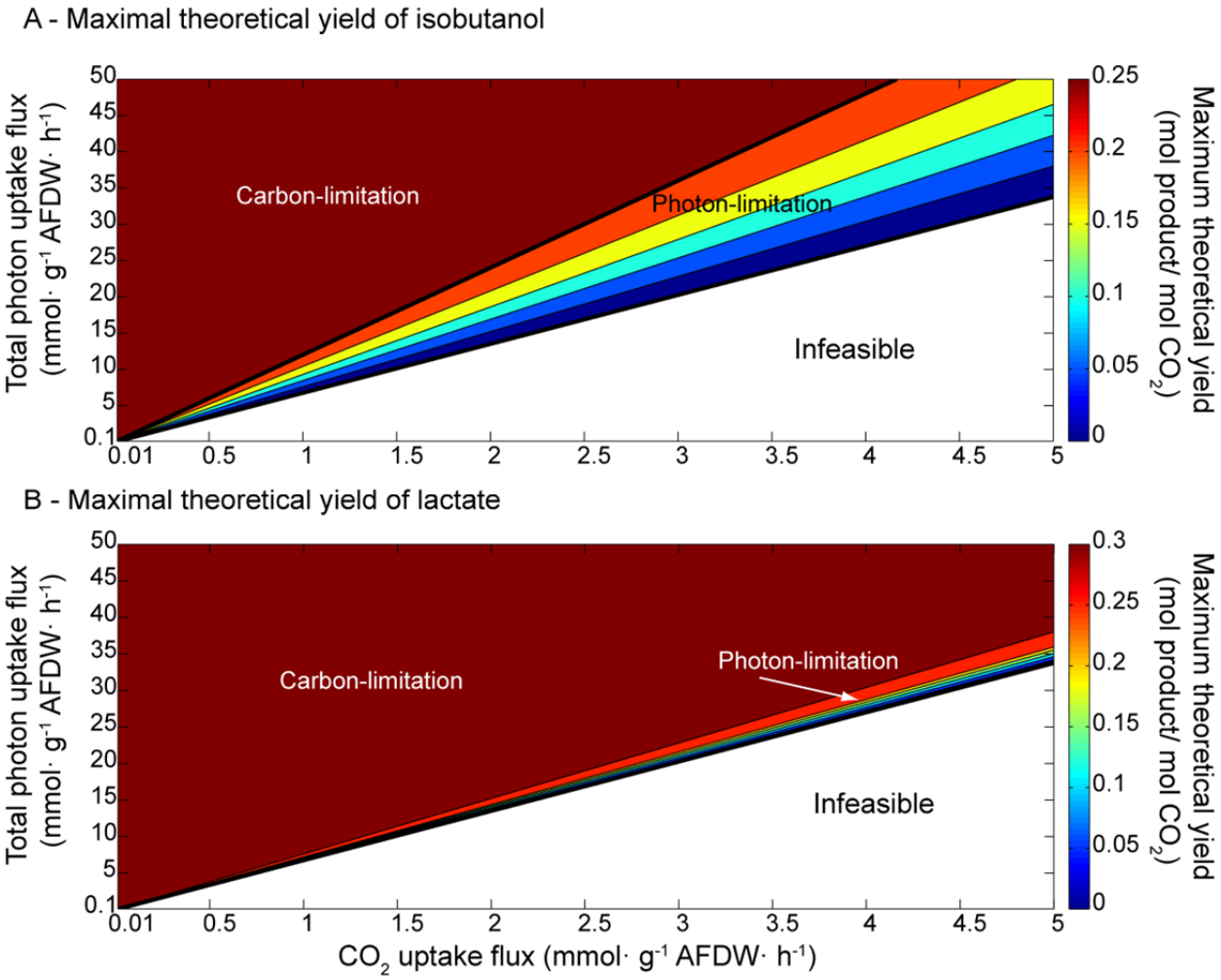

Figure 4.7: Model-predicted maximum theoretical yields of isobutanol and lactate produced by *Synechococcus* 7002 at different photon and CO_2 uptake fluxes
2-D phase plane displaying maximum theoretical yields of A) isobutanol and B) lactate for different values of CO_2 and photon uptake fluxes. The phase plane has 3 distinct regions: carbon-limitation region (CO_2 is limiting the yield, photon is in excess), photon-limitation region (photon is limiting the yield, CO_2 is in excess) and infeasible regions (photon is too limited for a fixed amount of CO_2).

The model predicted that most products required 8 – 12 photons and 0.5 – 1.5 water molecules per carbon atom incorporated into the product, with the biofuel products requiring close to 12 photons per carbon atom. For ethylene, it was significantly higher and required ~22.4 photons per carbon atom incorporated into ethylene. This was due to the co-production of guanidine since *Synechococcus* 7002 has no pathway to degrade this side product.

Table 4.2: Predicted photon and water requirements of *Synechococcus* 7002 grown under photoautotrophic conditions

Products	Photon required [a)]	H_2O required [b)]
Acetate	16	2
Alanine	24	2
Lactate	24	3
Succinate	32.75	3
Fattyacid	208	18
Isobutanol	48	5
Isoprene	56	4
2-Methyl-1-butanol	60	6
1-Butanol	48	5
3-Methyl-1-Butanol	60	6
1-Propanol	36	4
2-Phenylethanol	80	5
Ethylene	44.75	1
Ethanol	24	3

a) Number of moles of photons and water required to make 1 mole of product
b) Number of electrons required to reduce 1 mole of carbon in product

Even though acetate and ethanol both contain two carbon atoms and had equal maximum theoretical yields (Table 4.1), they have different photon (8 and 12 photons per carbon, respectively) and water requirements (1 and 1.5 water per carbon, respectively). The same was observed for 2-methyl-1-butanol and isoprene. The light and yield calculations were similar to recent estimates for some of these products in another cyanobacterium, *Synechocystis* 6803 [127]. These yield calculations and assessment of light and water requirements are useful for identifying conditions to optimize biochemical production (e.g., dark anoxic versus dark oxic), and for conducting economic and life cycle analyses of biochemical production using cyanobacteria.

Our predictions of single gene deletion mutant phenotypes using MOMA and RELATCH showed that for most products, MOMA tended to predict more mutants would have increased chemical production than RELATCH. Since MOMA's objective is to minimize the total

difference between wildtype and mutant's fluxes for all reactions in the network, the flux magnitude of each reaction in mutant strains will tend to be close to that in the wildtype strain [68]. When non-native pathways for a particular product were introduced into the model, MOMA tended to divert flux through reactions in these pathways in response to gene deletions, thereby leading to product formation. In contrast, RELATCH tries to minimize relative flux changes and latent pathway activation with limits on flux increases [116], and thus results in different secretion profiles. While RELATCH has been shown to make more accurate flux predictions for gene knockout mutants [116], it is not clear which method will be more accurate when non-native pathways are additionally incorporated into the mutants. In addition, the predictive ability of MOMA and RELATCH is sensitive to the reference flux distribution [68, 116]. Since the photoautotrophic reference flux distribution of *Synechococcus* 7002 was estimated by fitting fluxes to ^{13}C MFA data of *Synechocystis* sp. PCC 6803 [126], differences in photoautotrophic fluxes may exist between the two strains and could affect the accuracy of the reference flux distributions. Based on genome annotations, these two cyanobacteria share common genes encoding enzymes involved in central metabolism including glycolysis, pentose phosphate pathway, TCA cycle, carbon fixation, glycolate cycle, and photorespiration. This provides a basis for using the MFA data from *Synechocystis* 6803 to generate a reference flux distribution in this study. Nevertheless, a thorough genomic, metabolic, physiological, and ecological comparison of the two organisms, which is beyond the scope of this study, may reveal more significant differences in photoautotrophic fluxes between the two organisms. When more experimental data for *Synechococcus* 7002 mutant phenotypes and ^{13}C MFA flux estimates become available, the predictions of both computational methods should be re-evaluated and compared to experimental results.

While RELATCH generally predicted that more adaptively-evolved mutants than unevolved mutants would produce the desired products, there were a few products for which this was not the case. For example, under the dark anoxic condition, RELATCH predicted that the unevolved ΔA0246 mutant (missing a glutamine synthetase) would produce ethanol, while an evolved ΔA0246 mutant would not produce this product. Further investigation revealed that *Synechococcus* 7002 has two genes encoding a glutamine synthetase, A0246 and A1630 with the former having higher gene expression and thus greater enzyme contribution to the glutamine synthetase flux [113]. RELATCH predicts that during evolution of the ΔA0246 mutant, the glutamine synthetase flux will recover as A1630 compensates for the loss of A0246, thereby decreasing production of chemicals that were produced before adaptive evolution.

MOMA and RELATCH predictions suggested that for non-native products (except for fatty acids), it was more likely to identify mutants with improved chemical production under dark anoxic conditions than under photoautotrophic conditions. As mentioned earlier, this was most likely due to the fact that some of the mutants predicted to produce chemicals under dark anoxic conditions were lethal under photoautotrophic conditions. As a result, these knockout mutants would be difficult to implement experimentally and would not be viable candidates for metabolic engineering. However, altering the regulation of these genes could enable cells to still grow photoautotrophically with down-regulation of the genes in response to dark anoxic conditions enabling chemical production.

While the gene deletions predicted by MOMA and RELATCH to increase chemical production under photoautotrophic conditions were product specific, the gene deletion strategies suggested by OptORF for different products appeared to share some common genes that encode enzymes involved in NAD(P)H consuming reactions (e.g.,, A0195, A0196, A0197, A0984,

A0985, and A0986). OptORF strategies suggest that in order to produce target chemicals, one needs to block pathways that produce byproducts and to preserve energy and reductant for producing desired compounds. For a particular product, the gene deletions predicted by MOMA, RELATCH and OptORF also had little overlap. For example, while MOMA and OptORF predicted that deleting one of the genes encoding NADH dehydrogenase (A0195, A0196, or A0197) would enable 1-butanol production, RELATCH (with tight parameters) identified a completely different set of non-intuitive single gene deletions (A0707, A1023, or A2508) that are involved in chlorophyll and heme biosynthesis. Therefore, it will be important to experimentally test the predictions from different approaches to evaluate their predictive power.

The OptORF calculations suggested that it was generally difficult to couple growth to chemical production under CO_2-limited photoautotrophic conditions. One reason for this is that light was not constrained in our simulations so there was never an excess of reductant, which is often the reason for product formation under anaerobic conditions in other microbes. For the products which had available OptORF strategies, at least 9 gene deletions were required in order to achieve significant coupling between growth and product formation. Interestingly, OptORF failed to identify any strategies for isobutanol production, which had been previously shown to be relatively easy to couple to growth in *Escherichia coli* [118]. In contrast, no 1-butanol strategies were found by OptORF for *E. coli* but strategies with production at ~ 34% of the maximum theoretical yield were found for *Synechococcus* 7002. These differences between *E. coli* and *Synechococcus* 7002 could be caused by differences in metabolism between the two organisms, differences in the number of gene deletions allowed, effects of transcriptional regulation (which was additionally used to constrain the *E. coli* model), and an unconstrained

source of ATP and reductant in the *Synechococcus* 7002 model under photoautotrophic conditions

4.3 Materials and Methods

4.3.1 Calculations for maximum theoretical yields for native and non-native products under different conditions

Synechococcus 7002 produces a variety of autofermentation products, including lactate, succinate, acetate, alanine, and hydrogen [12]. Therefore, we added transport and exchange reactions for succinate, alanine, and lactate to the model (transport and exchange reactions for acetate and hydrogen were already in *i*Syp611 model). In addition, biosynthetic pathways for a range of biofuels and biofuel precursors were added to the model as needed to enable their production [20, 21, 31]. It should be noted that octadecanoic acid was chosen to represent a non-native free fatty acid produced by *Synechococcus* 7002 in this study. Maximum theoretical yield calculations were carried out in photoautotrophic, dark anoxic (i.e., autofermentation), and dark oxic conditions. In the photoautotrophic simulations, the CO_2 uptake flux was limited to 10 $mmol \cdot g^{-1}$ $AFDW \cdot h^{-1}$ and photon uptake fluxes were not limited. FBA [67] (Chapter 2) was then used to maximize production of each of the desired products. We calculated the maximum theoretical yield of each product as the ratio of the maximum production flux to the CO_2 consumption rate.

For the dark simulations, we performed FBA in two steps. In the first step, glycogen production flux was maximized with the same constraints as the photoautotrophic simulations. In the second step, we allowed glycogen to be the only carbon source (by setting the lower limits for CO_2 and other carbon containing exchange fluxes to zero) and set the photon uptake fluxes to zero to simulate dark conditions. Because *Synechococcus* 7002 can grow in a 12/12 light/dark

cycle [28], we assumed that the amount of glycogen accumulated during the light period is equal to that degraded during the dark period. Therefore, the glycogen degradation rate was constrained to the maximum photoautotrophic glycogen production rate found in the first step, and the O_2 exchange flux was either constrained to be zero (dark anoxic condition) or unconstrained (dark oxic condition). The maximum production flux of each desired product was then determined under dark conditions. The maximum theoretical yield under dark conditions was then calculated as the ratio of the maximum production flux to the CO_2 consumption rate, the latter of which was obtained from the first step where glycogen production was maximized under photoautotrophic conditions.

4.3.2 Predicting phenotypes of gene deletion mutants under photoautotrophic and dark-anoxic conditions

We used MOMA and RELATCH [68, 116] to predict flux distributions in single gene knockout mutants. Since both methods require a reference flux distribution, we first used expression data and either ^{13}C metabolic flux analysis (MFA) data for photoautotrophic conditions, or measured external flux values (substrate uptake and product secretion fluxes) for dark anoxic conditions to estimate a reference flux distribution ($\boldsymbol{w}$) [116]. We slightly modified the method described in [116] to obtain the reference flux distribution, in which the single optimization problem was broken into two optimization problems, which are shown below (Eq. 4.1 – 4.8).

Step 1: $$\min \quad \sum_{j \in J_{EXP}} \left(\frac{w_j^{meas} - w_j}{\sigma_j^{err}} \right)^2 \qquad \text{(Eq. 4.1)}$$

$$s.t. \quad \sum_j S_{ij} w_j = 0 \qquad \forall i \in I \qquad \text{(Eq. 4.2)}$$

$$w_j \geq 0 \qquad \forall j \in J_{irr} \qquad \text{(Eq. 4.3)}$$

$$w_j^{meas} - \sigma_j^{err} \leq w_j \leq w_j^{meas} + \sigma_j^{err} \qquad \forall j \in J_{EXT} \qquad \text{(Eq. 4.4)}$$

Step 2: $\min \quad \sum_{j \in J_{GPR}} \sum_{n \in N(j)} \frac{\left(W_{j,n}^{enz}\right)^2}{E_n}$ (Eq. 4.5)

$s.t.$ (Eq. 2) − (Eq. 4)

$w_j = w_j^{Step\ 1} \qquad \forall j \in J_{EXP}$ (Eq. 4.6)

$-\sum_{n \in N(j)} W_{j,n}^{enz} \leq w_j \leq \sum_{n \in N(j)} W_{j,n}^{enz} \qquad \forall j \in J_{GPR}$ (Eq. 4.7)

$W_{j,n}^{enz} \geq 0 \qquad \forall j \in J_{GPR}, n \in N(j)$ (Eq. 4.8)

In the first optimization problem (Step 1), the sum of squared differences between the measured fluxes values ($\boldsymbol{w^{meas}}$) and the flux variables ($\boldsymbol{w}$) was minimized and weighted by the reciprocal of measured flux errors ($\boldsymbol{\sigma^{err}}$) (Eq. 4.1). This objective function was defined for a subset of reactions ($\boldsymbol{J_{EXP}}$) where experimental data regarding flux values (^{13}C MFA or external flux data) was available. Since ^{13}C MFA data was not available for *Synechococcus* 7002 under photoautotrophic or dark anoxic conditions, we used ^{13}C MFA data recently reported for *Synechocystis* sp. PCC 6803 grown under photoautotrophic conditions [126] to predict the reference flux distribution for *Synechococcus* 7002 under photoautotrophic conditions. For the dark anoxic condition, we used previously reported secretion rates for alanine, hydrogen, acetate, and lactate for *Synechococcus* 7002 obtained under a dark anoxic condition [12] to constrain the exchange fluxes for the secreted products (See 4.3.2). The first optimization problem included steady-state mass balance (Eq. 4.2) and reaction reversibility (Eq. 4.3) constraints to define a feasible solution space. The measured external flux values and errors were used to limit the flux variables for reactions ($\boldsymbol{J_{EXT}}$) whose external flux data were available (Eq. 4), and these constraints (Eq. 4.4) were only used to predict the reference flux for *Synechococcus* 7002 under dark anoxic condition.

4.3.2.1 Estimating MOMA's and RELATCH's reference flux distributions under photoautotrophic and dark anoxic conditions

For the photoautotrophic condition, we used the ^{13}C metabolic flux analysis (MFA) data for *Synechocystis* 6803 grown under photoautotrophic condition [126] to estimate a reference flux distribution for *Synechococcus* 7002 using the two-step optimization procedure above. The MFA flux distribution, which was normalized to a CO_2 uptake flux of 100 mmol·g^{-1} AFDW· h^{-1}, was scaled down to a CO_2 uptake flux of 8.62 mmol·g^{-1} AFDW·h^{-1} so that the predicted growth rate would be ~ 0.20 h^{-1}, which is equivalent to the physiological doubling time of *Synechococcus* 7002 (~ 3.5 h [113]). Some fluxes reported in the MFA dataset were not considered in the fitting procedure because the reactions were not present in the *i*Syp708 model. The MFA data that was used in our calculations were shown in Table 4.3 below.

Table 4.3: ^{13}C Metabolic flux analysis (MFA) data from *Synechocystis* 6803

Reaction [a]	MFA data [b]	Scaled data [c]	Reaction [a]	MFA data [b]	Scaled data [c]
PGI	-19 ± 5	-1.64	TAL	-1 ± 8	-0.09
G6PDHy	16 ± 5	1.38	FBA3	-36 ± 8	-3.10
FBP	60 ± 7	5.17	SBP	36 ± 7	3.10
FBA	-60 ± 7	-5.17	PDH	11.8 ± 0.3	1.02
TPI	-95 ± 5	-8.19	CS	3.2 ± 0.2	0.28
GAPD	-228 ± 9	-19.65	ACONT	3.2 ± 0.2	0.28
PGM	23.2 ± 0.4	2.00	ICDHy	3 ± 0	0.26
ENO	23.6 ± 0.2	2.03	SUCD1i	0.2 ± 0.2	0.02
PYK	9.5 ± 1.6	0.82	FUMH	1.8 ± 0.2	0.16
RPE	-75.9 ± 0.4	-6.54	MDH	-3.6 ± 1.7	-0.31
RPI	35.6 ± 0.2	3.07	ME2	5.3 ± 1.6	0.46
PRKIN	127 ± 5	10.95	PPC	11.6 ± 1.7	1.00
RBC	127 ± 5	10.95	RBO	0.4 ± 0.4	0.03
TKT2	-38.5 ± 0.2	-3.32	PGLYCP	0.4 ± 0.4	0.03
TKT1	-37.3 ± 0.2	-3.22	GLYCTDH	0.4 ± 0.4	0.03

[a] Reactions' abbreviations are the same as those used in *i*Syp611 [61].
[b] Normalized MFA data (in mmol·g^{-1} AFDW·h^{-1}) for *Synechocystis* 6803 [126]
[c] The scaled flux data was calculated by taking the normalized MFA data and multiplying it by 0.0862.

For dark anoxic conditions, we used measured secretion rates for *Synechococcus* 7002 [12] to estimate the reference flux distribution. All measured fluxes except for glycogen degradation rate (DM_glycogen) shown below in Table 4.4 were used in the objective function of the first step (Eq. 4.1). The measured glycogen degradation rate was not included in the objective function, but instead used to fix the glycogen uptakeflux (glycogen uptake flux was fixed to 0.118 mmol·g^{-1} AFDW·h^{-1}). We also excluded the hydrogen data from the set of external fluxes ($\boldsymbol{J_{EXT}}$) (Eq. 4.4) to ensure the model's feasibility, but the hydrogen data was still used in the objective function (Eq. 4.1). In addition, we converted the measured external rates and errors from the reported unit (mol·10^{-17} cells·day^{-1}) to a unit compatible with the model (mmol·g^{-1} AFDW·h^{-1}) using the reported cell weight (5 mg/ 10^{11} cells), assuming it was ash-free dry weight. The measured secretion rates and errors that were used to estimate the reference flux distribution under dark anoxic condition were shown in Table 4.4 below.

Table 4.4: Measured external fluxes and errors for *Synechococcus* 7002 under dark anoxic condition

External fluxes	mol·10^{-17} cells·day^{-1}	mmol·g^{-1} AFDW·h^{-1}
Ex_h2_e (*)	2.7 ± 0.4	0.023 ± 0.003
Ex_ala-L_e	3.6 ± 1.2	0.03 ± 0.01
Ex_lac-D_e	21.6 ± 12	0.18 ± 0.1
Ex_ac_e	3.0 ± 3	0.025 ± 0.007
DM_glycogen	14.1 ± 5	0.118 ± 0.04

(*) Hydrogen was excluded from the set of measured external fluxes ($\boldsymbol{J_{EXT}}$) in equation 4.4

In the second optimization problem (Step 2), fluxes through experimentally measured reactions in set ($\boldsymbol{J_{EXP}}$) were fixed to the values found in Step 1 $\left(\boldsymbol{w_j^{Step\ 1}}\right)$ while the sum of squared enzyme contribution variables $\left(\boldsymbol{W_{j,n}^{enz}}\right)$, weighted by the reciprocal of enzyme expression values ($\boldsymbol{E_n}$), was minimized (Eq. 4.5 – 4.6). The sum of the enzyme contribution variables over all isozymes **(n)** for a particular reaction was used to place limits on fluxes for reaction that have gene-protein-reaction (GPR) associations ($\boldsymbol{J_{GPR}}$) (Eq. 4.7). The gene expression value ($\boldsymbol{E_n}$) for

multi-component enzymes was determined by summing all the gene expression values for genes encoding enzyme subunits. The expression data for *Synechococcus* 7002 grown under photoautotrophic and dark anoxic conditions were used from previously reported RNA-seq experiments [113]. The reference flux distribution was the solution to the second optimization problem and was used for MOMA and RELATCH predictions of mutant fluxes [68, 116].

For all native products, we simulated single-gene deletions by applying MOMA and RELATCH [68, 116] on the metabolic network of *i*Syp708. For each non-native product, we added the biosynthetic pathway to *i*Syp708 and applied MOMA and RELATCH [68, 116] on the modified model. For each single-gene deletion, we predicted the production of target chemicals under photoautotrophic and dark anoxic conditions. We used growth rates and production rates of different *Synechococcus* 7002 mutants to identify sets of tight and relaxed parameter values used by RELATCH. These tight and relaxed parameter values are used by RELATCH to predict the non-adapted (i.e., unevolved) and adapted (i.e., evolved) states, respectively, after genetic perturbations (See 4.3.3).

4.3.2.2 Determining values of alpha (α) and gamma (γ) for RELATCH

We performed sensitivity analyses for the parameters α and γ that were used in the RELATCH algorithm following the procedure described in [116] (See 4.3.4). Since MFA data were not available for *Synechococcus* 7002 mutants under either photoautotrophic or dark anoxic conditions, we used the predicted growth rate or product secretion rates to determine these parameter values. For the photoautotrophic condition, we identified 6 un-evolved mutant strains of *Synechococcus* 7002 whose growth rates have been reported either qualitatively or quantitatively to be similar to the wildtype strain [12, 28, 62, 128]. We ran RELATCH for these knockouts separately with [101x101] pairs of different α and γ values, and evaluated the

predicted growth rates of each of these mutants. RELATCH predicted that growth variation was more sensitive to α and γ for all tested mutants, where a lower α value corresponded to higher growth rate (Figure 4.8). From these results we chose $(\alpha = 3.16,\ \gamma = 10)$ and $(\alpha = 0.316, \gamma = \infty)$ for the tight and relaxed parameters to predict the un-evolved and evolved states of *Synechococcus* 7002 mutants a under photoautotrophic condition.

For dark anoxic condition, we chose the parameters for which RELATCH's predictions of product secretion fluxes of alanine, acetate, hydrogen and succinate in the lactate dehydrogenase mutant strain (Δ*ldhA*) were most consistent with experimental data [12]. Since the wildtype *Synechococcus* 7002 does not produce succinate experimentally, we first looked at the effects of α and γ on RELATCH's prediction for succinate production in a *ldhA* knockout mutant (Figure 4.9). Our result showed that there was a region of α and γ values for which succinate secretion flux was non-zero (Figure 4.9A). To evaluate the differences between model predictions and experimental data, we also calculated the sum of squared errors per flux (SSE) value for this mutant (Figure 4.9B). The SSE values were smallest in a similar region of α and γ values, indicating that the predicted production rates of alanine, hydrogen, and acetate were closest to the experimental data. Based on these results we used $(\alpha = 10,\ \gamma = 1)$ and $(\alpha = 1, \gamma = \infty)$ for the un-evolved and evolved states of *Synechococcus* 7002 mutants under a dark anoxic condition.

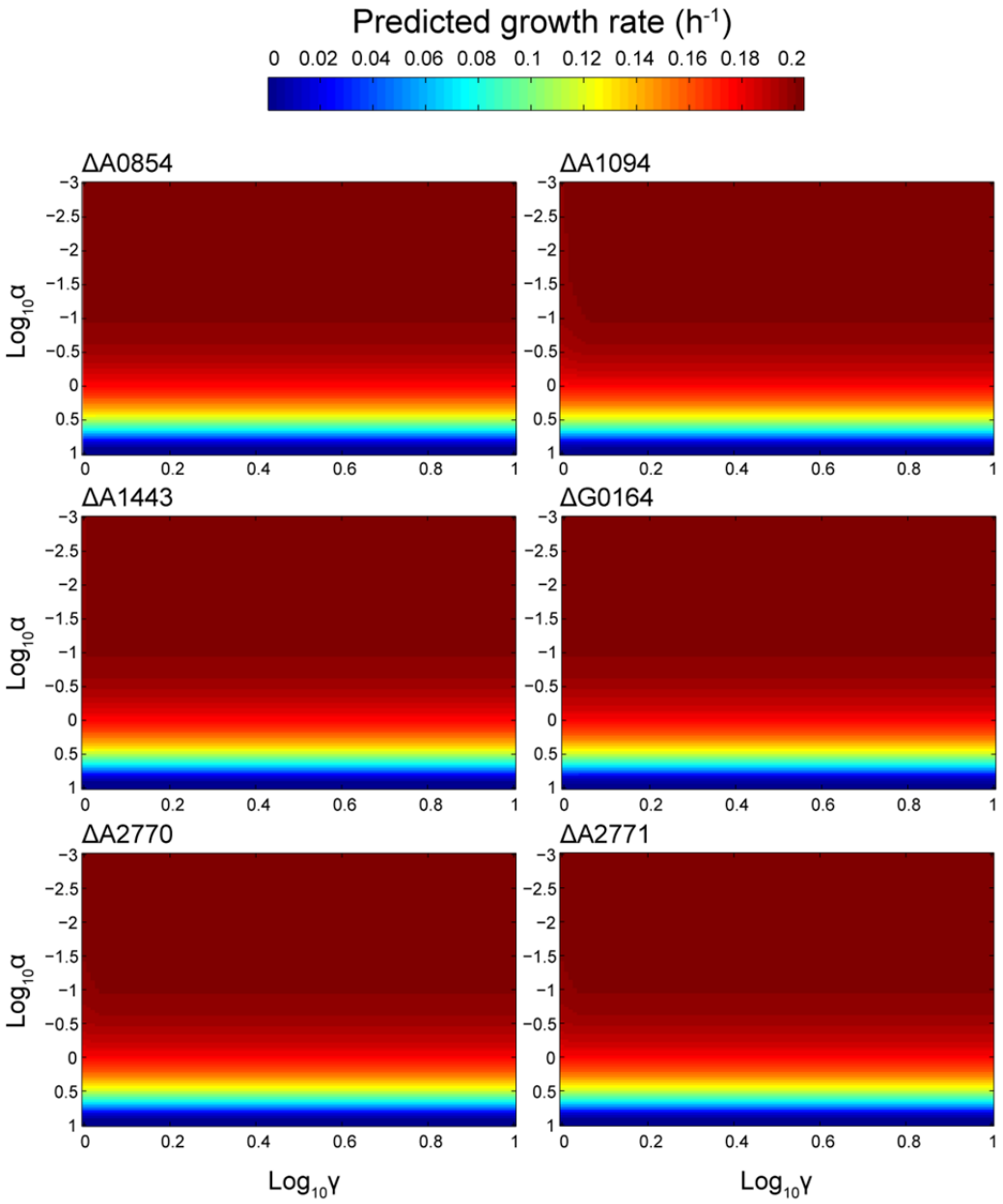

Figure 4.8: Sensitivity analyses of RELATCH parameters on predicted growth rates of *Synechococcus* 7002 mutants under photoautotrophic condition

Each heat map plot shows the growth rate (in h^{-1}) for a *Synechococcus 7002* mutant predicted by RELATCH for different pairs of α and γ values.

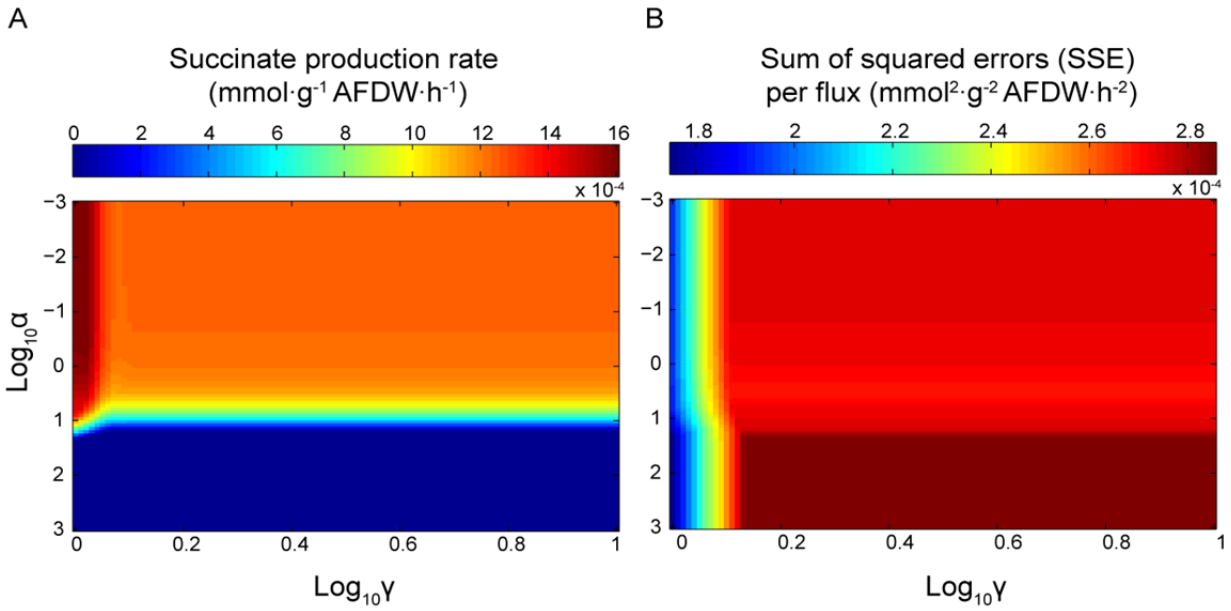

Figure 4.9: Sensitivity analyses of RELATCH parameters for succinate production rate and sum of squared errors per flux of *Synechococcus* 7002 *ldhA* mutant under dark anoxic conditions
A) Heat map plot shows the predicted succinate production rate (in mmol·g^{-1} AFDW·h^{-1}) for the *ldhA* mutant under dark anoxic conditions. B) Heat map plot shows the sum of squared error (SSE) per flux (in mmol^{2}·g^{-2} AFDW·h^{-2}) between measured and predicted rates for acetate, alanine, succinate, and lactate in *ldhA* mutant under dark anoxic conditions.

4.3.3 Identifying strain design strategies for production of target chemicals during photoautotrophic growth using OptORF

While MOMA and RELATCH predict a given mutants' capability to secrete target chemicals, the predicted gene deletion strategies often result in production of other byproducts besides the chemicals of interest. In addition, to find strategies involving larger numbers of deletions all possible combinations of higher order knockouts would need to simulated using these approaches, which is computationally difficult [119]. We thus used the OptORF algorithm [118] without regulatory constraints (since these are not known for *Synechococcus* 7002) to identify gene knockout mutants with up to 10 gene deletions that can produce target chemicals while growing maximally under photoautotrophic conditions. Briefly OptORF is a bi-level programing algorithm that looks for gene deletions that would produce chemicals of interest if cells maximized their growth [118]. We ran OptORF to identify growth-coupled gene deletion

strategies for both native and non-native products under photoautotrophic conditions, since adaptive evolution under this condition would lead to improved production rates.

Chapter 5

Development of a co-culture model for *Synechoccoccus* sp. PCC 7002 and *Shewanella* sp. W3-18-1

Microorganisms in nature tend to live in a community where more than one species cohabitate the same environment, and therefore they interact with one another. The term 'symbiosis' refers to the interaction between two different species living in proximity and the interaction can be beneficial for one or both organisms [129]. The interactions between pair of species living in the same microbial consortia can be classified based on the effects of metabolic exchange between the two species. Mutualisms refer to the bidirectional exchange of metabolites between pair of organisms such that the exchange is beneficial to both species [130]. Commensalism is the unidirectional exchange of metabolites such that only one organism benefits from the interaction [130]. If one species benefits from the interaction while the other is harm, the interaction is called parasitism [130]. In addition, an interaction can be neutralism if neither species benefits from the other [130].

Organisms that can live symbiotically are very diverse with microorganism are the dominant species. A famous example of symbiotic relationship is the light-emitting symbiosis between the marine bacteria *Vibrio fischeri* and the Hawaiian bobtail squid *Euprymna scolopes.* In this relationship, the light emitted from the bacteria help the squid in several activities such as prey attraction, and predator evasion and in return, the squid provide nutrients for bacterial growth [131]. Another example of symbiosis occurs between plants and *Rhizobia* in which the bacteria *Rhizobia* induce the formation of a specialized organ called nodule on roots of the host

plant; the bacteria fix nitrogen in this organ and provide the plant with ammonia in return for starch and sugars [132]. The benefits that *Rhizobia* offer to the host plants have motivated studies on characterization and isolation of different *Rhizobia* strains that can tolerate different soil conditions and improve crop yield [133]. Symbiotic microorganisms are also subjects for studying the biosynthesis of secondary metabolites which are metabolites that are not essential for the growth of the microorganisms and can be pharmaceutically valuable; hence they often are potential candidates for the development and discovery of new drugs [134]. The associations between sponge and marine microorganisms have produced many natural products such as swindholide A (a cytotoxic agent), and mycalamide A (an antitumor compound) [135].

Among symbiotic microorganisms, cyanobacteria are considered advantageous symbionts (organisms that are in symbiotic relationship) because of their ability to fix carbon dioxide and nitrogen, which can provide significant carbon and nitrogen sources in usable forms for the organisms that they are associated with [135]. In addition, cyanobacteria can naturally produce various secondary metabolites such as anatoxin-a (a neurotoxin produced by *Anabaena flos-aquae*), and borophycin (a cytotoxin produced by *Nostoc linckia*) [136], or can be engineered to produce various non-native or native chemicals (Chapter 1). Moreover, cyanobacteria can form symbiotic associations with many organisms including plants, animals, and bacteria [137]. Although symbiosis has been studied for many years, the mechanisms of how metabolites are exchanged between symbiotic organisms are not fully understood. Therefore, it is desirable to develop a co-culture model that can represent a synthetic microbial interaction between cyanobacteria and other microorganism, and reflect such interactions and thus help fill in the knowledge gaps that have not been known before. Such a model can also be

used to study the biosynthesis of valuable natural products produced from synthetic symbiotic association.

This chapter reviews recent research development in modeling microorganism community from the literature. It also reports the development of a co-culture model for *Synechococcus* sp. PCC 7002 (*Synechococcus* 7002) and *Shewanella* sp. W3-18-1 (*Shewanella* W3181) using genome-scale metabolic models that have been developed for these organisms (Chapter 2). The co-culture was analyzed to study the metabolite exchange between the two organisms, and to see if the co-culture can sustain growth under different nutrients. The result of this study provides insights into microbial interactions and bacteria's growth in a community.

5.1 Literature survey of recent development in modeling community of microorganism

The symbiotic relationship between *Rhizhobia* and host plant has been modeled to study the first step in the interaction between the two organisms which was the nitrogen fixation within the nodule of the host plants'roots [138]. This model was able to qualitatively predict the pathways being utilized during nitrogen fixation stage and the predictions were in good agreement with the literature. The relationship between a methanogen archae *Methanococcus maripaludis* and a hydrogen-producing bacteria *Desulfovibrio vulgaris* was also modeled as symbiotic association [139] and this model not only captured some of the physiological aspects of symbiotic growth between the two organisms but also determined that hydrogen was more essential as an electron carrier than formate, which cannot be easily determined using classical biochemical and genetic techniques. Modeling interactions among different cell types or tissues from the same organism (eukaryotes, brain tissues, human whole-body) also provided insights into methodology of modeling complex communities for which each cell type or species in the

community was treated as a separate compartments in the combined network [140-142]. Computational algorithms associated with multi-species models were also developed to analyze the models. Notably, Klitgord *et. al.*, have developed an algorithms that search for metabolites that can be exchanged between pairs of microorganisms, and also for media that can induce specific types of interaction between two species [130]. The results of this work helped identify synthetic interactions among species through genetic and environmental perturbations, which can potentially be useful in metabolic engineering.

Most of the methods discussed above involve a single optimization problem. In contrast, OptCom, a computational algorithm that has been developed to study microbial interactions, was formulated as a bilevel optimization problem extended from flux balance analysis (FBA) algorithm ([67], Chapter 2). In this formulation, it was assumed that the microbial community maximizes the total biomass while each species in the community also tries to maximize its own biomass [143]. Although OptCom predictions for the interaction between *M. maripaludis* and *D. vulgaris* agreed with experimental data, OptCom overpredicted the growth rate ratio between two species in a three-species photoautotrophic microbial community [143, 144]. In addition, OptCom assumes that the biomass concentration of each species in the community is equal, which might not be practical.

5.2 Development and analysis of a co-culture model for *Synechococcus* 7002 and *Shewanella* W3181.

Cyanobacteria often possess oxygen-sensitive hydrogenases which are responsible for the production of hydrogen [3]. In *Synechococcus* 7002, hydrogen was only produced via dark-fermentation to avoid hydrogenases being inhibited by oxygen evolved during photosynthesis

[12] (Chapter 1). This issue in theory can be alleviated by coupling the oxygen evolution in *Synechococcus* 7002 with oxygen consumption in *Shewanella* W3181 (Figure 5.1).

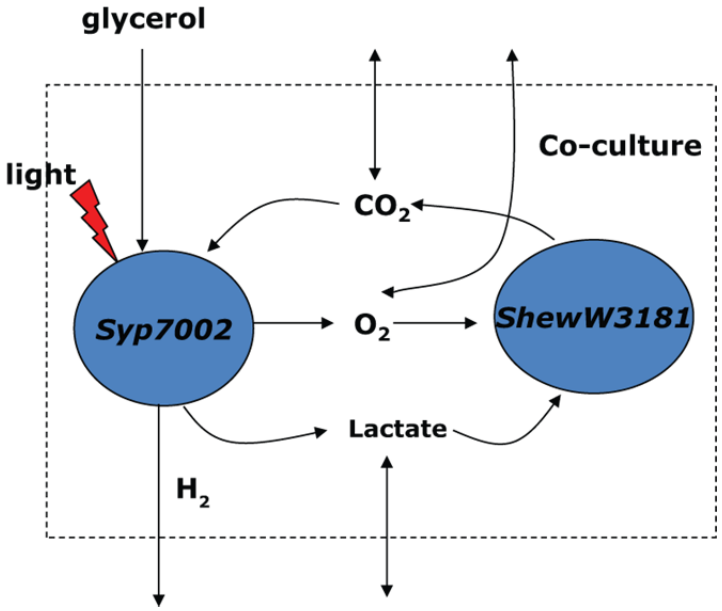

Figure 5.1: Schematic illustration of the metabolite exchanges between *Synechococcus* 7002 and *Shewanella* W3181
Synechococcus 7002 (Syp702) can grow photoautotrophically using light energy and CO_2, or photoheterotrophically with light and glycerol. Oxygen evolved from photosynthesis of Syp7002 can be consumed by *Shewanella* W3181 (ShewW3181) as an electron acceptor. ShewW3181 can grow on a number of carbon sources (e.g., lactate, acetate), and secreted CO_2 as by product, which can be consumed by Syp7001 via carbon fixation. The metabolites (e.g., H_2, CO_2, lactate, O_2) exchanged from each species in the co-culture model can be exchanged to the environment.

As shown in Figure 5.1 oxygen and organic carbon compound (e.g., lactate) produced from *Synechococcus* 7002 grown photoautotrophically are consumed by *Shewanella* W3181. In return, *Shewanella* supplies carbon dioxide for *Synechococcus* 7002. Like other *Shewanella* species, this organism can utilize a wide range of carbon sources (Figure 5.2) and electron acceptors (such as oxygen, iron, etc) [145].

The co-culture was developed to model growth of *Synechococcus* 7002 and *Shewanella* W3181 in chemostats in which growth rates can be controlled by monitoring the dilution rate. To illustrate the reconstruction of the co-culture model from the genome-scale models of each species, a toy-network was constructed (Figure 5.3). In this model, the media and the two

organisms are treated as three separate compartments that are linked to each other via transport reactions.

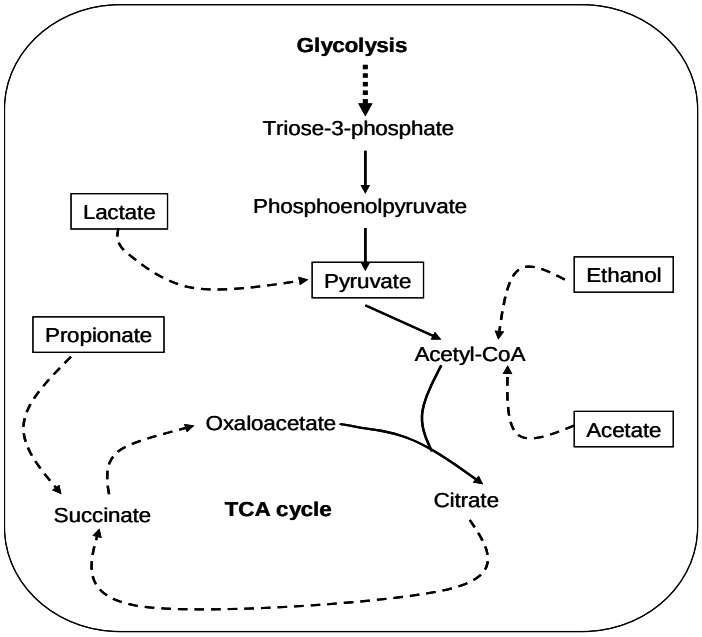

Figure 5.2: Example of intermediary carbon metabolism in *Shewanella*
Carbon sources of different chain length that can be metabolized by *Shewanella*: C2 (ethanol, acetate), C3 (lactate, propionate, pyruvate). Ethanol, acetate and lactate are known to be secreted by *Synechococcus 7002*. *Synechococcus* does not produce propionate naturally but can produce its precursor 2-oxobutanoate which can be converted to propionate by *Shewanella* enzyme.

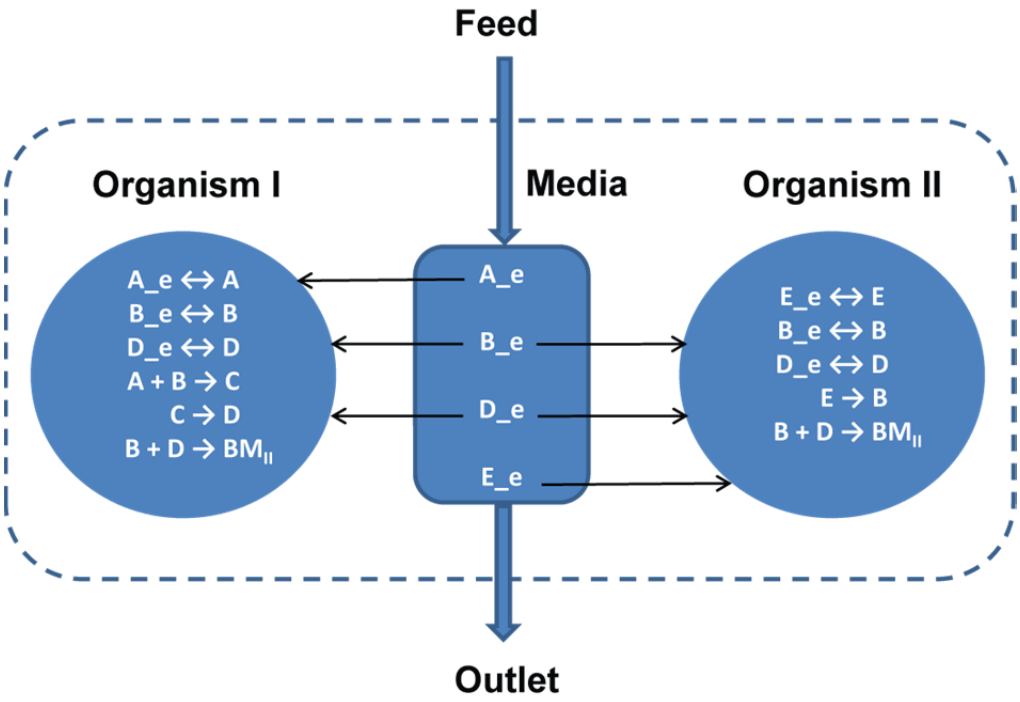

Figure 5.3: Toy network illustrating the co-culture model of two organisms in chemostat
A, B, C, D, and E represent the metabolites in the metabolic network of each organism I and II. BMI and BMII represent the biomass of organism I, and II, respectively. A_e, B_e, D_e, and E_e are external metabolites, which can be transported in/out of the cells and of the reactor (dot lines) via Feed and Outlet streams. B and D can be exchanged between the two organisms, while A can only be consumed by organism I, and E can only be consumed by organism II.

We then used constraint-based modeling (Chapter 2) to model the co-culture as an optimization problem, in which the objective is to maximize the total biomass concentration of the community instead of total growth rates as described in OptCom [143] (Eq. 5.1). In this formulation, $X_I \; and \; X_{II}$ represent the biomass concentrations of organisms I and II, respectively while $f_I \; and \; f_{II}$ represent the exchange fluxes of external metabolites (e.g., A_e, B_e, D_e, E_e) that are exchanged between the media and organism I and II, respectively. Like FBA (Chapter 2), we also used mass balance and flux limit to constrain the solution space (Eq. 5.2 – 5.7). In the chemostat, the growth rates of each organism are equal to the dilution rate *D* (Eq. 5.8). The last equation is the reactor balance equation in which Feed, Outlet, and Media are the flow rates of external metabolites in the feed, outlet and in the media. The variables in this optimization are the biomass concentrations and fluxes through all reactions in each metabolic network.

$$max \text{ or } min \quad X_I + X_{II} \qquad \text{(Eq. 5.1)}$$

$$s.t. \quad S_I \cdot v_I = 0, \forall \text{ internal metabolites in species I} \qquad \text{(Eq. 5.2)}$$

$$S_I \cdot v_I = f_I, \forall \text{external metabolites in species I} \qquad \text{(Eq. 5.3)}$$

$$LB_I \leq v_I \leq UB_I \qquad \text{(Eq. 5.4)}$$

$$S_{II} \cdot v_{II} = 0, \forall \text{ internal metabolites in species II} \qquad \text{(Eq. 5.5)}$$

$$S_{II} \cdot v_{II} = f_{II}, \forall \text{external metabolites in species II} \qquad \text{(Eq. 5.6)}$$

$$LB_{II} \leq v_{II} \leq UB_{II} \qquad \text{(Eq. 5.7)}$$

$$\mu_I = \mu_{II} = D \qquad \text{(Eq. 5.8)}$$

$$Feed - Outlet + f_I.X_I + f_{II}.X_{II} + Media = 0, \forall \text{external metabolites} \qquad \text{(Eq. 5.9)}$$

We applied the above formulation to the genome-scale metabolic networks for *Synechococcus* 7002 (*i*Syp708), and *Shewanella* W3181 (*i*SW3181_794) that have been developed in Chapter 2. We solved the optimization problem for three chemostat scenarios with a dilution rate of $0.05h^{-1}$ and the flow rates of inorganic carbon or lactate were fixed to 10 $mmol \cdot L^{-1} \cdot h^{-1}$. Simulation results were shown in Table 5.1 below. The co-culture model was able to sustain growth in all three conditions. The total cell concentration was predicted to be increased with the supplement of organic carbon source (lactate).

Table 5.1: Predicted maximum total cell concentration of the co-culture in different chemostat experiments

Scenarios	Maximum total cell concentration (g AFDW/L)
Feed only inorganic C	5.14
Feed only lactate	15.4
Feed inorganic C and lactate	20.6

It should be noted that the above optimization is a non-linear programing (NLP) problem, and therefore the solution obtained is not guaranteed global. When solving this NLP optimization problem, it is often critical to have reasonable initial values for the variables. In this case, obtaining a set of proper initial values was challenging since there were too many unknowns in the models. We addressed this issue by first identifying a feasible solution to the problem and perturbed the variables from the current level of the feasible solution and resolving the problem. We found that in all cases, the total biomass concentration remained the same but only individual cell concentration changed. In addition, the biomass concentration of *Synechococcus* 7002 was predicted to be always higher than that of *Shewanella* W3181, which implied that to maintain the co-culture at a certain dilution rate, one would need to have more cyanobacteria cells in the co-culture.

Chapter 6

Conclusions and future directions

In this work, we have focused on the development and genome-scale metabolic model for cyanobacteria. We have demonstrated the applications of genome-scale metabolic models in three main contexts. Firstly, *Cyanothece* 51142 metabolic model was analyzed in the context of understanding the production and partitioning of energy and reductant in the complex electron transport systems of the cyanobacteria. Secondly, we used the genome-scale metabolic model of *Synechococcus* 7002 to address the feasibility of the biochemical production of various compounds under different conditions, which we believe to be useful in strain-designing cyanobacteria for production of high-value products. Lastly, we have started to obtain initial results from analyzing the co-culture model of the cyanobacterium *Synechococcus* 7002 and the 'metal reducing' bacterium *Shewanella* W3181 such that the co-culture can sustain growth under different nutrient conditions. The result of this work will contribute more knowledge to symbiotic relationship between microorganisms.

6.1 Future directions

6.1.1 Validate and further improve genome-scale models

Model building is an iterative process and therefore discrepancies between model predictions and experimental data can be used to refine and improve model predictions. The predictive power of the genome-scale metabolic models developed in this work need to be verified by comparing model predictions with experimental data. While the predictions of *Cyanothece* 51142 model have been qualitatively and quantitatively validated with growth

experiments (Chapter 3), we have not yet verified model predictions by *Synechococcus* 7002 and *Shewanella* W3181 due to the lack of experimental data available for these organisms. Looking forward, since the genetic modification tools have been developed for *Synechococcus* 7002, we can verify model prediction against gene essentiality data if available. In addition, advances in metabolic flux analysis have enabled the quantification of flux distribution in photoautotrophic bacteria, which would be helpful to verify our cyanobacteria model-predicted flux distribution under photoautotrophic conditions. The MFA data will not only allow us to verify the predicted flux distribution, but also to redo the calculations for the reference flux distributions that were used to predict mutant phenotypes in MOMA and RELATCH algorithms (Chapter 4). Additionally, since *Shewanella* W3181 can grow on various carbon sources we can test *Shewanella* W3181 model predictions on the ability to grow on different carbon source using carbon utilization experiments.

Similarly, the co-culture predictions also need to be verified with experimental data when they become available. In addition, alternative methods for solving non-linear programming problem efficiently and different objective functions in modeling the co-culture should be explored to analyze the co-culture.

6.1.2 Validate predicted gene-deletion strategies

In Chapter 4, we have used MOMA [68], RELATCH [116], and OptORF [118] algorithms to obtain the strain-designing predictions that are pertaining to metabolic engineering *Synechococcus* 7002 for producing chemicals. The strategies suggested by these algorithms are not always overlapped and therefore it would be interesting to verify these predictions in the laboratory. Since there are many equivalent predicted gene-deletions strategies, it is necessary to further manually examine each of these strategies to narrow down the list of candidates. As an

example, one may make use of additional data (if available) such as gene/protein expression data or thermodynamic data to evaluate the feasibility and enzymatic activity of the reactions and proteins associated with deleted genes to eliminate some of the candidates from the list of proposed strategies.

Most of the engineering strategies suggested in the literature for improving productivity of non-native chemicals involved overexpression of non-native pathways instead of knocking out native genes. One can apply OptORF to identify metabolic engineering strategies that involved gene overexpression if a regulatory network is available [118]. Knowledge of transcriptional factor and regulatory interactions of genes in cyanobacteria is still limited and thus there is a need to develop regulatory networks from available expression data. In addition, it would be interesting to identify a set of reactions that can be added to the model that would improve chemical production.

6.2 Concluding remarks

Looking back over the years I spent learning how to develop metabolic models for cyanobacteria, I am convinced that models are useful as they provided us a systematic way to represent living organisms, and to see how materials are connected. As complicated as living organisms are, it is amazing how simple models such as the *Cyanothece* 51142 model that include only ~ 30% of the genome can make accurate growth predictions. Analyzing the metabolic network of photosynthetic bacteria (cyanobacteria) was more challenging than that of other microorganisms such as *E. coli*, or yeast for the cyanobacteria metabolic network is dependent on not only the carbon sources but also the energy from light. I also learned that models have limitations but these limitations open doors for many research areas in the fields of

systems biology and synthetic biology that together really make a significant improvement on quality of life.

Appendix 1: Biomass composition of *Cyanothece* 51142 measured in ammonia- and light-limited chemostat

			Ammonia-limited	Light-limited	Average
	Components	**Molecular weight (g/mol)**	**Content (mmol/gAFDW)**	**Content (mmol/gAFDW)**	**Content (mmol/gAFDW)**
Biomass composition (*)	Protein	1124316.142	1.868E-04	3.531E-04	2.699E-04
	Carbohydrates	162.14	3.096E+00	1.018E+00	2.057E+00
	Lipids	38616.4121	3.677E-03	3.496E-03	3.587E-03
	RNA	31953.742	3.036E-03	4.757E-03	3.896E-03
	DNA	30788.548	1.072E-03	1.104E-03	1.088E-03
	CGP	562.604	3.555E-03	1.724E-01	8.798E-02
	Chl	892.5353	1.569E-02	2.241E-02	1.905E-02
Soluble pool (**)	10fthf (#)	471.438	9.222E-04	1.229E-03	1.076E-03
	mlthf (#)	455.438	9.222E-04	1.229E-03	1.076E-03
	accoa	805.548	3.214E-03	4.257E-03	3.736E-03
	amp	345.216	1.475E-03	1.954E-03	1.715E-03
	coa	763.512	7.377E-03	9.769E-03	8.573E-03
	nad	662.428	1.370E-02	1.814E-02	1.592E-02
	nadh	663.436	4.374E-04	5.792E-04	5.083E-04
	nadp	740.39	1.107E-05	1.465E-05	1.286E-05
	nadph	741.398	6.324E-04	8.374E-04	7.349E-04
	succcoa	862.576	1.212E-03	1.605E-03	1.408E-03
	FAD	783.548	8.958E-04	1.186E-03	1.041E-03
	udpg	564.286	1.317E-02	1.744E-02	1.531E-02

(*) Biomass concentration was measured in this study

(**) The soluble pool concentration was obtained from Bennett and Rabinowitz NatChemBiol (2009), 5(8):593-599 (ref. 26 in the article has L/gDW conversion = 0.0023)

(#) The concentration of 10fthf and mlthf was obtained from Harvey and Dev. Advances in Enzyme Regulation (1975), 13: 97-124. (Total folates concentration = 2 to 5x10-4 M, average of 3.5x10^-4M)

Appendix 2: Genes added to the model *i*Syp708

Locus tag	Gene product
SYNPCC7002_A0028	D-glycerate 3-kinase, plant type, GlyK
SYNPCC7002_A0108	5-methylthioadenosine phosphorylase, MtaP
SYNPCC7002_A0162	6-phosphofructokinase, PfkA
SYNPCC7002_A0210	light-dependent protochlorophyllide oxidoreductase, Por
SYNPCC7002_A0271	L-arginine deiminase, ArcA
SYNPCC7002_A0308	alternative dihydrofolate reductase, FolM
SYNPCC7002_A0395	high affinity urea ABC transport system, ATPase protein, UrtD
SYNPCC7002_A0396	high affinity urea ABC transporter, permease component, UrtC
SYNPCC7002_A0397	high affinity urea ABC transporter, permease component, UrtB
SYNPCC7002_A0398	high affinity urea ABC transporter, substrate binding component, UrtA
SYNPCC7002_A0430	S-adenosylmethionine decarboxylase, SpeD
SYNPCC7002_A0481	glycogen phosphorylase, GlgP
SYNPCC7002_A0492	5' nucleotidase, SurE family
SYNPCC7002_A0521	integral membrane phospholipid phosphatase
SYNPCC7002_A0526	putative glycerol-3-phosphate acyltransferase, PlsX
SYNPCC7002_A0553	methionine salvage pathway protein E-2/E-2", MtnD
SYNPCC7002_A0554	ribulose-5-phosphate 4-epimerase
SYNPCC7002_A0560	NADH dehydrogenase, subunit L inorganic carbon tranpsort protein, NdhL
SYNPCC7002_A0569	NADH dehydrogenase, subunit M, NdhM
SYNPCC7002_A0603	homoserine kinase type II
SYNPCC7002_A0636	3',5'-cyclic-nucleotide phosphodiesterase, CpdA
SYNPCC7002_A0642	nicotinamidase, PncA
SYNPCC7002_A0643	cyanate hydratase, CynS
SYNPCC7002_A0670	2-methyl-6-phytyl-1,4-benzoquinone cyclase
SYNPCC7002_A0706	inositol monophosphatase family protein
SYNPCC7002_A0736	F0 ATP synthase complex, B subunit, AtpF
SYNPCC7002_A0737	F0 ATP synthase complex, B' subunit, AtpFb
SYNPCC7002_A0738	F0 ATP synthase complex, subunit C, AtpE
SYNPCC7002_A0750	F1 ATP synthase complex, epsilon subunit, AtpC
SYNPCC7002_A0792	thioredoxin reductase, TrxB
SYNPCC7002_A0849	cytosine-specific methyltransferase, Dcm
SYNPCC7002_A0863	cyanate hydratase, CynS2
SYNPCC7002_A0869	S-formylglutathione hydrolase

Locus tag	Gene product
SYNPCC7002_A0874	UDP-N-acetylmuramyl tripeptide synthase
SYNPCC7002_A0887	sucrose-phosphate phosphatase, SppA
SYNPCC7002_A0909	adenylate cyclase
SYNPCC7002_A0918	1-acyl-sn-glycerol-3-phosphate acyltransferase, PlsC
SYNPCC7002_A0985	nicotinamide nucleotide transhydrogenase, alpha chain part 2, PntC
SYNPCC7002_A1074	bifunctional proline dehydrogenase / 1-pyrroline-5-carboxylate dehydrogenase, PutA
SYNPCC7002_A1143	NADH dehydrogenase, subunit N, NdhN
SYNPCC7002_A1154	lysine decarboxylase
SYNPCC7002_A1188	cytosine-specific methyltransferase, beta subunit, AquIB
SYNPCC7002_A1206	orotate phosphoribosyltransferase, PyrE
SYNPCC7002_A1256	DNA polymerase III, delta subunit, HolB
SYNPCC7002_A1263	DNA polymerase III, beta subunit, DnaN
SYNPCC7002_A1315	mercuric reductase, MerA
SYNPCC7002_A1381	FeIII ABC transport system, permease component, FutB
SYNPCC7002_A1408	DNA polymerase III, alpha subunit, DnaE-C
SYNPCC7002_A1417	7,8-didemethyl-8-hydroxy-5-deazariboflavin synthase, subunit 1, CofG
SYNPCC7002_A1437	5'-nucleotidase, SurE family
SYNPCC7002_A1460	transaldolase, Tal
SYNPCC7002_A1567	DNA polymerase III, delta subunit, HolA
SYNPCC7002_A1653	methionyl-tRNA formyltransferase, Fmt
SYNPCC7002_A1733	manganese ABC transport system, permease component, MntB
SYNPCC7002_A1735	manganese ABC transport system, substrate binding component, MntC
SYNPCC7002_A1791	gamma-tocopherol C-methyltransferase
SYNPCC7002_A1895	phosphate uptake ABC transporter, ATPase component, PstB
SYNPCC7002_A1906	adenylate cyclase, CyaD
SYNPCC7002_A1920	7,8-didemethyl-8-hydroxy-5-deazariboflavin synthase, subunit 2, CofH
SYNPCC7002_A1991	Heme oxygenase 2, Hox
SYNPCC7002_A2015	acetyl-CoA synthetase, ADP forming alpha subunit, Acs
SYNPCC7002_A2022	sucrose phosphorylase, ScrP
SYNPCC7002_A2188	aspartate aminotransferase
SYNPCC7002_A2228	phycocyanobilin:ferredoxin oxidoreductase, PcyA
SYNPCC7002_A2284	phosphate ABC transporter, substrate binding component, PstS
SYNPCC7002_A2285	phosphate ABC transporter, permease component, PstC

Appendix 2, continued

Locus tag	Gene product
SYNPCC7002_A2286	phosphate ABC transporter, permease component, PstA
SYNPCC7002_A2302	pantothenate kinase type III, CoaX
SYNPCC7002_A2308	methylthioribose-1-phosphate isomerase, MtnA
SYNPCC7002_A2318	copper transporting CPx-type ATPase, PacS
SYNPCC7002_A2324	biliverdin reductase, BvdR
SYNPCC7002_A2384	DNA polymerase III, alpha subunit, DnaE
SYNPCC7002_A2422	catalase/peroxidase HPI, KatG
SYNPCC7002_A2438	glucokinase, Glck
SYNPCC7002_A2496	zinc-dependent gamma carbonic anhydrase, CagA
SYNPCC7002_A2497	secondary thiamine-phosphate synthase, YbjQ
SYNPCC7002_A2499	zinc ABC transporter, inner membrane permease protein, ZntB
SYNPCC7002_A2500	zinc ABC transport system, ATPase component, ZntC
SYNPCC7002_A2501	Zinc ABC transport system, substrate binding component, ZntA
SYNPCC7002_A2507	FeIII ABC transporter, substrate binding component, FutA
SYNPCC7002_A2508	heme oxygenase decyclizing, Hox
SYNPCC7002_A2610	copper-transporting P-type ATPase, CtaA
SYNPCC7002_A2710	4-hydroxyphenylpyruvate dioxygenase
SYNPCC7002_A2739	2-methyl-6-phytyl-1,4-benzoquinone methyltransferase / 2-methyl-6-solanyl-1,4-benzoquinone methyltransferase
SYNPCC7002_A2750	NADPH-quinone oxidoreductase chain C, NdhJ
SYNPCC7002_A2841	glucosylglycerol 3-phosphatase, GgpP
SYNPCC7002_A2863	diguanylate cyclase
SYNPCC7002_E0012	cytosine-specific methyltransferase
SYNPCC7002_F0006	bifunctional nicotinamide-nucleotide adenylyltransferase / ribosylnicotinamide kinase, NadR
SYNPCC7002_G0014	ABC transport system, ATP-binding component, nitrate/nitrite/cyanate uptake family
SYNPCC7002_G0015	ABC transport system, substrate binding component, nitrate/nitrite/cyanate uptake family
SYNPCC7002_G0016	ABC transport system, permease component, nitrate/nitrite/cyanate uptake family
SYNPCC7002_G0024	L-2,4-diaminobutyrate decarboxylase
SYNPCC7002_G0054	high affinity ATP-driven K+ transport system, subunit D, KdpD
SYNPCC7002_G0057	high affinity ATP-driven K+ transport system, G subunit, KdpG
SYNPCC7002_G0113	FeIII ABC transport system, substrate binding component, FbpA
SYNPCC7002_G0114	FeIII ABC transport system, permease component, FbpB
SYNPCC7002_G0146	N-type Na+ translocating ATPase complex locus membrane protein, AtpQ
SYNPCC7002_G0147	N-type Na+ translocating ATPase complex locus membrane protein, AtpR

Appendix 3: Reaction and GPR changes in *i*Syp708 (updated from *i*Syp611)

Abrreviation	Name	Equation	Gene	EC
CAT	Catalase	[c]: (2) h2o2 --> (2) h2o + o2	A2422	1.11.1.7 1.11.1.6
DAGPYP_SYN	diacylglycerol pyrophosphate phosphatase (*Synechococcus*)	[c]: h2o + (0.02) pa_SYN --> (0.02) 12dgr_SYN + pi	A0521	3.1.3.4
G3POA_SYN	glycerol-3-phosphate O-acyltransferase (*Synechococcus*)	[c]: (0.07) 9hdeACP + glyc3p + (0.01) linlnACP + (0.01) ocdACP + (0.18) ocdcyaACP + (0.17) octeACP + (0.56) palmACP --> (0.02) 1ag3p_SYN + ACP	A0526	2.3.1.15
ASPTA1	aspartate transaminase	[c]: akg + asp-L <==> glu-L + oaa	A0512 or A2188	2.6.1.1
ATPS4r	ATP synthase (four protons for 1 ATP)	adp[c] + (4) h[e] + pi[c] <==> atp[c] + (3) h[c] + h2o[c]	((A0739 and A0738 and A0736) and (A0734 and A0749 and A0733 and A0735 and A0750)) or ((A0739 and A0738 and A0737) and (A0734 and A0749 and A0733 and A0735 and A0750)) or ((G0148 and G0150 and G0149) and (G0151 and G0144 and G0152 and G0145 and A0735))	3.6.3.14
DHFR	dihydrofolate reductase	[c]: dhf + h + nadph <==> nadp + thf	A0498 or A0308	1.5.1.3
DNASYN_SYN	DNA replication, *Synechococcus* PCC 7002	[c]: (100) atp + (26) datp + (24) dctp + (24) dgtp + (26) dttp + (100) h2o --> (100) adp + dna_SYN + (100) h + (100) pi + (100) ppi	(A1280) or (A2384 and A1263 and A1567), or (A2384 and A1263 and A1256) or (A1408 and A1263 and A1567) or (A1408 and A1263 and A1256)	2.7.7.7
GLCP	glycogen phosphorylase	[c]: glycogen + pi --> g1p	A2139 or A0481	2.4.1.1 2.4.1.1
HCO3E	carbonate dehydratase (HCO3 equilibration reaction)	[c]: co2 + h2o <==> h + hco3	A1997 or A1096 or A2496	4.2.1.1
HGR	mercuric reductase	[c]: hg2 + nadph <==> h + hg0 + nadp	A1315 or A1295	1.16.1.1
HSK	homoserine kinase	[c]: atp + hom-L --> adp + h + phom	A0091 or A0603	2.7.1.39
Kabc	potassium transport via ABC system	atp[c] + h2o[c] + k[e] --> adp[c] + h[c] + k[c] + pi[c]	G0055 and G0060 and G0059 and G0057 and G0054	3.6.3.12
LYSDC	lysine decarboxylase	[c]: h + lys-L --> 15dap + co2	A1159 or A1154	4.1.1.18

Appendix 3, continued

Abrreviation	Name	Equation	Gene	EC
MNabc	manganese transport via ABC system	atp[c] + h2o[c] + mn2[e] --> adp[c] + h[c] + mn2[c] + pi[c]	A1734 and A1733 and A1735	3.6.3.35
NTD1	5'-nucleotidase (dUMP)	[c]: dump + h2o --> duri + pi	A1437 or A0492	3.1.3.5
NTD10	5'-nucleotidase (XMP)	[c]: h2o + xmp --> pi + xtsn	A1437 or A0492	3.1.3.5
NTD2	5'-nucleotidase (UMP)	[c]: h2o + ump --> pi + uri	A1437 or A0492	3.1.3.5
NTD3	5'-nucleotidase (dCMP)	[c]: dcmp + h2o --> dcyt + pi	A1437 or A0492	3.1.3.5
NTD4	5'-nucleotidase (CMP)	[c]: cmp + h2o --> cytd + pi	A1437 or A0492	3.1.3.5
NTD5	5'-nucleotidase (dTMP)	[c]: dtmp + h2o --> pi + thymd	A1437 or A0492	3.1.3.5
NTD6	5'-nucleotidase (dAMP)	[c]: damp + h2o --> dad-2 + pi	A1437 or A0492	3.1.3.5
NTD7	5'-nucleotidase (AMP)	[c]: amp + h2o --> adn + pi	A1437 or A0492	3.1.3.5
NTD8	5'-nucleotidase (dGMP)	[c]: dgmp + h2o --> dgsn + pi	A1437 or A0492	3.1.3.5
NTD9	5'-nucleotidase (GMP)	[c]: gmp + h2o --> gsn + pi	A1437 or A0492	3.1.3.5
ORPT	orotate phosphoribosyltransferase	[c]: orot5p + ppi <==> orot + prpp	A0610 or A1206	2.4.2.10
PHETA1	phenylalanine transaminase	[c]: akg + phe-L <==> glu-L + phpyr	A0512 or A2188	2.6.1.1
PIabc	phosphate transport via ABC system	atp[c] + h2o[c] + pi[e] --> adp[c] + h[c] + (2) pi[c]	(A1364) or (A2286 and A2285 and A2284 and A1895)	3.6.3.27
PRO1z	proline oxidase	[c]: fad + pro-L --> 1pyr5c + fadh2 + h	A1071 or A1074	1.5.1.12 1.5.99.8
TAL	Transaldolase	[c]: g3p + s7p <==> e4p + f6p	A2558 or A1460	2.2.1.2
THD5	NAD transhydrogenase	[c]: nad + nadph --> nadh + nadp	A0984 and A0986 and A0985	1.6.1.2
TMPPP	thiamine-phosphate diphosphorylase	[c]: 2mahmp + 4mpetz + h --> ppi + thmmp	A0279 or A2497	2.5.1.3
TYRTA	tyrosine transaminase	[c]: akg + tyr-L <==> 34hpp + glu-L	A0512 or A2188	2.6.1.1
UAAGDS	UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimelate synthetase	[c]: 26dap-M + atp + uamag --> adp + h + pi + ugmd	A0874 or A0121	6.3.2.13
UREAtabc	Urea transport via ATP	atp[c] + h2o[c] + urea[e] <==> adp[c] + h[c] + pi[c] + urea[c]	A0395 and A0396 and A0397 and A0398 and A1627	3.6.3.-
ACHBS	2-aceto-2-hydroxybutanoate synthase	[c]: 2obut + h + pyr --> 2ahbut + co2	A2704 and A2797	
ACLS	acetolactate synthase (Also catalyzes ACHBS)	[c]: h + (2) pyr --> alac-S + co2	A2704 and A2797	4.1.3.18

Abrreviation	Name	Equation	Gene	EC
NADHPQ9	NADH dehydrogenase (plastoquinone-9 & 4 protons)	(5) h[c] + nadh[c] + pq9[c] --> (4) h[e] + nad[c] + pq9h2[c]	((A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A0172 and A0173 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A0854 and A0173 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A1805 and A0173 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A0172 and A1806 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A0854 and A1806 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A1805 and A1806 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A0172 and A1973 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A0854 and A1973 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A1805 and A1973 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A0172 and A2000 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or(A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A0854 and A2000 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A1805 and A2000 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A0172 and A2327 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A0854 and A2327 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197) or (A2748 and A2749 and A2750 and A2541 and A0926 and A0925 and A0924 and A0923 and A1805 and A2327 and A2547 and A0560 and A0569 and A1143 and A0195 and A0196 and A0197))	1.6.5.3

Appendix 3, continued

Abrreviation	Name	Equation	Gene	EC
2OGDC (*)	2-oxoglutarate decarboxylase	[c]: akg + h --> sucsal + co2	A2770	
ACS3 (*)	acetyl-CoA synthetase, ADP forming	[c]: ac + atp + coa <==> accoa + adp + pi	A2015	6.2.1.13
ADNCYC (*)	Adenylate cyclase	[c]: atp --> camp + ppi	A0909 or A1906	4.6.1.1
AMETDC (*)	adenosylmethionine decarboxylase	[c]: amet + h <==> ametam + co2	A0430	4.1.1.50
ARD (*)	acireductone dioxygenase (Require Ni2+)	[c]: 12dh5mp + o2 <==> 3mtpa + for + co + (2) h	A0553	1.13.11.54
ARD1 (*)	acireductone dioxygenase (Require Fe2+)	[c]: 12dh5mp + o2 <==> 4mtoba + for + (2) h	A0553	1.13.11.54
ARGD (*)	arginine deaminase	[c]: arg-L + h2o --> citr-L + nh4	A0271	3.5.3.6
BLVDR (*)	biliverdin reductase	[c]: biliverdin + nadph + h --> bilirubin + nadp	SYNPCC7002_A2324	1.3.1.24
Cut1 (*)	Copper export via ATPase	atp[c] + cu2[c] + h2o[c] --> adp[c] + cu2[e] + h[c] + pi[c]	A2318 or A2610	3.6.3.-
CYNS (*)	Cyanate hydratase	[c]: cyn + hco3 + h <==> car + co2	SYNPCC_A0643 or SYNPCC_A0863	4.2.1.104
CYTOM (*)	cytosine 5-methyltransferase	[c]: amet + csn <==> 5mcsn + ahcys + h	E0012 or A1188 or A0849	2.1.1.37
DABA (*)	diaminobutyrate decarboxylase	[c]: l24dabut + h <==> 13dap + co2	G0024	4.1.1.86
DGC (*)	Diguanylate cyclase	[c]: (2) gtp --> cdigmp + ppi	A2863	2.7.7.65
DTCPHMT (*)	delta-tocopherol C-methyltransferase	[c]: amet + dtcph --> ahcys + h + btcph	A1791	2.1.1.95
FE3abc (*)	iron (III) transport via ABC system	atp[c] + fe3[e] + h2o[c] --> adp[c] + fe3[c] + h[c] + pi[c]	(G0114 and G0113) or (A1381 and A2507)	3.6.3.-
FMETTRS (*)	Methionyl-tRNA formyltransferase	[c]: mettrna + 10fthf <==> thf + fmettrna	A1653	2.1.2.9
FOS (*)	4-hydroxyphenylpyruvate:5-amino-6-(D-ribitylamino)uracil 4-hydroxybenzyltransferase	[c]: 4ohphpyr + 4r5au +(2) amet + h2o --> 78d8h5drib + (2) met-L + (2) dad-5 + oxa + nh4 + h	A1417 and A1920	2.5.1.77
GLYRK (*)	glycerate kinase	[c]: atp +glyc-R <==> 3pg + adp + h	A0028	2.7.1.31
GSG3P (*)	glucosylglycerol 3-phosphatase	[c]: bglug3p + h2o <==> bgluglyc + pi	A2841	3.1.3.69
GTCPHMT (*)	gamma-tocopherol C-methyltransferase	[c]: amet + gtcph --> ahcys + h + atcph	A1791	2.1.1.95
GUNCYC (*)	Guanylate cyclase	[c]: gtp --> cgmp + ppi	A0909 or A1906	4.6.1.1
HEX1 (*)	hexokinase (D-glucose:ATP)	[c]: atp + glc-D --> adp + g6p + h	A2438	2.7.1.2

(*) New reactions

Appendix 3, continued

Abrreviation	Name	Equation	Gene	EC
HMOX (*)	heme oxygenase	[c]: pheme + (3) o2 + (3) nadh + (4) h --> biliverdin + co + fe2 + (3) nad + (3) h2o	A1991 or A2508	1.14.99.3
HPPDO1 (*)	4-hydroxyphenylpyruvate dioxygenase	[c]: 34hpp + o2 --> co2 + hgentis	A2710	1.13.11.27
HPT (*)	tocopherol phytyltransferase	[c]: h + hgentis + phdp <==> co2 + mtpq + ppi	A0703	
MDRPI (*)	methylthioribose 1-phosphate isomerase	[c]: 5mdr1p <==> 5mdru1p	A2308	5.3.1.23
MPTQMT (*)	Methyl phytyl quinone methyl transferase	[c]: amet + mtpq --> ahcys + h + dmptq	A2739	2.1.1.
MSBQMT (*)	Methyl solanyl benzoquinone methyl transferase	[c]: amet + msbq --> ahcys + h + pq9h2	A2739	2.1.1.
MTAP (*)	5'-methylthioadenosine phosphorylase	[c]: 5mta + pi --> 5mdr1p + ade	A0108	2.4.2.28
MTRUPDHR (*)	methylthioribulose 1-phosphate dehydratase	[c]: 5mdru1p <==> 5m23dp + h2o	A0554	4.2.1.109
Naabc (*)	Na+-transporting two-sector ATPase	atp[c] + na1[e] + h2o[c] --> adp[c] + na1[c] + h[c] + pi[c]	G0146 and G0147	3.6.3.15
NCTNA (*)	Nicotinamidase	[c]: ncam + h2o --> nac + nh4	A0642	3.5.1.19
NO3abc (*)	Nitrate transport via ABC system	atp[c] + h2o[c] + no3[e] --> adp[c] + h[c] + pi[c] + no3[c]	G0014 and G0015 and G0016	3.6.3.26
NRK (*)	ribosylnicotinamide kinase	[c]: atp + rnam <==> adp + nmn + h	F0006	2.7.1.22 2.7.7.1
PCBOR (*)	phycocyanobilin:ferredoxin oxidoreductase	[c]: biliverdin + (4) fdxr-2-2 + (4) h --> phycocyanobilin + (4) fdxo-2-2	A2228	1.3.7.5
PDE1 (*)	3',5'-cyclic-nucleotide phosphodiesterase	[c]: camp + h2o --> amp + h	A0636	3.1.4.17
PDE2 (*)	3',5'-cyclic-nucleotide phosphodiesterase	[c]: cgmp + h2o --> gmp + h	A0636	3.1.4.17
PFK (*)	Phosphofructokinase	[c]: atp + f6p --> adp + fdp + h	A0162	2.7.1.11
SFGH (*)	S-formylglutathione hydrolase	[c]: sfgthrd + h2o <==> gthrd + for	A0869	3.1.2.12
SHGD (*)	S-hydroxymethylglutahione dehydrogenase	[c]: shngthrd + nad <==> sfgthrd + nadh + h	A0868	1.1.1.284
SUCPPT (*)	sucrose-phosphate phosphatase	[c]: suc6p + h2o --> suc + pi	A0887	3.1.3.24
SUCPR (*)	sucrose phosphorylase	[c]: suc + pi <==> fru + g1p	A2022	2.4.1.7
VTE1a (*)	tocopherol cyclase	[c]: dmptq --> gtcph	A0670	
VTE1b (*)	tocopherol cyclase	[c]: mptq --> dtcph	A0670	
ZN2abc2 (*)	zinc transport in via ABC system	atp[c] + h2o[c] + zn2[e] --> adp[c] + h[c] + pi[c] + zn2[c]	A2500 and A2501 and A2499	3.6.3.-
Cot (*)	CO transport out via diffusion	co[e] <==> co_e		

(*) New reactions

Abrreviation	Name	Equation	Gene	EC
ALA_Lt6 (*)	L-alanine transport in/out via proton symport	ala-L[e] + h[e] <==> ala-L[c] + h[c]		
D-Lact2 (*)	D-lactate transport	lac-D[e] <==> lac-D[c]		
Fe2t (*)	iron (II) transport	fe2[e] <==> fe2[c]		
SUCCt2b (*)	Succinate efflux via proton symport	h[c] + succ[c] --> h[e] + succ[e]		
SUC_PQ (*)	Succinate dehydrogenase (plastoquinone)	[c]: fadh2 + pq9 --> fad + pq9h2		
EX_fe2_e (*)	iron (II) exchange	fe2[e] ==>		
EX_lac-D_e (*)	D-lactate exchange	lac-D[e] -->		
EX_succ_e (*)	Succinate exchange	succ[e] ==>		
EX_ala-L_e (*)	L-alanine exchange	ala-L[e] ==>		
EX_cu2_e (*)	copper exchange	cu2[e] ==>		
EX_zn2_e (*)	zinc exchange	zn2[e] ==>		

(*) New reactions

Appendix 4: Biomass composition of *Synechococcus* 7002 grown under nitrogen-, carbon-, and light-limited conditions

Composition	Raw data [a)]			Scaled data [b)]		
	N-limited	C-limited	Light-limited	N-limited	C-limited	Light-limited
Total Protein	0.329	0.58	0.673	0.327	0.709	0.742
Carbohydrates	0.606	0.097	0.118	0.602	0.119	0.130
RNA	0.032	0.040	0.048	0.032	0.049	0.053
DNA	0.003	0.003	0.003	0.003	0.004	0.003
Phycocyanobilins	0.001	0.009	0.010	0.001	0.011	0.011
Chlorophyll	0.003	0.009	0.011	0.003	0.011	0.012
Lipids	0.035	0.079	0.045	0.034	0.097	0.050

a) Biomass composition measured in g/ g AGDW.

b) Biomass composition (in g/ g AFDW) after rescaling so that the total composition of biomass components equals 1.

Appendix 5: Reactions and GPR in *i*W3181_794 that are modified from *i*SO783

Abbreviations	Equations	Protein-Gene Associations	Notes
GAPD	[c] : g3p + nad + pi <==> 13dpg + h + nadh	(GapA2 (W3181_2119)) or (GapA3 (W3181_2117)) or (W3181_2923 (W3181_2923))	Change GPR
ARSRD2	[c] : arsna + (2) gthrd + (2) h --> arsni2 + gthox + h2o	(ArsC3 (W3181_2331)) or (ArsC1 (W3181_1657)) or (W3181_2950 (W3181_2950))	Change GPR
CAT	[c] : (2) h2o2 --> (2) h2o + o2	(KatG2 (W3181_0269)) or (W3181_3582 (W3181_3582))	Change GPR
TDPDRR	[c] : dtdp6dm + nadp <==> dtdpddm + h + nadph	W3181_1465 (W3181_1465)	Change GPR
CYTBD	(2) h[c] + (0.5) o2[c] + ubq8h2[c] --> (2) h[e] + h2o[c] + ubq8[c]	(CydA (W3181_1367 and W3181_1366)) or (CydA2 (W3181_3246 and W3181_3247))	Change GPR
HEMEOS	[c] : frdp + h2o + pheme --> h + hemeO + ppi	(CtaB (W3181_0126)) or (CyoE (W3181_0091))	Change GPR
GPDDA2	[c] : g3pe + h2o --> etha + glyc3p + h	(UgpQ (W3181_0649)) or (W3181_4006 (W3181_4006))	Change GPR
GPDDA4	[c] : g3pg + h2o --> glyc + glyc3p + h	(UgpQ (W3181_0649)) or (W3181_4006 (W3181_4006))	Change GPR
GTHS	[c] : atp + glucys + gly --> adp + gthrd + h + pi	(GshB (W3181_0781)) or (W3181_2944 (W3181_2944))	Change GPR
GLU5K	[c] : atp + glu-L --> adp + glu5p	(ProB (W3181_0971)) or (W3181_3995 (W3181_3995))	Change GPR
G5SD	[c] : glu5p + h + nadph --> glu5sa + nadp + pi	(ProA (W3181_0972)) or (W3181_3996 (W3181_3996))	Change GPR
GMPS2	[c] : atp + gln-L + h2o + xmp --> amp + glu-L + gmp + (2) h + ppi	(GuaA (W3181_1362)) or (W3181_0925 (W3181_0925))	Change GPR
TDPDRE	[c] : dtdpddg <==> dtdpddm	(RfbC (W3181_1476)) or (RfbC2 (W3181_1466))	Change GPR
PFL	[c] : coa + pyr <==> accoa + for	(PflA (W3181_1631 and W3181_1630)) or (PflD (W3181_0426 and W3181_0427))	Change GPR
FUM	[c] : fum + h2o <==> mal-L	(FumB (W3181_2218)) or (W3181_3574 (W3181_3574)) or (W3181_0198 (W3181_0198))	Change GPR
METGL	[c] : h2o + met-L --> 2obut + ch4s + nh4	(MdeA (W3181_2594)) or (W3181_3444 (W3181_3444))	Change GPR

Appendix 5, continued

Abbreviations	Equations	Protein-Gene Associations	Notes
Cut1	atp[c] + cu2[c] + h2o[c] --> adp[c] + cu2[e] + h[c] + pi[c]	(CopA (W3181_2693)) or (W3181_1110 (W3181_1110))	Change GPR
ICL	[c] : icit --> glx + succ	(AceA (W3181_2865)) or (W3181_3788 (W3181_3788))	Change GPR
NADH14	[c] : h + mmqn7 + nadh --> mmql7 + nad	(Ndh (W3181_3093)) or (Ndh2 (W3181_2862))	Change GPR
NADH12	[c] : h + nadh + ubq8 --> nad + ubq8h2	(Ndh (W3181_3093)) or (Ndh2 (W3181_2862))	Change GPR
NADH4	[c] : h + mqn7 + nadh --> mql7 + nad	(Ndh (W3181_3093)) or (Ndh2 (W3181_2862))	Change GPR
TTTNTR3	mql7[c] + tttnt[e] --> (2) h[e] + mqn7[c] + (2) tsul[e]	TtrABC (W3181_3510 and W3181_3511 and W3181_3512)	Change GPR
TTTNTR4	mmql7[c] + tttnt[e] --> (2) h[e] + mmqn7[c] + (2) tsul[e]	TtrABC (W3181_3510 and W3181_3511 and W3181_3512)	Change GPR
GLUR	[c] : glu-D <==> glu-L	(MurI (W3181_3676)) or (W3181_0949 (W3181_0949))	Change GPR
NTR4	mql7[c] + no3[e] --> h2o[e] + mqn7[c] + no2[e]	(CymA (W3181_3916) and NapABGH (W3181_0792 and W3181_0793 and W3181_0794 and W3181_0795)) or (CymA (W3181_3916) and NapA2B2DH (W3181_2105 and W3181_2104 and W3181_2107 and W3181_0792))	Change GPR
NTR5	mmql7[c] + no3[e] --> h2o[e] + mmqn7[c] + no2[e]	(CymA (W3181_3916) and NapABGH (W3181_0792 and W3181_0793 and W3181_0794 and W3181_0795)) or (CymA (W3181_3916) and NapA2B2DH (W3181_2105 and W3181_2104 and W3181_2107 and W3181_0792))	Change GPR
NTRIRmmq2	(2) h[e] + (3) mmql7[c] + no2[e] --> (2) h2o[e] + (3) mmqn7[c] + nh4[e]	(CymA (W3181_3916) and NrfA (W3181_3486)) or (CymA (W3181_3916) and NrfA2 (W3181_3743))	Change GPR
NTRIRmq2	(2) h[e] + (3) mql7[c] + no2[e] --> (2) h2o[e] + (3) mqn7[c] + nh4[e]	(CymA (W3181_3916) and NrfA (W3181_3486)) or (CymA (W3181_3916) and NrfA2 (W3181_3743))	Change GPR
FDH10	for[e] + (2) h[c] + mmqn7[c] --> co2[e] + h[e] + mmql7[c]	(Fdh1 (W3181_3871 and W3181_3872 and W3181_3873)) or (Fdh2 (W3181_3875 and W3181_3876 and W3181_3877))	Change GPR
FDH9	for[e] + (2) h[c] + mqn7[c] --> co2[e] + h[e] + mql7[c]	(Fdh1 (W3181_3871 and W3181_3872 and W3181_3873)) or (Fdh2 (W3181_3875 and W3181_3876 and W3181_3877))	Change GPR
Kt2i	h[e] + k[e] --> h[c] + k[c]	(TrkA1 (W3181_4032 and W3181_4031)) or (TrkH (W3181_0016)) or (TrkH2 (W3181_0020))	Change GPR

Abbreviations	Equations	Protein-Gene Associations	Notes
ADNt2	adn[e] + h[e] --> adn[c] + h[c]	(NupC (W3181_1187)) or (NupC2 (W3181_0452))	Change GPR
CYTDt2	cytd[e] + h[e] --> cytd[c] + h[c]	(NupC (W3181_1187)) or (NupC2 (W3181_0452))	Change GPR
DADNt2	dad-2[e] + h[e] --> dad-2[c] + h[c]	(NupC (W3181_1187)) or (NupC2 (W3181_0452))	Change GPR
DCYTt2	dcyt[e] + h[e] --> dcyt[c] + h[c]	(NupC (W3181_1187)) or (NupC2 (W3181_0452))	Change GPR
DURIt2	duri[e] + h[e] --> duri[c] + h[c]	(NupC (W3181_1187)) or (NupC2 (W3181_0452))	Change GPR
THMDt2	h[e] + thymd[e] --> h[c] + thymd[c]	(NupC (W3181_1187)) or (NupC2 (W3181_0452))	Change GPR
URIt2	h[e] + uri[e] --> h[c] + uri[c]	(NupC (W3181_1187)) or (NupC2 (W3181_0452))	Change GPR
PUNP1	[c] : adn + pi <==> ade + r1p	(DeoD (W3181_1192)) or (DeoD3 (W3181_2368))	Change GPR
PUNP2	[c] : dad-2 + pi <==> 2dr1p + ade	(DeoD (W3181_1192)) or (DeoD3 (W3181_2368))	Change GPR
PUNP3	[c] : gsn + pi <==> gua + r1p	(DeoD (W3181_1192)) or (DeoD3 (W3181_2368))	Change GPR
PUNP4	[c] : dgsn + pi <==> 2dr1p + gua	(DeoD (W3181_1192)) or (DeoD3 (W3181_2368))	Change GPR
PUNP5	[c] : ins + pi <==> hxan + r1p	(DeoD (W3181_1192)) or (DeoD3 (W3181_2368))	Change GPR
PUNP6	[c] : din + pi <==> 2dr1p + hxan	(DeoD (W3181_1192)) or (DeoD3 (W3181_2368))	Change GPR
PUNP7	[c] : pi + xtsn <==> r1p + xan	(DeoD (W3181_1192)) or (DeoD3 (W3181_2368))	Change GPR
NAt3	h[e] + na1[c] --> h[c] + na1[e]	(NhaB (W3181_2343)) or (NhaC (W3181_0874)) or (NhaD (W3181_3296))	Change GPR
PIabc	atp[c] + h2o[c] + pi[e] --> adp[c] + h[c] + (2) pi[c]	PstABCS (W3181_2805 and W3181_2665 and W3181_2664 and W3181_2663)	Change GPR
Kt3	h[e] + k[c] --> h[c] + k[e]	(KefC (W3181_2868)) or (W3181_0693 (W3181_0693))	Change GPR
ALA_Dt4	ala-D[e] + na1[e] --> ala-D[c] + na1[c]	(W3181_3113 (W3181_3113)) or (AgcS (W3181_0800))	Change GPR
ALAt4	ala-L[e] + na1[e] --> ala-L[c] + na1[c]	(W3181_3113 (W3181_3113)) or (AgcS (W3181_0800))	Change GPR
GLYt4	gly[e] + na1[e] --> gly[c] + na1[c]	(W3181_3113 (W3181_3113)) or (AgcS (W3181_0800))	Change GPR
MNOR	(2) h[e] + mn4o[e] + mql7[c] --> (2) h2o[e] + mn2[e] + mqn7[c]	(CymA (W3181_3916) and MtrAso (W3181_2624) and MtrCso (W3181_2623))	Change GPR
MNOR2	(2) h[e] + mmql7[c] + mn4o[e] --> (2) h2o[e] + mmqn7[c] + mn2[e]	(CymA (W3181_3916) and MtrAso (W3181_2624) and MtrCso (W3181_2623))	Change GPR
ACCOAC	[c] : accoa + atp + hco3 --> adp + h + malcoa + pi	Acc (W3181_0788)	Change GPR
METS	[c] : 5mthf + hcys-L --> h + met-L + thf	MetH (W3181_0982)	Change GPR
HMGDx	[c] : hmglut-S + nad --> fglut-S + h + nadh	FrmA (W3181_1835)	Change GPR
ASNS1	[c] : asp-L + atp + gln-L + h2o --> amp + asn-L + glu-L + h + ppi	AsnB (W3181_2430)	Change GPR
UDPG4E	[c] : udpg <==> udpgal	GalE (W3181_2717)	Change GPR
L-LACt2	h[e] + lac-L[e] <==> h[c] + lac-L[c]	LctP (W3181_2834)	Change GPR

Appendix 5, continued

Abbreviations	Equations	Protein-Gene Associations	Notes
D-LACt2	h[e] + lac-D[e] <==> h[c] + lac-D[c]	LctP (W3181_2834)	Change GPR
GLYCLTt2r	glyclt[e] + h[e] <==> glyclt[c] + h[c]	LctP (W3181_2834)	Change GPR
AHCYSNS	[c] : ahcys + h2o --> ade + rhcys	Mtn (W3181_3030)	Change GPR
MTAN	[c] : 5mta + h2o --> 5mtr + ade	Mtn (W3181_3030)	Change GPR
5DOAN	[c] : dad-5 + h2o --> 5drib + ade	Mtn (W3181_3030)	Change GPR
UDCPDP	[c] : h2o + udcpdp --> h + pi + udcpp	BacA1 (W3181_3052)	Change GPR
THRt3	h[e] + thr-L[c] --> h[c] + thr-L[e]	ThrT (W3181_3476)	Change GPR
H2Ot5	h2o[e] <==> h2o[c]		Change GPR
UAG4E	[c] : uacgam --> uacgala		Change GPR
SO3De	(2) ficytcc[c] + h2o[e] + so3[e] --> (2) focytcc[c] + (2) h[e] + so4[e]	SorA (SO0714 and SO0715 and SO0716 and SO0717)	Remove GPR
NADH11	(5) h[c] + nadh[c] + ubq8[c] --> (4) h[e] + nad[c] + ubq8h2[c]	Nuo (SO1009 and SO1010 and SO1011 and SO1012 and SO1013 and SO1014 and SO1015 and SO1016 and SO1017 and SO1018 and SO1019 and SO1020 and SO1021)	Remove GPR
NADH13	(5) h[c] + mqn7[c] + nadh[c] --> (4) h[e] + mql7[c] + nad[c]	Nuo (SO1009 and SO1010 and SO1011 and SO1012 and SO1013 and SO1014 and SO1015 and SO1016 and SO1017 and SO1018 and SO1019 and SO1020 and SO1021)	Remove GPR
NADH16	(5) h[c] + mmqn7[c] + nadh[c] --> (4) h[e] + mmql7[c] + nad[c]	Nuo (SO1009 and SO1010 and SO1011 and SO1012 and SO1013 and SO1014 and SO1015 and SO1016 and SO1017 and SO1018 and SO1019 and SO1020 and SO1021)	Remove GPR
TMAOR3e	h[e] + tmao[e] + ubq8h2[c] --> h2o[e] + tma[e] + ubq8[c]	(TorA (SO1232) and TorC (SO1233))	Remove GPR
HGt	hg2[e] <==> hg2[c]	MerPT (SO1406 and SO1407)	Remove GPR
GLCGSD	[c] : glycogen + h2o --> glc-D	(Cga (SO2459)) or (GlcA1 (SO2213))	Remove GPR
MOBDabc	atp[c] + h2o[c] + mobd[e] --> adp[c] + h[c] + mobd[c] + pi[c]	ModABC (SO3863 and SO3864 and SO3865)	Remove GPR

Abbreviations	Equations	Protein-Gene Associations	Notes
FHLe	[e] : for + h --> co2 + h2	(Fdh3 (SO3922) and HydAB (SO3920 and SO3921))	Remove GPR
DMSOR3e	dmso[e] + mql7[c] --> dms[e] + h2o[e] + mqn7[c]	(CymA (W3181_3916) and DmsAB (SO1429 and SO1430) and DmsE (SO1427) and DmsF (SO1428))	Remove GPR
DMSOR4e	dmso[e] + mmql7[c] --> dms[e] + h2o[e] + mmqn7[c]	(CymA (W3181_3916) and DmsAB (SO1429 and SO1430) and DmsE (SO1427) and DmsF (SO1428))	Remove GPR
CMLDC	[c] : 2c25dho + h --> 5odhf2a + co2	PcaC2 (SO0083)	Remove GPR
ARSt1	arsni2[c] --> arsni2[e]	Arst (SO0534)	Remove GPR
MALTAT	[c] : accoa + malt --> acmalt + coa	Maa (SO1961)	Remove GPR
PHYT3e	[e] : h2o + inoshp --> inospp1 + pi	PhyS (SO2524)	Remove GPR
FORt	for[e] <==> for[c]	FocA (SO2911)	Remove GPR
UAGDH	[c] : h2o + (2) nad + uacgala --> (3) h + (2) nadh + uagnr	WbpO (SO3190)	Remove GPR
CLAT	[c] : accoa + cm <==> cmac + coa	ClaT (SO4299)	Remove GPR
ASPt2_3	asp-L[e] + (3) h[e] --> asp-L[c] + (3) h[c]	DcuB (SO4417)	Remove GPR
FUMt4	fum[e] + (3) h[e] --> fum[c] + (3) h[c]	DcuB (SO4417)	Remove GPR
MALt2_3	(3) h[e] + mal-L[e] --> (3) h[c] + mal-L[c]	DcuB (SO4417)	Remove GPR
SUCCt2_3	(3) h[e] + succ[e] --> (3) h[c] + succ[c]	DcuB (SO4417)	Remove GPR
SUCFUMtdc	fum[e] + succ[c] <==> fum[c] + succ[e]	DcuB (SO4417)	Remove GPR
FRD10	[c] : fum + mql7 --> mqn7 + succ	(FrdABCD (W3181_0201 and W3181_0202 and W3181_0203 and W3181_0204)) or (FrdABC1C2 (W3181_0396 and W3181_0397 and W3181_0398 and W3181_0399))	New GPR
FRD11	[c] : fum + mmql7 --> mmqn7 + succ	(FrdABCD (W3181_0201 and W3181_0202 and W3181_0203 and W3181_0204)) or (FrdABC1C2 (W3181_0396 and W3181_0397 and W3181_0398 and W3181_0399))	New GPR

Abbreviations	Equations	Protein-Gene Associations	Notes
CYTBO3	(2.5) h[c] + (0.5) o2[c] + ubq8h2[c] --> (2.5) h[e] + h2o[c] + ubq8[c]	Cyo (W3181_0092 and W3181_0093 and W3181_0094 and W3181_0095)	New GPR
MANpts	man[e] + pep[c] --> man6p[c] + pyr[c]	ManXYZ (W3181_0432 and W3181_0431 and W3181_0430)	New GPR
MTHPTGHM	[c] : 5mthglu + hcys-L --> met-L + thglu	(MetE (W3181_0771)) or (MetE2 (W3181_1250))	New GPR
Kabc	atp[c] + h2o[c] + k[e] --> adp[c] + h[c] + k[c] + pi[c]	(Kdp1 (W3181_0862 and W3181_0863 and W3181_0864)) or (Kdp2 (W3181_0862 and W3181_0863 and W3181_0865))	New GPR
ArstATPase	arsni2[c] + atp[c] + h2o[c] --> adp[c] + arsni2[e] + h[c] + pi[c]		New GPR
OAD	[c]: oaa + h --> pyr + co2	OadABG (W3181_3135 and W3181_3134 and W3181_3136)	New GPR
RBLK	[c] : atp + rbl-L <--> adp + h + ru5p-L	AraB (W3181_1944)	New GPR
RBP4E	[c] : ru5p-L <==> xu5p-D	AraD (W3181_1945)	New GPR
ARAI	[c] : arab-L <==> rbl-L	AraA (W3181_1946)	New GPR
EX_asp-L(e)	L-Aspartate exchange		Remove exchange reaction
EX_hg2(e)	Mercury (charged +2) exchange		Remove exchange reaction
EX_dmso(e)	Dimethyl sulfoxide exchange		Remove exchange reaction
EX_mobd(e)	Molybdate exchange		Remove exchange reaction
EX_inoshp(e)	myo-Inositol hexakisphosphate exchange		Remove exchange reaction
EX_inospp1(e)	1D-myo-inositol 1,3,4,5,6-pentakisphosphate exchange		Remove exchange reaction
EX_dms(e)	Dimethyl sulfide exchange		Remove exchange reaction
EX_man(e)	Manose exchange		New exchange reaction

Bibliography

[1] Crabtree, G. W., Dresselhaus, M. S., and Buchanan, M. V. The hydrogen economy. *Phys Today*. 2004, *57*, 39-44.

[2] Edwards, P. P., Kuznetsov, V. L., and David, W. I. F. Hydrogen energy. *Philos Trans A Math Phys Eng Sci*. 2007, *365*, 1043-1056.

[3] Tamagnini, P., Leitao, E., Oliveira, P., Ferreira, D, *et al.* Cyanobacterial hydrogenases: diversity, regulation and applications. *FEMS Microbiol. Rev.* 2007, *31*, 692-720.

[4] Josetespardellier, F., Astier, C., Evans, E. H., and Carr, N. G. Cyanobacteria grown under photoautotrophic, photoheterotrophic, and heterotrophic regimes - sugar metabolism and carbon-dioxide fixation. *FEMS Microbiol. Lett.* 1978, *4*, 261-264.

[5] Vermaas, W. F. Photosynthesis and respiration in cyanobacteria. *eLS*. http://www.els.net/ John Wiley & Sons, Ltd. 2001, 1-7.

[6] Johnson, C. H., and Golden, S. S. Circadian programs in cyanobacteria: adaptiveness and mechanism. *Annu. Rev. Microbiol.* 1999, *53*, 389-409.

[7] Beck, C., Knoop, H., Axmann, I. M., and Steuer, R. The diversity of cyanobacterial metabolism: genome analysis of multiple phototrophic microorganisms. *BMC Genomics*. 2012, *13*, 56.

[8] Parmar, A., Singh, N. K., Pandey, A., Gnansounou, E., and Madamwar, D. Cyanobacteria and microalgae: a positive prospect for biofuels. *Bioresour. Technol.* 2011, *102*, 10163-10172.

[9] Quintana, N., Van der Kooy, F., Van de Rhee, M. D., Voshol, G. P., and Verpoorte, R. Renewable energy from cyanobacteria: energy production optimization by metabolic pathway engineering. *Appl. Microbiol. Biotechnol.* 2011, *91*, 471-490.

[10] Angermayr, S. A., Hellingwerf, K. J., Lindblad, P., and de Mattos, M. J. T., Energy biotechnology with cyanobacteria. *Curr. Opin. Biotechnol.* 2009, *20*, 257-263.

[11] Bandyopadhyay, A., Stockel, J., Min, H., Sherman, L. A., and Pakrasi, H. B. High rates of photobiological H_2 production by a cyanobacterium under aerobic conditions. *Nat Commun.* 2010, *1*, 139.

[12] McNeely, K., Xu, Y., Bennette, N., Bryant, D. A., and Dismukes, G. C. Redirecting reductant flux into hydrogen production via metabolic engineering of fermentative carbon metabolism in a cyanobacterium. *Appl. Environ. Microbiol.* 2010, *76*, 5032-5038.

[13] Machado, I. M., and Atsumi, S. Cyanobacterial biofuel production. *J. Biotechnol.* 2012, *162*, 50-56.

[14] Hallenbeck, P. C., and Benemann, J. R. Biological hydrogen production; fundamentals and limiting processes. *Int J Hydrogen Energy*. 2002, *27*, 1185-1193.

[15] Atsumi, S., Higashide, W., and Liao, J. C. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat. Biotechnol.* 2009, *27*, 1177-1180.

[16] Lan, E. I., and Liao, J. C. ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 2012, *109*, 6018-6023.

[17] Shen, C. R., and Liao, J. C. Photosynthetic production of 2-methyl-1-butanol from CO_2 in cyanobacterium *Synechococcus elongatus* PCC7942 and characterization of the native acetohydroxyacid synthase. *Energy Environ Sci*. 2012, *5*, 9574-9583.

[18] Li, H., and Liao, J. C. Engineering a cyanobacterium as the catalyst for the photosynthetic conversion of CO_2 to 1,2-propanediol. *Microb. Cell Fact.* 2013, *12*.

[19] Oliver, J. W. K., Machado, I. M. P., Yoneda, H., Atsumi, S. Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. *Proc. Natl. Acad. Sci. U.S.A.* 2013, *110*, 1249-1254.

[20] Lindberg, P., Park, S., and Melis, A. Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. *Metab. Eng.* 2010, *12*, 70-79.

[21] Ungerer, J., Tao, L., Davis, M., Ghirardi, M., *et al.* Sustained photosynthetic conversion of CO_2 to ethylene in recombinant cyanobacterium *Synechocystis* 6803. *Energy Environ Sci*. 2012, *5*, 8998-9006.

[22] Angermayr, S. A., Paszota, M., and Hellingwerf, K. J. Engineering a cyanobacterial cell factory for production of lactic acid. *Appl. Environ. Microbiol.* 2012, *78*, 7098-7106.

[23] Dexter, J., and Fu, P. C. Metabolic engineering of cyanobacteria for ethanol production. *Energy Environ Sci*. 2009, *2*, 857-864.

[24] Liu, X., Sheng, J., Curtiss, R., 3rd. Fatty acid production in genetically modified cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 2011, *108*, 6899-6904.

[25] Tan, X. M., Yao, L., Gao, Q. Q., Wang, W. H., *et al.* Photosynthesis driven conversion of carbon dioxide to fatty alcohols and hydrocarbons in cyanobacteria. *Metab. Eng.* 2011, *13*, 169-176.

[26] Lan, E. I., and Liao, J. C. Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. *Metab. Eng.* 2011, *13*, 353-363.

[27] Stal, L. J., and Moezelaar, R. Fermentation in cyanobacteria. *FEMS Microbiol. Rev.* 1997, *21*, 179-211.

[28] McNeely, K., Xu, Y., Ananyev, G., Bennette, N., *et al. Synechococcus* sp. strain PCC 7002 *nifJ* mutant lacking pyruvate:ferredoxin oxidoreductase. *Appl. Environ. Microbiol.* 2011, *77*, 2435-2444.

[29] Ducat, D. C., Way, J. C., and Silver, P. A. Engineering cyanobacteria to generate high-value products. *Trends Biotechnol.* 2011, *29*, 95-103.

[30] Trinh, C. T., Unrean, P., and Srienc, F. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl. Environ. Microbiol.* 2008, *74*, 3634-3643.

[31] Atsumi, S., Hanai, T., and Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*. 2008, *451*, 86-89.

[32] Rodriguez, G. M., and Atsumi, S. Isobutyraldehyde production from *Escherichia coli* by removing aldehyde reductase activity. *Microb. Cell Fact.* 2012, *11*.

[33] Atsumi, S., Cann, A. F., Connor, M. R., Shen, C. R., *et al.* Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.* 2008, *10*, 305-311.

[34] Cann, A. F., and Liao, J. C. Production of 2-methyl-1-butanol in engineered *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 2008, *81*, 89-98.

[35] Maeda, T., Sanchez-Torres, V., and Wood, T. K. Hydrogen production by recombinant *Escherichia coli* strains. *Microb Biotechnol*. 2012, *5*, 214-225.

[36] Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., *et al.* EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* 2013, *41*, D605-612.

[37] Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., *et al. Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* 2012, *40*, D700-705.

[38] Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., *et al.* CyanoBase: the cyanobacteria genome database update 2010. *Nucleic Acids Res.* 2010, *38*, D379-381.

[39] Davidsen, T., Beck, E., Ganapathy, A., Montgomery, R., *et al.* The comprehensive microbial resource. *Nucleic Acids Res.* 2010, *38*, D340-345.

[40] Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., *et al.* The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 2005, *33*, 5691-5702.

[41] Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., *et al.* The NCBI BioSystems database. *Nucleic Acids Res.* 2010, *38*, D492-496.

[42] Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012, *40*, D109-114.

[43] Schomburg, I., Chang, A., Placzek, S., Sohngen, C., *et al.* BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.* 2013, *41*, D764-772.

[44] Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 2012, *40*, D742-753.

[45] Ren, Q., Chen, K., and Paulsen, I. T. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.* 2007, *35*, D274-279.

[46] Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., *et al.* High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* 2010, *28*, 977-982.

[47] Karp, P. D., Paley, S. M., Krummenacker, M., Latendresse, M., *et al.* Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief. Bioinformatics*. 2010, *11*, 40-79.

[48] Kumar, V. S., Dasika, M. S., and Maranas, C. D. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics*. 2007, *8*, 212.

[49] Orth, J. D., and Palsson, B. Gap-filling analysis of the *i*JO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions. *BMC Syst Biol*. 2012, *6*, 30.

[50] Risso, C., Van Dien, S. J., Orloff, A., Lovley, D. R., and Coppi, M. V. Elucidation of an alternate isoleucine biosynthesis pathway in *Geobacter sulfurreducens*. *J. Bacteriol.* 2008, *190*, 2266-2274.

[51] Wu, B., Zhang, B., Feng, X., Rubens, J. R., *et al.* Alternative isoleucine synthesis pathway in cyanobacterial species. *Microbiology*. 2010, *156*, 596-602.

[52] Latendresse, M., Krummenacker, M., Trupp, M., Karp, P. D. Construction and completion of flux balance models from pathway databases. *Bioinformatics*. 2012, *28*, 388-396.

[53] Vu, T. T., Stolyar, S. M., Pinchuk, G. E., Hill, E. A., *et al.* Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacteriu*m Cyanothece* sp ATCC 51142. *PLoS Comput. Biol.* 2012, *8.*

[54] Welsh, E. A., Liberton, M., Stockel, J., Loh, T., *et al.* The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proc. Natl. Acad. Sci. U.S.A.* 2008, *105*, 15094-15099.

[55] Edwards, J. S., and Palsson, B. O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U.S.A.* 2000, *97*, 5528-5533.

[56] Shastri, A. A., and Morgan, J. A. Flux balance analysis of photoautotrophic metabolism. *Biotechnol. Prog.* 2005, *21*, 1617-1626.

[57] Sun, J., Sayyar, B., Butler, J. E., Pharkya, P., *et al.* Genome-scale constraint-based modeling of *Geobacter metallireducens. BMC Syst Biol*. 2009, *3*, 15.

[58] Harvey, R. J., and Dev, I. K. Regulation in the folate pathway of *Escherichia coli*. *Adv. Enzyme Regul.* 1975, *13*, 99-124.

[59] Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* 2009, *5*, 593-599.

[60] Meeks, J. C., and Castenholz, R. W. Growth and photosynthesis in an extreme thermophile, *Synechococcus lividus* (Cyanophyta). *Arch Mikrobiol*. 1971, *78*, 25-41.

[61] Hamilton, J. J., and Reed, J. L. Identification of functional differences in metabolic networks using comparative genomics and constraint-based models. *PLoS ONE*. 2012, *7*.

[62] Zhang, S., and Bryant, D. A. The tricarboxylic acid cycle in cyanobacteria. *Science*. 2011, *334*, 1551-1553.

[63] Vu, T. T., Hill, E. A., Kucek, L. A., Konopka, A. E., *et al.* Computational evaluation of *Synechococcus* sp. PCC 7002 metabolism for chemical production. *Biotechnol J*. 2013, *8*, 619-630.

[64] Pinchuk, G. E., Hill, E. A., Geydebrekht, O. V., De Ingeniis, J., *et al.* Constraint-based model of *Shewanella oneidensis* MR-1 metabolism: a tool for data analysis and hypothesis generation. *PLoS Comput Biol*. 2010, *6*, e1000822.

[65] Palsson, B. Ø. *Systems biology: properties of reconstructed networks*. Cambridge University Press, New York. 2006.

[66] Westerhoff, H. V., Kolodkin, A., Conradie, R., Wilkinson, S. J., *et al.* Systems biology towards life in silico: mathematics of the control of living cells. *J Math Biol*. 2009, *58*, 7-34.

[67] Orth, J. D., Thiele, I., Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* 2010, *28*, 245-248.

[68] Segre, D., Vitkup, D., and Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.* 2002, *99*, 15112-15117.

[69] Edwards, J. S., and Palsson, B. O. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* 2000, *16*, 927-939.

[70] Mahadevan, R., and Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* 2003, *5*, 264-276.

[71] Reddy, K. J., Haskell, J. B., Sherman, D. M., and Sherman, L. A. Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus *Cyanothece*. *J. Bacteriol.* 1993, *175*, 1284-1292.

[72] Schneegurt, M. A., Sherman, D. M., and Sherman, L. A. Growth, physiology, and ultrastructure of a diazotrophic cyanobacterium, *Cyanothece* sp. strain ATCC 51142, in mixotrophic and chemoheterotrophic cultures. *J. Phycol.* 1997, *33*, 632-642.

[73] Nakamura, Y., Takahashi, J., Sakurai, A., Inaba, Y., *et al.* Some cyanobacteria synthesize semi-amylopectin type alpha-polyglucans instead of glycogen. *Plant Cell Physiol.* 2005, *46*, 539-545.

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

[75] Stockel, J., Welsh, E. A., Liberton, M., Kunnvakkam, R., *et al.* Global transcriptomic analysis of *Cyanothece* 51142 reveals robust diurnal oscillation of central metabolic processes. *Proc. Natl. Acad. Sci. U.S.A.* 2008, *105*, 6156-6161.

[76] Toepel, J., Welsh, E., Summerfield, T. C., Pakrasi, H. B., and Sherman, L. A. Differential transcriptional analysis of the cyanobacterium *Cyanothece* sp. strain ATCC 51142 during light-dark and continuous-light growth. *J. Bacteriol.* 2008, *190*, 3904-3913.

[77] Oberhardt, M. A., Palsson, B. O., and Papin, J. A. Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 2009, *5*, 320.

[78] Price, N. D., Reed, J. L., and Palsson, B. O. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2004, *2*, 886-897.

[79] Knoop, H., Zilliges, Y., Lockau, W., and Steuer, R. The metabolic network of *Synechocystis* sp. PCC 6803: systemic properties of autotrophic growth. *Plant Physiol*. 2010, *154*, 410-422.

[80] Montagud, A., Navarro, E., Fernandez de Cordoba, P., Urchueguia, J. F., and Patil, K. R. Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. *BMC Syst* Bio.l 2010, *4*, 156.

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

[82] Saha, R., Verseput, A. T., Berla, B. M., Mueller, T. J., *et al.* Reconstruction and comparison of the metabolic potential of cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803. *PLoS ONE*. 2012, *7*, e48285.

[83] Imam, S., Yilmaz, S., Sohmen, U., Gorzalski, A. S., *et al.* *i*Rsp1095: a genome-scale reconstruction of the *Rhodobacter sphaeroides* metabolic network. *BMC Syst Biol*. 2011, *5*, 116.

[84] Boyle, N. R., and Morgan, J. A. Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC Syst Biol*. 2009, *3*, 4.

[85] Chang, R. L., Ghamsari, L., Manichaikul, A., Hom, E. F., *et al.* Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol. Syst. Biol.* 2011, *7*, 518.

[86] Feng, X., Bandyopadhyay, A., Berla, B., Page, L., *et al.* Mixotrophic and photoheterotrophic metabolism in *Cyanothece* sp. ATCC 51142 under continuous light. *Microbiology*. 2010, *156*, 2566-2574.

[87] Aryal, U. K., Stockel, J., Krovvidi, R. K., Gritsenko, M. A., *et al.* Dynamic proteomic profiling of a unicellular cyanobacterium *Cyanothece* ATCC 51142 across light-dark diurnal cycles. *BMC Syst Biol*. 2011, *5*, 194.

[88] Stockel, J., Jacobs, J. M., Elvitigala, T. R., Liberton, M., *et al.* Diurnal rhythms result in significant changes in the cellular protein complement in the cyanobacterium *Cyanothece* 51142. *PLoS ONE*. 2011, *6*, e16680.

[89] Schluchter, W. M., and Bryant, D. A. Molecular characterization of ferredoxin-NADP+ oxidoreductase in cyanobacteria: cloning and sequence of the *petH* gene of *Synechococcus* sp. PCC 7002 and studies on the gene product. *Biochemistry*. 1992, *31*, 3092-3102.

[90] Teusink, B., Wiersma, A., Molenaar, D., Francke, C., *et al.* Analysis of growth of *Lactobacillus plantarum* WCFS1 on a complex medium using a genome-scale metabolic model. *J. Biol. Chem.* 2006, *281*, 40041-40048.

[91] Shlomi, T., Cabili, M. N., Herrgard, M. J., Palsson, B. O., and Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* 2008, *26*, 1003-1010.

[92] Allahverdiyeva, Y., Ermakova, M., Eisenhut, M., Zhang, P., *et al.* Interplay between flavodiiron proteins and photorespiration in *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* 2011, *286*, 24007-24014.

[93] Bailey, S., Melis, A., Mackey, K. R., Cardol, P., *et al.* Alternative photosynthetic electron flow to oxygen in marine *Synechococcus*. *Biochim. Biophys. Acta.* 2008, *1777*, 269-276.

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

[95] Shikanai, T. Cyclic electron transport around photosystem I: genetic approaches. *Annu Rev Plant Biol*. 2007, *58*, 199-217.

[96] Munekage, Y., Hashimoto, M., Miyake, C., Tomizawa, K., *et al*. Cyclic electron flow around photosystem I is essential for photosynthesis. *Nature*. 2004, *429*, 579-582.

[97] Battchikova, N., and Aro, E. M. Cyanobacterial NDH-1 complexes: multiplicity in function and subunit composition. *Physiol Plant*. 2007, *131*, 22-32.

[98] Iwai, M., Takizawa, K., Tokutsu, R., Okamuro, A., *et al.* Isolation of the elusive supercomplex that drives cyclic electron flow in photosynthesis. *Nature.* 2010, *464*, 1210-1213.

[99] Battchikova, N., Eisenhut, M., and Aro, E. M. Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles. *Biochim. Biophys. Acta.* 2011, *1807*, 935-944.

[100] Reed, J. L., Patel, T. R., Chen, K. H., Joyce, A. R., *et al.* Systems approach to refining genome annotation. *Proc. Natl. Acad. Sci. U.S.A.* 2006, *103*, 17480-17484.

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

[102] Raghunathan, A., Reed, J., Shin, S., Palsson, B., and Daefler, S. Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. *BMC Syst Biol*. 2009, *3*, 38.

[103] Becker, S. A., and Palsson, B. O. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol.* 2005, *5*, 8.

[104] Jensen, P. A., and Papin, J. A. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics*. 2011, *27*, 541-547.

[105] Min, H., and Sherman, L. A. Genetic transformation and mutagenesis via single-stranded DNA in the unicellular, diazotrophic cyanobacteria of the genus *Cyanothece*. *Appl. Environ. Microbiol.* 2010, *76*, 7641-7645.

[106] Niederholtmeyer, H., Wolfstadter, B. T., Savage, D. F., Silver, P. A., and Way, J. C. Engineering cyanobacteria to synthesize and export hydrophilic products. *Appl. Environ. Microbiol.* 2010, *76*, 3462-3466.

[107] Dellomonaco, C., Clomburg, J. M., Miller, E. N., and Gonzalez, R. Engineered reversal of the [bgr]-oxidation cycle for the synthesis of fuels and chemicals. *Nature*. 2011, *476*, 355-359.

[108] Nomura, C. T., Sakamoto, T., and Bryant, D. A. Roles for heme–copper oxidases in extreme high-light and oxidative stress response in the cyanobacterium *Synechococcus* sp. PCC 7002. *Arch. Microbiol.* 2006, *185*, 471-479.

[109] Batterton, J. C., Jr., and Van Baalen, C., Growth responses of blue-green algae to sodium chloride concentration. *Arch Mikrobiol*. 1971, *76*, 151-165.

[110] Sakamoto, T., and Bryant, D. A. Synergistic effect of high-light and low temperature on cell growth of the Delta12 fatty acid desaturase mutant in *Synechococcus* sp. PCC 7002. *Photosyn. Res.* 2002, *72*, 231-242.

[111] Frigaard, N. U., Sakuragi, Y., and Bryant, D. A. Gene inactivation in the cyanobacterium *Synechococcus* sp. PCC 7002 and the green sulfur bacterium *Chlorobium tepidum* using in vitro-made DNA constructs and natural transformation. *Methods Mol. Biol.* 2004, *274*, 325-340.

[112] Jacobsen, J. H., Rosgaard, L., Sakuragi, Y., and Frigaard, N. U. One-step plasmid construction for generation of knock-out mutants in cyanobacteria: studies of glycogen metabolism in *Synechococcus* sp. PCC 7002. *Photosyn. Res.* 2011, *107*, 215-221.

[113] Ludwig, M., and Bryant, D. A. Transcription profiling of the model cyanobacterium *Synechococcus* sp. strain PCC 7002 by Next-Gen (SOLiD) sequencing of cDNA. *Front Microbiol*. 2011, *2*, 41.

[114] Ludwig, M., and Bryant, D. A. Acclimation of the global transcriptome of the cyanobacterium *Synechococcus* sp. strain PCC 7002 to nutrient limitations and different nitrogen sources. *Front Microbiol.* 2012, *3*, 145.

[115] Shlomi, T., Berkman, O., and Ruppin, E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Natl. Acad. Sci. U.S.A.* 2005, *102*, 7695-7700.

[116] Kim, J., and Reed, J. L. RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. *Genome Biol*. 2012, *13*, R78.

[117] Burgard, A. P., Pharkya, P., and Maranas, C. D. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 2003, *84*, 647-657.

[118] Kim, J., and Reed, J. L., OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst Biol*. 2010, *4*, 53.

[119] Kim, J., Reed, J. L., and Maravelias, C. T. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *PLoS ONE*. 2011, *6*, e24162.

[120] Pharkya, P., Burgard, A. P., and Maranas, C. D. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.* 2004, *14*, 2367-2376.

[121] Ranganathan, S., Suthers, P. F., and Maranas, C. D. OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.* 2010, *6.*

[122] Fong, S. S., Burgard, A. P., Herring, C. D., Knight, E. M., *et al.* In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol. Bioeng.* 2005, *91*, 643-648.

[123] Burgard, A. P., and Van Dien, S. J. Methods and organisms for the growth-coupled production of succinate**.** *Patent* 2007. WO/2007/030830.

[124] Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., *et al.* Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* 2011, *7*, 445-452.

[125] Tamagnini, P., Axelsson, R., Lindberg, P., Oxelfelt, F., *et al.* Hydrogenases and hydrogen metabolism of cyanobacteria. *Microbiol. Mol. Biol. Rev.* 2002, *66*, 1-20.

[126] Young, J. D., Shastri, A. A., Stephanopoulos, G., and Morgan, J. A. Mapping photoautotrophic metabolism with isotopically nonstationary C-13 flux analysis. *Metab. Eng.* 2011, *13*, 656-665.

[127] Kamarainen, J., Knoop, H., Stanford, N. J., Guerrero, F., *et al.* Physiological tolerance and stoichiometric potential of cyanobacteria for hydrocarbon fuel production. *J. Biotechnol.* 2012, *162*, 67-74.

[128] Schluchter, W. M., Zhao, J., and Bryant, D. A. Isolation and characterization of the *ndhF* gene of *Synechococcus* sp. strain PCC 7002 and initial characterization of an interposon mutant. *J. Bacteriol.* 1993, *175*, 3343-3352.

[129] Overmann, J., and Schubert, K. Phototrophic consortia: model systems for symbiotic interrelations between prokaryotes. *Arch. Microbiol.* 2002, *177*, 201-208.

[130] Klitgord, N., and Segre, D. Environments that induce synthetic microbial ecosystems. *PLoS Comput Biol*. 2010, *6.*

[131] Jones, B. W., and Nishiguchi, M. K. Counterillumination in the Hawaiian bobtail squid, *Euprymna scolopes* Berry (Mollusca : Cephalopoda). *Mar. Biol.* 2004, *144*, 1151-1155.

[132] Oke, V., and Long, S. R. Bacteroid formation in the *Rhizobium*-legume symbiosis. *Curr. Opin. Microbiol.* 1999, *2*, 641-646.

[133] Broughton, W. J., Zhang, F., Perret, X., and Staehelin, C. Signals exchanged between legumes and *Rhizobium*: agricultural uses and perspectives. *Plant Soil.* 2003, *252*, 129-137.

[134] Lee, Y. K., Lee, J. H., and Lee, H. K. Microbial symbiosis in marine sponges. *J. Microbiol.* 2001, *39*, 254-264.

[135] Konig, G. M., Kehraus, S., Seibert, S. F., Abdel-Lateff, A., and Muller, D. Natural products from marine organisms and their associated microbes. *Chembiochem*. 2006, *7*, 229-238.

[136] Moore, B. S. Biosynthesis of marine natural products: microorganisms and macroalgae. *Nat Prod Rep.* 1999, *16*, 653-674.

[137] *CRC Handbook of Symbiotic Cyanobacteria*, CRC Press, Inc. 1990.

[138] Resendis-Antonio, O., Reed, J. L., Encarnacion, S., Collado-Vides, J., and Palsson, B. O. Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. *PLoS Comput Biol.* 2007, *3*, 1887-1895.

[139] Stolyar, S., Van Dien, S., Hillesland, K. L., Pinel, N., *et al.* Metabolic modeling of a mutualistic microbial community. *Mol. Syst. Biol.* 2007, *3*, 92.

[140] Dobson, P. D., Smallbone, K., Jameson, D., Simeonidis, E., *et al.* Further developments towards a genome-scale metabolic model of yeast. *BMC Syst Biol*. 2010, *4*.

[141] Bordbar, A., Feist, A. M., Usaite-Black, R., Woodcock, J., *et al.* A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC Syst Biol.* 2011, *5*, 180.

[142] Lewis, N. E., Schramm, G., Bordbar, A., Schellenberger, J., *et al.* Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat. Biotechnol.* 2010, *28*, 1279-U1291.

[143] Zomorrodi, A. R., and Maranas, C. D. OptCom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput Biol*. 2012, *8*.

[144] Taffs, R., Aston, J. E., Brileya, K., Jay, Z., *et al.* In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. *BMC Syst Biol.* 2009, *3*.

[145] Rodionov, D. A., Yang, C., Li, X. Q., Rodionova, I. A., *et al.* Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus. *BMC Genomics.* 2010, *11*.