DISSERTATION

GENOME-SCALE METABOLIC MODELING OF CYANOBACTERIA: NETWORK

STRUCTURE, INTERACTIONS, RECONSTRUCTION AND DYNAMICS

Submitted by

Chintan Jagdishchandra Joshi

Department of Chemical and Biological Engineering

In partial fulfilment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2016

Doctoral Committee:

    Advisor: Ashok Prasad

    Christie A. M. Peebles
    Kenneth Reardon
    Graham Peers

ABSTRACT

GENOME-SCALE METABOLIC MODELING OF CYANOBACTERIA: NETWORK STRUCTURE, INTERACTIONS, RECONSTRUCTION AND DYNAMICS

Metabolic network modeling, a field of systems biology and bioengineering, enhances the quantitative predictive understanding of cellular metabolism and thereby assists in the development of model-guided metabolic engineering strategies. Metabolic models use genome-scale network reconstructions, and combine it with mathematical methods for quantitative prediction. Metabolic system reconstructions, contain information on genes, enzymes, reactions, and metabolites, and are converted into two types of networks: (i) gene-enzyme-reaction, and (ii) reaction-metabolite. The former details the links between the genes that are known to code for metabolic enzymes, and the reaction pathways that the enzymes participate in. The latter details the chemical transformation of metabolites, step by step, into biomass and energy. The latter network is transformed into a system of equations and simulated using different methods. Prominent among these are constraint-based methods, especially Flux Balance Analysis, which utilizes linear programming tools to predict intracellular fluxes of single cells. Over the past 25 years, metabolic network modeling has had a range of applications in the fields of model-driven discovery, prediction of cellular phenotypes, analysis of biological network properties, multi-species interactions, engineering of microbes for product synthesis, and studying evolutionary processes. This thesis is concerned with the development and application of metabolic network modeling to cyanobacteria as well as *E. coli*.

Chapter 1 is a brief survey of the past, present, and future of constraint-based modeling using flux balance analysis in systems biology. It includes discussion of (i) formulation, (ii) assumption, (iii) variety, (iv) availability, and (v) future directions in the field of constraint based modeling. Chapter 2, explores the enzyme-reaction networks of metabolic reconstructions belonging to various organisms; and finds that the distribution of the number of reactions an enzyme participates in, i.e. the enzyme-reaction distribution, is surprisingly similar. The role of this distribution in the robustness of the organism is also explored. Chapter 3, applies flux balance analysis on models of *E. coli*, *Synechocystis sp.* PCC6803, and C. reinhardtii to understand epistatic interactions between metabolic genes and pathways. We show that epistatic interactions are dependent on the environmental conditions, i.e. carbon source, carbon/oxygen ratio in *E. coli*, and light intensity in *Synechocystis sp.* PCC6803 and *C. reinhardtii*.

Cyanobacteria are photosynthetic organisms and have great potential for metabolic engineering to produce commercially important chemicals such as biofuels, pharmaceuticals, and nutraceuticals. Chapter 4 presents our new genome scale reconstruction of the model cyanobacterium, *Synechocystis sp.* PCC6803, called iCJ816. This reconstruction was analyzed and compared to experimental studies, and used for predicting the capacity of the organism for (i) carbon dioxide remediation, and (ii) production of intracellular chemical species. Chapter 5 uses our new model iCJ816 for dynamic analysis under diurnal growth simulations. We discuss predictions of different optimization schemes, and present a scheme that qualitatively matches observations.

ACKNOWLEDGEMENTS

I started my journey in the field of metabolic modeling about seven years ago, while I was a Professional Science Masters' (PSM) student at Oregon State University. Little did I know that modeling of MAPK pathway in bioprocess control systems, a class taught by Dr. Ganti Murthy, will send me down a path to pursue doctorate, an year later.

Now, after 6 years as I come to the final steps of my doctorate, I want to take an opportunity to express my gratitude to my advisor, co-advisor, committee members, professors, colleagues, department secretaries, friends, and family. These 6 years would not have been as productive, if it was not for these thoughtful, brilliant, dedicated, and hardworking people. I have truly come to understand the meaning of the phrase, "It takes a village…"

I deeply express my heartfelt gratitude to my advisor, Dr. Ashok Prasad, for his consistent faith in me since the inception of our student-advisor relationship. I highly commend his scientific enthusiasm, to allow me free rein at the choice of projects, which a young student can only dream about. Though he challenged me to do my best work, he also made sure that I remain on track rather than pursue tangents. I find myself indebted to his insurmountable patience, and support, during these 6 years of my learning in matters both professional and personal.

I am grateful to my co-advisor, Dr. Christie Peebles, who enhanced my learning with regular discussions on experimental biology of *E. coli* and cyanobacteria. Our collaborations on topics (included in this thesis) has greatly helped me in thinking about my work from various different aspects. I found my interactions with her during group and personal meetings as very enlightening.

Lastly, and most importantly, I am grateful to my parents and my brother for their patience, encouragement, and love which gives me strength to do better work each day. Their support has made me who I am.

DEDICATION

To my parents and brother

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1.   CONSTRAINT BASED MODELING OF METABOLIC NETWORKS IN SYSTEMS BIOLOGY

1.   SYSTEMS BIOLOGY

Systems biology is a field of study involving "the study of interaction between components of biological systems, and how these interactions give rise to the function and behavior of that system" (Snoep & Westerhoff, 2005). This field of study, in part, is based on the recognition that the knowledge of properties of single biological parts outside the system alone, is insufficient to satisfactorily explain the behavior of the whole biological system they belong to. The emergence of this field was also a reaction against what was thought to be excessive reductionism in biology. Reductionism claims that all complex biological entities can be explained by the sum of its parts; while Holism claims that complex biological entities are inherently greater than the sum their parts (Gilbert & Sarkar, 2000). To capture a holistic view of the whole, an iterative approach is prescribed in systems biology. This iterative approach involves generation of models and testing them by conducting experiments. The model changes as and when the model predictions do not match the experiments. This represents one of the most fundamental paradigm shift to holistic from reductionist approach, where the knowledge flows from component to system analysis (Palsson, 2006). Therefore, systems biology studies the idea of holism (in context of biology) best described by Aristotle: *"The whole is greater than the sum of its parts."*

1.1.   PARTS

In context of systems biology, parts refer to cellular components such as: (i) genes; (ii) proteins, formed due to genes; (iii) intracellular chemical species, formed by reactions caused by proteins/enzymes; (iv) cellular/intra-cellular membranes, which compartmentalize cellular

functions; and (v) intracellular fluids, which provide medium for cellular functions. During the late 20th century, biology was practiced with reductionist approaches, i.e. gaining the knowledge about properties of a cellular component (such as a gene, or a protein) by isolating it from the cell (Palsson, 2006). It should be noted that earlier studies in systems biology were supported by data generated using these reductionist approaches. Therefore, these early reductionist studies played a vital role in understanding the importance and need for systems biology.

1.2.   SUM OF ITS PARTS

Isolating a single part and studying it outside the system can be slow, less efficient, and time consuming. During the mid-1990's, the first complete genome sequences of three organisms (*Haemophilus influenza*, *Saccharomyces cerevisiae*, and *Methanococcus jannaschii*) belonging to three domains of life were released (Bult et al., 1996; Fleischmann et al., 1995; Goffeau et al., 1996). These developments ushered in the age of "-omics": genomics (DNA), transcriptomics (RNA), proteomics (proteins), and metabolomics (chemical species or metabolites). Using the "-omics" approaches, data on a single class of cellular component (DNA, RNA, protein, or chemical species) belonging to an organism can be gathered. New high-throughput "-omics" methods are being developed faster than ever (Gomez-Cabrero et al., 2014). As of this year, more than 9000 organisms have been completely sequenced and more than that are underway (Reddy et al., 2015). The "-omics" data are being turned into genome-scale models which can be computationally simulated and analyzed. Currently, a large part of the field is involved in simulating individual "-omics" parts. A large amount of efforts are focused on understanding "the sum of its parts"; i.e. studying protein-protein interactions (Lv et al., 2015; Schoenrock et al., 2014; Wuchty et al., 2014), gene interactions (D'Souza, Waschina, Kaleta, & Kost, 2015; He, Qian, Wang, Li, &

Zhang, 2010; Joshi & Prasad, 2014; Phillips, 2008), organismal metabolic networks (Chang et al., 2011; Feist et al., 2007; Förster, Famili, Fu, Palsson, & Nielsen, 2003; N Jamshidi, Edwards, Fahland, Church, & Palsson, 2001; Knoop et al., 2013), and genome-wide RNA expression levels (Camas & Poyatos, 2008; Kochanowski, Sauer, & Chubukov, 2013; Kopf et al., 2014). Not only have the models been created, but modeling in systems biology have also facilitated predicting gaps in knowledge which can be later identified (Satish Kumar, Dasika, & Maranas, 2007).

1.3.   THE WHOLE

The whole, in systems biology, may refer to a set of proteins/genes/chemical species, a biological process, an organism, an ecosystem, or something as grandeur as biological life (Balaram, 2003; Xavier, Patil, & Rocha, 2014). The ability to model an organism, in its entirety, would involve analysis of all its "-omics" parts simultaneously. Systems biology is still in its infancy; taking all of its parts into consideration is computationally expensive and time consuming, and we are still learning about new parts (e.g. discovery of long noncoding RNAs) (Mattick & Rinn, 2015) and new functions of old parts (e.g. peroxisomes are involved in biotin biosynthesis) (Maruyama, Yamaoka, Matsuo, Tsutsumi, & Kitamoto, 2012). However, increasing number of models with multi-omics approach are being released and are underway (D. R. Hyduke, Lewis, & Palsson, 2013; Kim & Lun, 2014). Further, the advances in systems biology has facilitated its interface with synthetic biology. In fact, design and synthesis of a minimal genome has been made possible from the understanding of essential cellular function (Xavier et al., 2014). Although, the genome has not been fully functionally characterized; the essentiality of all the 473 genes in *M. mycoides* JCVI-syn3.0 is qualitatively understood (Hutchison et al., 2016). This minimal genome can further facilitate understanding of essentiality of cellular functions.
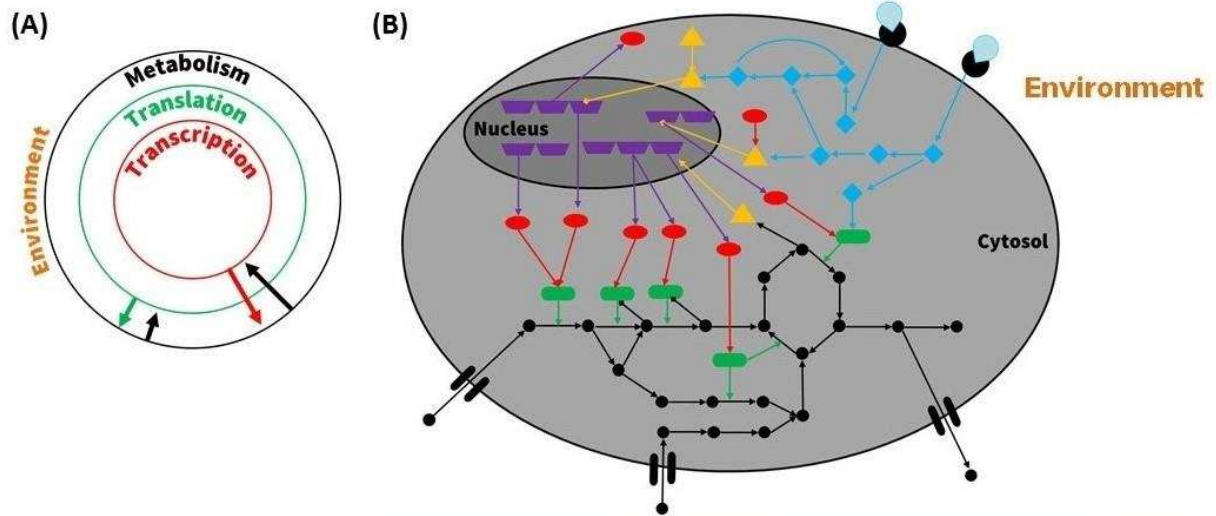
FIGURE 1.1: A SIMPLISTIC VIEW OF REGULATION BY EXCHANGE OF INFORMATION
(A) Cartoon depicting information exchange between different layers of cellular operations. Black line represents metabolism, green line represents translation, red line represents transcription, and arrows represent exchange of information. **(B)** Cartoon depicting how various cellular operations exchange information. Purple trapezoids represent genes, red ovals represent mRNA, green ovals represent enzymes/proteins, black dots represent metabolites, yellow triangles represent transcription factors, blue rhombi represent signaling cascades, and light blue drops represent environmental stimuli.

A simplistic view of understanding the structure of systems biology can be related to a simplistic view of understanding intracellular regulation (Figure 1.1). Each layer of intracellular regulation not only interacts with other layers but also leads to exchange of information. For example genes (genomics) are transcribed to form RNA; RNA (transcriptomics), then, forms an enzyme; an enzyme (proteomics) may in turn catalyze a reaction; then, resulting in conversion of one metabolite (metabolomics) to another; and preparing the cell to divide and secrete metabolites into the environment (Figure 1.1B). However, cells from different species result in various different phenotypes. This is done by exchange of information between different layers via environmental stimuli, signaling cascades, and transcriptional factors (Figure 1.1A). The cellular metabolism as an important interface with the environment. Therefore, understanding organismal

systems biology requires understanding of metabolism as part of the system, the organism; which can best be accomplished by modeling organismal metabolism.

2.   METABOLIC MODELING LANDSCAPE

Metabolism can be defined as the set of processes that allow the cell to maintain itself and to grow. Therefore, the two primary tasks of metabolism are to enable (i) the maintenance of energy, redox, and storage machinery and (ii) the growth of the cell, e.g. to produce metabolites and biomass required by daughter cells after division. These tasks are accomplished by a metabolic network, which is a set of chemical interconversions from various nutrient uptakes to cellular biomass and energy via a set of enzymes produced within a cell. Therefore, modeling of metabolism is primarily concerned with reconstructing and simulating the metabolic network of the organism for different environmental stimuli. Metabolic models are primarily growth models of the organism. Microbial growth models can be categorized at (i) intracellular level as structured (multi-component system) or unstructured (single component system), based on the treatment of intracellular molecules; and (ii) multicellular level as segregated (heterogeneous) or unsegregated (homogeneous), based on the treatment of cell population (Figure 1.2). Modeling organismal growth began with simplified unstructured and unsegregated models like Monod equation, which expressed the growth rate as a function of nutrient uptakes without going into how nutrient was assimilated into biomass and how growth was taking place (Monod, 1949). Though such models captured growth kinetics fairly well, they lacked any information on intracellular state, mainly because this information was limited at best during that time. Therefore, development of metabolic modeling has largely been an effort to develop structured models which capture the intracellular states of various components (metabolites and reactions) within the cell. Most of the efforts have

Unstructured

All molecules are <u>same and have same</u> <u>properties</u>, for e.g. no difference between **pyruvate** and **ATP**.

Structured

All molecules are <u>different and have</u> <u>different properties</u>, for e.g. **ATP** transfers *chemical energy*, and **pyruvate** *interfaces glycolysis and TCA cycle.*

Unsegregated

Homogeneous: All <u>cells</u> within a growth culture are <u>same</u>, for e.g. **biomass composition** of all cells is same.

Segregated

Heterogeneous: All <u>cells</u> within a growth culture are <u>different.</u>, for e.g. **biomass composition** of all cells is different.

FIGURE 1.2: CATEGORIES OF MICROBIAL GROWTH MODELS
Cartoon depiction of categories of microbial growth models at (i) intracellular level – structured or unstructured, and (ii) multicellular level – segregated or unsegregated. Colorful shapes with the cyan circle (cell) depict intracellular molecules.

been driven towards unsegregated metabolic models. However, research is underway to capture heterogeneity in cellular biomass composition to make segregated metabolic models (Personal communication with Dr. Maciek Antoniewicz at University of Delaware). We will discuss this in later sections.

Metabolic modeling has proved highly useful over the years. However, as mentioned earlier, the state of the mathematical models of metabolism have been limited by the amount of data that could be (or has been) experimentally verified or measured. Other limitations to the field also include lack of kinetic information about the various intracellular reactions. However, more quantitative information about intracellular reactions/genes/enzymes is becoming available. This

information proves highly useful in creating highly accurate models. Metabolism, as we know it, acts under different time scales, with different reaction rates, and shows different kinetic behavior at different levels of regulation. Hence, most of the progress has been made in the field of pseudo steady state models that ignore the kinetic information and regulation; and are solved using steady state mass-balance equations, without taking time into consideration (Song & Ramkrishna, 2009b). However, new improvements have facilitated implementation of transcriptional information with genome-scale metabolic models.
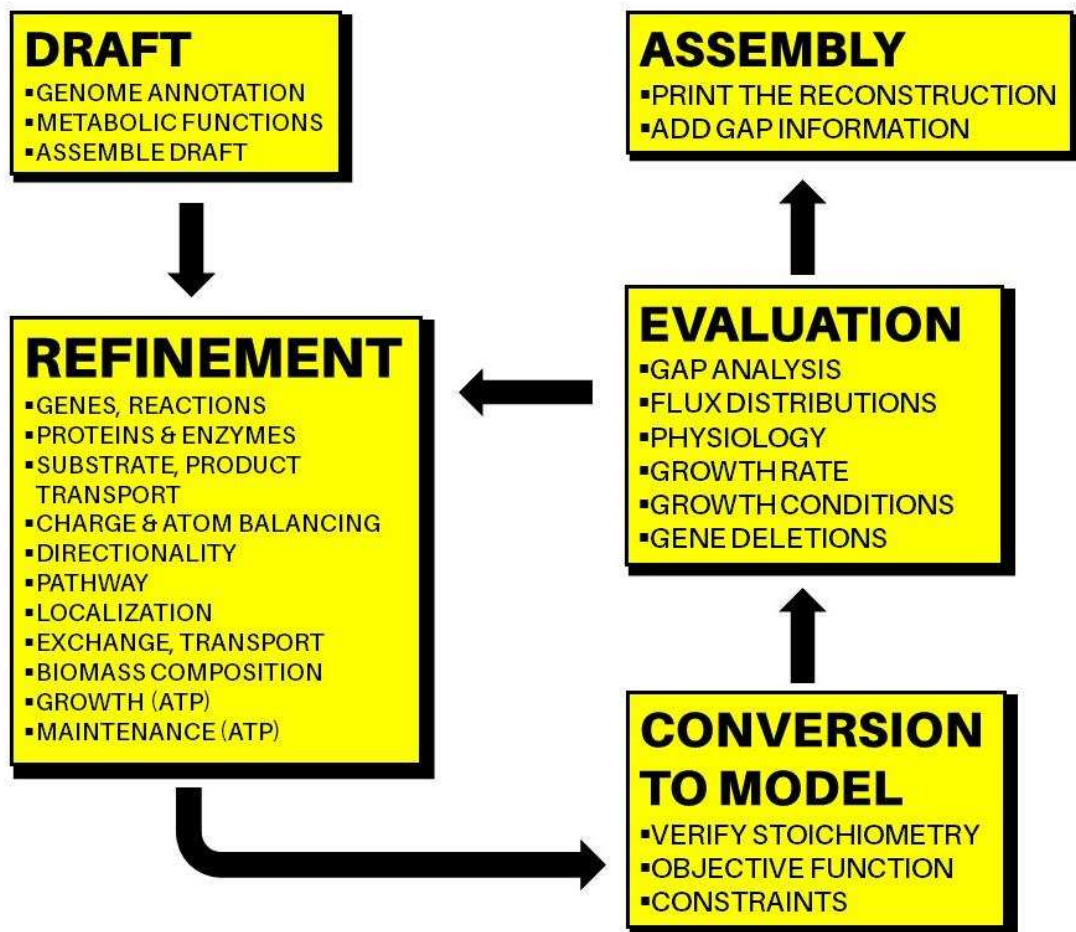
FIGURE 1.3: PARADIGM IN PREPARING GENOME-SCALE NETWORK RECONSTRUCTIONS
The flow elaborates how genome-scale metabolic network reconstructions are built. It involves 5 parts: (i) draft, (ii) refinement, (iii) conversion to model, (iv) evaluation, and (v) assembly.

2.1. METABOLIC NETWORK RECONSTRUCTIONS

A precursor to all genome-scale metabolic models is a genome-scale metabolic network reconstruction. To date, genome-scale network reconstructions are available for 69 different organisms and strains (Table A1) (King et al., 2016). A wide-array of tools and databases are available for preparing metabolic reconstructions. These include genome databases (Reddy et al., 2015), biochemical databases (Minoru Kanehisa et al., 2014; Scheer et al., 2011; Y. Wang et al., 2009), organism-specific databases (Keseler et al., 2009), protein localization databases (N. Y. Yu et al., 2010), reconstruction packages (Paley & Karp, 2006), simulation environments (D. Hyduke et al., 2011; Klamt, Saez-Rodriguez, & Gilles, 2007; Klamt, Stelling, Ginkel, & Gilles, 2003; R. Luo, Liao, Zeng, Li, & Luo, 2006), and visualization packages (Maarleveld, Boele, Bruggeman, & Teusink, 2014). A simplified version of the current paradigm in preparing reconstructions involves five steps: (i) drafting, (ii) refinement, (iii) conversion to mathematical model, (iv) evaluation, and (v) assembly and dissemination (Figure 1.3) (Thiele & Palsson, 2010). Reconstruction is subjected to iterations of refinement and evaluation (ii-iv) until predictions match well with organism phenotype. Though many tools and databases are available, the process of reconstruction is semi-automated at best. This can be attributed to two main reasons: (i) varied objectives of each reconstruction, and (ii) availability of physiological data.

2.2. MATHEMATICAL MODEL

Once the reconstruction is built, the process of converting it into a mathematical model may be automated. Mathematical models are usually condition-specific, which involve invoking constraints and defining system boundary (Thiele & Palsson, 2010). The network itself is represented as a stoichiometric matrix of reactions and metabolites within the mathematical model;

and it is derived from the reconstruction. Each element within the matrix represents the stoichiometry of a metabolite in a reaction. A negative value in the matrix represents consumption of a metabolite, while a positive value represents production of a metabolite. As mentioned earlier, the reconstruction, and hence, the network is only representation of metabolism under balanced growth assumptions (Palsson, 2006). However, most biological networks are underdetermined; therefore, multiple flux solutions which satisfy the intracellular mass balance exist. A linear programming (LP) problem is formulated by removing the kinetics out of the system and making it a time-invariant mass-balance problem. To be able to reduce the solution space, we choose a cellular objective to optimize and apply constraints (Feist & Palsson, 2010; Price, Reed, & Palsson, 2004). Therefore, constraints are a crucial part of the mathematical model, making the models condition-specific, which is also why these models are also known as constraint based models. The main assumption or hypothesis behind constraint based models is that the organism optimizes some cellular objective function. These predicted fluxes are the solution of the metabolic model and can be then compared with experimental results. The constraints, the objective function and the solution in context of constraint based modeling are discussed in more detail below.

3.   CONSTRAINT BASED MODELING

The field of constrained based modeling, most notably flux balance analysis (FBA), has become one of the most important tools in genome-scale metabolic flux analyses. Currently, we have a wide array of FBA models available for many organisms; amongst the most well-known are *E.coli* (Feist et al., 2007), *S. cerevisiae* (Förster et al., 2003), *Clostridium thermocellum* (Roberts, Gowen, Brooks, & Fong, 2010), *Arabidopsis thaliana* (Poolman, Miguet, Sweetlove, &

Fell, 2009), *Synechocystis sp.* PCC6803 (Knoop et al., 2013; J. Nogales, Gudmundsson, Knight, Palsson, & Thiele, 2012), and *C. reinhardtii* (Chang et al., 2011).

3.1.   CONSTRAINTS

In constraint based modeling, constraints belong to four different categories: physico-chemical, topobiological, regulatory and environmental constraints; and are applied as bounds and balances (Price et al., 2004). Physico-chemical constraints are dependent on free energies of biochemical reactions (Hamilton, Dwivedi, & Reed, 2013), diffusion rates (Weisz, 1973), enzyme turnover, and confinement of molecules (Lew & Bookchin, 1986). Topobiological constraints are dependent on molecular crowding, and number of molecules of metabolites. Regulatory constraints are dependent on transcriptional, translational or enzymatic regulation within the cell and are hypothesized to eliminate suboptimal cellular states. Lastly, environmental constraints are dependent on concentration of nutrients and are subject to change with interaction of metabolic network and its environment. These constraints determine the solution space (all possible phenotypes) within which the solution lies. Constraints can be implemented in various ways: (i) reaction reversibility – whether the enzyme catalyzing the reaction is reversible, (ii) reaction bounds, and (iii) biomass composition. Reaction reversibility can be determined by checking whether the free energy of a given reaction is negative, positive, or zero. If negative, the reactions must be allowed in only forward direction; if positive, the reaction can only carry negative flux; and if zero, the reaction must be allowed in both directions. Reaction bounds refer to upper, lower, or both bounds of flux through the reaction and are applied when diffusion rates of metabolites, uptake of nutrients, or regulatory control of enzyme is known; for e.g. flux through Ribulose-1,5-bisphopshate oxygenase activity is shown to be approximately 3-5% of the total Ribulose-1,5-

bisphosphate carboxylase activity (Timm & Bauwe, 2013; Vermaas, 2001). Biomass composition applies topobiological constraints on the model and is implemented by changing the equation representing the requirements of various metabolites for the cell to grow. As mentioned previously, the heterogeneity of biomass equation for a given microorganism is still under active research. Due to the large size of genome-scale metabolic models, even after applying the constraints, it is likely that a unique solution is not possible. Therefore, the solution often, refers to the feasible solution space rather than a unique solution.

3.2.   SOLUTION SPACE

The solution space refers to all the possible solutions to the problem of determining metabolic fluxes given the model and the constraints. Flux space can be visualized as a region in n-dimensional space. The n coordinates on each point represent the number of reaction fluxes in the metabolic model. The space we refer to here is bounded only by constraints, discussed above, applied on the model and forms a solid. This solid is further shaved off by drawing out mass balance equations resulting from the stoichiometric network. Therefore, the true solution space is actually a polytope (surface) in n-dimensional space (Figure 1.4). For the linear mathematical programming problem to be feasible, it is important that the polytope is convex, and not concave. The polytope would be characterized by m intersection points of the entire stoichiometric network and all the constraints. It should be noted that the case referred to here is that of an under-determined system. As can be visualized from the above description, a unique solution may not exist. Therefore, the next step is to invoke an objective function of the metabolic network.
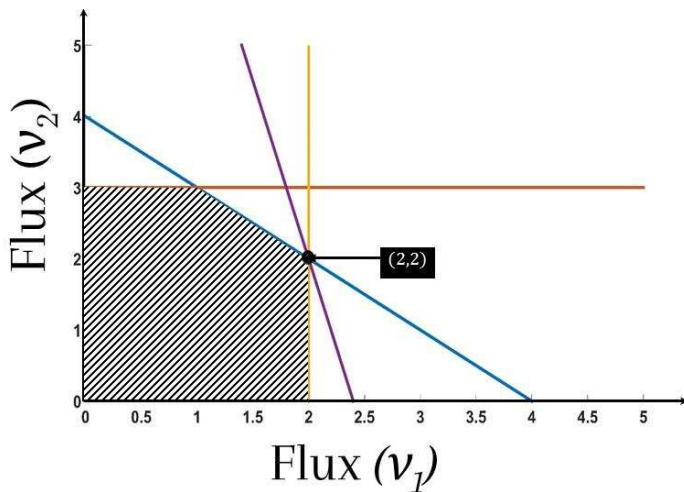
Objective function:

$Maximize\ Z = 5v_1 + v_2$

$f = \max_{v(t)} \sum \alpha_j . v_j(t)$

Stoichiometry of the Network:

$L_1: v_1 + v_2 \leq 4$

$S.v = 0$

Constraints:

$0 \leq v_1 \leq 2$

$0 \leq v_2 \leq 3$

$v_{j,min} \leq v_j \leq v_{j,max}$

FIGURE 1.4: AN EXAMPLE OF A SOLUTION SPACE GIVEN BY THE PROBLEM ON THE RIGHT.
The problem is defined on the right hand side. The plot on the left represents the solution of the problem. The color of the line corresponds to the equation of that color on the problem side. The shaded region indicates all possible values of the objective function (Z). In this case, the solution is (2, 2), which corresponds to the maximum value of Z and satisfied by bounds.

3.3. CELLULAR OBJECTIVE

To further constrain the solution space, it is required to hypothesize that the metabolic network maximizes or minimize a cellular function. This is often referred to as objective function. The purpose of objective function is (i) to explore the solution space (phenotypic space), (ii) to determine physiological state that best represents a physiological function of the organism such as growth or ATP production, and (iii) to determine fitness of engineered strains such as secretion of a desired product (Price et al., 2004). Exploration of solution space and choosing a fitness function for engineered strains are interrogative purposes and vary based on the goal of the study. From the description of solution space, an objective is described as a function of other intracellular fluxes; and in an n-dimensional space, it can be visualized as intersecting with the polytope. This solution space corresponding to maximum or minimum value of the objective can be found by substituting

allowable values of various fluxes in the equation of the objective. Therefore, if an objective function is to be maximized or minimized, the solution lies on the vertices of the intersection of the polytope and the objective function.

The choice of a best representative physiological function for the organism has long been in debate. However, as the size of the metabolic networks increase, it is possible to not get a single unique solution but rather a number of solutions. Often, the entire range of flux values, corresponding to the optimum value of objective function, is reported to facilitate analysis.

3.4. *SIMULATION ENVIRONMENT*

There are number of interfaces/environments available in the field to solve resulting linear equations; some of them include General Algebraic Modeling System (GAMS), MATLAB (various Toolboxes like SimBiology, COBRA (Schellenberger et al., 2011), SBML (Keating, Bornstein, Finney, & Hucka, 2006)), OptKnock (Burgard, Pharkya, & Maranas, 2003) etc. All these can optimize the objective function for the organism's metabolic network formulated as a linear/quadratic/mixed-integer linear/non-linear programming problem.

4. FBA PARADIGM

As mentioned previously, in FBA, a set of reactions is prepared that leads to production and consumption of each of the chemical compounds within a metabolic network, upon which constraints are imposed, and a cellular objective is chosen (Jeremy S. Edwards, Covert, & Palsson, 2002; Feist & Palsson, 2010; K. J. Kauffman, Prakash, & Edwards, 2003; Orth, Thiele, & Palsson, 2010). On the basis of this hypothesis and the constraints, standard optimization techniques such as linear programming are applied, that yield a vector of fluxes that optimize the objective and

satisfy all of the constraints. Therefore, mathematical frameworks such as FBA make it possible to calculate and analyze the flow of metabolites through a metabolic network and allow making predictions of growth and/or biotechnologically relevant products (Orth et al., 2010). Eventually, the cell metabolite pool is experimentally tested for various compounds; based on the set of chemical species involved in the model to calculate intracellular fluxes. The steps to constructing an FBA model are: (i) defining the system, (ii) obtaining reaction stoichiometry, (iii) defining biologically relevant objective functions and adding constraints, and (iv) solving the resulting linear equations (Raman & Chandra, 2009). We have discussed other parts of construction in detail except the details of understanding objective functions.

An obvious question still remains as to what cellular objective to choose. A simple answer to this question would be to find the experimentally obtained solution within the phenotypic space and use mathematical programming to identify the biochemical reaction state that maximizes network function experimentally obtained solution points at. However, as previous studies have noted, that this method only works for wild-type organisms (Robert Schuetz, Kuepfer, & Sauer, 2007; Segrè, Vitkup, & Church, 2002). The issue of choice of objective function was also recognized in the first study conducted by Savinell and Palsson (Savinell & Palsson, 1992a, 1992b), where systematically four different objective functions were tested for *E. coli* metabolism including minimization of ATP production, minimization of nutrient uptake (in moles), minimization of nutrient uptake (in mass), and minimization of NADH production. The study revealed that no single objective function captured the cell behavior accurately. However, later it was realized that a growth objective performed best in predicting cell behavior. It should be noted that *E. coli* strains have been growing long enough to evolve in laboratory conditions and have acquired optimal growth phenotype. This is also evident from experiments where after 700

generations under growth selection pressure, *E. coli* growing on glycerol shifted from sub-optimal growth rate to optimal growth rate predicted using *in silico* model (Ibarra, Edwards, & Palsson, 2002). This is a classic scenario where environmental perturbations drive the genetic perturbations such that the organism evolves to exhibit optimal growth phenotype.

However, sometimes it could be of interest to learn the intermediate state while the organism is transitioning to optimality. For such cases, a different objective function was introduced, minimization of metabolic adjustment (MOMA), as an extension to FBA. MOMA hypothesizes that gene deletion mutants undergo a minimal metabolic adjustment with respect to wild-type metabolic state (Segrè et al., 2002). Another objective function which models gene deletion mutants has also been introduced, regulatory on/off minimization (ROOM). It follows the similar assumptions of minimal adjustment from wild-type phenotypic state, but also hypothesizes that new phenotypic state of the mutant is reached through transient metabolic changes by the regulatory network which is minimized (Shlomi, Berkman, & Ruppin, 2005).

The search for a global objective function did not end there. Subsequently, many different objective functions have been tested by various groups. These include maximization of ATP per flux unit, minimization of overall flux, maximization of biomass per unit flux, minimization of reaction steps, maximization of ATP production (Robert Schuetz et al., 2007), and an objective function selector using Bayesian-based technique (Knorr, Jain, & Srivastava, 2007). A multidimensional approach was also able to gain limited success in forming an objective function which was a combination of maximal biomass yield, ATP yield, and minimization of sum of fluxes (R. Schuetz, Zamboni, Zampieri, Heinemann, & Sauer, 2012). In this approach wide variety of organisms growing under different environmental conditions were mapped on to a Pareto-optimal surface. This allowed them to make predictions; and show that evolution shapes metabolic fluxes

in microorganisms' environmental context by (i) optimal flux distribution under one given condition, and (ii) minimizing the adjustment between any two conditions. Among the above mentioned studies, there have been cases where contrary objectives such as maximization of ATP per flux unit was a better predictor of experimental data than biomass (Robert Schuetz et al., 2007). It should be noted that most of the work mentioned above was done in *E. coli* or yeast.

The evidence so far suggests that growth objective is the most consistent among all the ones evaluated. There could be conditions where the growth objective is not appropriate, such as for organisms in a nutrient limited environment, organisms undergoing physical stress, etc. Most scientists that have critiqued this hypothesis in publications have argued that if growth rate is maximized, it is also necessary for the organism to maintain an appropriate level of expression for protein synthesis or ribosome expression (Bonven & Gulløv, 1979; Forchhammer & Lindahl, 1971).

5.   DYNAMIC FLUX BALANCE ANALYSIS

FBA utilizes a static optimization framework, yielding a solution of flux vectors that do not change with time. However metabolism is dynamic and changes with environmental conditions. There have been several attempts to incorporate dynamics within the FBA framework, called dynamic FBA. Advances in dynamic-FBA (dFBA) (Radhakrishnan Mahadevan, Edwards, & Doyle, 2002; Varma & Palsson, 1994) have shown that given some insights into the substrate uptake kinetics, a time variant problem can be solved for batch kinetics as a function of rate of reactions. Dynamic-FBA includes information about dynamics of a certain metabolite under batch kinetics or under time-dependent processes allowing interaction of the metabolic network with the environment (Jared L. Hjersted & Henson, 2006; Radhakrishnan Mahadevan et al., 2002). DFBA

provides a structured model of biochemical process where intracellular pathways interact with the environmental conditions, which is represented by functional dependency of the substrate. There are two most used versions of dFBA; dynamic optimization approach (a non-linear programming problem, that optimizes the fluxes over the entire time, DOA) and static optimization approach (a linear programming problem, that instantaneously optimizes over small time intervals to make up the entire time; and updates concentration after each time interval, SOA) (J L Hjersted & Henson, 2009; Jared L. Hjersted & Henson, 2006; Radhakrishnan Mahadevan et al., 2002). In addition to environmental interactions, regulatory changes due to the environment can be included in the model. A third dynamic FBA method involves embedding a linear program within a system of kinetic equations representing the exchange fluxes (Gomez, Höffner, & Barton, 2014). DFBA is increasingly becoming more efficient for larger models, moving from 43 metabolites and 38 reactions of a monoculture to more than 3000 reactions from a co-culture simulation of *C. reinhardtii* (iRC1080) and yeast (iND750) (Table A2) (Höffner, Harwood, & Barton, 2013).

6.   OTHER MODELING FRAMEWORKS

Cybernetic models of microbial growth involve taking into account metabolic regulation, enzymatic regulation, and a single substrate uptake kinetics, and couples them with the metabolic network (Kompala, Ramkrishna, & Tsao, 1984). Unlike FBA, which requires uptake rates of multiple substrates to be specified; cybernetic modeling needs information from only one substrate, and the uptake rates of other substrates is estimated. It was first applied to predictions of diauxic growth patterns in multiple substrate bacterial cultures (Kompala, Ramkrishna, Jansen, & Tsao, 1986; Kompala et al., 1984). There have been some advances in Cybernetic modeling as well (Song & Ramkrishna, 2011; J. Young, Henne, Morgan, Konopka, & Ramkrishna, 2004).

However, there is little known about it in large complex networks. Cybernetic Modeling has heavily relied upon lumping of metabolic pathways (Song & Ramkrishna, 2009a, 2011). Applying this technique over large metabolic network is highly constrained due to lack of experimentally determined parameters. Further, due to lumping of metabolic pathways, many interesting genetic changes cannot be predicted. However, recent successors of this technique such as Lumped Hybrid Cybernetic Models (L-HCM) and Lumped-Elementary Mode (L-EM) have been applied to *E. coli* and *S. cerevisiae* (Papin et al., 2004; Schwartz & Kanehisa, 2006; Song & Ramkrishna, 2011; J. Young et al., 2004). There are other tools that have evolved on similar lines and given us some insights into the network structure such as elementary mode analysis (EMA) (Zanghellini, Ruckerbauer, Hanscho, & Jungreuthmayer, 2013), and lumped kinetic modeling (LKM) (Nikolaev, 2010). The field of quantitative metabolic modeling has been on constant progress for the past 20 years now and continues to grow. As the field develops, new methods are emerging in parameter identification, metabolite kinetics, and other fields that might involve more sophisticated model formulations.

7.   APPLICATIONS

As mentioned previously, the number of constraints based genome scale models have been rising consistently. Our analysis of Biochemical Genetic and Genomic (BiGG) databases suggests more than 60 models are in existence (King et al., 2016). Their usefulness in learning more about the phenotypic space has been elaborated in the previous sections. Therefore, here, we will focus on the biotechnological contributions that led (i) to understanding more about the systemic behavior, and (ii) to improve commercial outcomes (biofuels, pharmaceuticals, or nutraceuticals).

7.1. APPLICATIONS IN NON-PHOTOSYNTHETIC BACTERIAL ORGANISMS

Bacterial models have demonstrated successful applications to production of industrially relevant chemicals. For example production of lactate has been modeled in *Lactobacillus plantarum*, *Lactococcus lactis*, *Streptococcus thermophilus*, and *Corynebacterium glutamicum*. *L. lactis* was also used to predict genetic modifications for improving production of diacetyl, which is a flavor compound in dairy products (Oliveira, Nielsen, Förster, & Forster, 2005). It has also been used to predict genetic modifications for synthesis of recombinant protein (Oddone, Mills, & Block, 2009). The resultant strains qualitatively enhanced GFP (a proxy for the recombinant protein) production by 15%. Genetic modifications in *Pseudomonas putida* were investigated for production of poly-3-hydroxyalkanes (PHA), which could be used to replace petrochemical-based plastics (Puchalka et al., 2008). The study demonstrated pools of acetyl-CoA, a precursor to PHAs, were increased by up to 26%. *E. coli* (Feist et al., 2007) and *C. acetobutylicum* (J. Lee, Yun, Feist, Palsson, & Lee, 2008; Salimi, Mandal, Wishart, & Mahadevan, 2010) models were used in making predictions about acetone-butanol-ethanol production systems. *Geobacter metallireducens* reduces Fe (III) and is used in bioremediation of radioactive elements. Its model was used to show that it grows inefficiently with complex electron donors and acceptors (Sun et al., 2009).

7.2. APPLICATIONS IN MAMMALIAN ORGANISMS

The first human genome scale model, Human Recon 1 (Duarte et al., 2007), was used to identify biomarkers of inborn errors of metabolism (Shlomi, Cabili, & Ruppin, 2009). This revealed a set of 233 metabolites whose concentration is predicted to deviate as a result of 176 possible dysfunctional enzymes. Another genome scale model reconstruction revealed the importance of systems modeling in human metabolism to aid drug discovery. Simulations and

predictions using genome scale models of NCI-60 cell lines have resulted in identification of a new objective function, as well as to study Warburg effect and identified metabolic targets for inhibiting cancer cell migration (Yizhak et al., 2014). Simulations of hybridoma cell line production of mAb in a genome scale model of *M. musculus* predicted growth and build-up of lactate and ammonia, known byproducts to cause cell death in mammalian cell culture (Sheikh, Forster, & Nielsen, 2005).

7.3.   APPLICATIONS IN E. COLI AND S. CEREVISIAE

These are some of the best studied microbial species to date. Their genome scale models have been equally well studied as well. Therefore, their applications have been far wider than any of the organisms previously mentioned. Some of the important contributions of *E. coli* genome scale models include increasing the production of lycopene (Alper, Jin, Moxley, & Stephanopoulos, 2005; Jin & Stephanopoulos, 2007), lactate (Burgard et al., 2003; Fong et al., 2005; Ibarra et al., 2002), ethanol (Pharkya & Maranas, 2006), hydrogen (Jones, 2008; Pharkya & Maranas, 2006), vanillin (Pharkya & Maranas, 2006), and 1,3-propanediol (Burgard et al., 2003). Similarly, *S. cerevisiae* has contributed to increasing production of succinate, glycerol, vanillin, and sesquiterpenes (Asadollahi et al., 2009; Patil, Rocha, Förster, & Nielsen, 2005).

7.4.   APPICATIONS IN PHOTOSYNTHETIC ORGANISMS

The most widely used and actively researched of all the photosynthetic organisms is *Synechocystis sp.* which can convert CO_2 to carbon based products. A recent study of a genome scale metabolic network of Synechocystis sp. analyzed the production of industrially relevant chemical compounds and growth trade-off (Knoop & Steuer, 2015) to find that shifts in

ATP/NADPH demand during autotrophic growth competed with product biosynthesis. A genome scale network reconstruction of Synechocystis sp. PCC6803 was also involved in studying epistatic maps within metabolic networks to elucidate that path of evolutionary adaptation is likely to be path dependent due to strong effect of the environment on epistasis (Joshi & Prasad, 2014). *Halobacter salinarium* can store energy using a high potassium gradients. Its genome-scale metabolic network was used to investigate aerobic essential amino acid degradation, energy generation, nutrient utilization, and biomass production (Gonzalez et al., 2008). A genome scale model of *C. reinhardtii* (Chang et al., 2011), an algae, was used to see the effects of co-culture with yeast (Gomez et al., 2014). It should be noted that research in modeling photosynthetic metabolism is still in its infancy compared to *E. coli* and yeast.

The applications listed here are by no means the only ones. The actual list may require a separate article of its own. However, it should be noted that as more genome scale models are published, the level of details within the reconstructions will also increase, resulting in even more applicability of genome scale metabolic models.

8.   THESIS OUTLINE

Both FBA and dynamic FBA depend upon a detailed understanding of the underlying metabolic network. As knowledge of this network grows, the models become better and better. An interesting question that sometimes arises is whether FBA is an idea, a hypothesis or a theory. In my view, our understanding of metabolism has progressed sufficiently that the underlying description of the metabolic network can be regarded as part of the biological theory of metabolism. FBA and dynamic FBA provide one method of estimating the internal fluxes of metabolites, based on the hypothesis of constrained optimality of some cellular objective.

However, there is significant evidence to suggest that this hypothesis may be true, at least under some conditions. In our view FBA is a self-consistent theoretical project that is a theory in the making: a theory of optimality in metabolism.

The metabolic network can be represented as a directed graph of nodes representing metabolites and directed edges representing reactions, along with an additional layer of complexity provided by the enzymes and transporters that participate in each reaction. In Chapter 2, we study the latter aspect of network structure, in particular the relationship between the number of enzymes and the number of reactions they participate in. We find that the distribution of the number of reactions an enzyme participates in, the enzyme-reaction distribution, is surprisingly similar across ten species. In six out of these ten species the distribution can be described by a power-law with statistical significance. We use flux balance analysis (FBA) to study the effect of the enzyme-reaction distribution on the robustness of two microorganisms, *E. coli* and *Synechocystis,* and based on a detailed study of gene deletions in both organisms we show that the form of this distribution plays an important and hitherto unappreciated role in robustness. Despite the similarity of the overall distribution of reactions among enzymes, we also uncover many differences in the specific details of this distribution between the two microorganisms, arising from their specific environmental niches. In particular, we discover that multifunctional enzymes play a major role in conferring lethality to many loss-of-function mutations in the photoautotrophic metabolism. Our analysis suggests that multifunctional enzymes may be contributing some unknown fitness benefits to the organism, by virtue of being multifunctional, that offsets their negative role in loss-of-function mutations, and that this may be especially important in photosynthetic metabolism. The similarity of the enzyme-reaction distribution between the ten species studied also strongly

suggests the existence of a shared design principle or evolutionary process (Joshi & Prasad, Structural and role of enzyme-reaction association in microbial metabolism. In preparation).

When the effect of the state of one gene is dependent on the state of another gene in more than an additive or neutral way, the phenomenon is termed epistasis. In particular, positive epistasis signifies that the impact of the double deletion is less severe than the neutral combination, while negative epistasis signifies that the double deletion is more severe. Epistatic interactions between genes affect the fitness landscape of an organism in its environment and are believed to be important for the evolution of sex and the evolution of recombination. In Chapter 3, we use large-scale computational metabolic models of microorganisms to study epistasis computationally using Flux Balance Analysis (FBA). We ask what the effects of the environment are on epistatic interactions between metabolic genes in three different microorganisms: the model bacterium *E. coli*, the cyanobacteria *Synechocystis* PCC6803 and the model green algae, *C. reinhardtii*. Prior studies had shown that in standard laboratory conditions epistatic interactions between metabolic genes are dominated by positive epistasis. We show here that epistatic interactions depend strongly upon environmental conditions, i.e. the source of carbon, the Carbon/Oxygen ratio, and for photosynthetic organisms, the intensity of light. By a comparative analysis of flux distributions under different conditions, we show that whether epistatic interactions are positive or negative depends upon the topology of the carbon flow between the reactions affected by the pair of genes being considered. Thus, complex metabolic networks can show epistasis even without explicit interactions between genes, and the direction and scale of epistasis are dependent on network flows. Our results suggest that the path of evolutionary adaptation in fluctuating environments is likely to be very history dependent because of the strong effect of the environment on epistasis (Joshi & Prasad, 2014).

Cyanobacteria are prokaryotes capable of performing oxygenic photosynthesis, making them attractive candidates for genetic engineering towards production of biofuel, pharmaceuticals, nutraceuticals, and other commercially important chemicals. In Chapter 4, we present and analyze a genome scale metabolic network reconstruction (iSynCJ816) of *Synechocystis sp*. PCC6803, the most widely studied cyanobacterium. This reconstruction consists of 816 genes, 1045 reactions, and 929 non-unique metabolites spanning across 7 compartments (extracellular, cytosol, cytosolic membrane, carboxysome, periplasm, thylakoid and thylakoid membrane). This updated model builds from previously published models, and develops them further by integrating an unconstrained photo-respiratory reaction mechanism. The model also includes various molecular mechanisms of electron transfer in three most important protein complexes of photosynthesis (photosystem I, photosystem II, and cytochrome b_6/f complex). We used Flux Balance Analysis (FBA) to calculate the flux distribution within iSynCJ816 and compare in silico predictions with values obtained by previous *in vivo* metabolic flux analyses in *Synechocystis sp.* PCC6803. We performed gene deletion analysis and qualitatively compared gene deletions of 167 genes with experimental studies to find an accuracy rate of ~80%. We used the model to estimate maximum theoretical yield of products using each metabolite as a precursor, as well as the feasibility of engineering Synechocystis to increase CO2 fixation. The model predicts that it may be possible to increase CO2 fixation by up to 35% from wild type levels (Joshi, Peebles, & Prasad, Modeling and analysis of bioproduct formation in Synechocystis sp. PCC6803 using a new genome-scale metabolic network reconstruction. Submitted).

To construct strains that not only grow optimally but also are efficient at the technology they are constructed for, it is important to understand intracellular metabolic regulation in these microorganisms in their full dynamic complexity. Photosynthetic organisms have an inherent dynamic complexity because in the natural habitat there are days and nights, seasons and the consequent changes in light intensity and composition. A variety of sustainable and green applications of metabolic engineering of cyanobacteria are ultimately possible only when translatable to utilization of the energy given out by the sun. In Chapter 5, we apply a direct method of dynamic flux balance analysis which involves imbedding a Linear Programming problem within a set of kinetic equations, and using hierarchical or "lexicographic" optimization to study diurnal objective functions and lexicographic priority of substrate exchange, biomass growth, ATP synthase, and ATP maintenance in *Synechocystis sp.* PCC6803.

CHAPTER 2.   STRUCTURE AND ROLE OF ENZYME-REACTION ASSOCIATION IN MICROBIAL METABOLISM

1.   SYNOPSIS

The metabolic network can be represented as a directed graph of nodes representing metabolites and directed edges representing reactions, along with an additional layer of complexity provided by the enzymes and transporters that participate in each reaction. Here we study the latter aspect of network structure, in particular the relationship between the number of enzymes and the number of reactions they participate in. We find that the distribution of the number of reactions an enzyme participates in, the enzyme-reaction distribution, is surprisingly similar across eighteen species and resembles a power law. In fifteen out of these eighteen species the power-law was found to be with statistically significant. We use Flux Balance Analysis to study the effect of the enzyme-reaction distribution on the robustness of the metabolic models of two microorganisms, *E. coli* and *Synechocystis,* and based on a detailed study of gene deletions in both models we show that the form of this distribution plays an important and hitherto unappreciated role in robustness. Despite the similarity of the overall distribution of reactions among enzymes, we also uncover many differences in the specific details of this distribution. In particular we discover that multifunctional enzymes play a major role in conferring lethality to many loss-of-function mutations in the current model of photoautotrophic metabolism. Our analysis suggests that multifunctional enzymes may be contributing some unknown fitness benefits to the organism, by virtue of being multifunctional, that offsets their negative role in loss-of-function mutations, and that this may be especially important in photosynthetic metabolism. The similarity of the enzyme-reaction distribution between the eighteen species studied also strongly suggests the existence of a shared design principle or evolutionary process.

2.   INTRODUCTION

Studies in *E. coli* report mutation rates as high as 10^{-3} per genome per generation (Kibota & Lynch, 1996; H. Lee, Popodi, Tang, & Foster, 2012; Perfeito, Fernandes, Mota, & Gordo, 2007), underlining the importance of robustness against deleterious mutations for microorganisms. Indeed, studies report that the majority of mutations appear neither beneficial nor deleterious, and deleterious mutations have mostly small fitness effects (H. Lee et al., 2012). Robustness to genetic mutations can arise because of a large variety of reasons. For example, genes could be coding for proteins that perform an inessential function, or proteins that are partly or entirely redundant because other proteins can carry out their functions. For genes that code for metabolic enzymes, network structure is an additional source of robustness, since multiple pathways exist for the synthesis of metabolites. To compensate for a gene deletion therefore, the organism can merely redistribute metabolic fluxes among surviving pathways (Segrè et al., 2002). This is easily seen by visualization of the metabolic network as a graph.

The most common representation of the metabolic network as a graph represents the metabolites as nodes, linked together by enzymatic or transport reactions that constitute the edges (Arita, 2004; Jeong, Tombor, Albert, Oltvai, & Barabási, 2000; Ravasz, Somera, Mongru, Oltvai, & Barabási, 2002). This may be called a "reaction-edge" graph (Light & Kraulis, 2004), and a gene deletion in this network can be represented by the removal of the edges corresponding to the reactions catalyzed by the enzyme that the deleted gene coded for. With the removal of an edge, the organism survives if an alternative pathway exists for the synthesis of the constituents of biomass. While in some cases, the alternative pathway may be too expensive and the organism fails to survive, the existence of an alternative pathway after removal of an edge is a topological

or a graph-theoretic property. Thus, the contribution of metabolic network topology to metabolic robustness is of great theoretical and practical significance.

Edges in the metabolic network are catalyzed by enzymes or represent specific protein mediated transport processes, and reaction-edge graphs therefore miss the contributions of these proteins to the metabolic network. Other representations of the metabolic network seek to correct this by including enzymes, in the form of "protein-centric" or "protein-vertex" graphs where the vertices are the proteins and the edges are the substrates that the proteins act on (Light & Kraulis, 2004). Alternatively, others have constructed "two-color" graphs where vertices represent reactions and there are two types of edges; one connects reactions that have a metabolite in common and the other is a weighted edge that represents genomic associations (Spirin, Gelfand, Mironov, & Mirny, 2006). However while both these studies are useful, they use very broad measures of protein associations or genomic associations, and thereby miss some crucial properties of the network. For example, metabolites like ATP link together almost all enzymes. Such broad measures of association are unlikely to be very useful since enzymes show high specificity for specific reactions that they catalyze. In this study, we specifically concentrate on the role that these proteins play, i.e. of catalyzing reactions (and transporting metabolites) in the metabolic network.

A single enzyme, typically, does not necessarily catalyze only a single reaction (or edge) of the network. Some gene products are isozymes, which catalyze the same reaction. Other genes constitute multifunctional enzymes that constrain more than one reaction in the network (Roy, 1999). In order to quantify the distribution of these two kinds of enzymes, we define the degree of multifunctionality, (k_e) of any enzyme. This is the number of reactions catalyzed by a particular enzyme, and we will call the distribution of k_e the enzyme-reaction distribution. Note that k_e encapsulates multiple kinds of enzyme promiscuity, including substrate promiscuity, i.e. being

28

able to perform the same function on multiple substrates and catalytic promiscuity, i.e. possessing multiple catalytic domains (Cheng et al., 2012).

Multifunctional enzymes have been studied for a long time, and it is believed that they may have played an important role in the evolution of life on the planet. It has been proposed on the basis of evolutionary arguments that precursor enzymes that catalyzed biochemical reactions when life emerged on earth are likely to have been multifunctional enzymes with broad substrate specificity, a hypothesis that has been called the patchwork hypothesis (Fani & Fondi, 2009; Jensen, 1976). In support of this argument, recent work has found that in E. coli, specialist enzymes, i.e. enzymes that catalyze a single reaction, are more likely to be essential, carry greater flux and are regulated to a greater extent than generalist enzymes that catalyze more than one reaction (Nam et al., 2012). Multifunctional enzymes that catalyze a sequence of reactions have definite advantages due to substrate channeling. However, many multifunctional enzymes are involved in reactions that are not sequential, and their persistence remains to be explained. Furthermore, we do not understand if the distribution of the degree of multifunctionality (which we will call the enzyme-reaction distribution) plays a physiological role in metabolic networks.

Since the enzyme-reaction, distribution forms a part of the structure or topology of metabolic networks and given the role of multifunctional enzymes in the debates over the evolution of life on earth, we asked ourselves how the degree of multifunctionality, k_e, was distributed in organisms. In order to calculate this we depended on the genome scale metabolic reconstructions being undertaken by many groups for the last decade (Oberhardt, Palsson, & Papin, 2009). We downloaded 21 publicly available models corresponding to 18 different species including both eukaryotes and prokaryotes. These models include details about how specific genes that code for enzymes and transporters map on to specific reactions in the model. Using the gene as a proxy for

the enzyme, we extracted this information and studied the enzyme-reaction distribution for each organism. For a more detailed analysis on the possible role of this distribution on fitness and other properties of the network, we focused on two organisms. One of these two organisms was *E. coli*, due to its very well studied and relatively comprehensive metabolic network reconstruction. The other organism we chose to look at was the model cyanobacterium *Synechocystis sp.* PCC6803, as a representative of the metabolic niche of photosynthetic microorganisms.

We used a recently constructed genome scale model, iJN678, of *Synechocystis sp. PCC6803* (Juan Nogales, Gudmundsson, Knight, Palsson, & Thiele, 2012) as well as a comprehensive metabolic model, iAF1260, of *E. Coli* MG1655 (Covert, Knight, Reed, Herrgard, & Palsson, 2004), and used these genome-scale models to calculate the distribution of k_e, or the enzyme-reaction distribution. In order to understand the role played by multifunctional enzymes, we used gene deletion analysis carried out using Flux Balance Analysis (FBA) (Orth et al., 2010). For the metabolic network, FBA can predict flux redistributions on single gene deletions with good accuracy (Reed & Palsson, 2004; Segrè, Deluna, Church, & Kishony, 2005; Segrè et al., 2002), with genome-scale *E. coli* models predicting gene lethality for example with an error of only about 8% (Covert et al., 2004).

We used large-scale metabolic models to ask whether there were any common patterns, or significant differences, in the structure of enzyme-reaction associations in both organisms. We used FBA to ask whether the observed enzyme-reaction distribution, which was similar for both organisms, could be contributing a fitness benefit to the organism. Finally, we used FBA to compare gene deletions between *Synechocystis* and *E. coli*.

3.   MATERIALS AND METHODS

3.1.   MODEL PREPARATION

For our analyses of enzyme-reaction distributions, we chose 21 genome scale reconstructions of 18 organisms. The organism names, model names, references and links to the SBML file of all eleven models can be found in Table B1 in *Supporting Information*. For FBA of *E. coli* and *Synechocystis sp.*, we also selected a growth condition, under which the organism was most studied. Growth conditions included in the analyses were autotrophic, heterotrophic, mixotrophic, and aerobic. Detailed analysis of network structure differences were performed only for *E. coli* and *Synechocystis*.

3.2.   ASSIGNING SUBSYSTEMS TO GENES AND COMPLEXES

For deeper analysis of the E. coli and Synechocystis, we made broad subsystems based on KEGG pathway analysis (M Kanehisa & Goto, 2000; Minoru Kanehisa, Goto, Sato, Furumichi, & Tanabe, 2012); e.g. Oxidative phosphorylation, Photosynthesis, Carbon fixation pathways, Methane metabolism, Nitrogen metabolism, and Sulfur metabolism were considered as Energy metabolism. Enzymes were assigned subsystems based on the reactions catalyzed by them. If the enzyme catalyzed reactions belonging to more than one subsystem, it was considered to be a part of all those subsystems. This was done to recognize that enzyme complexes can be highly multifunctional and might catalyze reactions that are far apart in the metabolic network and belong to completely different coarse-grained subsystems. A detailed account, of how the actual subsystems were distributed among the coarse grained subsystem for both organisms is presented in *Table S2* in *Supporting Information.* Note that since metabolic models usually report the genes associated with each reaction, the enzymes in our data are labeled by their gene names.

3.3. GENE ASSOCIATION WITH REACTIONS AND EFFECTIVE GENE DELETION OR SINGLE ENZYME DELETION

Each metabolic reconstruction contains a matrix of gene-reaction associations. This matrix contains the information about the reactions catalyzed by an enzyme which either wholly, or partially, is coded for in a gene. Many possible associations between a reaction, the involved enzymes and the genes that code for them exist, such as:

(i)   Two or more proteins are required to make a single enzyme. Each protein will be coded by a different gene, and the genes share an "AND" relationship.

(ii)   Two or more enzymes, each coded for by a different gene, catalyze exactly the same set of reactions. Here the genes share an "OR" relationship.

In order to distinguish between the role of isozymes and multifunctional enzymes we pick only one of any set of genes in either an "OR" or an "AND" relationship with each other. This ensures that all multi-subunit enzymes are treated as a single unit, and isozymes are treated separately. The list of genes remaining after those coding for isozymes have been removed we call the unique gene list, and the list of enzymes corresponding to it the unique enzymes. Only enzymes that show an exact overlap of reactions catalyzed are treated as isozymes. Enzymes that show partial overlap of reactions constrained are not treated as isozymes and are retained in the unique gene list. When a gene (or an enzyme) is deleted, we delete every reaction that it can catalyze except those catalyzed by another enzyme. Such a strategy ensures the selection of every reaction that has a gene associated to it.

3.4. POWER-LAW ANALYSIS

Power law analysis was carried out by two methods. The first is a linear fit to the log-log plot using built-in Matlab fit (`polyfit`; Figure B1). However except for a few generic transporter proteins, proteins catalyze at most a few tens of reactions. With this small decadal span of the data, a linear fit on a log-log plot is very likely to give a false positive for a power law. We therefore analyzed the data using the Maximum Likelihood Estimators of Clauset et. al. (Clauset, Shalizi, & Newman, 2009) and using the Matlab script files made available by them. For the power law described by $p(x) \sim x$ valid for some x x_{min} we used the MLE estimate and a goodness of fit measure to estimate the parameters , x_{min}. The plausibility of the power-law fit was then estimated by the publically available code from Ref. (Clauset et al., 2009) which samples synthetic data sets from the true power-law distribution multiple times and measures the Kolmogorov-Smirnov (KS) statistic for the synthetic data with respect to its best power-law fit. The fraction of times the KS statistic is larger for the synthetic data than the KS statistic for the empirical data is the p-value that estimates the probability that the empirical data comes from the fitted power law. Note that a higher p-value represents support for the hypothesis. We follow the commonly used benchmark of assuming insufficient support of the power-law hypothesis when $p \leq 0.1$. More detailed descriptions are provided in the reference cited.

3.5. DISTRIBUTION ANALYSIS

In addition to power-law analysis, we also used (i) Two sample Kolmogorov-Smirnov test, and (ii) Mann-Whitney U-test (a.k.a. Wilcoxon rank-sum test).

Kolmogorov-Smirnov test (two sample) is a non-parametric test used to test whether two underlying one-dimensional probability distributions differ. The KS statistic quantifies the

distance between the distribution functions of two samples. It should be noted that for power-law analysis, we used KS statistic between corresponding power-law fit and a given organismal enzyme-reaction distribution; however, here we used calculated KS statistic between any given two organismal enzyme-reaction distribution. Here, $p \leq 0.05$ suggests that null hypothesis can be rejected. The null hypothesis, here, is that data from both organism belong to the same distribution. The statistic was calculated using the Matlab R2014b built-in function, "`kstest2`".

Mann-Whitney U-test (a.k.a. Wilcoxon rank-sum test), similar to KS test, is a non-parametric test used to test whether two organismal data come from the same distribution. Here, $p \leq 0.05$ suggests that null hypothesis can be rejected. The statistic was calculated using the Matlab R2014b built-in function, "`ranksum`".

3.6.   FLUX BALANCE ANALYSIS (FBA)

Flux Balance Analysis (FBA) is a mathematical framework used to calculate the flow of the metabolites through the metabolic network at steady state (Orth et al., 2010). FBA was performed using COBRA Toolbox (Schellenberger et al., 2011) with Gurobi 4.6.1 on MATLAB R2011b. Briefly, each available metabolic reconstruction that we make use of involves the construction of a *(M-by-N)* stoichiometric matrix, *S* for the metabolic reactions and a table of gene associations for each reaction (here *M* is the number of metabolites and *N* is the number of reactions). S_{ij} represents the stoichiometric coefficient for i^{th} metabolite in j^{th} reaction. In steady state a solution to the flux distribution in the organism is found under the condition that the growth rate reaction is maximized, making it a linear programming problem. Certain constraints are then imposed to find a unique solution to the under-determined system. The most important constraint arises from the reaction network at steady state:

$$\sum_{i=1}^{N} S_{ij}v_j = 0 \tag{2.1}$$

Here, v_j is a vector of reaction flux. The growth rate reaction is described as:

$$\sum_{i=1}^{M} c_i m_i \rightarrow \mu \tag{2.2}$$

Two other types of constraints arise:

1) Constraints on uptake and secretion rates of metabolites.

2) Limits on the upper and lower bounds of each reaction flux, i.e.

$$\alpha_j \leq v_j \leq \beta_j \tag{2.3}$$

α_j and β_j are respectively the lower and upper limits placed on each reaction flux, v_j. Reversible reactions can take either negative or positive values of fluxes, while irreversible reactions are constrained to take only positive values. Further, if any reactions were turned off, inactivated or deleted, the flux through the reaction is set to zero, i.e. $v_j = 0$.

3.7. *SIMULATION OF GROWTH CONDITIONS IN VARIOUS ORGANISMS*

Aerobic growth of *E. coli* (model name: iAF1260) was simulated by applying following constraints: (i) maximum glucose uptake rate (EX_glc(e)) was set to 8 mmoles/gDW/h (Feist et al., 2007); (ii) maximum oxygen uptake rate (EX_o2(e)) was left unconstrained (Feist et al., 2007); (iii) all the other sources of carbon uptake rates were set to zero; and (iv) the constraints applied on each of the reactions were same as reported in original article where the model was published. Heterotrophic growth of *Synechocystis* sp. PCC6803 (model name: iJN678) was simulated by applying following constraints: (i) maximum glucose uptake rate (EX_glc(e)) was set to 0.85

mmoles/gDW/h (C. Yang, Hua, & Shimizu, 2002); (ii) maximum oxygen uptake rate was left unconstrained; (iii) all the other sources of carbon uptake rates and photon uptake (incident on the organism) rate were set to zero; and (iv) the constraints applied on each of the reactions were same as reported in original article where the model was published. Autotrophic growth of *Synechocystis* sp. PCC6803 was simulated by applying following constraints: (i) maximum carbon dioxide uptake was set to 3.7 mmoles/gDW/h (Shastri & Morgan, 2005); (ii) minimum photon uptake uptake rate corresponding to maximum growth rate was calculated and subsequently set to 54.0948 mmoles/gDW/h; and (iii) no excretion of carbon dioxide was allowed through carbonate exchange for autotrophic growth (EX_hco3(e)) (Juan Nogales et al., 2012); (iv) all the other sources of carbon uptake rates were set to zero; and (v) all the other constraints were used from the original article where the model was published. The default constraints on flux of a reversible reaction was [-1000, 1000], and of an irreversible reaction was [0,1000], unless specified here or in the original article where these models are published. These constraints are a norm in the field and have been used in numerous FBA studies (Feist et al., 2007; Juan Nogales et al., 2012; Shastri & Morgan, 2005). All the other models used in this study were used only to extract the information about genes and reactions. The list of papers where the model was published, and the link to the SBML file of the model is provided in Table B1.

3.8. *SIMULATION OF ENZYME DELETIONS, AND ESSENTIAL ENZYMES*

To simulate deletions of enzyme complexes FBA was used. Enzyme deletions were simulated by setting the flux of every reaction catalyzed by that enzyme to zero. Essential genes are genes that lead to a growth rate of 0.1 or less of the wild-type value. This is normally regarded

as a lethal deletion. The percentage of essential effective genes (or enzymes) is the percentage of lethal genes to total number of genes in the group.

3.9. LETHAL COMPARATIVE MAPPING ANALYSIS (LCMA) AMONGST DIFFERENT MODELS

The following steps describe this technique:

Step 1: Find the E.C. numbers of reactions constrained by a lethal gene for one organism.

Step 2: Find the same E.C. numbers for the other organism.

Step 3: Find the reactions that are associated with these E.C. numbers in the other organism.

Step 4: Find all the genes involved, not necessarily constraining, in those reactions.

Step 5: Perform a single gene deletion analysis, on all those genes found.

Step 6: Classify these genes as being lethal or non-lethal.

Step 7: Analyze the flux distributions to understand the reasons for escape in case the gene deletions are non-lethal.

For example, deletion of sll0300 constrains riboflavin synthase in Synechocystis proves lethal and has the reaction EC number 2.5.1.9; this generates the gene b0415 in E. coli, also is lethal. Similarly, deletion of sll0290 constrains diphosphate/monophosphate kinase in Synechocystis proves lethal and has the reaction EC numbers 2.7.4.1; this generates the gene b2501 in E. coli, but is non-lethal.

4.   RESULTS

4.1.   THE NUMBER OF REACTIONS CATALYZED BY AN ENZYME FALLS OFF AS A POWER-LAW

To obtain a global picture of the distribution of enzyme-reaction associations, enzymes were classified according to the number of reactions they constrain. The resulting histogram looks linear on a log-log scale, and simple fitting yielded a power law relationship:

$$n(k_e) \sim k_e^{-\alpha} \tag{2.4}$$

where $n(k_e)$ is the number of enzymes with degree of multifunctionality k_e (Fig B1). For *E. coli* the exponent α was about 1.8 while for *Synechocystis* it was about 2.2, while for *Chlamydomonas reinhardtii* (Chang et al., 2011) it was 1.3 and for *Saccharomyces cerevisiae* (Duarte, Herrgard, & Palsson, 2004) it was about 1.9. Multicellular organisms showed a similar behavior, with the exponent ranging from 1.4 to 1.7 for *Homo sapiens*, *Mus musculus* and *Arabidopsis thaliana*. In all the cases for the simple linear fits, the R^2-values were close to 1, except *Arabidopsis thaliana* ($R^2 = 0.69$) and *Synechocystis sp.* PCC6803 model: iSyn816CJ ($R^2 = 0.83$) and the exponent lay between 1.5 and 2.5).

To test whether this power law was statistically plausible we used the Maximum Likelihood method of Clauset et. al. (Clauset et al., 2009) to both estimate the exponent and the goodness of fit. We found that the gene reaction association of *E. coli* was fit very well with an exponent of either 2.5 or 2.6, x_{min} of 2 and a p-value of 0.25 or 0.73 depending on which model we chose (Table B1 and Figure 2.1). Note that when using this method higher p-values mean stronger support for the power-law hypothesis, with statistical significance being normally associated with $p \geq 0.1$. Of the 15 models that we analyzed for prokaryotes, only four of them did

not show a statistically significant fit. These were *Bacillus subtilis*, the iSyn816CJ model of *Synechocystis*, as well as *Yersinia pestis* and *Mycobacterium tuberculosis*. The exponent α as estimated by this method was systematically higher than the simple fits and ranged from 2.5 to 3.5.
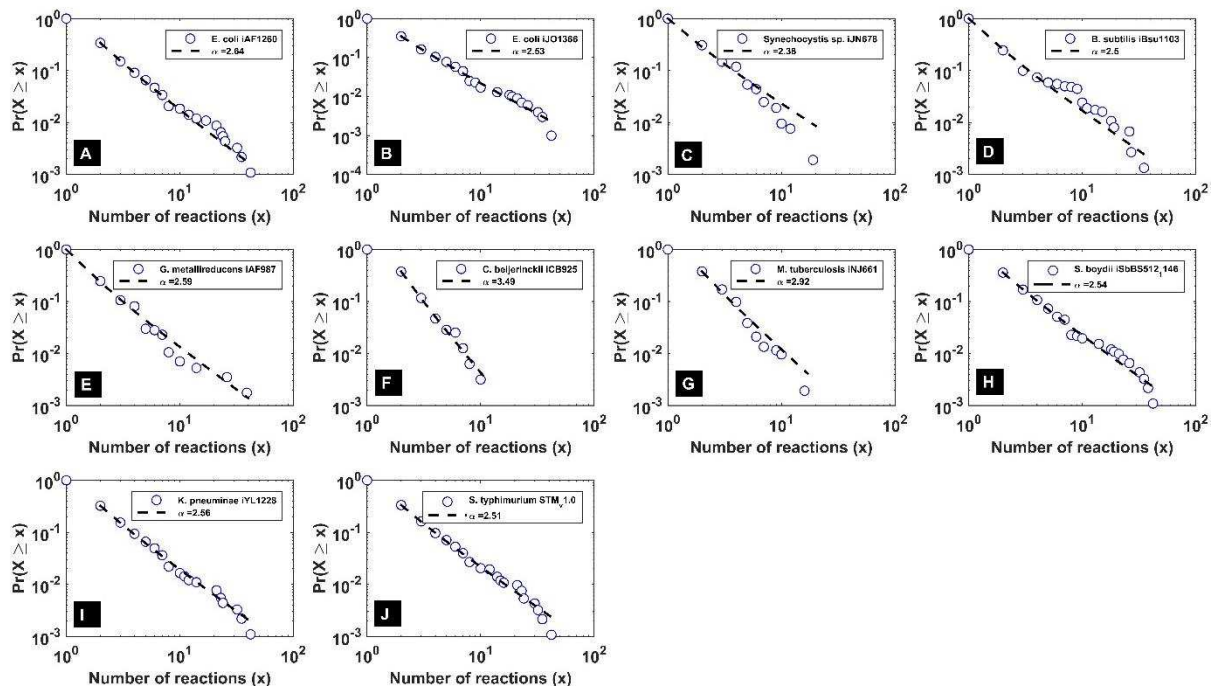
FIGURE 2.1: THE POWER-LAW FIT USING MAXIMUM LIKELIHOOD METHODS FOR TEN MODELS (NINE PROKARYOTIC SPECIES).
The x-axis is the number of reactions that each enzyme constrains. The y-axis in this case is the cumulative probability distribution of the number of enzymes. The black line represents the power law fit of Clauset et al. [23] for each organism. The cumulative probability distribution is more robust against data variations in the tail and is thus a better object to fit. The details of the power-law fits are stated in the text. (A) *Escherichia coli* iAF1260, (B) *Escherichia coli* iJO1366, (C) *Synechocystis sp*. iJN678, (D) *Bacillus subtilis* iBSu1103, (E) *Geobacter metallireducens* iAF987, (F) *Clostridium beijerinckii* iCB925, (G) *Mycobatcerium tuberculosis* iNJ661, (H) *Shigella boydii* iSbBS512_1146, (I) *Klebsiella pneumoniae* iYL1228, and (J) *Salmonella typhimurium* STM_v1.0

Analysis of the distribution of the five eukaryotes that we studied revealed a similar picture (Figure 2.2). In this case, all the models were statistically significant except for the smaller of the two models for *Saccharomyces cerevisiae* that we analyzed. This may just be because the larger model, which showed statistical significance for a power-law relationship, had improved coverage of the genome. The exponents for all the organisms were very similar too, ranging from about 2.4 to 2.7. Details of all the fits are presented in Table 2.1.

Based on the results of our analysis of these eighteen different species, it is therefore possible to hypothesize that the distribution of the degree of multifunctionality of enzymes in the metabolic network is similar across species and kingdoms ($\alpha \approx 2.5$, 3 exceptions) (Table 2.1), and in many species may be well described by a power-law.
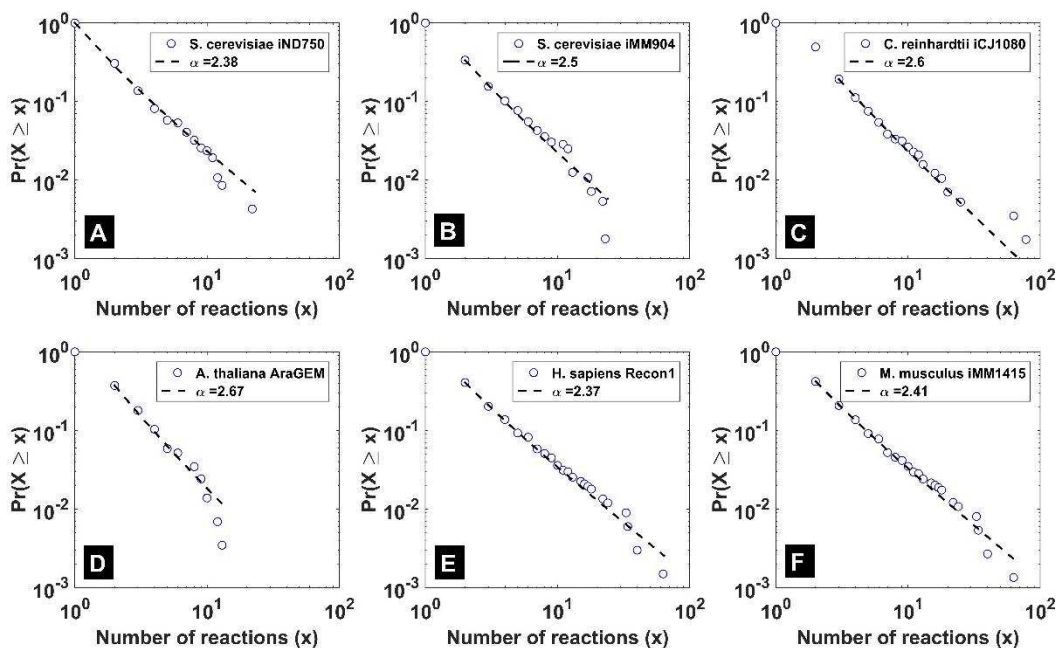
FIGURE 2.2: THE POWER-LAW FIT USING MAXIMUM LIKELIHOOD METHODS FOR SIX MODELS (FIVE EUKARYOTIC SPECIES).
The x-axis is the number of reactions that each enzyme constrains. The y-axis in this case is the cumulative probability distribution of the number of enzymes. The black line represents the power law fit of Clauset et al. [23] for each organism. The cumulative probability distribution is more robust against data variations in the tail and is thus a better object to fit. The details of the power-law fits are stated in the text. (A) *Saccharomyces cerevisiae* iND750, (B) *Saccharomyces cerevisiae* iMM904, (C) *Chlamydomonas reinhardtii* iRC1080, (D) *Arabidopsis thaliana* AraGEM, (E) *Homo sapiens* Recon1, (F) *Mus musculus* iMM1415.

Note that a mere random assignment of reactions to enzymes will produce a multinomial distribution and cannot produce a power law distribution. However, it should be noted that there has been debates on existence of power laws with low decadal spans. Therefore, here we also used non-parametric tests described in section 2.5 to compare the actual distribution amongst different models, as presented in Figure 2.3. We tested each pair of models for similarity of their enzyme-reaction distributions. *Y. pestis* (iAN818m) was similar to thirteen models (highest) in the study, and the distributions for *A. thaliana* (AraGEM) and *C. reinhardtii* (iCJ1080) did not match well

with any other model in the study. Interestingly, *Y. pestis* and *Shigella* genus models were the only

two pathogenic, facultative anaerobic, bacteria that compared well with eukaryotic, *M. musculus*

and *H. sapiens*, models. We also found that different models of the same organism did not give

the same results when compared with different organisms. For example, *E. coli* iAF1260 matched

12 models and iJO1366 matched 10 models while *S. cerevisiae* iND750 matched 4 models and

iMM904 matched 11 models. It should also be noted that some of the models were developed

using a different organismal model as a template. *Shigella* genus models were developed using

each other, and therefore, have excellent comparability ($p < 0.05$) with respect to each other.

Similarly, we also know that *M. musculus* (iMM1415) was developed from the model of *H. sapiens*

(Recon 1), and therefore, agree with each other.
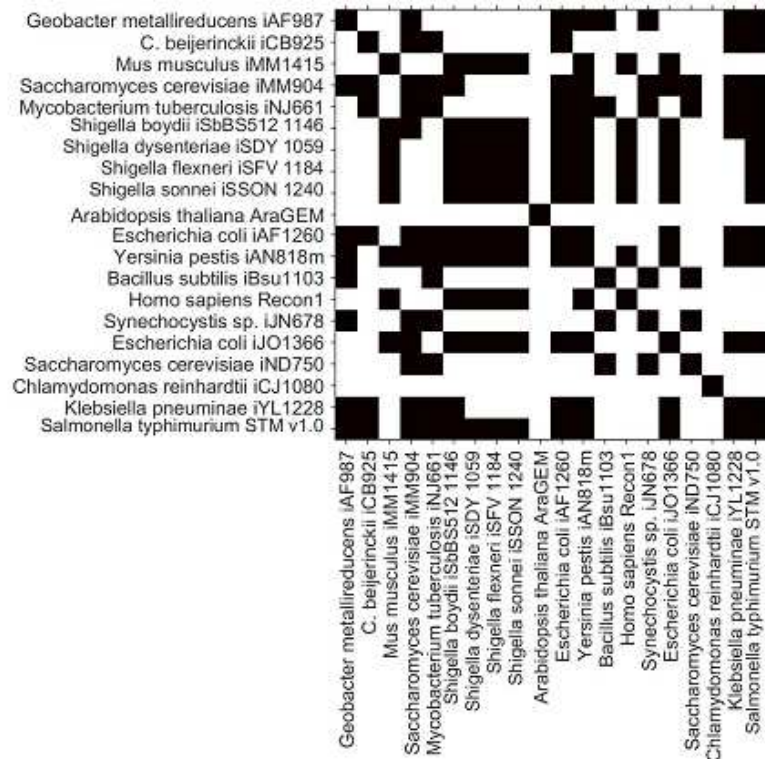
FIGURE 2.3. COMPARISON BETWEEN ANY TWO MODELS USING TWO-SAMPLE
KOLMOGOROV-SMIRNOV TEST.
The x-axis and y-axis represent the organism and model name. The colors represent
whether the null hypothesis could be rejected. Black represents that the null hypothesis
could not rejected ($p > 0.05$), and white represents that the null hypothesis can be rejected
($p < 0.05$), for comparison between the given two models.

The data suggests that there is a remarkable quantitative similarity between the enzyme-reaction distributions between these 18 organisms. This analysis thus begs the question as to whether deeper principles exist behind this similarity. In order to answer this question, we concentrated on models of two organisms, *E. coli* and *Synechocystis*. The choice of these two organisms was in part because they are both very well studied microbes, with specific metabolic differences, and are important targets for metabolic engineering. They also show very stark differences in their behavior under gene deletions (Juan Nogales et al., 2012; Orth et al., 2011). We were interested in seeing whether any of these differences (discussed below) were due to differences in the enzyme-reaction association between the two models. Note that since metabolic reconstructions are still incomplete, our conclusions regarding any particular organism are going to be subject to these limitations. Our analysis thus should be seen as an analysis of the properties of the two models that give some insight into the physiological role of the enzyme-reaction association.

TABLE 2.1: PARAMETERS OF THE POWER-LAW FITS OF THE ENZYME-REACTION ASSOCIATION OF THE SEVEN ADDITIONAL MODELS TESTED, USING THE MAXIMUM LIKELIHOOD METHOD OF (CLAUSET ET AL., 2009) (DETAILS IN *MATERIALS AND METHODS*).
The first column is the number of genes in the model, the second is the value of the exponent, α. The power-law was assumed to hold from a minimum value of k_e that we call x_{min}. The p-value reports the statistical significance of the fitted exponent and the x_{min} as a descriptor of the data. Note that for these fits a higher p-value is better; our criteria for significance is $p \geq 0.1$.

Organism	Genes in model	α	p-value (1000 runs)	x_{min}
BACTERIA				
Bacillus subtilis	**1103**	2.5	0.005	1
Escherichia coli	**1366**	2.53	0.238	2
Escherichia coli	**1260**	2.64	0.723	2
Klebsiella pneumoniae	**1228**	2.56	0.449	2
Salmonella typhimurium	**1270**	2.51	0.001	2
Yersinia pestis	**818**	2.25	0.019	1

Synechocystis sp.	**678**	2.53	0.25	2
Synechocystis sp.	**816**	2.03	0.001	1
Geobacter metallireducens	**987**	2.59	0.107	1
Clostridium beijerinkii	**925**	3.49	0.715	2
Mycobacterium tuberculosis	**661**	2.92	0.004	2
Shigella boydii	**1147**	2.54	0.264	2
Shigella dysenteriae	**1059**	2.48	0.829	2
Shigella flexneri	**1184**	2.59	0.281	2
Shigella sonnei	**1240**	2.60	0.439	2
EUKARYOTES	-			
Arabidopsis thaliana	**1419**	2.67	0.364	2
Mus musculus	**1375**	2.41	0.809	2
Homo sapiens	**1,496**	2.37	0.463	2
Saccharomyces cerevisiae	**905**	2.5	0.25	2
Saccharomyces cerevisiae	**750**	2.38	0.055	1

4.2. DELETION ANALYSIS SUGGESTS FITNESS BENEFITS OF MULTIFUNCTIONAL ENZYMES

One way of understanding the role of the enzyme-reaction distribution is to think of enzymes as flux controllers. Different types of enzyme-reaction distributions represent different types of enzymatic control of network fluxes. At one extreme lies perfectly distributed control, where a unique enzyme controls every reaction. For a network with essential reactions in parallel, one would expect perfectly distributed control to result from maximization of robustness against loss of enzyme function due to gene deletions. Thus, the proportion of essential enzymes and essential reactions should be the same, since a unique enzyme controls each essential reaction.
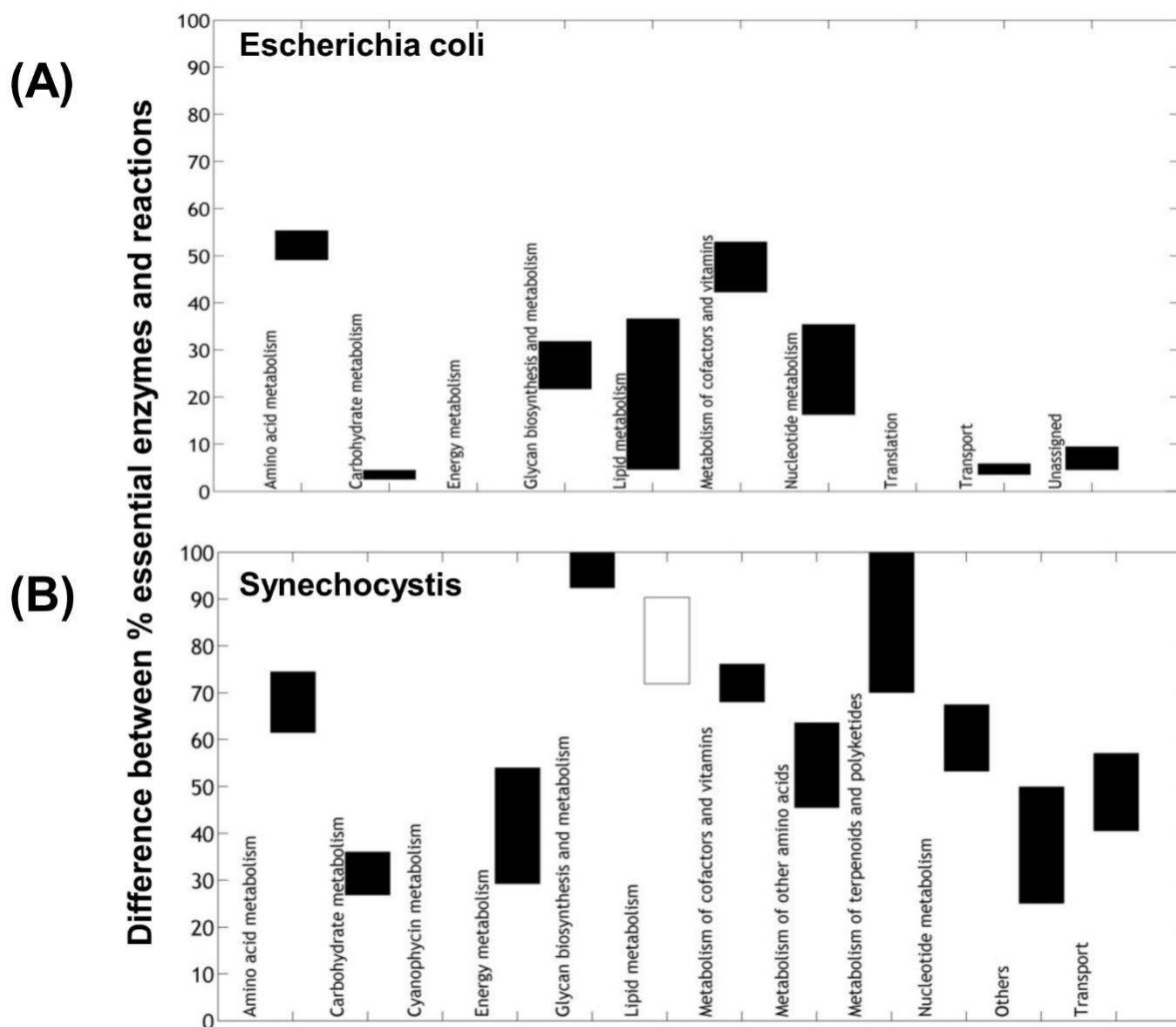
FIGURE 2.4: COMPARISON BETWEEN ESSENTIAL REACTIONS AND ESSENTIAL COMPLEXES, AND DISTRIBUTION OF COMPLEXES IN ENERGY METABOLISM.
Comparison of the proportion essential enzyme complexes among all enzyme complexes and essential reactions among all reactions by subsystem. The black bars denote a positive difference while unfilled bar represents a negative difference. **(A)** Difference in the proportion of essential genes and essential reactions by subsystem in *E. coli*, and **(B)** *Synechocystis*.

For a network with all essential reactions in a linear chain, there is no advantage to having different enzymes catalyze different reactions since every reaction is essential. Maximization of genetic robustness would result in the proportion of essential genes being less than the proportion of essential reactions. Therefore, for both these extreme cases, maximization of genetic robustness would require that the proportion of essential genes is either equal to or less than the proportion of essential reactions. Thus, the degree by which the proportion of essential genes *exceeds* that of

essential reactions is a measure of the fitness cost of the multifunctional enzymes. We can therefore argue that this is in fact a measure of the possible fitness contribution of these complex control motifs, on the assumption that their retention over evolutionary time scales must be conferring a yet to be understood benefit.

We first converted the 55 subsystems in *Synechocystis* iJN678 and 38 subsystems in *E. coli* iAF1260*,* corresponding to the KEGG (M Kanehisa & Goto, 2000; Minoru Kanehisa et al., 2012) classification, into 15 and 10 coarse-grained subsystems respectively (details in *Table S2*). We performed single gene deletions as detailed in the *Materials and Methods* section and also performed single reaction deletions by setting the flux of each reaction to be zero. We calculate the difference between the proportions of lethal reactions to the proportion of lethal genes for all subsystems of both organisms in Figure 2.4A & 2.4B.

Lipid metabolism in *Synechocystis* is the only subsystem where the proportion of lethal reactions exceeds that of lethal genes. This can be explained by the fact that it is characterized by multifunctional enzymes catalyzing linear chains of essential reactions. On the other hand, in most other subsystems, especially in Energy Metabolism and Amino Acid Metabolism (Figure 2.4A & 2.4B), the proportion of lethal genes is significantly higher than that of lethal reactions. Therefore, the persistence of multifunctional enzymes carries a fitness cost, which could be dramatically reduced by gene duplications creating a unique gene for each reaction.

The above analysis also suggests that despite the global similarity of the enzyme-reaction distribution, there are subtler differences in the metabolic network of these two organisms that may be a consequence of their specific environmental niche and evolutionary history. To investigate this further we employed the tool of gene deletions, in tandem with our classification of enzymes into specialist and multifunctional.

4.3. MULTIFUNCTIONAL ENZYMES ARE MORE ESSENTIAL IN SYNECHOCYSTIS SP. PCC6803
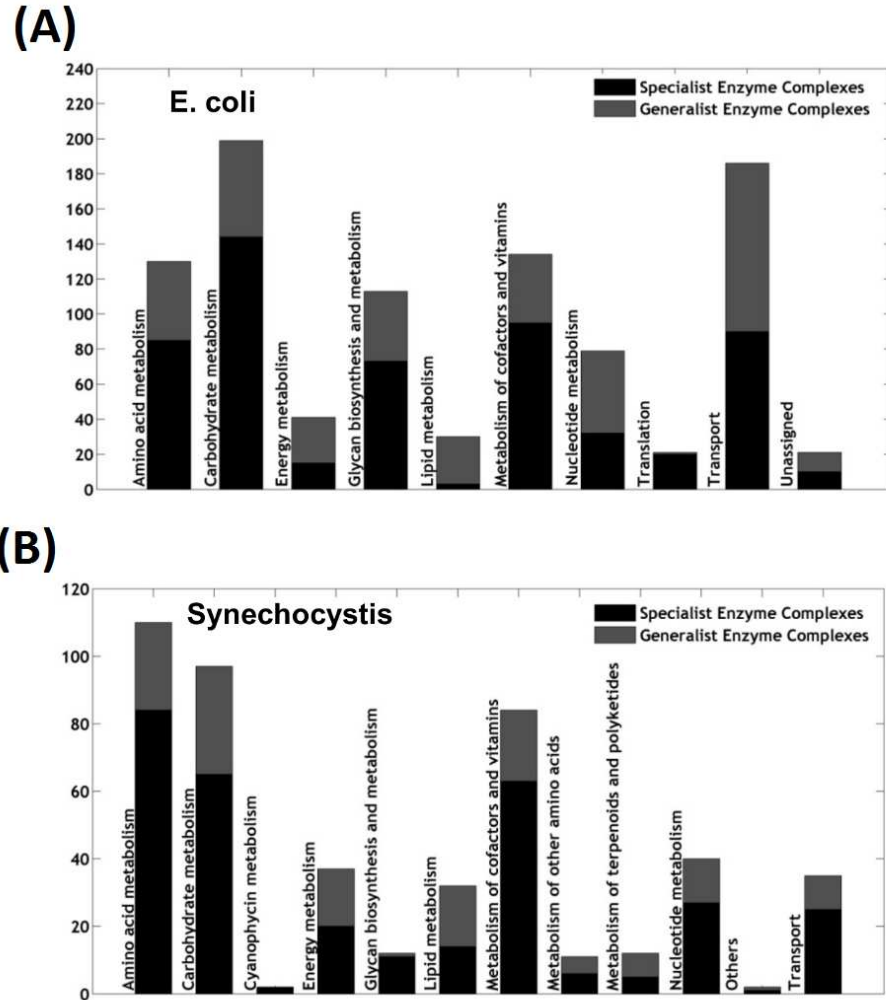
FIGURE 2.5: THE DISTRIBUTION OF ESSENTIAL/LETHAL GENE DELETIONS, AND OF SPECIALIST AND GENERALIST ENZYME COMPLEXES WITHIN METABOLIC SUBSYSTEMS. The total number of metabolic enzymes in **(A)** *E. coli,* and **(B)** *Synechocystis sp.* PCC6803 are shown distributed in each subsystem and further divided into specialist and generalist enzymes. Genes or enzymes are defined as functional protein complexes as detailed in *Materials and Methods*.

Of the 1260 genes in the *E. coli* model, 513 are multifunctional, or generalists, i.e. are associated with more than one reaction, of which 99 are lethal (19%), while 168 out of the remaining 747 specialists (22.5%) are lethal on deletion. In *Synechocystis* however, of the 677 genes in the model, there are 213 generalists of which 151 are lethal on deletion (71%) while there are 464 specialists of which 292 are lethal (63%). Thus generalists seem to be more essential in

Synechocystis. However this analysis does not do full justice to the role of enzyme complexes, which are often made of subunits coming from different genes. As described in *Materials and Methods*, we constructed a list of unique enzymes by picking only one out of any set of enzymes that catalyze exactly the same set of reactions. This removes all isozymes and combines all multi-unit enzymes into a single species. The distribution of these unique enzymes is shown in Figure 2.5A & 2.5B and the difference between the unique list and the complete list can be gleaned by a
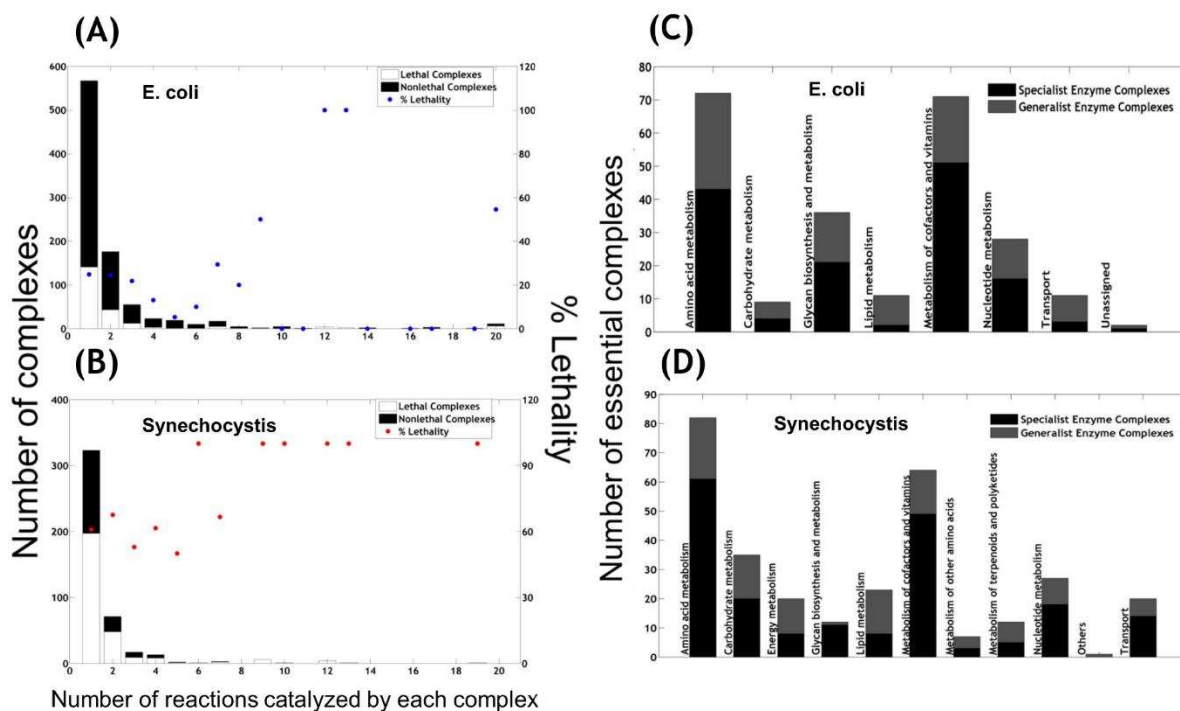
FIGURE 2.6: THE DISTRIBUTION OF ESSENTIAL COMPLEXES AND LETHALITY OF BOTH ORGANISMS.
The left panel shows genes arranged by the number of reactions they constrain (bars) as well as the percentage of lethal genes while the right panel shows the division of the essential genes in each subsystem into specialist and generalist. (A) Essential and nonessential genes are classified by the number of reactions they constrain in *E. coli* MG1655, and (B) in *Synechocystis sp.* PCC6803. (C) Essential genes broken up into specialist and generalist enzymes by subsystem in *E. coli* and (D) *Synechocystis*.

comparison with Figure B2 in *Supporting Information*. As Figure 2.5A & 2.5B and Figure B3 (*Supporting Information*) shows, the composition of the enzymes in the two organisms is similar, though *Synechocystis* has a larger percentage of specialist enzymes (67%) compared to *E. coli* (53%). In the major subsystems, the percentage of generalist complexes is somewhat higher in E.
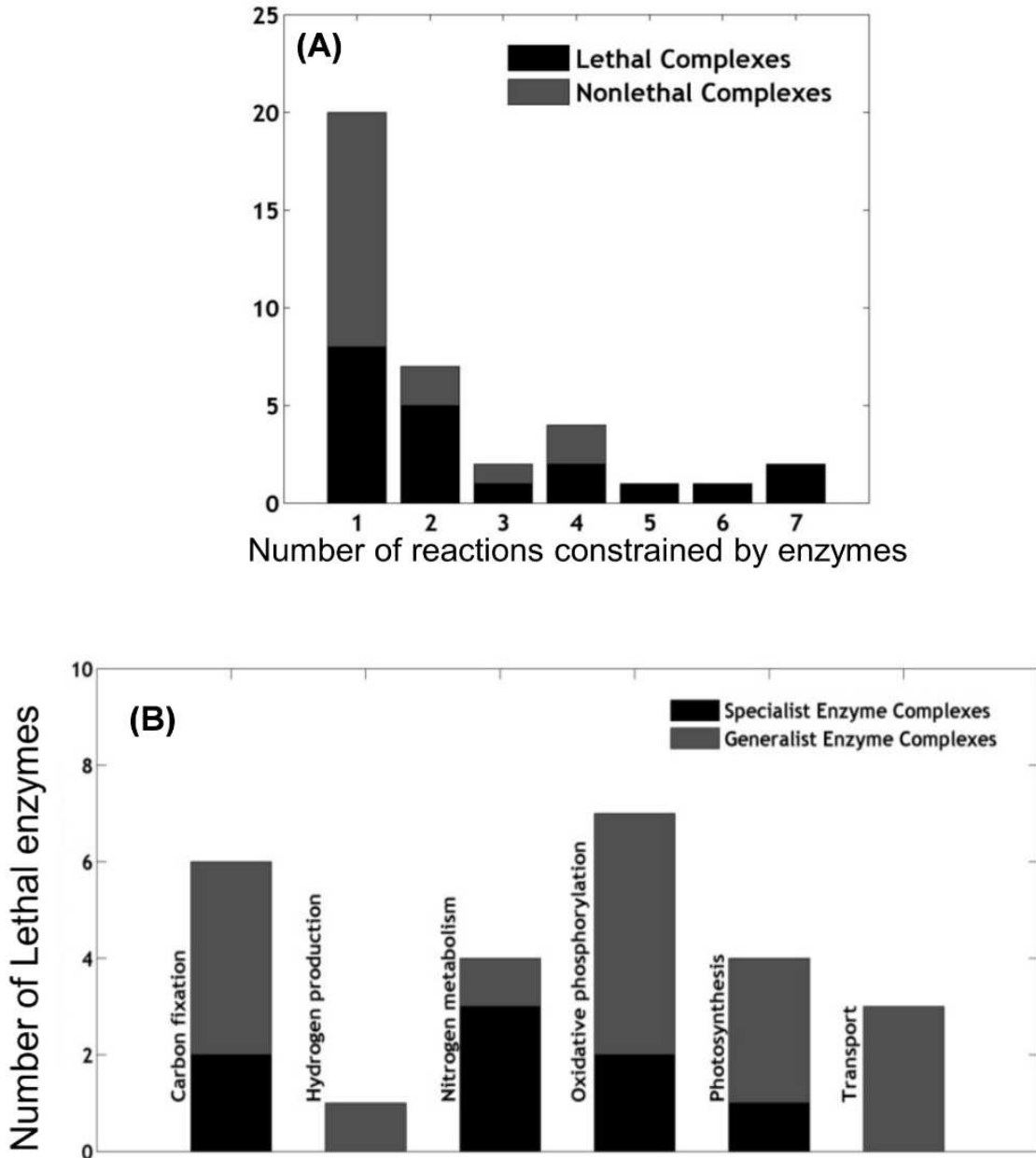
FIGURE 2.7: A SNAPSHOT OF MULTIFUNCTIONAL ENZYMES IN ENERGY METABOLISM IN SYNECHOCYSTIS.
(A) The distribution of lethal and nonlethal complexes by the number of reactions they constrain in energy metabolism in Synechocystis. **(B)** The distribution of specialist and generalist enzyme complexes in all lethal complexes in Energy Metabolism in Synechocystis, organized according to subsystem.

coli than in *Synechocystis* and in some subsystems like Lipid metabolism, the differences are quite

large.

Using FBA we then calculated the percentage of deletions of the list of unique enzyme complexes that are lethal. *E. coli* has 904 total unique enzyme species, or which 337 are generalists, and about 23% of those (79 in number) turn out to be lethal deletions. The *Synechocystis* model on the other hand has 443 unique enzyme species of which 120 are generalists out of which 68% (82) are lethal deletions. The differences are more dramatic for enzymes with more multitasking; the proportion of essential reactions in *Synechocystis* roughly increases as the number of reactions they constrain increases (Figure 2.6A), while in *E. coli* the same percentage decreases (Figure 2.6B). These trends were approximately similar even in case of heterotrophic and mixotrophic growth conditions (*Figure B3* in *Supporting Information*). Most major subsystems in *Synechocystis* stand out in comparison with *E. coli* in the proportion of essential generalist enzymes (Figure 2.6C & 2.6D). This is even more so for the subsystem that makes *Synechocystis* different -- Energy metabolism -- which has no lethal genes in *E. coli*, but 40% of its specialist and 71% of its generalist enzymes are essential in *Synechocystis*. A detailed analysis of Energy metabolism in *Synechocystis* reveals that lethality increases with the number of reactions catalyzed by each complex (Fig. 2.7A). In contrast to the rest of *Synechocystis*, most lethal genes in energy metabolism are generalist complexes rather than specialist ones (Fig. 2.7B).

These results suggest that despite the similarity of the gene-reaction associations in the two organisms, there nevertheless exist topological differences between the metabolic networks of the two organisms. In order to investigate these differences further, we decided to compare the functional impact of similar gene deletions between the two organisms.

4.4.   COMPARATIVE LETHAL DELETIONS ANALYSIS SHOWS THAT E. COLI HAS A

        GREATER DEGREE OF DISTRIBUTED CONTROL IN THE METABOLIC NETOWRK

        COMPARED WITH SYNECHOCYSTIS SP. PCC6803

The distribution of lethal complexes among generalist and specialist enzymes differs among smaller components of other subsystems too. To test if similar differences can be found even in functionally similar parts of the network, we mapped all the lethal genes from one organism to the other to look for equivalent mutations that were lethal to one but not the other (described in *Materials and Methods*). All genes do not have comparable analogs in the other organism due to metabolic differences. As shown in Table 2.2, the 187 lethal genes in *E. coli* mapped on to 130 genes in *Synechocystis*, of which 30 coded for isozymes. The remaining genes coded for 100 unique enzymes that were associated with 340 reactions. Of these 100 enzymes only 8 were nonlethal in *Synechocystis*. Thus most gene deletions that were lethal in *E. coli* were also lethal in *Synechocystis*.

TABLE 2.2: COMPARATIVE MAPPING OF LETHAL GENE DELETIONS FROM ONE ORGANISM TO ANOTHER.
Mapping lethal gene deletions from *Synechocystis sp.* PCC6803 to *E. coli* MG1655 and vice versa. Only reactions that had EC numbers were compared. Unique genes were obtained by removal of isozymes from the list, and lethality was assessed as described in the *Materials and Methods* section.

	Syn. to E. coli	E. coli to Syn.
Number of **lethal genes in source**	350	187
Total **reactions constrained** combined (**reactions with no EC #** associated to it)	406 (32)	340 (100)
Number of **mapped genes in target** (unique phenotype)	207 (112)	130 (100)
Lethal phenotypes in target	73	92
Non-Lethal phenotypes in target	39	8
Percentage of non-lethal phenotypes	~35%	8%

The 350 lethal genes in *Synechocystis* mapped on to 207 genes in *E. coli*, but only 112 of the enzymes they coded for were unique, of which as many as 39 deletions (i.e. 35%) were non-lethal. Thus *E. coli* escaped from a significant fraction of cognate gene deletions that were lethal in *Synechocystis*.

We manually analyzed all these 39 deletions by mapping them back to Synechocystis and comparing the flux distributions in the two organisms after gene deletion. We found four specific reasons (see *Table S3* in *Supporting Information* for details of all deletions, and *Table S4 & S5* in the same file for a summary) why *E. coli* was able to survive these gene deletions and *Synechocystis* was not. Two of these four reasons are related to differences in biomass composition and to environmental conditions. However, the remaining two reasons, which also accounted for the overwhelming majority (24 out of 39 cases), were related to network structure. In 12 cases, the lethal gene mutation in *Synechocystis* was a gene that constrained more than one reaction, at least one of which was essential, while the corresponding mutant in *E. coli* survived because either there were isozymes available or the gene constrained a single non-lethal reaction. For example, adenine phosphate transferase (ADPT) is constrained by b0469 in *E. coli*, while in *Synechocystis*, the gene (sll1430) constraining ADPT also constrains the reaction that makes 5-phospho-ribose 1-diphosphate (PRPP).

In another 12 cases, *E. coli* possesses multiple pathways for the synthesis of the same crucial metabolite. For example, slr1722 that codes for Inosine Monophosphate (IMP) dehydrogenase, required for xanthosine 5-phosphate (XMP) synthesis, is lethal in *Synechocystis*. However b2508 that similarly constrains IMP dehydrogenase is non-lethal in *E. coli* because there are multiple pathways that can make XMP, for example by the enzyme xanthosine pyrophosphate transferase from xanthosine, and nucleotide pyro phosphatase from xanthosine triphosphate

(XTP). Another interesting example is slr0492 which codes for O-succinylbenzoate-CoA ligase (SUCBZL), a precursor step to making demethylmenaquinol-8, and which can be regenerated by only one other enzyme, 1,4-dihydroxy-2-naphthoate phytyltransferase in *Synechocystis*. In *E. coli*, SUCBZL is produced by the gene b2260, but demethylmenaquinol-8 can be made or regenerated by many other reactions catalyzed by enzymes such as fumarate reductase, and NADH dehydrogenase. The alternative pathways for the formation or regeneration of demethylmenaquinol-8 rescue E. coli from other deletions as well in the same linear pathway. The *Synechocystis* to *E. coli* mappings for these rescues are: sll0603 to b2264, sll1127 to b2262, slr0817 to b2265. More information about all the 39 rescues, including these mutations, are presented in Table S3 of *Supporting Information*.

The comparative deletion analysis thus shows that even in those parts of the network that are functionally equivalent *E. coli* appears to possess a higher degree of distributed control among essential reactions as compared with *Synechocystis*, which appears to give a greater weight to generalist enzymes among the essential enzymes, especially in autotrophic energy metabolism. Organisms occupying different evolutionary niches therefore may be characterized by enzyme-reaction distributions that confer different properties to the metabolic network, even though the distribution itself appears to be quite similar. These differences are quite interesting from a metabolic engineering standpoint too. Organisms that have high robustness due to alternative pathways are very versatile, but may not be optimal for some kinds of genetic engineering such as flux re-routing. At the same time organisms with fewer alternative pathways are more likely to show adverse growth effects on gene deletions.

5.   DISCUSSION

There has been significant previous work and debate on the existence of power-laws in the degree distribution of metabolic networks. Most of the attention has concentrated on the degree distribution of the metabolic network, envisioned as a network of nodes representing metabolites and edges representing reactions. The central debate in the field is whether this metabolic network possesses a power-law degree distribution of the edges, and if so, what its possible implications could be. The first demonstration of this phenomenon was in a paper by Barabási's group (Jeong et al., 2000) that analyzed the metabolic information available at that time in the WIT database (Overbeek et al., 2000) for 43 different organisms and showed that scale-free networks describe metabolic networks for all organisms studied. This claim has attracted significant debate, with some papers claiming that statistical analysis using Maximum Likelihood Methods (such as those used in this paper) do not support the existence of power law behavior in any biological networks at all (Khanin & Wit, 2006), though metabolic networks were not re-analyzed in this study. Later studies however found no convincing evidence of power-laws in the degree distribution of the *E. coli* metabolic network (Clauset et al., 2009). The original finders of the power-law behavior have not yet directly addressed these claims. Nevertheless, despite critical voices (Stumpf & Porter, 2012) the consensus appears to be that many biological networks are characterized by a scale-free degree distribution, and some effort has been devoted to finding possible reasons why that could be so, and how such a network architecture could have arisen (Barabási, 2009). Both questions are as yet unclear, though many intriguing ideas have been put forward recently (Papadopoulos, Kitsak, Serrano, Boguna, & Krioukov, 2012).

While most papers have focused on representing the metabolic network as composed of metabolites attached to each other by reaction-edges, a couple of papers have noticed that this

representation leaves the enzymes out of the picture (Light & Kraulis, 2004; Spirin et al., 2006). These papers have incorporated enzymes either by making "protein-vertex" graphs or "two-color" graphs that explicitly incorporate the enzymes. However due to the emphasis on large-scale network properties like network interconnectivity and small world properties, these papers used very broad measures of enzyme connectivity, coming up with fully connected networks of enzymes through common substrates or genomic similarity. These overly broad measures of connectivity, as has been pointed out earlier, ignore the real functional roles of enzymes and metabolites, and create artificial links in the network (Lima-Mendez & van Helden, 2009) that may have no functional significance.

There also has been considerable work on the existence of power-laws of another kind in biology, the celebrated allometric scaling laws like the 3/4 power-law for metabolic rates with body mass in animals (West, Brown, & Enquist, 1997). These scaling laws are generally better established and theoretical arguments can be made for their existence. The situation of power-laws in biological networks is however more murky, as we have pointed out above.

In this paper, we report the existence of a novel regularity in metabolic networks, the distribution of multifunctionality, or the enzyme-reaction distribution. Our analysis above has shown that the enzyme-reaction association distribution in metabolic networks shows a power law relationship between the number of enzymes and the number of reactions they catalyze, or the degree of multifunctionality of enzymes, for fifteen out of eighteen species analyzed, at a statistically significant level. The three species that did not show statistical significance using maximum likelihood, nevertheless showed a very similar distribution that was significant using simple linear fits (on a log-scale). It should be noted however that there are statistical issues with fitting power-laws to data with such a small decadal span that we are fully aware of. In fact, we

would argue that the power-law fits we present should not be taken as much more than evidence of the similarity of this distribution between species. This similarity, we argue, suggests that there may exist a universal principle leading to this relationship. However the functional role of the enzyme-reaction distribution is unlikely to be related to any of the putative benefits of scale-free degree distributions such as robustness against random node removal (Barabási, 2009).

Does this empirical distribution in fact confer a fitness benefit? We conducted a detailed analysis of the impact of this power-law distribution on the robustness of two models that show very different degrees of robustness against gene deletions: *E. coli* iAF1260 and *Synechocystis* iJN678 against loss-of-function mutations using Flux Balance Analysis. We show that the impact of the distribution on this genetic robustness depends upon the distribution of essential genes between specialist and generalist complexes. If all generalist enzymes were nonlethal, it would imply that fitness benefits of multifunctional enzymes are marginal (for example minimization of the energy spent on transcription). The analysis presented above however very strongly suggests that this is not the case, and suggests that the observed enzyme-reaction distribution must carry fitness benefits.

This is even more so in energy metabolism in autotrophic microorganisms. Previous *in silico* analysis shows (Juan Nogales et al., 2012), and our computational results corroborate that *Synechocystis* appear to be characterized by a low level of genetic robustness. We show here that a significant reason for this low genetic robustness is in fact the structure of the metabolic network, in particular the weight of generalist enzyme complexes among lethal deletions. Gene duplications could have reduced the percentage of lethal gene deletions here by about half, suggesting that the generalist nature of many enzymes must be conferring a yet to be understood significant fitness benefit in photosynthesis. It is intriguing that photosynthetic organisms are also characterized by

a multiplicity of photosynthetic genomes, whether it is the high polyploidy (Griese, Lange, & Soppa, 2011) in *Synechocystis* or the multiple chloroplasts in algae and plant cells, and one could speculate that this may be related to the low robustness of photosynthesis against loss-of-function mutations. Note, however, that this result is based on the present availability of knowledge about the autotrophic metabolism. It is possible that new isozymes may be discovered in these organisms that could change these conclusions. However, the theoretical conclusions regarding the different kinds of metabolic control and their effect on fitness are robust against this new knowledge.

We do not know yet what benefits could be resulting from the observed enzyme-reaction association distribution. It is interesting to speculate whether such a power law structure arises from an optimization principle such as minimization of total energy. It is possible however, that multifunctional enzymes play a role in rapid flux balancing among different parts of a network. For example, a generalist enzyme catalyzing two reactions that are in parallel to each other, would set up a negative feedback between the two reactions, whereby a flux increase through one, due to increase in substrate availability for example, would rapidly decrease the flux through the other. This could lead to a more rapid response to changing environmental conditions than dependence upon gene regulation. Multifunctional enzymes could exist in several different network motifs with the reactions they catalyze, and different motifs could be playing different roles in controlling network flux at much faster timescales than gene regulation could provide. In this context it is intriguing that a recent paper discovered that transcriptional regulation could in fact explain only a small part of substrate-induced flux changes in *Bacillus subtilis* (Chubukov et al., 2013).

The distribution of multifunctionality we discover in eighteen different species is also very interesting from an evolutionary standpoint. As we mentioned earlier, a prominent theory of the evolution of metabolism, called the patchwork hypothesis, has held that the earliest life was largely

dominated by enzymes with a broad specificity that formed a metabolic network with a high degree of interconnection. Specialist enzymes arose from these due to gene duplications, and positive selection following mutations that increased enzymatic efficiency (Fani & Fondi, 2009). However alternative hypotheses, such as the "retrograde hypothesis" exist, that argue that early life probably involved minimal biosynthesis, and metabolic networks arose through the necessity of finding substitutes for final metabolic products (Fani & Fondi, 2009). The challenge for both of these hypotheses is to find an explanation for the global structure of the enzyme-reaction distribution that we discover in this paper.

These results is also interesting in the context of the another theory for the origin of life. This theory suggests that, prior to life, chemical reactions in some primordial soup resulted in formation of an autocatalytic set, i.e., a set of molecules and reactions such that all molecules are formed from within this set, and all catalysts are also present in the same set (S. A. Kauffman, 1986). Computer models for systems of random chemical reactions have shown that autocatalytic sets can emerge by chance with a non-zero probability, and that once an autocatalytic set emerges, it invariably grows and increases in connectivity and complexity (Jain & Krishna, 1998). Theoretical computational analysis of the emergence of autocatalytic sets in the case of power-law distributed catalysis has shown that under this assumption autocatalytic sets emerge even faster in the computational model, approaching a probability of 1 with a fairly low average percentage of catalysts and with the exponent α between 2.2 and 2.7 (Hordijk, Hasenclever, Gao, Mincheva, & Hein, 2014). It is therefore also possible that the existence of power-law appearing catalysis in the metabolic networks that we see represent signatures of prebiotic metabolism in Earth's primordial soup.

These results have implications for genome scale engineering of microorganisms too. The current level of small-scale genetic engineering suffers from the disadvantage of low efficiencies, primarily because it involves tinkering with an organism that has evolved to work under a very different set of objectives than what metabolic engineers desire. Efficient bioengineering of microorganisms for commercial purposes therefore requires large-scale genetic engineering of their metabolic networks (Esvelt & Wang, 2013). These attempts would have to keep in mind the possible functional roles played by the global structure of the enzyme-reaction association distribution.

CHAPTER 3.   EPISTATIC INTERACTIONS AMONG METABOLIC GENES DEPEND UPON ENVIRONMENTAL CONDITIONS

1. SYNOPSIS

When the effect of the state of one gene is dependent on the state of another gene in more than an additive or neutral way, the phenomenon is termed epistasis. In particular, positive epistasis signifies that the impact of the double deletion is less severe than the neutral combination, while negative epistasis signifies that the double deletion is more severe. Epistatic interactions between genes affect the fitness landscape of an organism in its environment and are believed to be important for the evolution of sex and the evolution of recombination. Here we use large-scale computational metabolic models of microorganisms to study epistasis computationally using Flux Balance Analysis (FBA). We ask what the effects of the environment are on epistatic interactions between metabolic genes in three different microorganisms: the model bacterium *E. coli*, the cyanobacteria *Synechocystis* PCC6803 and the model green algae, *C. reinhardtii*. Prior studies had shown that in standard laboratory conditions epistatic interactions between metabolic genes are dominated by positive epistasis. We show here that epistatic interactions depend strongly upon environmental conditions, i.e. the source of carbon, the Carbon/Oxygen ratio, and for photosynthetic organisms, the intensity of light. By a comparative analysis of flux distributions under different conditions, we show that whether epistatic interactions are positive or negative depends upon the topology of the carbon flow between the reactions affected by the pair of genes being considered. Thus, complex metabolic networks can show epistasis even without explicit interactions between genes, and the direction and scale of epistasis are dependent on network flows. Our results suggest that the path of

evolutionary adaptation in fluctuating environments is likely to be very history dependent because of the strong effect of the environment on epistasis.

2. INTRODUCTION

One of the central problems in biology is that of understanding the mapping between genotype and phenotype. It is now clear that a simple list of active genes do not sufficiently explain phenotype since genes interact in myriad intricate ways. The word "epistasis" has come to suggest the multiple deviations from mere additive effects displayed by genes in an organism. It was first coined by Bateson in 1909 as one genetic variant masking the effect of another (Bateson, 1909). Broadly speaking, when the effect of the state of one gene is dependent on the state of another gene in more than an additive or neutral way, the phenomenon is termed epistasis (Breen, Kemena, Vlasov, Notredame, & Kondrashov, 2012; J. a. G. M. de Visser, Cooper, & Elena, 2011). Epistatic interactions have been classified in multiple ways. For example directional or mean epistasis, also called magnitude epistasis, occurs when both mutations are either deleterious or beneficial, and may be further classified as either aggravating (negative) or buffering (positive) (J. a. G. M. de Visser et al., 2011). Aggravating, or negative, interactions between two genes lead to a reduction in fitness of the double mutation that is greater than that expected by the two single mutations acting independently. Buffering, or positive, interactions occur when one mutation masks the effect of the other mutation (Segrè et al., 2005). Sign epistasis on the other hand occurs when the effect of one of the mutations changes sign in the background of the other mutation. Finally, the situation when both the mutations are separately deleterious but beneficial when they happen together has been named reciprocal sign epistasis (Dawid, Kiviet, Kogenaru, de Vos, & Tans, 2010; de Vos, Poelwijk, Battich, Ndika, & Tans, 2013; Poelwijk, Tănase-Nicola, Kiviet, & Tans, 2011).

Epistasis is evolutionarily important since epistatic effects can affect the shape of the evolutionary fitness landscape, or the adaptive landscape, that maps gene mutations to fitness. Arguments about the importance and role of epistatic effects played a major role in the debate between Sewall Wright and R. A. Fisher in the 1930s (Brodie III, 2000). Epistasis is believed to be necessary for the evolution of sex and recombination (J. A. G. M. de Visser & Elena, 2007). Adaptive landscapes are typically thought of as rugged, with multiple fitness peaks and valleys, and it has been shown that a rugged fitness landscape requires the existence of reciprocal sign epistasis (Dawid et al., 2010; Poelwijk et al., 2011).

However, epistatic effects are hard to uncover experimentally. Cellular metabolism is one arena of research that lends itself easily to the analysis of some kinds of epistatic interactions since it is relatively well understood, and genome-scale constraint based models using Flux Balance Analysis (FBA) do a reasonably good job in predicting intracellular fluxes (R. Schuetz et al., 2012; Robert Schuetz et al., 2007) as well as the effect of perturbations (Fong & Palsson, 2004; Segrè et al., 2002). A small but significant body of literature has emerged that uses these computational methods to search for epistatic interactions via gene deletions (He et al., 2010; Segrè et al., 2005; Snitkin & Segrè, 2011; L. Xu, Barker, & Gu, 2012). An advantage of computational methods is their ability to analyze all putative epistatic interactions; however, the framework of FBA limits the analysis to mean or magnitude epistasis, and sign epistasis cannot be studied. Using these methods, it has been shown that metabolic networks of yeast and *E. coli* are characterized by the dominance of small positive epistatic interactions (He et al., 2010; Segrè et al., 2005). Epistatic interactions were shown to be largely either positive or negative between metabolic subsystems, allowing a redefinition of modularity between functional modules of cellular metabolism (Segrè et al., 2005). A key insight in this work was that epistasis in the context of FBA-based

61

computational approaches is a consequence of network structure, with linearly connected pathways likely to show positive epistasis with each other, and branched pathways likely to show negative epistasis (Segrè et al., 2005). In the same work, it was also shown that positive epistasis is higher amongst functionally unrelated genes while negative epistasis was higher among functionally related genes. It has also been shown in later work that epistatic interactions are not absolute, but depend upon the effect being considered. For metabolic models this is most often a function representing "fitness", and thus epistatic interactions depend upon the particular definition of fitness being used (Snitkin & Segrè, 2011), in other words, different fitness functions capture different aspects of functional relationships between genes.

However, a phenotype constitutes the observable characteristic of a genotype in a particular environment. Relatively few experimental studies have analyzed the effect of changing environments on epistatic interactions. One paper analyzed a small set of 18 mutations showed that about a third of mutations analyzed exhibited the joint effect of both the environment and the genetic background (Remold & Lenski, 2004). Much more recently, a set of 5 beneficial mutations in *E. coli* were analyzed by constructing 32 double mutations and studying them in 1920 different environments. The effect of both the single mutations as well as epistatic interactions were found to be environmentally dependent (Flynn, Cooper, Moore, & Cooper, 2013). Another equally recent experiment studied three variations of the well-studied Lac operon in *E. coli*, each of which contained three to six point mutations, in the presence or absence of IPTG, and again found strong dependence of epistatic effects on the environment (de Vos et al., 2013).

These results suggest that despite their limitations, computational studies of epistasis in different environments could yield significant insight into the impact of fluctuating environments on the evolutionary process. However, to date no such studies using metabolic models have been

carried out. This paper seeks to fill that gap by using constraint-based models of metabolism to delineate the effects of the environment on epistatic interactions between metabolic genes in three different microorganisms, the model bacterium *E. coli*, the cyanobacteria *Synechocystis* PCC6803 and the model green algae, *C. reinhardtii*. Computational studies of epistasis have concentrated on yeast and *E. coli*. We therefore also present here the first computational analysis of epistatic interactions in photosynthetic organisms. We also perform a comparative analysis of epistasis in the central carbon metabolism between *E. coli* and *Synechocystis*.

Epistatic analysis is performed using double gene deletions on these three organisms under various different growth conditions. Our analysis throws up a number of novel conclusions. Prior work had indicated that magnitude epistasis in metabolism is dominated by positive interactions in both yeast and *E. coli*. We show that while this remains true in an aerobic environment, epistasis in anaerobic conditions is dominated by negative epistasis. More generally we show that the increase in the C/O ratio leads to disappearance of large number of positive interactions. We find both differences and similarities in the epistatic interactions of similar genes between *E. coli* and *Synechocystis* under heterotrophic conditions, and show that these arise out of differences in network flows. We show therefore that epistatic interactions are not so much determined by network structure as they are by network flows, and *E. coli* under different carbon sources has different epistatic interactions. We find that under photoautotrophic conditions, the C/photon ratio affects epistatic interactions in the same way as the C/O ratio did in *E. coli*, and under conditions of unlimited light both *Synechocystis* and *Chlamydomonas* are characterized by the relative disappearance of positive interactions between metabolic genes. We thus show that the epistatic interactions uncovered by the computational analysis are not only dependent on the organization of the metabolic network, but also on the environmental conditions.

3.   RESULTS ABD DISCUSSION

3.1.   DIFFERENT CARBON SOURCES LEAD TO DIFFERENT PATTERNS OF FLUXES AND EPISTATIC INTERACTIONS

To analyze the effect of environmental conditions on fluxes, we calculated flux distributions in the 174 different carbon sources (substrates) under which *E. coli* could grow, according to the previous model predictions of the model iAF1260 (Feist et al., 2007). The reactions were assigned ranks in each growth conditions based on the absolute value of flux (Figure 3.1A). We find that fluxes indeed change drastically across different growth conditions as shown by the change in the rankings of reactions (Figure 3.1A). The magnitude of the change can be seen by the wide variations in the coefficient of variation of the ranks (Figure 3.1B and Figure C1). A number of reactions also showed flux reversal under different environment conditions (Figure
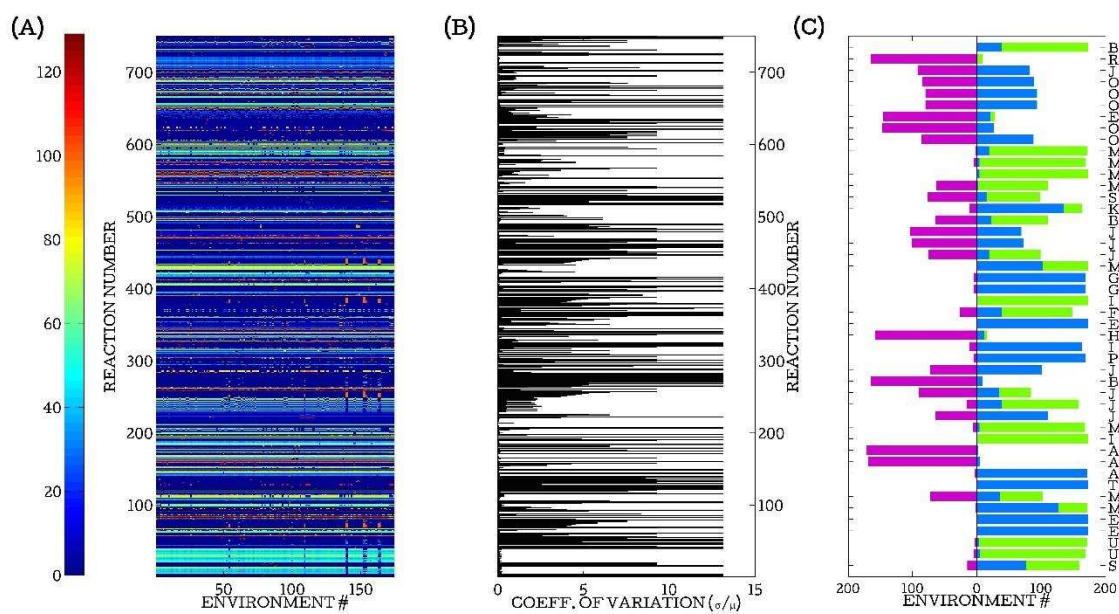
FIGURE 3.1: FLUXES CHANGE DEPENDING UPON GROWTH CONDITIONS.
(A) Flux ranks associated with reactions corresponding to the 174 growth (environmental) conditions; the color axis represents the rank of the reaction in any particular environment. (B) Coefficient of variation (s/m) of the rank calculated across 174 growth conditions for each reaction. (C) Histogram of reactions in E. coli, X-axis represents the number of environmental conditions in which a given reaction has positive (blue), negative (magenta), or zero (green) flux; the letters (in parentheses) on the Y-axis correspond to subsystems as listed in Table C2. NOTE: for (C), we had to use reversible model, while (A) and (B), reactions must be irreversible.

3.1C). Taken together the data suggest that the topology of carbon flows can change significantly under different carbon substrates.
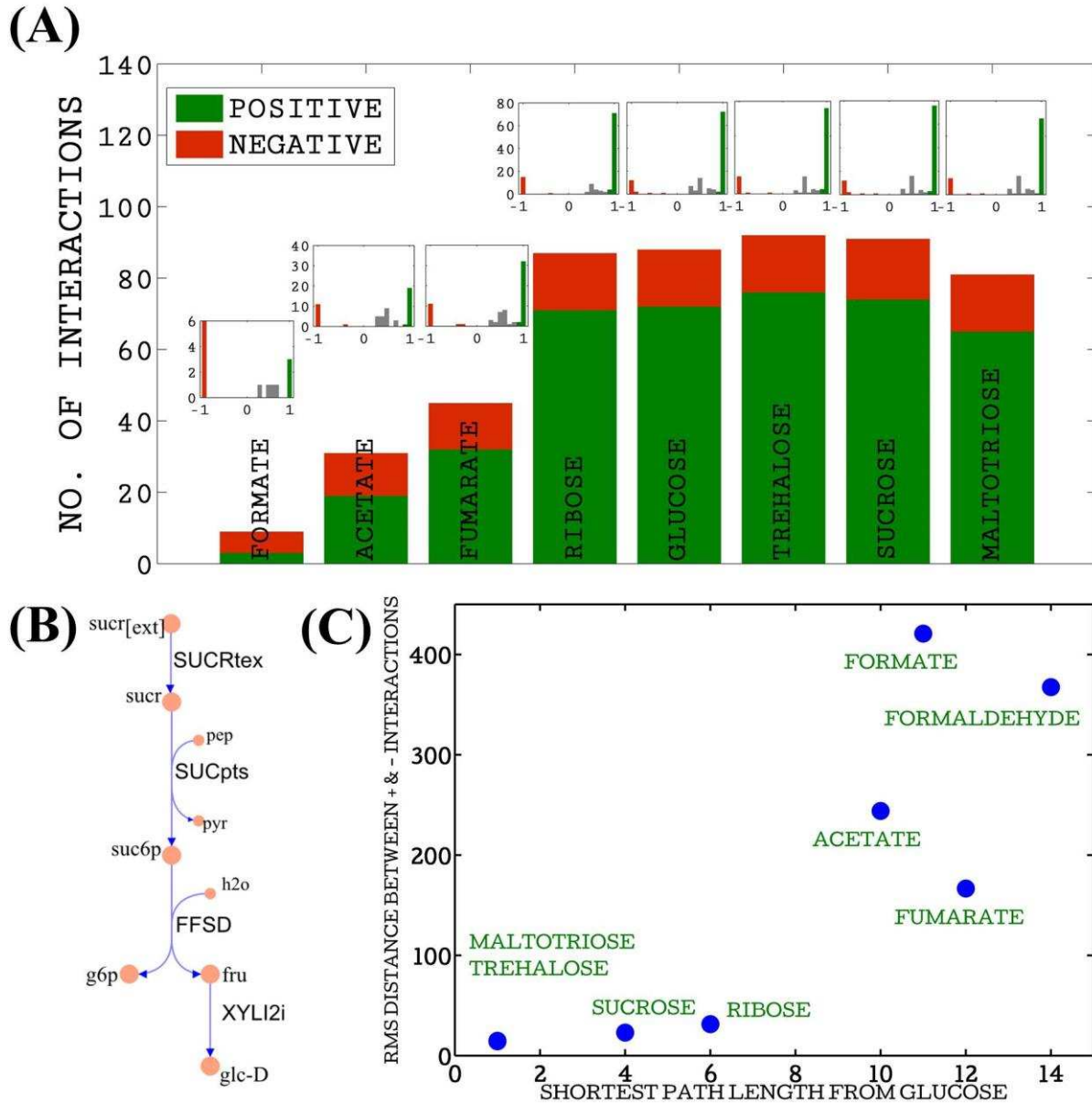
FIGURE 3.2: EPISTASIS UNDER VARIOUS DIFFERENT CARBON SOURCES.
(A) Number of positive and negative interactions for E. coli, when grown under different carbon sources. The histograms above each bar represents the distribution of epistasis values in each of the growth conditions. Pairs with no interaction are not shown for proper visualization of data; these pairs do peak at epistasis value of 0 (giving rise to a trimodal distribution). Red represents negative interactions, green represents positive interactions; (B) four reactions defining the path from sucrose to glucose; (C) RMS distance between positive and negative interactions varies based on the shortest path length between carbon source mentioned in green and glucose.

We chose 9 different carbon substrates to analyze the effect of environmental conditions on epistasis in greater depth in *E. coli*. After removing isozymes we generated a list of 93 genes (Table C1), which were non-lethal and generated a flux perturbation in presence of at least one carbon source, and constructed 4278 (93 times 92 times half) double deletion mutations. We find that substrates with more carbon atoms generally result in a greater number of non-zero interactions (Figure 3.2A). The only exception is Maltotriose which has fewer total interactions than Glucose. Note that a core set of negative interactions remain conserved under every examined carbon source, thus what is changing are positive interactions. The smallest total number of positive interactions was observed when the organism was grown in presence of formate. During its metabolism formate is used for formylation of tetrahydrofolate (THF) and converted to methylene-tetrahydrofolate (MLTHF). Methylene in MLTHF enriches glycine to serine. Serine is then sequentially converted to phosphoenol-pyruvate (PEP), which feeds gluconeogenic and citric acid pathway. Metabolically speaking therefore, formate is quite different from glucose. It requires that a larger number of anabolic reactions be turned on, compared with glucose which requires decomposition to smaller molecules like pyruvate to form higher carbon derivatives. To quantify the metabolic distance between glucose and formate, we introduce the idea of a "metabolic path length", which is defined as the minimum number of steps required to form one carbon source from another. We, therefore, hypothesized that average path length between two substrates is proportional to the difference between numbers of positive interactions under different substrates. Consider the case of glucose and sucrose. These two metabolites differ by four reactions (Figure 3.2B), SUCRtex (sucrose transporter), SUCptspp (sucrose phosphate), FFSD (β-fructofuranosidase), and (XYLI2: a hexose isomerase).

66

We manually calculated the shortest path length from glucose to the other substrates like formate, formaldehyde, acetate, fumarate, ribose, sucrose, trehalose, and maltotriose. To quantify the notion of difference between number of interactions, taking both positive and negative interactions into account, we calculated the Root Mean Square difference (or the Euclidian distance) between the two dimensional vector representing the number of positive and the number of negative interactions respectively for pairs of growth conditions. In agreement with our hypothesis we find that increase in path length leads to an increase in difference of number of interactions (Figure 3.2C). We find that interactions observed under glucose did not change drastically from interactions observed under sucrose, trehalose and maltotriose (Figure C2). Thus the short RMS distance between interactions in glucose and sucrose is due to the metabolic path length of just four reactions mentioned above that separate the two metabolites.

Amongst these 4278 pairs, 150 pairs interacted positively and 22 interacted negatively, in at least one growth condition. Out of these 172 pairs with non-zero interactions, interestingly, only one gene pair changed sign in different environmental growth conditions, b2779 (Enolase, ENO) – b3956 (Phosphoenolpyruvate (PEP) carboxylase, PPC). This pair interacted positively in presence of sugars (trehalose, ribose, glucose, sucrose, and maltotriose), interacted negatively in presence of aldehyde (formaldehyde) and did not interact in presence of carboxylate (formate, acetate, and fumarate). Analysis of flux distributions revealed that the positive interactions are the result of forward flux through ENO (catalyzing dehydration of 2-phosphoglycerate (2PG) to PEP). The product of ENO is PEP, which is a substrate for PPC (catalyzing carboxylation of PEP to oxaloacetate (OAA)). Thus, this linear chain of reaction results in positive interactions. However, in presence of formaldehyde, ENO has a backward flux (hydrolysis of PEP to 2PG). The burden of PEP utilization to make important cellular biomass components results in a synthetic lethal.

67

Thus, under these conditions, the pathway bifurcation occurring due to flow of carbon results in negative interaction. In presence of carboxylates (formate, fumarate, and acetate), PPC carries no flux because OAA is made by TCA cycle. Thus, no interaction occurs between these two genes with carboxylates as substrates.

Interestingly, we find that none of the gene pairs interacted positively in all 9 growth conditions, but 5 gene pairs interacted negatively in all 9 growth conditions. Further, we find that out of these 150 positively interacting pairs, 46% of the interactions occur in either one of the growth condition, while only 18% of the negatively interacting gene pairs occur in either one of the growth condition. Our results indicate that negative interactions, in general, are more likely to persist than positive interactions. The list of 93 genes is presented in the Appendix C (Table C1).

3.2. *POSITIVE EPISTASIS DOMINATES AEROBIC GROWTH OF E. COLI AND SYNECHOCYSTIS SP. PCC6803*

It has previously been shown that metabolic epistatic interactions uncovered by flux balance analysis in *E. coli* and in yeast are dominated by positive or buffering interactions (He et al., 2010). Photosynthetic organisms have not been previously analyzed for epistatic interactions. We therefore performed a similar analysis on *Synechocystis* under heterotrophic aerobic growth on glucose, to compare genetic interactions between metabolic genes of *Synechocystis* with *E. coli* under similar environmental conditions (epistasis in autotrophic conditions is discussed later). For completeness, and to validate our method, we repeated the exercise for *E. coli*. A single gene deletion was performed to find all essential genes, and a double gene deletion was performed on the remaining set of non-essential genes. An essential gene, in our simulations, is defined as the gene that leads to a growth rate of less or equal to 10% of the wild-type growth rate. Epistasis

values were calculated as shown in *Methods section*. The histogram of scaled epistasis values showed that *E. coli* and *Synechocystis*, under aerobic growth on glucose, is dominated by positive interactions, which were about 5-fold (Figure 3.3A) and 2.5-fold (Figure 3.3C) more than the negative interactions, respectively. When these deletions were categorized as a deletion in a particular subsystem of the organism, oxidative phosphorylation and glycolysis had the highest number of interactions with other subsystems in both the organisms. (Figure 3.3B & 3.3D). Note that a previous study on epistasis in metabolic genes (He et al., 2010) reported a much larger
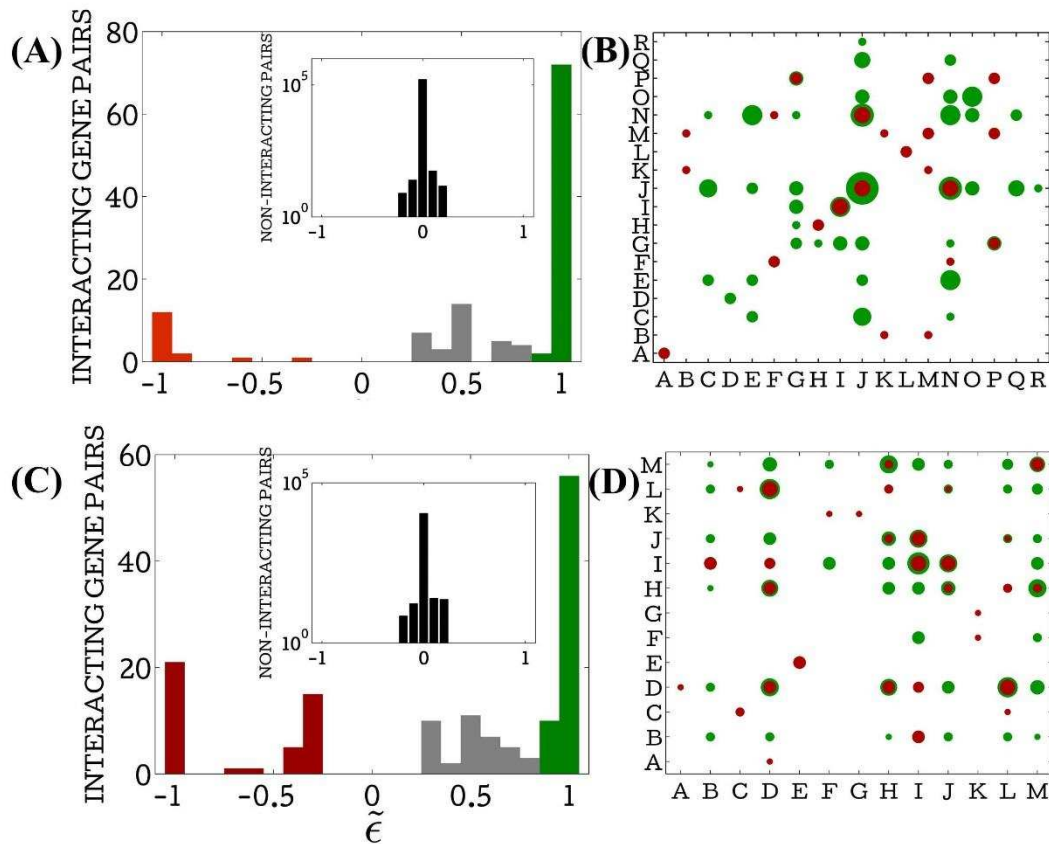
FIGURE 3.3: EPISTATIC INTERACTIONS MAPS RELATIVE TO AEROBIC GROWTH ON GLUCOSE FOR *SYNECHOCYSTIS SP.* PCC6803 AND *E. COLI*.
(A) Histogram of epistasis for E. coli under aerobic growth with glucose; (B) epistatic map for E. coli; (C) histogram of epistasis for Synechocystis under aerobic growth with glucose (heterotrophic); and (D) epistatic map for *Synechocystis*. Red represents negative interactions, green represents strong positive interactions, and grey represents weak positive interactions. The inset graph in (A) and (C) represents non-interacting pairs (black). The subsystems corresponding to letters is present in Tables C2 and C3. The size of the dots is proportional to the number of interactions (this convention is followed in all of these types of plots in the chapter).

number of positive interactions since they were reporting epistasis due to partial deletion of reactions rather than total deletion of genes.

3.3. MAXIMUM NUMBER OF POSITIVE INTERACTIONS CORRESPONDS TO MAXIMUM RESPIRATORY CAPACITY IN E. COLI

We next varied the ratio of glucose to oxygen uptake ratio (C/O ratio) in *E. coli* and repeated the epistasis analysis (Figure 3.4A). We varied the glucose to oxygen uptake ratios by changing the glucose uptake rate from 8 mmole/gDW/h to 64 mmole/gDW/h. Using experimentally determined specific glucose uptake rate of 8 mmoles/gDW/h (Fischer, Zamboni, & Sauer, 2004), we calculated the maximum specific oxygen uptake rate (18.2 mmoles/gDW/h) required by the wild-type cell. We call this rate the maximum respiratory capacity of the wild-type cell. The simulated C/O ratio (0.4395) under the nominal conditions fell within the range of experimentally determined C/O ratios from various different experiments, 0.35 – 0.49 (Kayser, Weber, Hecht, & Rinas, 2005). We find that as the C/O ratio is increased total number of buffering or positive interactions dramatically decrease, while the number of negative interactions remain approximately constant (Figure 3.4A). We also noticed that most of the negative interactions remained robust throughout different C/O ratios.

The increase in the C/O ratio is analogous to the organism shifting from aerobic growth to anaerobic growth. This inspired us to evaluate the anaerobic condition, which corresponds to a scenario where C/O ratio goes to infinity. As *E. coli* is a facultative anaerobe, it is also able to grow under anaerobic conditions. We find that under anaerobic condition, the number of positive interactions almost vanish, while the negative interactions are unaffected (Figure 3.4B). Thus environmental conditions resulting in excess of carbon substrate (in this case, glucose) help

mutations that would otherwise be deleterious under maximum respiratory conditions to grow at optimal growth rates. In the presence of excess carbon, many positively interacting (under nominal conditions) double gene deletions do not interact with each other, leading to their disappearance.
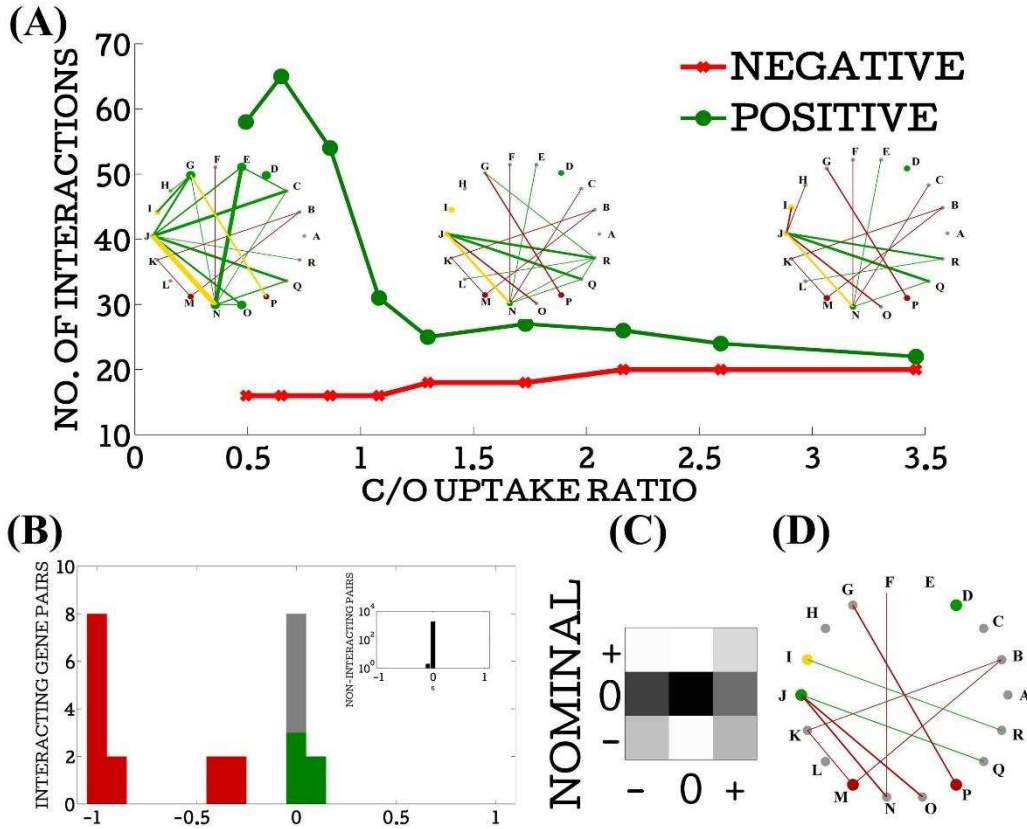
FIGURE 3.4: EPISTASIS UNDER VARYING GLUCOSE-TO-OXYGEN UPTAKE RATIOS.
(A) Number of positive (green) and negative (red) interactions corresponding to each v_{gluc}/v_{O2} (C/O specific uptake) ratio. The clock diagrams shown in insets represent interactions amongst subsystem at (left to right) C/O uptake ratio = 0.4918, C/O uptake ratio = 1.7297, and C/O uptake ratio = 3.4595. The letters correspond to subsystems as listed in Table C2. The size of the dots pertaining to each subsystem indicates the number of epistatic interactions, with green = positive; red = negative and yellow = mixed (both negative and positive). All clock diagrams shown in the paper follow these conventions. (B) Histogram of scaled epistasis of E. coli for anaerobic growth with glucose. Histogram is read as distribution of scaled epistasis based on the non-scaled epistasis value of the interacting pair. Red represents negative interactions, green represents strong positive interactions, and gray represents weak positive interactions. The inset graph represents non-interacting pairs (black). (C) Difference between aerobic growth of E. coli (nominal) and anaerobic growth of E. coli, represented by a grey map where darker the value higher is proportion of total gene interaction in that category. A perfect black corresponds to all pairs having same type of interaction; a perfect white corresponds to no pairs in the region. (D) Clock diagram representing interaction between genes belonging to various subsystems in E. coli under anaerobic growth conditions.

3.4. DOMINANCE OF NEGATIVE EPISTASIS UNDER HIGH LIGHT CONDITIONS IN

SYNECHOCYSTIS SP. PCC6803

In order to study the effect of varying light conditions on epistatic interactions, we simulated autotrophic growth of *Synechocystis* under very low to high light conditions. As before non-lethal genes (174 in number) which constrained at least one reaction were included in the analysis. We find that as the photons absorbed increase from 50 mmoles/gDW/h (Figure C3E) to 60 mmoles/gDW/h (Figure C3F), the number of positive interactions decrease, and under high or unconstrained light conditions, the positive interactions disappear entirely (Figure 3.5B). Further, our analysis showed that except for one weakly interacting pair in low light, negative interactions remained unchanged, irrespective of the amount of light available to the organism (Figure C3A-I).
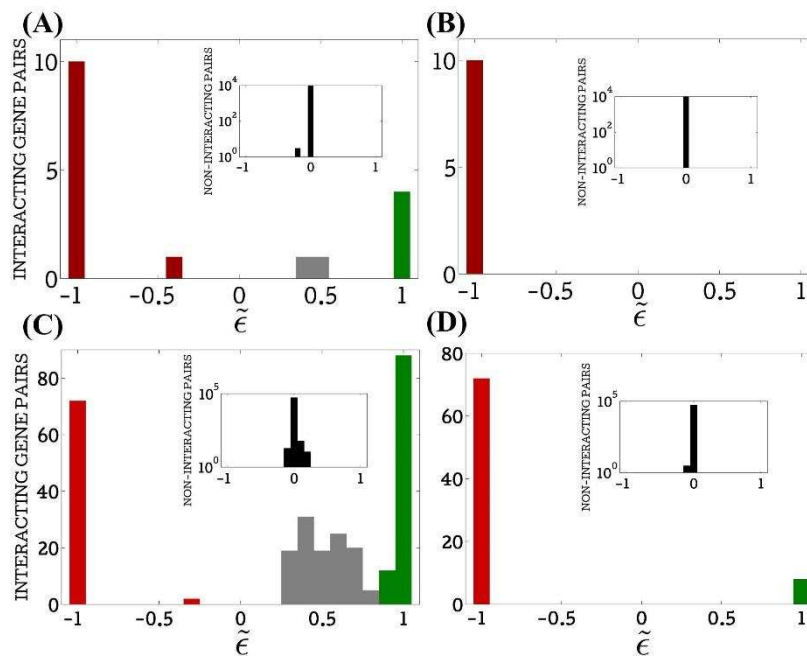
FIGURE 3.5: HISTOGRAMS OF SCALED EPISTASIS FOR PHOTOAUTOTROPHIC ORGANISMS UNDER LIMITED LIGHT AND HIGH LIGHT CONDITIONS.
For Synechocystis sp. PCC6803, (A) limited light, (B) high light; for C. reinhardtii (C) limited light, (D) high light. Red represents negative interactions, green represents strong positive interactions, and grey represents weak positive interactions. The inset graph represents non-interacting pairs (black).

We find that fluxes through reactions belonging to the following subsystems increase under high light conditions: Oxidative phosphorylation, photosynthesis, nitrogen metabolism, glyoxylate metabolism, and pyrimidine metabolism.

To test whether the disappearance of positive interactions is a more general property of photoautotrophic metabolism, a similar analysis was performed on another single cell photosynthetic organism, *C. reinhardtii* (iRC1080). We find results to be similar for *C. reinhardtii* as for *Synechocystis*. Under limited light conditions, number of positive interactions and negative interactions were comparable (Figure 3.5C); while under high light, the number of positive interactions reduced considerably (~90%) and number of negative interactions remained same (Figure 3.5D).

Under high light conditions, autotrophic organisms suffer from reduced growth rate (Allahverdiyeva et al., 2011; Demmig-Adams, 1992; Hackenberg et al., 2009; Kopecná, Komenda, Bucinská, & Sobotka, 2012). Three main changes that occur during such an environmental condition are: (i) reduction in growth rate owing to increase in damage and *de novo* synthesis of photosynthetic proteins (Allahverdiyeva et al., 2011; Demmig-Adams, 1992), (ii) increase in the photo-respiratory flux (Hackenberg et al., 2009) and (iii) decrease in carbon fixation (Hackenberg et al., 2009). This reduction in growth rate is not captured in our model due to absence of pathways for damage of photosynthetic proteins. However, it must be noted that the model does correctly predict an increase in the photo-respiratory flux (Juan Nogales et al., 2012). Since the negative effects of high light cannot be properly accounted for under the current model framework, we cannot comment on how realistic the results of the FBA optimization under high light are. However they do correspond with the case of a high C/O ratio in *E. coli*. Thus similar to excess nutrients, excess light too, leads to a reduction in the number of buffering interactions. Note that both *E. coli*

growing under formate, essentially a 1-carbon source, and *Synechocystis* growing under CO_2; show the dominance of negative interactions.

Chlamydomonas and *Synechocystis* had 2 gene pairs and 1 gene pair respectively which were weakly negatively interacting under low light and non-interacting in high light. In Chlamydomonas, the 2 pairs belonged to acetyl-CoA transport across various compartments and the other gene pair belonged to energy production via ATPase in thylakoid membranes. In *Synechocystis*, the weakly interacting genes belonged to ferredoxin oxidoreductase and Glutamate dehydrogenase.

There are three main types of molecules absolutely required for a mutant to grow even at sub-optimal growth rates: (i) ATP, (ii) electron carriers and (iii) carbon. We hypothesize that these mutants were limited by electron carriers and/or ATP when under limited light. However, under high light, there would be a relative excess of these electron carriers and/or ATP. This enrichment of electron carriers and/or ATP under high light helps the organism to grow at optimal growth rate. In presence of high light, the carbon fixation efficiency ($v_{(CO2\ fixed)}/v_{(hv\ utilized)}$) reduces, as a consequence of which mutations tend to be less deleterious and are able to achieve optimal growth rate, resulting in no interaction between genes.

3.5. *EPISTATIC INTERACTIONS ARE DEPENDENT ON CARBON FLOW IN THE NETWORK*

If epistasis in metabolic genes depends on carbon flows in the network, identical genes in two organisms should display mostly similar epistatic interactions, while the differences should be attributable to differences in carbon flow patterns. We compared scaled epistasis amongst gene pairs that constrained identical reactions in both organisms. Of 74 such gene pairs, we found that

54 have identical types of epistasis (positive or negative). Out of these 15 are negatively interacting, while 39 are positively interacting in both the organisms. Some interactions which are common to both are as follows: positive interactions among Glycolysis and TCA cycle, positive and negative interactions within Glycolysis, positive interactions among Pentose phosphate pathway and oxidative phosphorylation, positive interactions among TCA cycle and oxidative phosphorylation (Figure 3.3B and 3.3D). However out of the remaining 20 gene pairs, 18 were mismatches. The mismatches (positive to negative) occur between succinate dehydrogenase (SUCDi), genes belonging to lower glycolysis, and NADH dehydrogenase (NADH); while mismatches (negative to positive) also occur amongst genes belonging to lower and middle glycolysis (Enolase (ENO), phosphoglycerate kinase (PGK), and triose phosphate isomerase (TPI)).

It was not possible to discern the reason for the mismatches from the flux distribution of the entire network, due to its complexity. We therefore decided to perform a reaction-wise epistasis analysis of a subnetwork consisting of reactions involved in glycolysis/gluconeogenesis, TCA cycle, and pentose phosphate pathway. Here by reaction-wise epistasis we mean the non-additive effects of deleting two reactions from the metabolic network. This is equivalent to assuming that each reaction is constrained by a different gene. In reality gene deletions may constrain more than one reaction, making their effect harder to interpret. This sub-network was made up of 33 reactions. It can be seen that large numbers of interactions remain the same in both the organisms (Figure 3.6). In *Synechocystis*, 25 and 39 reaction pairs interacted positively and negatively, respectively. Comparing the positively interacting reaction pairs in *Synechocystis* to *E. coli*, we found that 18 reaction pairs interacted positively, and 7 pairs did not interact in *E. coli*. However

on comparing the negatively interacting reaction pairs in *Synechocystis* to *E. coli*, we found that

25 reaction pairs interacted negatively but 14 pairs interacted positively in *E. coli*.
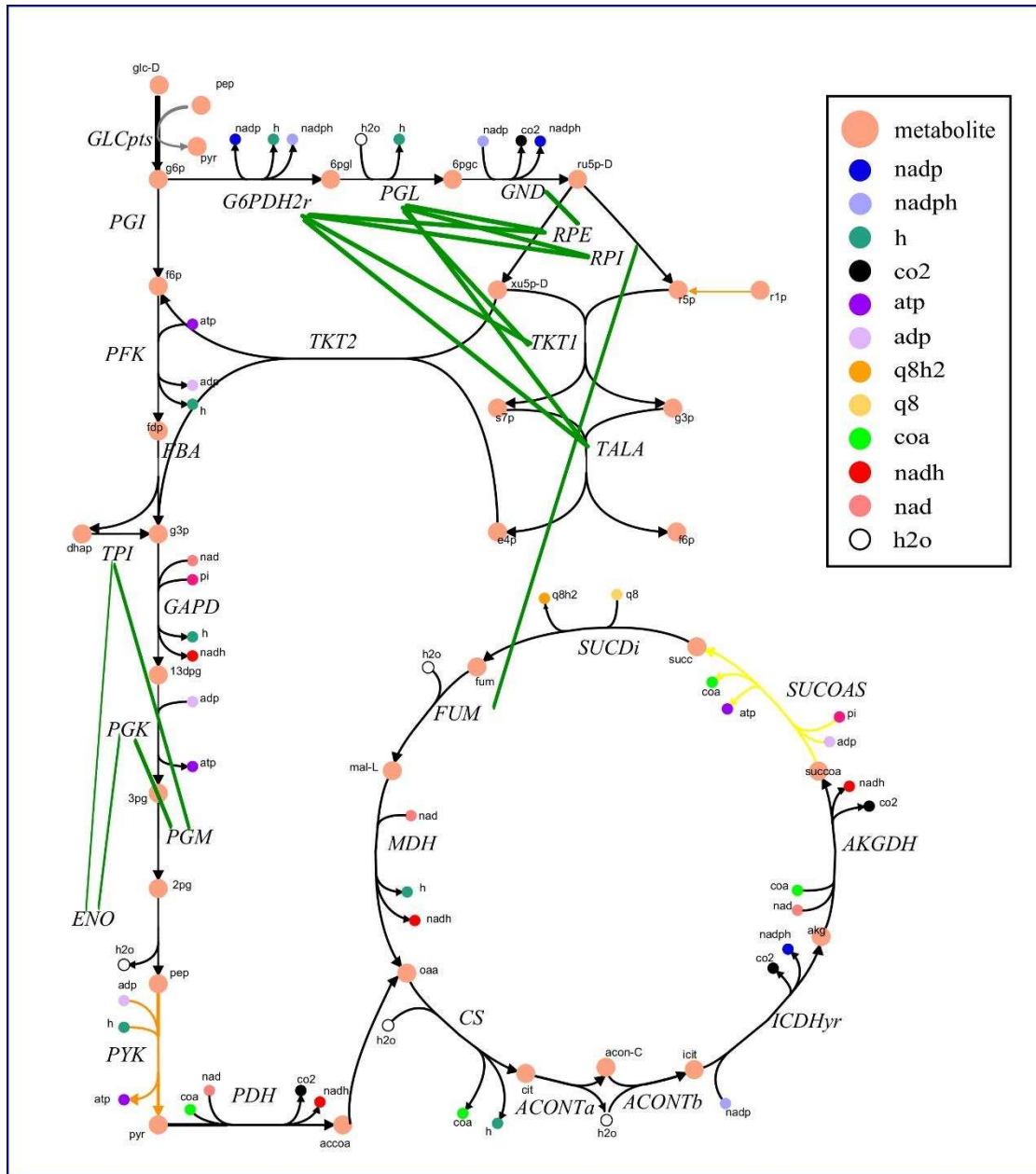
FIGURE 3.6: EPISTASIS INTERACTIONS AMONGST REACTIONS BELONGING TO THREE COMPARTMENTS GLYCOLYSIS, CITRATE CYCLE, AND PENTOSE PHOSPHATE PATHWAY FOR CELLS GROWN AEROBICALLY WITH GLUCOSE.

The overall picture represents the flow of carbon in *E. coli*. The black arrows indicate the direction of flow in *E. coli* and *Synechocystis*. The yellow arrow indicates reaction operating in reverse direction in *Synechocystis*. The grey arrow indicates significantly less (<10% of proportion of reaction flux through *E. coli*) flux through the reaction in *Synechocystis*. The orange arrows indicates significantly less (<10% of proportion of reaction flux through *Synechocystis*) flux through the reaction in *E. coli*. The green lines indicate the differences in epistasis which was negative for *Synechocystis* but positive for *E. coli*.

We analyzed these 14 mismatches manually and determined that they arise due to differences in carbon flow. In *E. coli*, reactions catalyzed by glucose 6-phosphate dehydrogenase (G6PDH2), 6-phosphogluconolactonase (PGL), and phosphogluconate dehydrogenase (GND) interact positively with each of the reactions catalyzed by ribose 5-phosphate isomerase (RPI), ribose 5-phosphate epimerase (RPE), Transketolase 1 (TKT1), and Transaldolase (TALA). In *E. coli*, any deletion in oxidative pentose phosphate pathway (G6PDH2r and PGL) results in the same metabolic flux redistribution. In the absence of oxidative pentose phosphate pathway, operation of TALA, TKT1 and TKT2 is reversed such that ribose 5-phosphate, xylulose 5-phosphate , and ribose 1-phosphate (R1P) is produced in both the organism. However, in *Synechocystis*, other than phosphopentomutase (PPM), R1P can only be produced by the decomposition of adenosine. There are many other reactions that can produce R1P, in *E. coli*. This is why interaction amongst oxidative pentose phosphate reactions and TKT1, RPI, RPE, and TALA is positive in *E. coli*. Similar reasons can be attributed to other 4 interactions occurring amongst reactions in glycolysis (Figure 3.3). The mismatches (epistasis sign-change) account for about 21% (14/64) of the total interactions which were positive or negative. Thus epistatic interactions are affected by metabolic flows, which are in turn affected by the environmental condition of an organism.

4.   EXPERIMENTAL

4.1.   *FLUX BALANCE ANALYSIS (FBA)*

Flux Balance Analysis (FBA) is a mathematical framework used to calculate the flow of the metabolites through the metabolic network at steady state (Orth et al., 2010). FBA was performed using the COBRA Toolbox (Schellenberger et al., 2011). In brief, FBA involves writing down an M by N stoichiometric matrix, S corresponding to the metabolic reactions for each

organism. Here M is the number of metabolites and N is the number of reactions. Under steady state conditions the system of differential equations representing the chemical reaction system become a system of linear equations in the fluxes,

$$\sum_{j=1}^{N} S_{ij} v_j = 0 \tag{1}$$

Here, $\mathbf{v}$ is a vector of reaction flux and S_{ij} represents the stoichiometric coefficient for i^{th} metabolite in j^{th} reaction. To find the fluxes an objective function is chosen that is believed to be optimized by the organism, such as its growth rate. This makes it a linear programming problem (LPP) that can be solved by standard techniques by imposing additional constraints, discussed below, in addition to Eq. 1 (Feist & Palsson, 2010; Orth et al., 2010). The objective function most commonly used for such models is an equation describing the growth rate of the organism. Growth rate reactions are described as:

$$\sum_{j=1}^{N} c_j v_j \rightarrow \mu \tag{2}$$

In the above equation, c_j and v_j refer to the weight in final biomass and the flux of the product of the jth reaction respectively, and $\mathbf{\mu}$ refers to the growth rate of the organism. Maximization of growth rate was used as the objective function for all the simulations conducted in this study. Additional constraints are constructed in the following way:

1. Incorporating measured or experimentally estimated uptake and secretion rates of metabolites.

2. Incorporating a global limit on the upper and lower bounds of each reaction flux.

$$\alpha_j \leq v_j \leq \beta_j \tag{3}$$

α_j and β_j are the lower and upper limits placed on each reaction flux, v_j, respectively. Reversible reactions can take either negative or positive values of fluxes, while irreversible values were constrained to take only positive values. Further, if any reactions were turned off, inactivated or

deleted, the flux through the reaction was set to zero: $v_j = 0.$ The linear programming problem was implemented using COBRA Toolbox with Gurobi 4.6.1 on MATLAB R2011b (Schellenberger et al., 2011).

4.2. *SIMULATION OF GROWTH CONDITIONS IN VARIOUS ORGANISMS*

For our analyses of different cellular metabolism, we chose genome scale models of *Escherichia coli* K12 MG1655 (iAF1260) (Feist et al., 2007), *Synechocystis sp.* 6803 (iJN678) (Juan Nogales et al., 2012), and *Chlamydomonas reinhardtii* (iRC1080) (Chang et al., 2011). Growth conditions included in the analyses were: *E. coli* (**Aerobic:** Formate, Formaldehyde, Acetate, Fumarase, Ribose, Glucose, Sucrose, Trehalose, Maltotriose, and **Anaerobic**), *Synechocystis* sp. (**Autotrophic:** high light, limited light, and **Aerobic:** Glucose), and *C. reinhardtii* (**Autotrophic:** high light, limited light). For simulation of different carbon sources in *E. coli*, we normalized carbon uptake to 8 mmoles for 6 carbon atoms in the molecule. For example, if 8 mmoles/gDW/h of glucose was used; then 12 mmoles/gDW/h of fumarate, a four carbon molecule, was used. Limited light conditions were simulated by setting the maximum light uptake to the optimal value calculated for wild-type cells. However, high light conditions were simulated by leaving light uptake unconstrained. Non-lethality criterion for a mutant was set to more than 10% (or 0.1 times) of wild-type growth rate, correct to first order of magnitude.

For aerobic growth of *E. coli* (model name: iAF1260) under various carbon sources, simulations were performed by applying the following constraints: (i) maximum uptake rate of the desired carbon substrate (EX_glc(e), EX_sucr(e), EX_for(e), EX_fum(e), EX_rib-D(e), EX_malttr(e), EX_tre(e), EX_fald(e), or EX_ac(e)) was set to 8 mmoles/gDW/h per 6 molecules of carbon in the substrate, while uptake rates of all other carbon sources were set to zero; maximum

oxygen uptake rate (EX_o2(e)) was left unconstrained (Feist et al., 2007); and all the other constraints were same as reported in the original article of the published model. Heterotrophic growth of *Synechocystis* sp. PCC6803 (model name: iJN678) was simulated by setting the maximum glucose uptake rate (EX_glc(e)) to 0.85 mmoles/gDW/h (C. Yang et al., 2002); leaving maximum oxygen uptake rate unconstrained; and setting the uptake rates of other sources of carbon and light to zero. Autotrophic growth of *Synechocystis* sp. PCC6803 was simulated by setting the maximum carbon dioxide uptake rate to 3.7 mmoles/gDW/h (Juan Nogales et al., 2012) and uptake rates of other carbon sources to zero; while minimum photon uptake uptake rate corresponding to maximum growth rate was calculated and subsequently set to 54.0948 mmoles/gDW/h; and all the other constraints were used from the original article where the model was published. Autotrophic growth of *Chlamydomonas reinhardtii* (model name: iRC1080) was simulated by utilizing the constraints from the original article where the model was published (Chang et al., 2011). The default constraints on flux of a reversible reaction was [-1000, 1000], and of an irreversible reaction was [0, 1000], unless specified here or in the original article where these models are published. These constraints are a norm in the field and have been used in numerous FBA studies (Chang et al., 2011; Feist et al., 2007; Juan Nogales et al., 2012).

4.3. RANKING OF FLUXES

Fluxes for each of the 174 conditions leading to growth as reported in the original publication of the E. coli (iAF1260) model (Feist et al., 2007) were simulated and ranked according to flux magnitudes. The directionality of reaction was ignored, in case of reversible reaction, because enzyme catalyzing the activity will be observed whether the reaction was operating in forward or reverse direction.

4.4. CALCULATION OF EPISTASIS

Firstly, a single gene deletion was performed to remove any essential genes. Then, a double gene deletion was performed on the remaining set of genes. Epistasis values were calculated as shown below (Segrè et al., 2005). The epistasis value for the interaction between gene X and gene Y is represented by ε. This value can be calculated by:

$$\varepsilon = W_{XY} - W_X W_Y \tag{4}$$

Here $W_X = \mu^{X-}/\mu^{wt}$, $W_{XY} = \mu^{XY-}/\mu^{wt}$ are the fitness values for the single mutant and the double mutant, and μ^{wt}, μ^{X-}, and μ^{XY-} are growth rates of wild-type, the mutant in gene X, and mutant in genes X and Y. While this is the absolute level of epistasis we need to establish a standard to compare it with. We follow (Segrè et al., 2005) and scale the epistasis value given by Equation (4) as follows:

$$\tilde{\varepsilon} = \frac{W_{XY} - W_X W_Y}{\left|\widetilde{W}_{XY} - W_X W_Y\right|}; \tag{5}$$

$$\widetilde{W}_{XY} = \begin{cases} \min(W_X, W_Y), \text{ for } W_{XY} > W_X W_Y \\ 0, \text{ otherwise} \end{cases} \tag{6}$$

The unscaled (ε) and scaled $(\tilde{\varepsilon})$ epistasis values can be then classified as shown in Table 1 below.

TABLE 3.1: CLASSIFICATION OF DIFFERENT RANGES OF UNSCALED AND SCALED EPISTASIS.

	Unscaled Epistasis	Scaled Epistasis
No epistasis	$\varepsilon = 0$	$\tilde{\varepsilon} \approx 0$
Aggravating	$\varepsilon < 0$	$\tilde{\varepsilon} \approx -1$
Buffering	$\varepsilon > 0$	$\tilde{\varepsilon} \approx 1$

The scaled epistasis ($\tilde{\varepsilon}$) was used to classify the interactions into buffering (green) at $\tilde{\varepsilon} > \theta_+$; aggravating (red), including synthetic lethal at $\tilde{\varepsilon} = -1$ and strong synthetic sick at $\tilde{\varepsilon} < \theta_-$; and no epistasis otherwise. Here we used $(\theta_-, \theta_+) = (-0.25, 0.85)$. It must be noted here that denominator in Eq. 5 is an absolute value and will not change the sign of the epistasis.

Note that here we characterize the phenotype by the growth rate of the organism. Growth rate makes a good choice of phenotype because of the role of epistasis in selection dynamics (Segrè et al., 2005), and it can be measured accurately using high throughput methods (Jakubowska & Korona, 2012; Martin, Elena, & Lenormand, 2007; Segrè et al., 2005). However, the mathematical framework of Flux Balance Analysis (FBA) used here to calculate growth rate requires maximization of growth rate. As a result, one is never able to calculate instances when the fitness of the mutants is higher than the fitness of the wild-type organism (Snitkin & Segrè, 2011). This is why sign epistasis cannot be studied using FBA. Therefore, we specify that our results are only relevant for epistatic interactions relative to growth rate.

4.5. MAPPING GENE PAIRS FROM ONE ORGANISM TO ANOTHER

For each of the genes involved in the pair, the E. C. numbers of the reactions constrained by the gene were found. These E. C. numbers were then searched for in the other organism. All the genes associated with the reactions with those E. C. numbers were found and pairs were created for the new organism, based on the pairs found in the source model. This mapping technique has been previously used to investigate the structure of enzyme-reaction association in microbial metabolism (Joshi & Prasad, 2016. Structure and role of enzyme-reaction association in microbial metabolism. In preparation).

4.6. CALCULATION OF RMS DIFFERENCE BETWEEN INTERACTIONS

The formula for calculating the root mean square (RMS) distance is:

$$D = \sqrt{\left(N_p - N'_p\right)^2 + (N_n - N'_n)^2} \tag{4}$$

The meaning of the symbols is as below:

D = RMS difference between interactions;

N_p, N_n = Number of positive and negative interactions in nominal case, respectively

N'_p, N'_n = Number of positive and negative interactions in growth condition, respectively.

5. CONCLUSIONS

Flux Balance Analysis of large-scale metabolic models is an attractive tool for studying epistatic interaction between genes computationally. It has been argued earlier that the sign of epistatic interactions between two genes gives us information about how the genes interact in the metabolic network. If the two genes belong to the same subsystem, a positive interaction suggests that they form a linear or sequential chain with respect to each other, while a negative interaction suggests that they are part of related pathways that form the same product (J. a. G. M. de Visser et al., 2011; Segrè et al., 2005). However, previous system-wide computational studies of epistasis have not considered the impact of environment conditions on predictions of epistatic interactions. Here, we systematically generated epistatic interaction network maps relative to growth rate for *E. coli*, *Synechocystis sp.*, and *C. reinhardtii* under various different environmental conditions, by which here is meant different substrates on which the organism grows. Analysis of these networks revealed that different environmental conditions yield different sets of epistatic interactions. Epistatic interactions therefore change with time as environmental conditions change.

We show that epistasis in anaerobic conditions is dominated by negative epistasis. More generally we show that the increase in the C/O ratio leads to disappearance of large number of positive interactions. We find both differences and similarities in the epistatic interactions of similar genes between *E. coli* and *Synechocystis* under heterotrophic conditions, and show that these arise out of differences in network flows. We find that under photoautotrophic conditions, the (CO2/photon) ratio affects epistatic interactions in the similar way as the C/O ratio did in *E. coli*, and under conditions of somewhat high light *Synechocystis* tends to have lower positive interactions, and in unlimited light both *Synechocystis* and *C. reinhardtii* are characterized by a sharp decline in positive interactions between metabolic genes.

We also analyze *E. coli* under different carbon sources and show that it has different set of epistatic interactions, governed primarily by the flow of the carbon within the metabolic network. We thus show that the epistatic interactions uncovered by the computational analysis are not only dependent on the organization of the metabolic network, but also on the environmental conditions. Our findings suggest that during adaptation in dynamically changing environment, the shape of the fitness landscape may be governed by the environmental history and the pattern of carbon flow in the current state of the metabolic network. Flux flows within similar parts of the metabolic network between two organisms under the same growth conditions gives rise to generally similar interactions. For example, the carbon flow through glycolysis in *E. coli* and *Synechocystis sp.*, under aerobic growth with glucose, will be similar (but not identical), and as a result the interactions occurring within the glycolysis pathway remain mostly similar as well. In both these organisms, under heterotrophic growth one molecule of glucose is catabolized to form two molecules of pyruvate, and is converted to acetyl-CoA, a precursor to the TCA cycle. However, *Synechocystis* grown photo-autotrophically will yield a different set of gene-gene interactions

within the glycolysis pathway because these conditions require the formation of glucose and pyruvate from 3-phosphoglycerate.

A previous study has stressed the significance of the finding, on the basis of FBA, that positive epistasis is highly abundant between functionally unrelated genes in both *E. coli* and *S. cerevisiae* (He et al., 2010). This study explained this phenomenon as occurring due to a second mutation having a relatively smaller effect than the first. However we show that while positive epistasis is highly abundant compared with negative epistasis in many environmental conditions, in many other conditions it is no longer abundant, and in some cases, disappears entirely. Negative interactions however, in particular synthetic lethals, tend to remain conserved under different conditions. As previously noted, (He et al., 2010) calculate epistasis differently from us, i.e. they perform reaction deletions rather than gene deletions and they constrain flux through each reaction to 50% of its wild-type value rather than setting it to zero. In this paper we consider only epistasis due to loss-of-function mutations in genes.

Previous work has shown that selection pressures exerted due to changing environmental background resulted in different fitness landscapes. Complementary to these findings, our analysis with different growth conditions for *E. coli* show that positive interactions are more likely to change/disappear, while negative interactions are likely to stay conserved.

Our analysis shows that epistasis among metabolic genes that is predictable by FBA methods depends upon network flows. Therefore positive epistasis is not simply the result of the network topology connecting two genes being linear, as suggested in previous work, but network flows between two genes forming a linear topology. Similarly if network flows between two genes constitute a branched topology with the two genes on separate branches, we get negative epistasis between their deletions. Since FBA models do not have any transcriptional regulation, or nonlinear

85

interactions between proteins, it is noteworthy that they show that epistatic effects can arise as a consequence of network structure alone.

We also show that excess nutrient uptake conditions result in a decrease in the number of positive interactions. Presumably, the excess of nutrient conditions result in enrichment of metabolites that under nominal conditions were limiting to the growth. This enrichment allows the organism to sustain the carbon, energy or electron flow in mutants, thereby changing deleterious mutants (under nominal conditions) to fit mutants (under excess of nutrient conditions). The behavior of *E. coli* under excess carbon mirrored the behavior of *Synechocystis* and *C. reinhardtii* under excess of light. In the latter case too, we found that formerly deleterious mutations become non-deleterious mutations as a result of which the most positive epistatic interactions between gene pairs vanish. Negative interactions that lead to synthetic lethality remain.

What is the importance of these epistatic predictions? An organism evolved in a specific niche should be, metabolically speaking, optimized to live in the niche. It should be expected therefore that loss-of-function mutations in metabolic genes are always accompanied by a decline in fitness. Given a single gene deletion that marginally decreases fitness; a second deletion with positive epistasis with the first is more likely to be selected in the population than one that further decreases fitness in a neutral way. This suggests that mutations during adaptation in varying environments are selectively directed by positive interactions occurring amongst deleterious mutations. This is in agreement with experimental studies that show that mutations that get fixed in populations undergoing environmental change, such as during the evolution of antibiotic resistance in bacteria, are deleterious in the background (X. Wang, Minasov, & Shoichet, 2002).

Since network flows can change if different substrates are being metabolized, epistatic interactions also change with change in substrate metabolized. This prediction agrees with

previous experimental studies on effect of environment on epistasis and fitness landscape (de Vos et al., 2013; Flynn et al., 2013). Environmental conditions that change the flows can dramatically change the set of epistatic interactions, and thus the adaptive fitness landscape of the population. The environmental dependence of epistasis makes the task of piecing together evolutionary history, and the role of epistasis in it, all the more difficult, since the specific evolutionary path followed by an organism, during adaption in a variable environment, would be therefore highly dependent upon the specific environmental fluctuations that it encountered in its evolutionary history.

CHAPTER 4.   MODELING AND ANALYSIS OF BIOPRODUCT FORMATION IN

SYNECHOCYSTIS SP. PCC6803 USING A NEW GENOME-SCALE METABOLIC

NETWORK RECONSTRUCTION

1. SYNOPSIS

Cyanobacteria are prokaryotes capable of performing oxygenic photosynthesis, making them attractive candidates for genetic engineering towards production of commercially important chemicals. However, harnessing this potential requires understanding of metabolic regulation in cyanobacteria under natural photoautotrophic conditions. Here we present an updated genome scale metabolic network reconstruction (iSynCJ816) of *Synechocystis sp*. PCC6803. This updated model, containing 816 genes and 1045 reactions, builds and develops on previously published models. New features include an unconstrained photo-respiratory reaction mechanism as well incorporation of a mechanism to account for changes in energy absorption from light of different wavelengths. We used Flux Balance Analysis (FBA) to calculate the flux distribution within iSynCJ816 and compare *in silico* predictions with values obtained by previous *in vivo* metabolic flux analyses in *Synechocystis sp.* PCC6803. A qualitative growth comparison of 167 gene-deletion mutants with experimental studies resulted in accuracy rate of ~80%. We used the model to estimate maximum theoretical yield of products using each metabolite as a precursor, as well as the feasibility of engineering *Synechocystis* to increase CO_2 fixation, which we found is possible to increase up to 35% from wild-type levels.

2. INTRODUCTION

Cyanobacteria are the only oxygenic prokaryotes capable of converting abundantly available carbon dioxide and sunlight, via photosynthesis, into chemical energy, which is stored

as biomass. As primary producers in aquatic environment, they play an important role in CO_2 assimilation and oxygen recycling. They are also primarily responsible for the presence of molecular oxygen in current atmosphere, a process that began approximately 3 billion years ago (Brocks, Logan, Buick, & Summons, 1999). Their overall contributions include nearly 30% of Earth's photosynthetic productivity (Rae et al., 2013). Their ability to photosynthesize has made them a target for genetic modifications to produce commercially important chemicals such as biofuels (Nozzi, Oliver, & Atsumi, 2013), pharmaceuticals (Vijayakumar & Menakha, 2015), and nutraceuticals (Gademann, 2011).

Among a gamut of cyanobacterial strains, Synechocystis sp. PCC6803 is widely and extensively studied for generating genomic, biochemical, and physiological data about photosynthetic organisms (Y. Yu et al., 2013). This makes the organism a model candidate to study a variety of different processes and phenomena, including photosynthesis. To appreciate the uniqueness of cyanobacteria and understand the biochemical conversion leading to carbon fixation, photosynthesis, and biomass conversion, there is a need to reconstruct a comprehensive and validated metabolic network. The first cyanobacterial metabolic network reconstruction was published in 2005 (Shastri & Morgan, 2005). Since then, there have been 9 more reconstructions (Fu, 2009; Hong & Lee, 2007; Knoop et al., 2013; Knoop, Zilliges, Lockau, & Steuer, 2010; Montagud et al., 2011; Montagud, Navarro, Fernández de Córdoba, Urchueguía, & Patil, 2010; J. Nogales et al., 2012; Saha et al., 2012; Shastri & Morgan, 2005; Yoshikawa et al., 2011). They have progressed from being qualitatively useful to quantitatively useful models. Like most modeling approaches, each model was more detailed and strain-specific than the previous one due to availability of better data. The discrepancies which existed between the gene annotations of the previous reconstructions were revised by the subsequent reconstructions. After obtaining a

comprehensive reconstruction, various sets of methods developed by computational system biologists can be applied to validate and make predictions. One such computational framework is flux balance analysis (FBA), a constraint-based modeling method used most often to calculate flow of the carbon through various metabolites to optimize reaction flux at steady state (Orth et al., 2010). In the past, FBA has been utilized to predict flux distribution through wild-type strain (Feist et al., 2007; Liao et al., 2011; R. Mahadevan & Schilling, 2003; Montagud et al., 2010; Oh, Palsson, Park, Schilling, & Mahadevan, 2007) and mutant strains (J S Edwards & Palsson, 2000a; J L Hjersted & Henson, 2009; Segrè et al., 2002), epistatic interactions (Joshi & Prasad, 2014; Phillips, 2008; Segrè et al., 2005; Segrè & Marx, 2010; Snitkin & Segrè, 2011), futile cycles (de Figueiredo, Gossmann, Ziegler, & Schuster, 2011), and gene essentiality analysis (Neema Jamshidi & Palsson, 2007; Rocha, Förster, & Nielsen, 2008; Suthers, Zomorrodi, & Maranas, 2009).

Here, we put forth a new metabolic network reconstruction of Synechocystis sp. PCC6803. This reconstruction identifies inconsistencies in the gene annotation within previously published reconstructions from careful inspection with databases such as KEGG (Minoru Kanehisa, 2002; Minoru Kanehisa et al., 2014, 2012) and Cyanobase (Fujisawa et al., 2014; Nakao et al., 2009). Molecular mechanisms of photosynthetic network around the thylakoid membrane have been included to facilitate better understanding of respiratory activities, implemented by only the most recent previous reconstruction (Vermaas, 2001). Our work significantly improves upon earlier reconstructions by including thermodynamic analysis of more than 500 reactions, which span more than the core metabolism of the cyanobacterial metabolic network. We also tested for thermodynamically infeasible loops (TILs), which identifies intracellular futile cycles. Inclusion of thermodynamic information leads us to remove or constrain reactions that participated in TILs.

To validate the model, we compare our model with fluxes which were previously obtained via metabolic flux analysis (MFA) (C. Yang et al., 2002; J. D. Young, Shastri, Stephanopoulos, & Morgan, 2011). We validate our model under two experimental conditions: autotrophic (growth on CO2 and light), and heterotrophic (growth on glucose). We also carried out a literature search for information on gene deletions and compared the experimentally reported results for 167 genes with the predictions of the model using gene essentiality analysis, the largest set used so far. Our predictions showed better accuracy than any of the previously published models.

3.   MATERIAL AND METHODS

3.1.   *MODEL RECONSTRUCTION AND ENHANCEMENT*

The initial draft of the *Synechocystis sp.* PCC6803 metabolic network was extracted from online databases. The genomic and pathway related information was extracted using MATLAB codes from online databases which include KEGG (Minoru Kanehisa, 2002; Minoru Kanehisa et al., 2014, 2012), Cyanobase (Fujisawa et al., 2014; Nakao et al., 2009), METACYC (Caspi et al., 2014), and annotated genome sequence (Kaneko et al., 1996). The information concerning enzymes was extracted from BRENDA (Scheer et al., 2011), and KEGG. The information specific to chemical species (metabolites) was taken from ChEBI (Degtyarenko et al., 2008), and PubChem (Y. Wang et al., 2009). We also took help from previously published metabolic reconstructions of *Synechocystis sp.* PCC6803 (Knoop et al., 2013; J. Nogales et al., 2012; Saha et al., 2012). The reaction-based information about photosynthetic machinery was adapted from previously published studies (Barber, 2014; Cooley & Vermaas, 2001; Latifi, Ruiz, & Zhang, 2009; Ma, Ogawa, Shen, & Mi, 2007) and cyanobacteria textbooks (Heldt & Piechulla, 2011; Lea & Leegood, 1999). To enhance the model, we took biochemical and genomic information from previously published work, the pathways enhanced using this method were photorespiration, serine synthesis,

electron transfer within and between light harvesting proteins, fatty acid synthesis, chlorophyll synthesis, amino acid metabolism, and purine and pyrimidine metabolism. Literature search also resulted in addition of a newly discovered light independent serine synthesis pathway (Klemke et al., 2015). We performed a homology search with BLASTp algorithm on NCBI (online)(Johnson et al., 2008) on these sequences against other genomes and accepted identity cut-off of 44% (slr1829, PHB synthetase), and accepted E-value of 10^{-50} (sll8012, phenylacetate-CoA ligase). In this way, the draft network was subjected to iterative manual gap-filling. This process of model building is consistent with the protocol described by Thiele and Palsson (Thiele & Palsson, 2010).

The metabolic reconstruction was made by storing all the above information in an SBML (.xml) file. The reconstruction is converted to a mathematical model in a software such as MATLAB. In our case, the model files are readable by SBML toolbox (Keating et al., 2006), and COBRA toolbox (D. Hyduke et al., 2011; Schellenberger et al., 2011) in MATLAB 2013b.

3.2. *MODELING AOF IMPORTANT PHOTOSYNTHETIC REACTIONS*

Photosystem II (PSII): Inclusion of thylakoid membrane as a separate compartment enabled us to include mechanisms associated with the photosystem II complex (PSIIa, PSIIb, PSIIc, & OEC) (Figure D1B). The following mechanisms were included: (i) electrons from the S-cycle utilizing oxygen evolving complex (OEC) are transferred to 4 P680 reaction center (OEC a-d), (ii) the uncharged reaction center is protonated by one photon and an electron is transferred to the PSII-bound plastoquinone (Q_A) on the stromal side (PSIIa), and (iii) in two consecutive reactions, electrons from Q_A ($2\ Q_A \rightarrow 2\ Q_A^-$) were transferred to an unbound Q_B ($Q_B \rightarrow Q_B^{2-}$) (PSIIb-c). This Q_B^{2-} gets protonated to PQH_2 to diffuse through thylakoid membrane. Protonation and unprotonation of P680 reaction center joins water-splitting to PSIIa; and hence creates a flux mode connecting PSII reactions and oxygen evolution (Lea & Leegood, 1999).

92

Cytochrome b₆/f (CBFC): The Q-cycle mechanism of operation of Cytochrome b_6/f (Figure D1D) was implemented: (i) the PQH_2 diffusing through thylakoid membrane, and regenerated at PSII, reduces the Rieske protein ([2Fe-2S] cluster) while releasing a proton in the lumen and a semiquinone on lumen side of the complex (CBFCua), (ii) the reduced form of Rieske protein reduces plastocyanin (PC) in lumen through cytochrome f (CBFCub), (iii) the semiquinone on lumen side transfers an electron to heme b_l (low state), while semiquinone is converted to quinone on stromal side (CBFCuc-d), (iv) the heme b_l then transfers an electron to heme b_h (high state) on stromal side and converts the quinone to a semiquinone on stromal side (CBFCue), and (v) the heme bl transfers an electron to heme bh on stromal side while converting the semiquinone on stromal side to fully reduced quinone (PQH2) on stromal side using protons from stroma (CBFCuf). The 1st half of Q-cycle utilizes the steps (i-iv), while only the 2nd half of utilizes the steps (i-iii, v) to regenerate PQ and PQH_2 which then diffuse within the membrane. The cyctochrome b6/f complex has been modeled in thylakoid membrane (CBFCua-uf) as well as cytosolic membrane (CBFCpa-pf). This process overall mediates transfer of 4 cytosolic protons to lumen side (Lea & Leegood, 1999).

Photosystem I (PSI): The PSI machinery (Figure D1C) is modeled as a two-step process: (i) the P700 reaction center upon absorption of a photon gets protonated and through a series of electron transfers reduces ferredoxin (PSIa), and (ii) the protonated P700 reaction center in PSI complex is reduced by plastocyanin in thylakoid (PSIb). The plastocyanin would be replaced by ferrocytochrome under alternate cytochrome b_6/f electron transfer (PSI_2) (Lea & Leegood, 1999).

Ferredoxin (NADP⁺) oxidoreductase: The ferredoxin reduced at PSI is then, used to regenerate $NADP^+$ to NADPH (FNOR) to be utilized by all other parts of cellular metabolic processes.

Alternate electron flow pathways: Other alternate electron transfer process have also been modeled: (i) Ferredoxin (PQ) reductase, catalyzes reduction of quinone (PQ → PQH$_2$) using ferredoxin via heme; (ii) NAD(P)H dehydrogenase complexes, (iii) the Mehler reaction (MEHLER), (iv) the cytochrome c oxidases (CYTBD), and (v) the quinone (PQ) oxidase (CYO1b). These alternate electron flow pathways have been modeled in both the membranes (cytosol and thylakoid) (Lea & Leegood, 1999).

3.3. THERMODYNAMICS – CALCULATION AND ADJUSTMENT OF $\Delta_R G^{\prime O}$ TO $\Delta_R G^{\prime M}$, $\Delta_R G^{\prime M}{}_{MIN}$, AND $\Delta_R G^{\prime M}{}_{MAX}$

The thermodynamic analysis of the metabolic reconstruction was performed according to the group contribution method developed by Jankowski and others (Jankowski, Henry, Broadbelt, & Hatzimanikatis, 2008). The $\Delta_r G^{\prime o}$ was calculated using $\Delta_f G'$ and $U_{f,est}$ extracted from previously published model of *E. coli* (iAF1260) (Feist et al., 2007). $\Delta_r G^{\prime o}$ refers to the reference state of 1M, which was adjusted to the reference state of 1mM ($\Delta_r G^{\prime m}$) using eq. (4.1-4.2).

$$\Delta_r G^{\prime o} = \sum_{i=1}^{m} n_i \Delta_f G_i^{\prime o} \tag{4.1}$$

$$\Delta_r G^{\prime m} = \sum_{i=1}^{m} n_i \Delta_f G_i^{\prime o} + RT \ln \left(\prod_{i=1}^{m} x_i^{n_i} \right) \tag{4.2}$$

where; n_i is the stoichiometry of the metabolite *i* (negative if metabolite is a reactant and positive if metabolite is a product), *R* is the gas constant in kcal K^{-1} mol^{-1}, *T* is temperature in K, and x_i is the metabolite activity as a proxy for metabolite concentration. Metabolite activity was set to 1mM for all metabolites except H$^+$, H$_2$O, H$_2$, and O$_2$. The reference concentrations of H$_2$ and O$_2$ were set to the saturation concentration for these metabolites in H$_2$O at 1 atm and 298.15 K. We also

take into account the proton gradient for all reactions involving H^+ transport across the biological membrane, which can be given by eq. (4.3-4.5).

$$\Delta_r G'^m = \Delta_r G'^m_{(transport)} + \Delta_r G'^m_{(intracellular)} \tag{4.3}$$

$$\Delta_r G'^m_{est(transport)} = \Delta_{\Delta pH} G \tag{4.4}$$

$$\Delta_{\Delta pH} G = -2.3 \, hRT \, (\Delta pH) \tag{4.5}$$

where; ΔpH is the pH difference across the membrane, h is the number of protons transported across the membrane, and $\Delta_r G'''_{(interacellular)}$ is the metabolic component of the transport reaction. *Synechocystis* sp. PCC6803 has been known to grow best between pH levels of 7-8.5. We assume that the ΔpH across the cytoplasmic membrane is 0.5. We do not take into account the pH difference across carboxysomes. All the $\Delta_r G'^o$ and $\Delta_r G'''$ used are provided in the supplementary information (Table D1).

We estimated variations resulting from two possible sources: (i) variation due to large changes in activities of metabolites, and (ii) variation from uncertainty due to the Gibbs free energy of formation of metabolites. The variation arising from both these sources can be given by the following eq. (4.6-4.8) (Jankowski et al., 2008).

$$U_{r,est} = \sqrt{\sum_{i=1}^{m} n_i^2 U_{f,i}^2} \tag{4.6}$$

$$\Delta_r G'^m_{max} = \sum_{i=1}^{m} n_i \Delta_f G'^o_i + RT \ln \left(\prod_{i=1}^{products} x_{max}^{n_i} \right) + RT \ln \left(\prod_{i=1}^{reactants} x_{min}^{n_i} \right) + U_{r,est} \tag{4.7}$$

$$\Delta_r G'^m_{min} = \sum_{i=1}^{m} n_i \Delta_f G'^o_i + RT \ln \left(\prod_{i=1}^{products} x_{min}^{n_i} \right) + RT \ln \left(\prod_{i=1}^{reactants} x_{max}^{n_i} \right) - U_{r,est} \tag{4.8}$$

where; x_{min} is minimal metabolite activity assumed to be 0.00001 M, and x_{max} is maximal metabolite activity assumed to be 0.02 M. As we did for calculation of $\Delta_r G'^m$, the physiological ranges for dissolved gases is lower than that of other metabolites, so we set x_{min} of H_2, O_2, and CO_2 as 10^{-8} M, and x_{max} for these dissolved gases was set to saturation concentration in water at 1 atm and 298.15 K which is 0.000034 M, 0.000055 M, and 0.0014 M, respectively. Eqs. (6-8) provide upper and lower limits for the thermodynamic estimates.

The reaction reversibility information was initially set from that used in a previously published model of *Synechocystis sp.* PCC6803 (Knoop et al., 2013). Reactions absent in the previously published model were assumed to be reversible. We then used the above calculated $\Delta_r G'_{min}$ and $\Delta_r G'_{max}$ to constrain directionality: (i) if the $\Delta_r G'_{max}$ was less than zero, the reaction was strictly irreversible; (ii) if the $\Delta_r G'_{min}$ was greater than zero, the reaction was thermodynamically infeasible in the forward direction; and (iii) if $\Delta_r G'_{max}$ was greater than zero and $\Delta_r G'_{min}$ was less than zero, the reaction was considered to be reversible. To detect the presence of thermodynamically infeasible loops, we coupled our analyses with flux variability analysis (*Materials and Methods, Section 3.6*). We checked if flux ranges were consistent with range of $\Delta_r G'^m$, else the directionality was modified.

The reference state for metabolite concentrations, on which our analyses is based ($\Delta_r G'^m$) (approx. 1mM), is different than the 1M ($\Delta_r G'^o$). However, intracellular metabolite concentrations are known to be significantly different (Bennett et al., 2009). To keep this in consideration, we calculated the Gibbs free energy change of formation ($\Delta_r G'_{min}$, $\Delta_r G'_{max}$) for all reactions in 0.00001 - 0.02 M concentration range, respectively. This range of calculated free energy of formation was used determine if the reaction could occur in either direction (forward and backward) or only one.

If the entire range including the uncertainty was positive or negative, the reaction was constrained to backward or forward direction respectively.

3.4. BIOMASS COMPOSITION

Biomass composition for some of the previously published *Synechocystis sp.* PCC6803 metabolic network computational studies (J. Nogales et al., 2012) have been extracted from other experimental studies (Shastri & Morgan, 2005). However, the deviations from the actual stoichiometric weightage of metabolites which contribute towards growth rate, amongst previously published reconstructions, only varies in the order of 10^{-6}. Hence, we chose to adopt the biomass objective equation from one such study, which had 114 biomass components (J. Nogales et al., 2012).

3.5. LIGHT COMPOSITION

To account for optical variation of photosystem I (PSI) and photosystem II (PSII) during the diel cycle, we resolved sunlight flux into fraction of photons interacting with PSI, and PSII. This was implemented according to the method proposed by Chang and others (Chang et al., 2011). Light from a given source can be decomposed into different wavelengths based on its spectral composition. To generate a spectral decomposition reaction of light, we define the spectral bandwidth that drove a particular reaction, in our case, Photosystem I, and II. These enzymes drive the photon-utilizing reactions. The general procedure to derive effective spectral bandwidth is shown as such. Absorption (activity) spectra for each reaction catalyzed by above mentioned enzymes can be obtained from previously published literature (Tomo et al., 2012; Watanabe et al., 2014). To define effective spectral bandwidth, we digitized the data from figures of absorption curves for each of these proteins. Further, the data was processed and analyzed within MATLAB

to get 1 nm resolution spectrum across the experimentally surveyed spectrum. The maximum reaction activity value within the interpolated data was identified to 1 nm precision and used to calculate the full-width half-maximum (FWHM) spectral bandwidth. This was considered to correspond to the spectral range of wavelength bound by the wavelengths at which half the maximum activity was observed. This range of wavelengths was considered the spectral bandwidth within which a given reaction would happen, in this case Photosystem I/II.

   a. *Photosystem I*: The absorbance spectrum of an unbound isolated Photosystem I (phycobilisome free) of Synechocystis sp. PCC 6803 was obtained from previous studies (Watanabe et al., 2014). Both red and blue wavelength ranges of light were found to be absorbed, these are treated separately by duplicating the reaction for blue light range and red light range. The FWHM was determined separately for both ranges. The effective spectral bandwidth for PSI were from 392 to 452 nm with peak absorbance at 438 nm for blue light, and 664 nm to 692 nm with peak absorbance at 680 nm for red light.

   b. *Photosystem II*: The absorbance spectrum of an unbound isolated Photosystem II (phycobilisome free) of Synechocystis sp. PCC 6803 was obtained from previous studies (Tomo et al., 2012). Photosystem II also shared similar absorption wavelengths to Photosystem I. Both red and blue wavelength ranges of light were found to be absorbed, these are treated separately by duplicating the reaction for blue light range and red light range. The FWHM was determined separately for both ranges. The effective spectral bandwidth for PSI were from 404 to 457 nm with peak absorbance at 441 nm for blue light, and 661 nm to 684 nm with peak absorbance at 675 nm for red light.

To convert spectral irradiance to composition of light, we used eq. (4.9) & (4.10).

$$L = \frac{E_\lambda}{EN_A} \qquad (4.9)$$

$$E = \frac{hc}{\lambda} \qquad (4.10)$$

where, E_λ is the spectral irradiance at wavelength (λ), L is the photon flux, N_A is the Avogadro's number, E is the photon energy, h is the Planck's constant, and c is the speed of light. Once we calculate the ratio of each of the wavelengths (we have resolved it up to 1 nm for each absorption spectra, see above). We, then, calculate the stoichiometry ($S_a{}^b$) of photons within the bandwidth region [a, b] for the spectral decomposition reaction of light using eq. (4.11)

$$S_a^b = \frac{\int_a^b L(\lambda).d\lambda}{\int_{380}^{750} L(\lambda).d\lambda} \qquad (4.11)$$

Photons were only modeled within the visible range of spectrum, such that photons interacting with photon dependent enzymes formed the components of photons. The remaining photons were considered to be non-interacting. Repeating for each of the five peaks for all the enzymes combined, we model the chemical reaction given by eq. (4.12)
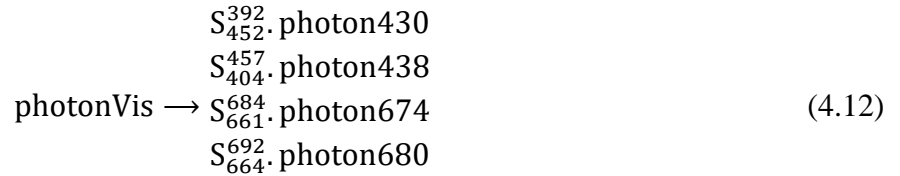
$$\text{photonVis} \longrightarrow \begin{array}{l} S_{452}^{392}.\text{photon430} \\ S_{404}^{457}.\text{photon438} \\ S_{661}^{684}.\text{photon674} \\ S_{664}^{692}.\text{photon680} \end{array} \qquad (4.12)$$

3.6.   FLUX BALANCE ANALYSIS (FBA)

Flux Balance Analysis (FBA) is a mathematical framework used to calculate the flow of the metabolites through the metabolic network at steady state (Feist & Palsson, 2010). A stoichiometric matrix S, of size M by N, corresponding to metabolic reactions (*N*) and metabolites (*M*) can be calculated based on the metabolic model. Using this matrix a system of differential equations can be written down for the rates of change of concentrations of each metabolite. Under

the assumption of steady state, reaction rates normalized to biomass production, a. k. a fluxes given by this system of differential equations becomes a system of linear equations which is set to zero, given by eq. (4.13).

$$\sum_{j=1}^{N} S_{ij} \cdot v_j = 0 \quad (OR) \quad S.v = 0 \tag{4.13}$$

Here, v is a vector of reaction flux and S_{ij} represents the stoichiometric coefficient for i^{th} metabolite in j^{th} reaction. To find the fluxes an objective function is chosen that is hypothesized to be optimized by the organism, such as its growth rate. This makes it a linear programming problem (LPP) that can be solved by standard techniques by imposing additional constraints, discussed below, in addition to Eq. 1. The objective function most commonly used for such models is an equation describing the growth rate of the organism. Growth rate reactions are described as given in eq. (4.14),

$$\max \sum_{j=1}^{N} c_j \cdot v_j \to \mu \quad (OR) \quad Max \ C^T.v \tag{4.14}$$

In the above equation, c_j and v_j refer to the weight in final biomass and flux of the product of the jth reaction respectively, and μ refers to the growth rate of the organism. Constraints can be applied to secretion/uptake fluxes of metabolites, or applying upper (α_j) and lower (β_j) limits to intracellular reaction fluxes (v_j), given by eq. (4.15),

$$\alpha_j \leq v_j \leq \beta_j \tag{4.15}$$

Reversible reactions can take either negative or positive values of fluxes, while irreversible values were constrained to take only positive values. Further, if any reactions were turned off, inactivated or deleted, the flux through the reaction was set to zero, as described by eq. (4.16),

$$v_j = 0 \tag{4.16}$$

The mathematical definition of the problem formed by equations above (13-16) is called Linear Programming (LP). The LP problem can be constrained by experimentally measured uptake or secretion fluxes of various exchange reactions for nutrients and/or products. Constraints of thermodynamic nature can be implemented by assigning reversibility of the reactions by calculating net change in Gibb's free energy. Variations to this problem are possible by using different choices of objective functions. The most common objective function used is microbial growth rate, which is also the one used here. A bi-level optimization problem can be formed by nesting one objective function into another (See section 3.7 of this chapter). The linear programming problem was implemented using COBRA Toolbox with Gurobi 4.6.1 on MATLAB R2014b (D. Hyduke et al., 2011).

3.7.   *FLUX VARIABILITY ANALYSIS (FVA)*

Genome-scale metabolic networks are usually under-determined systems that have many different possible solutions. Solutions are usually chosen based on optimizing one or more objective functions. Even so it is very common to find multiple solutions that yield the same value of the objective function under the same set of constraints. To investigate this flux solution space, approaches have been developed in previous studies (R. Mahadevan & Schilling, 2003; Reed & Palsson, 2004), called alternate optima analyses. Here, we implemented one of the basic types of alternate optima analyses called flux variability analysis (FVA).

In FVA we formulate a bi-level optimization LP problem as follows. We start by determining the optimal value (Z_{obj}) of the linear objective function using equations (4.13-4.16) given above. Now, we use this value to constrain the original objective function (for which the optimal value was determined) and perform N (number of variables in the problem; for metabolic networks, number

of reactions) iterations, where in each iteration a new objective function is assigned. This new objective function is flux through one of the reactions. This flux is maximized (eq. (4.17a)) and minimized (eq. (4.17b)), while keeping the rate of growth at its optimal value. This procedure after 2N iterations (maximizing and minimizing each reaction flux), generates a range of flux values for each reaction within which exactly the same value of the original objective function can be achieved. The set of ranges therefore define the boundaries of the optimal solution space.

$$\text{max } v_j \qquad\qquad\qquad\qquad \text{min } v_j$$
$$\text{s.t.} \quad S.v = 0 \qquad\qquad\qquad \text{s.t.} \quad S.v = 0$$
$$C^T.v = Z_{obj} \qquad\qquad\qquad C^T.v = Z_{obj}$$
$$\alpha_j \le v_j \le \beta_j \quad \text{for } j = 1 \dots n \quad (4.17a) \qquad \alpha_j \le v_j \le \beta_j \quad \text{for } j = 1 \dots n \qquad (4.17b)$$

3.8. GROWTH CONDITIONS AND SINGLE GENE DELETIONS

Autotrophic growth of Synechocystis sp. PCC6803 was simulated by setting the uptake rates of all the sources of carbon except carbon dioxide (CO_2) to zero. The CO_2 uptake was set to 3.7 mmoles/gDW/h (Shastri & Morgan, 2005). The light uptake rates were first determined by the growth condition being carbon limited or light limited. For our simulations at the given CO_2 uptake flux, the minimum amount of light required for maximum growth was calculated by dual maximization; i.e. calculating maximum growth rate (leaving light uptake flux unconstrained) using maximization of growth rate as objective function and then minimizing the light uptake (constraining growth rate equation to the value calculated in previous step) using minimization of light uptake as objective function. The light above this minimum value is referred to as extra light. To simulate heterotrophic growth, we shut off the reaction for light absorption and set the lower bounds of glucose uptake to 0.85 mmoles/gDW/h, as per previously published studies (C. Yang et

al., 2002). The carbon dioxide (CO_2) exchange with bulk was set to secretion by setting the lower bound to zero.

To simulate gene deletions, methods have been developed previously (Fong & Palsson, 2004; Segrè et al., 2002; Z.-X. Xu, 2008). LP and QP optimization based frameworks have been proposed in the past (Segrè et al., 2002). However, a study has shown, for *E. coli* K-12, *in silico* FBA predicted mutant and wild-type flux distributions are more accurate (Fong & Palsson, 2004). Hence, for our single gene deletion analysis, we used FBA to simulate gene deletions. To implement it, we apply constraint given by eq. (8) to all the reactions that are associated to the single gene deletion being simulated.

4. RESULTS AND DISCUSSION

4.1. IMPROVEMENTS IN NETWORK RECONSTRUCTION

A genome-scale metabolic reconstruction of *Synechocystis* (iSyn816CJ) was developed, based on existing genomic and biochemical information using previously established protocol (Thiele & Palsson, 2010), as detailed in the *Materials and Methods* section. The initial reconstruction contains 816 genes (Figure 4.1A), 925 metabolites (Figure 4.1B), 1060 reactions (metabolic, transport, GPR and non-GPR based reactions) distributed over 56 subsystems and 7 compartments (extracellular, cytoplasm, periplasmic space, periplasmic membrane, thylakoid lumen, thylakoid membrane, and carboxysomes) (Figure 4.1C). This is larger than the previously largest reported reconstruction by 142 genes. The reconstruction was built on top of previous reconstructions. A network analyses of previously built reconstructions (Knoop et al., 2013; J. Nogales et al., 2012; Saha et al., 2012) resulted in a net difference of 103 genes and 256 reactions. We also added genes and reactions from previous models, if they were missing from our reconstruction.

Reactions which have not been previously modeled belong to tRNA biosynthesis, amino acid and amino-sugar metabolism, sugar metabolism, nucleotide metabolism, nitrogen metabolism, terpenoid metabolism, fatty acid metabolism, photosynthesis, biosynthesis of secondary metabolism, and ion metabolism. We found that there were many inconsistencies in the gene reaction association amongst the two previously published models, iHK677 (Knoop et al., 2013) and iJN678 (J. Nogales et al., 2012). In iSyn160CJ we fixed these inconsistencies; for e.g.
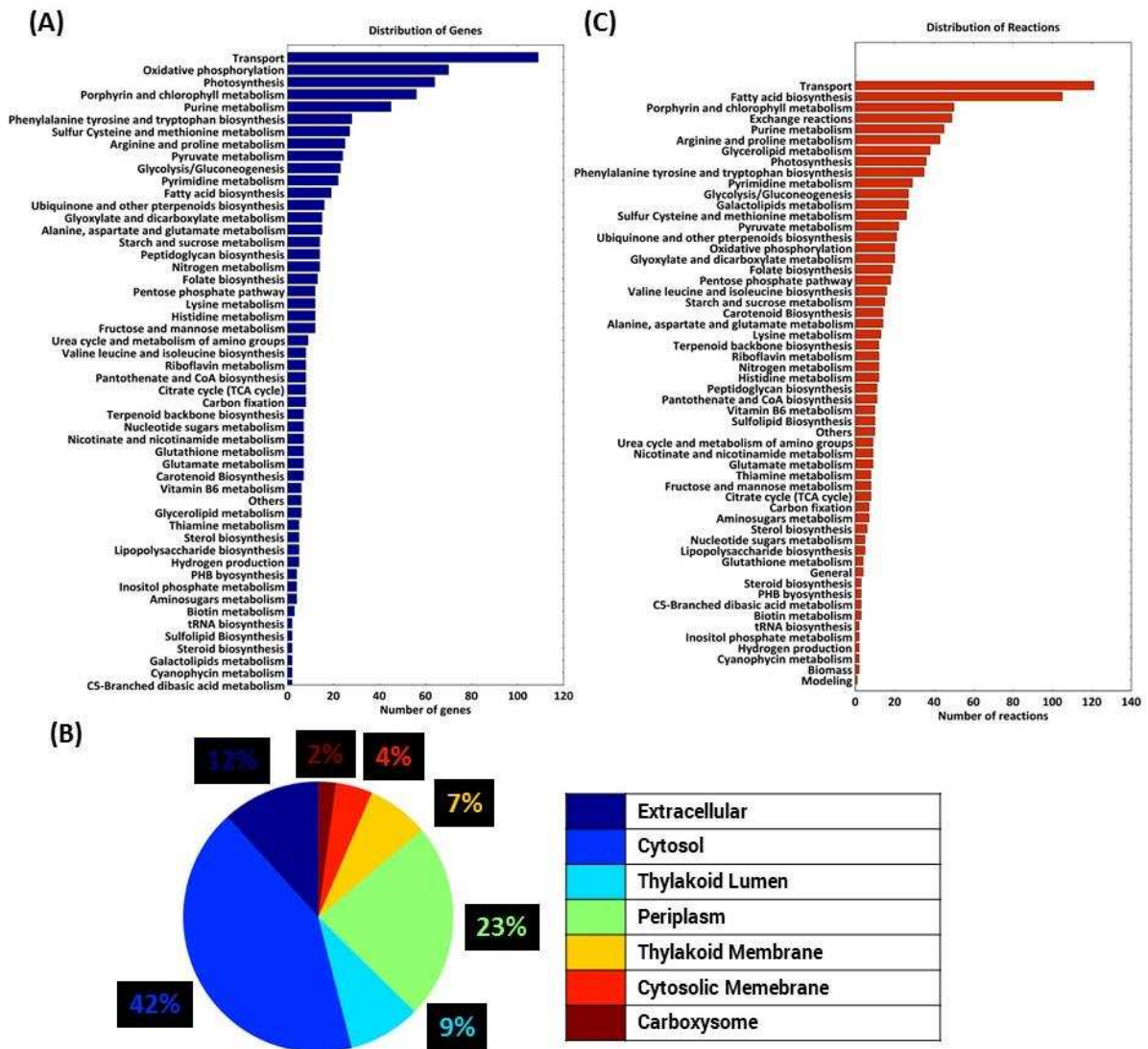
FIGURE 4.1: PROPERTIES OF THE METABOLIC NETWORK RECONSTRUCTION OF SYNECHOCYSTIS SP. PCC6803.
(A) Distribution of genes amongst various subsystems, (B) Distribution of metabolites amongst various compartments within the reconstruction, and (C) Distribution of reactions amongst various subsystems.

sll1392 (Fe2+ transporter) in iJN678, while slr1392 in iHK678; and sll1510 (G3P acyltransferase) in iJN678, while slr1510 in iHK677. We fixed 30 such instances of inconsistencies occurring in gene-reaction association, which describes whether the protein translated has a subunit or an isozyme. Several missing EC numbers were also added to facilitate construction of enzymatic information from within the model.

The photosynthetic electron transfer machinery was enhanced by inclusion of electron transfer mechanism through photosystems I and II (PSI and PSII), and electron transfer mechanism within cytochrome b6/f. To systemically study the photosynthetic process, we include thylakoid and cytosolic membranes as separate compartments. The photosynthetic alternate electron flow pathways (Battchikova, Eisenhut, & Aro, 2011; Cooley & Vermaas, 2001; Howitt, Udall, & Vermaas, 1999; Matsuo, Endo, & Asada, 1998) were modeled taking account of separate thylakoid and cytosolic membrane compartments. This implies that plastoquinone (PQ) involved in the electron transfer from PSII to cytochrome b6/f is spatially limited to thylakoid membrane, while plastocyanin (PC) and ferricytochrome involved in the electron transfer from cytochrome b6/f to PSI is spatially limited to thylakoid lumen. It is essential to include detailed photosynthetic and photo-oxidative machinery as it consists of proton-pumping reactions and electron transfer amongst various electron carriers such plastoquinol, semi-quinone, NADPH, NADH, ferricytochrome, and plastocyanin across the membranes; which determine the amount of light needed for the organism to grow, as shown by a previous model (J. Nogales et al., 2012). For details on how the modeling of photosynthetic reactions was implemented, please refer to *Materials and Methods section.* Once all the desired information about genes, reactions, and metabolites was assembled, we performed charge balancing based on charge on each metabolite

105

as informed by various databases. After charge balancing, we looked at the distribution of major electron carriers, ATP, protons, and H_2O (Figure D2).

Gap analysis was performed using GapFind (D. Hyduke et al., 2011) on COBRA toolbox on MATLAB finds the blocked metabolites and assesses network connectivity. We find that there are 89 intracellular downstream gaps and 47 intracellular root gaps, i.e. 136 blocked metabolites (Table D2). Gaps represent gaps in our knowledge, and we expect that with time these gaps would be filled. None of the gaps were in essential metabolites, and hence they did not affect the simulations of the model. We have a larger number of gaps than previous models, which may have arisen because of incorporation of more genes, reactions and compartments.

Our reconstruction iSyn816CJ improves upon the previous reconstruction by inclusion of multistep reaction cascades such as photosystems, glycine cleavage system, pyruvate dehydrogenases, cytochrome b6/f, etc. and expands upon the electron carrier promiscuity by inclusion of 15 peroxidases and NADH dehydrogenase catalyzed reactions.

4.2. *THERMODYNAMIC ANALYSIS CORRECTS REACTION DIRECTIONALITY AND IDENTIFIES UNFAVORABLE CYCLES*

The standard Gibbs free energy change of formation, $\Delta_f G'^o$, and reaction, $\Delta_r G'^o$, were estimated for compounds which occurred in both iAF1260 (*E. coli* model) (Feist et al., 2007) and our model. A total of 416 unique metabolites (~45%), common to both iAF1260 and iSynCJ816, are taken into account. Using these we were able to calculate $\Delta_r G'^o$ for 504 reactions (~49%) present in iSynCJ816. All $\Delta_r G'^o$ values were calculated using the contribution method implemented previously (Jankowski et al., 2008). The conversion from $\Delta_r G'^o$ to $\Delta_r G'^m$ is shown in *Materials and Methods, Section 2.3*. Our analyses indicates that $\Delta_r G'^m$ for 392 (~78% out of the

total calculated) reactions is less than or equal to zero. It should be noted that these numbers include the changes we made to the model.

In Figure 4.2 we show a comparison of $\Delta_rG'^o$ (magenta), $\Delta_rG'^m$ (green), and the range $\Delta_rG'_{min}$ - $\Delta_rG'_{max}$ (black). It can be seen that the range is quite significant reflecting uncertainties in the concentration of metabolites in the cell. We considered the reactions to be forward feasible, if $\Delta_rG'_{min}$ for the reaction is less than or equal to zero. Using this as a guide, we ensured that, the range, $[\Delta_rG'_{min}, \Delta_rG'_{max}]$ for every reaction overlapped with $\Delta_rG' < 0$.

Among the reactions previously considered reversible, we found 16 reactions which had the entire range ($\Delta_rG'_{min}$, $\Delta_rG'_{max}$) including uncertainty ($U_{r,est}$) less than zero; and hence, were
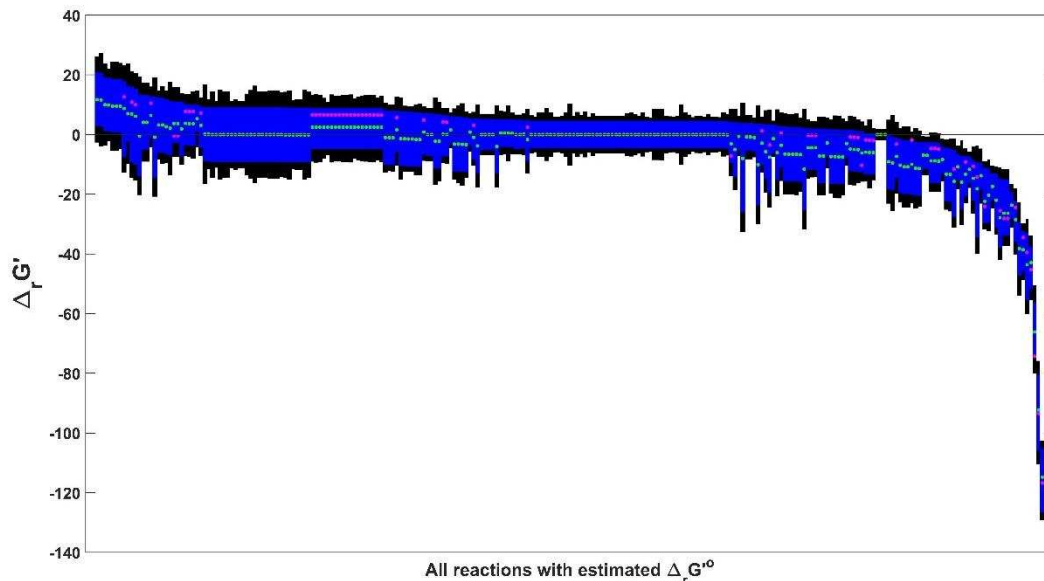
FIGURE 4.2: THERMODYNAMIC PROPERTIES OF THE REACTIONS FOR WHICH Δ_RG' WAS CALCULATED.
The range of possible Δ_rG' values (kcal/mol) for the reactions in descending order. Δ_rG^o (magenta), Δ_rG^m (green), range of Δ_rG^m (without uncertainty, blue), and range of Δ_rG^m (with uncertainity, black).

restricted to forward direction. Among the reactions previously considered irreversible, we found 4 reactions in which ($\Delta_rG'_{min}$, $\Delta_rG'_{max}$) including uncertainty ($U_{r,est}$) was positive; hence, the directionalities for these reactions were changed. The list of these reactions can be found in supplementary information (Table D1). These changes did not turn out to affect growth rates in

autotrophic or heterotrophic growth. We also compared the free energy values with calculated flux distribution to ensure that fluxes were predicted in the correct direction for reversible reactions. For reversible reactions, if the range, $[\Delta_rG'_{min}, \Delta_rG'_{max}]$ is entirely in the negative region, the calculated flux must be positive. On the other hand, if the above range is entirely in the positive region, the calculated flux must be negative. We changed the directionality of reactions for which the product of flux and Δ_rG' resulted in a positive value.

We next identified thermodynamically unfavorable cycles within the model using flux variability analysis. A cycle was considered thermodynamically unfavorable if flux through one of the reactions in the cycle achieved maximum possible flux by any given reaction. A reaction can achieve highest possible flux (1000 mmol/gDW/h), if it forms a futile cycle with some other set of reactions which reverses this reaction (Figure D3). After making thermodynamic corrections to the model, we identified a total of 13 unbounded reactions under heterotrophic growth
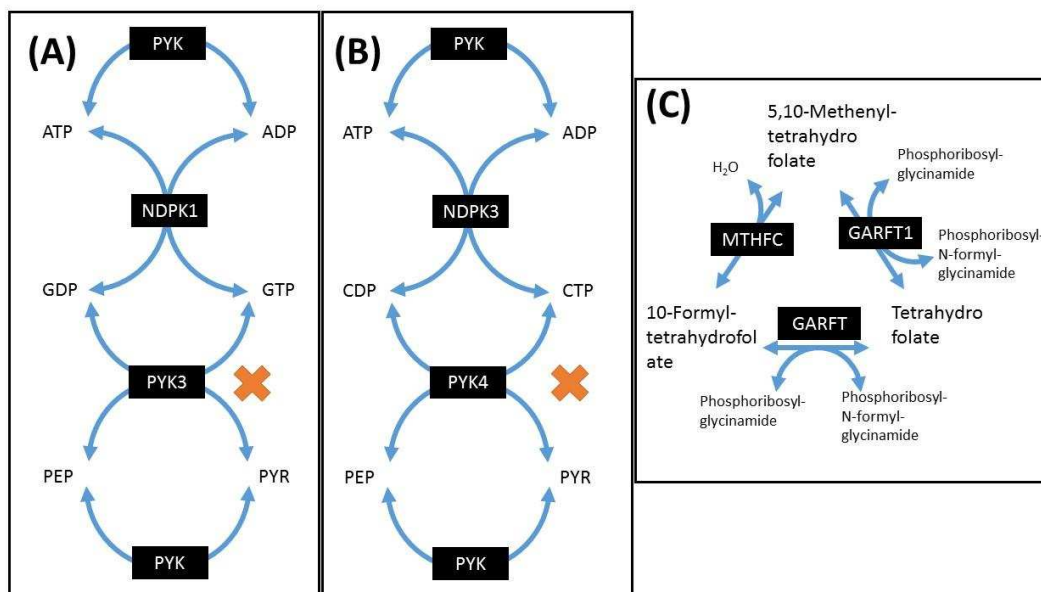
FIGURE 4.3: EXAMPLES OF THERMODYNAMICALLY INFEASIBLE CYCLES OR FUTILE CYCLES IDENTIFIED BY OUR ANALYSIS.
(A) Pyruvate kinase/Nucleoside diphosphate kinase (GDP utilizing)/Pyruvate kinase (GTP utilizing), (B) Pyruvate kinase/Nucleoside diphosphate kinase (CDP utilizing)/Pyruvate kinase (CTP utilizing), and (C) Interconversion of 5,10-Methenyltetrahydrofolate/10-tetrahydrofolate/tetrahydrofolate.

conditions. Out of the 13 reactions, a reversible reaction nucleoside-diphosphate kinase: ATP-GDP (NDPK1, sll1852), which regenerates GDP to GTP by transferring the phosphate from ATP, led to an infeasibility when forced in backward direction. This forms a futile cycle in combination with PYK and PYK3, both pyruvate kinase utilizing ATP and GTP respectively (Figure 4.3A). PYK3 was also identified as a thermodynamically unfavorable (range of Δ_rG' was entirely positive) in direction which utilizes GTP. Another similar futile cycle identified was NDPK3-PYK-PYK4. This futile cycle was based on interaction between pyruvate phosphorylation (PYK4), nucleoside-diphosphate kinase: ATP-CDP (NDPK3) activity, and ATP utilizing pyruvate kinase (PYK) (Figure 4.3B). These were resolved by removing pyruvate kinase reactions which utilize GTP and CTP. We chose to remove these reactions because the reactions were identified to be thermodynamically unfavorable towards utilizing GDP or CDP as substrates. However, other thermodynamically favorable reactions of pyruvate kinase exist within the reconstruction which utilize IDP, ADP, and UDP. We constrained the PPK1r to only forward direction. We identified a set of reversible reactions which were present from our reconstruction phosphoribosylglycinamide formyltransferase, which can utilize either 10-formyltetrahydrofolate (GARFT) or 5, 10-methenyltetrahydrofolate (GARFT1) (Figure 4.3C). To resolve the thermodynamically unfavorable cycle involving these two reactions, we removed GARFT1 to disrupt the thermodynamically unfavorable loop. The other two thermodynamically unfavorable cycles have been identified in one of the previous reconstructions (Saha et al., 2012) and we deal with them in the following way. The glycine cleavage system was initially modeled as single step reaction, but reverse reaction was modeled separately with only one (slr1096) of the four genes associated with the forward reaction. However, this reaction is a multistep reaction utilizing lipoylprotein. The enzyme commission numbers (EC numbers) of the net reaction involves decarboxylation

(GLYCLa, EC 1.4.4.2), deamination of glycine via utilization of tetrahydrofolate (GLYCLb, EC 2.1.2.10), and a lipoylprotein oxidoreductase (GLYCLc, EC 1.8.1.4, only reversible reaction within the multistep reaction cascade). We introduced the above reactions as mentioned. Finally, we identified leucine transaminase (an irreversible version of the reaction, LEUTAi) which transfers an amine group from L-leucine to produce 4-methyl-2-oxopentanoate. We remove the irreversible version because the reaction was calculated to be thermodynamically feasible in both directions. The thermodynamic information calculated for this reaction, LEUTA: $\Delta_r G'$ range = (-11.0990, 11.0990) kcal mol^{-1}, $\Delta_r G'^o$ = $\sim 10^{-14}$ kcal mol^{-1}, and $\Delta_r G'^m$ = 10^{-14} kcal/mol.

Our final reconstruction after taking thermodynamic feasibility into account contained 816 genes and 1045 reactions. While many enzymes are specific to electron carriers too, there are those which are able to utilize more than one electron carrier. Electron carrier promiscuity has been taken into account in our model and the distribution of reactions among various electron carriers is illustrated in Supplementary information (Figure D2). Previous studies (Knoop et al., 2013) have reported absence of a pathway which produces malate from isocitrate via glyoxylate, hence called glyoxylate shunt; and hence, has been left out of models. As shown by previously published results (Zhang & Bryant, 2011), a fully functional TCA cycle exists within *Synechocystis sp.* PCC6803, and this has been included in the reconstruction. The reconstruction includes an additional carboxysome compartment, where three main set of reactions in fixing carbon occurs: (i) carbonate dehydrogenase: conversion of HCO_3^- to CO_2, (ii) ribulose 1,5-bisphosphate carboxylase/oxygenase: CO_2 fixation and formation of 2-phosphoglycolate via oxygenase activity, and (iii) transport of by-products back to cytosol.

4.3. *ELECTRON TRANSFER IN THYLAKOID MEMBRANE*

The core electron transfer processes in the thylakoid membrane are well understood (Lea & Leegood, 1999). PSI and PSII are charged due to absorption of photons. Electrons originating within PSII reaction center flow through Q_A and Q_B to PQ located in thylakoid lumen. This PQ interacts with cytochrome b_6/f. The charged PSII returns back to original state via S-cycle of oxygen evolving complex. Electrons originating at PSI reduce ferredoxin. This reduced ferredoxin interacts with ferredoxin (NADPH) reductase. The charged PSI returns back to original state via oxidation of plastocyanin (reduced state). However photosynthetic and respiratory electron transfer processes include a number of other reactions that are less well understood, and model simulations may provide some insight into their possible physiological role. Below we discuss some of the key observations regarding the electron transfer machinery based on simulations of the model.

We also characterized various respiratory pathways according to maximum flux achievable at maximum growth rate (Table D3). We find that under carbon limited (CL) conditions, the achievable flux of succinate dehydrogenase (SUCD) was lowest amongst all the electron processing machinery, and that of superoxide dismutase (SOR) was highest amongst all the electron processing machinery. Previously published experimental studies have reported higher superoxide activities under carbon limited conditions as compared with light limited conditions (Badger, von Caemmerer, Ruuska, & Nakano, 2000), as suggested by our simulations. However, our simulations suggest lower flux through SUCD, while experimental studies suggest a higher flux through SUCD (Cooley & Vermaas, 2001). After carrying out simulations by fixing SUCD flux in thylakoid manually, we find that SUCD has a growth rate component associated to it (Figure D4). This means that higher flux at constant light and C_i through SUCD compromises

cellular growth objective as it pulls flux from 2-oxoglutarate, which is an important precursor for production of amino acids, which are in turn crucial for growth. As photosynthetic and oxidative machinery are fairly well characterized in cyanobacterial metabolism and given the experimental observations of high SUCD activity, we suspect that, the succinate dependency of the growth rate maybe more flexible than what is used in the model. Our flux variability simulations confirm that for same growth rate values, SUCD can have major portion of its flux directed towards its role in electron transfer under high light conditions (Figure 4.4).

We find that, in absence of electron transfer through cytoplasmic membrane, a non-zero minimal flux is required through CO_2 transporting NADPH dehydrogenase (NDH1_3u) and SUCD. Further, our simulations suggest that in absence of NDH1_3u, cytoplasmic membrane may be undergoing significant electron transfer. Absence of NADPH dehydrogenase in thylakoid (NDH1_3u), pushes the electron transfer through the cytoplasmic variation of NADPH
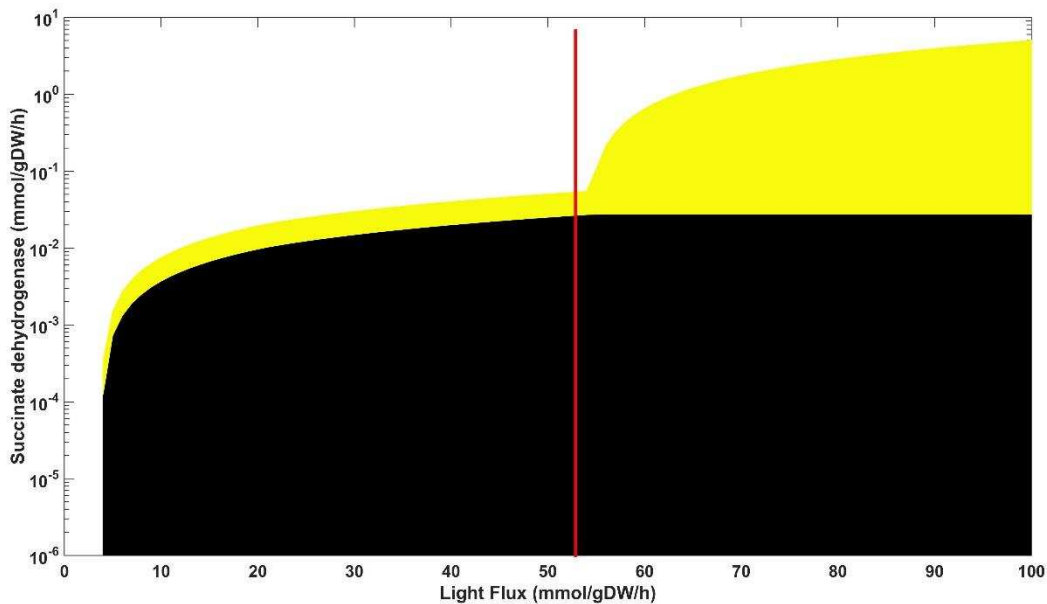
FIGURE 4.4: FLUX VARIABILITY OF SUCCINATE DEHYDROGENASE UNDER VARIOUS LIGHT UPTAKE CONDITIONS.
Red line indicates the light uptake under wild-type simulations, black region indicates the succinate dehydrogenase flux necessary for growth, and yellow region indicates electron transfer flux through succinate dehydrogenase.

dehydrogenase (NDH1_4pp). In presence of both variations, NADPH dehydrogenase could be active in either or both these variations. We did not find any experimental evidence that suggests any specific situation. Therefore, this could be possible due to modes of operation of photosynthesis and oxidative pathways, which are present in both, thylakoid and cytoplasmic membrane. It is also known that more than one electron transfer and oxidative pathways may be operational under any given environmental condition (Vermaas, 2001).

Previously published experimental studies have reported higher superoxide activities under CL conditions as compared with light limited (LL) conditions (Badger et al., 2000), as suggested by our simulations. However, a previous study in *Synechococcus* (Bailey et al., 2008) suggested a different alternative candidate for electron transfer, cytochrome oxidase. Given high super oxide production rate, it can be argued that higher localized oxygen concentration will generate higher superoxide dismutase to cytochrome oxidase activity under CL conditions than under optimal (or LL) conditions.

4.4.   *RUBISCO OXYGENASE AND LIGHT-INDEPENDENT SERINE PRODUCTION*

Recently, a light-independent serine production pathway has been characterized in *Synechocystis* (Klemke et al., 2015). Previously published metabolic reconstructions have shown that, in presence of photorespiration and absence of light-independent serine production pathway, all of serine is produced via the photorespiratory pathway (Juan Nogales et al., 2012). However, it is known that both serine production pathways are active during photosynthetic growth (Klemke et al., 2015). It is also known that RuBisCO oxygenase (RBCh) flux (which forms the photorespiratory precursor: 2-phosphoglycolate) forms 3%-5% of RuBisCO carboxylase (RBPC) flux (Huege et al., 2011; Knoop et al., 2013). Our simulations in presence of the newly identified light-independent serine production pathway predict that all of serine is produced via the light-

independent pathway and that photorespiration becomes dispensable. Photorespiration causes loss of carbon dioxide, which is the also the only source of carbon for autotrophic conditions. In presence of light-independent serine production pathway, the experimentally observed non-zero flux through photorespiration can be achieved via two mechanisms: (i) directly constraining RuBisCO oxygenase flux, resulting in catabolism of 2-phosphoglycolate, or (ii) inclusion of a growth requirement on 2-phosphoglycolate catabolism. Previous metabolic models have dealt with this issue by constraining RuBisCO oxygenase flux to 5% (J. Nogales et al., 2012) or 3% (Knoop et al., 2013). However, this appears to be an ad hoc fix and there is little discussion on the balance of photorespiratory and light-independent serine production pathway.

Experimental evidence suggests that photorespiration via RuBisCO oxygenase is an indispensable process under atmospheric conditions (Allahverdiyeva et al., 2011; Bauwe, Hagemann, Kern, & Timm, 2012; Eisenhut et al., 2008; Hackenberg et al., 2009; Hagemann, Weber, & Eisenhut, 2016). However, because there is no growth-associated demand (except L-serine) for photorespiration; in presence of light-independent serine production pathway, the FBA prediction for flux through photorespiration will be zero. A non-zero flux through the photorespiratory pathway requires a growth-associated demand because fluxes calculated using FBA rely on the composition of objective function (in this case, biomass growth equation). It should be noted that the composition of biomass is determined experimentally. Given that (i) photorespiration is indispensable (Eisenhut et al., 2008), (ii) biomass-related metabolites made by photorespiration can be made by another pathway (Klemke et al., 2015), and (iii) non-zero photorespiratory flux has been observed in previous experiments (Huege et al., 2011), photorespiration maybe playing roles which are beyond just meeting metabolic requirements of biomass growth. Previously experiments suggest that, indeed, photorespiration may play many

roles (Allahverdiyeva et al., 2011; Bauwe et al., 2012; Eisenhut et al., 2008; Hagemann et al., 2016; Knoop et al., 2010; Juan Nogales et al., 2012). Therefore the overall metabolic role of photorespiration cannot be captured adequately, if photorespiration is manually fixed to a certain (3% or 5%) value, under all environmental and genetic perturbations. It is important to leave photorespiration relatively unconstrained and allow the model to generate predictions on photorespiratory flux.

In order to generate a growth demand for photorespiratory flux, we devised a scheme where certain level of serine production occurs through light-independent serine production pathway while the remaining serine must be produced via photorespiration. Firstly, we assume a basal RBCh activity as 3% of total RuBisCO. Then, using this as a constraint, we calculated the viable flux of light-independent pathway. Knowing the flux through light-independent pathway, we calculated the light independent serine pathway flux as a fraction of net serine production flux required, under optimal conditions. We found that 3% of total RuBisCO flux is directed towards its oxygenase flux, and light-independent serine production was calculated to be 41.72% of the original flux through light-independent pathway. The simulation scheme devised here assumes that light-independent serine production may have an enzymatic bottleneck which still requires certain photorespiratory flux (3% of RuBisCO carboxylase flux) to meet growth associated serine demand, under light limited conditions (laboratory conditions). Our scheme is a plausible metabolic hypothesis about the role that photorespiration may be playing in the system, that can be tested through measurements of flux through the light-independent serine production pathway under different conditions. By setting an upper limit to light-independent serine production pathway, we were able to leave photorespiration unconstrained and capture effects on

photorespiratory flux under conditions when there is a demand flux associated with a given metabolite (see section 4.5).
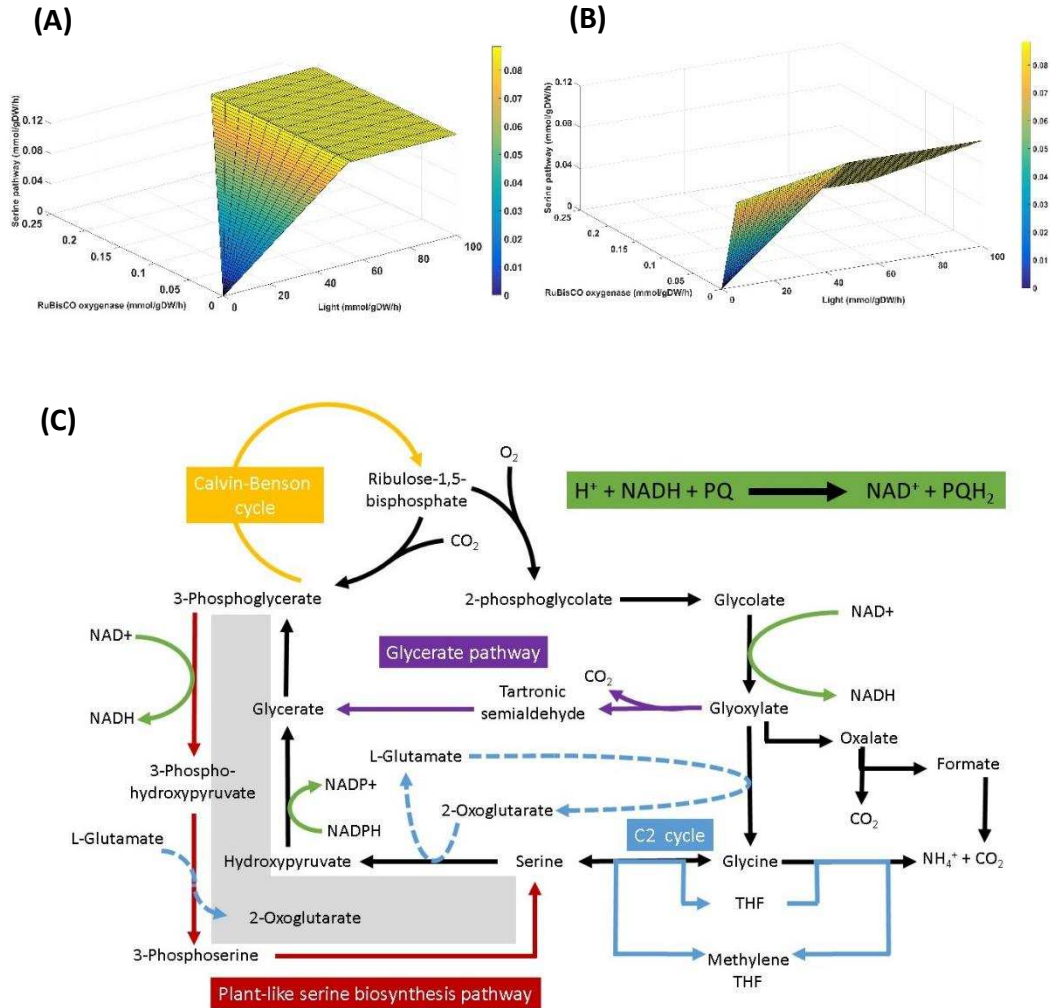
FIGURE 4.5: SERINE PRODUCTION VIA LIGHT-INDEPENDENT PATHWAY AND PHOTORESPIRATORY PATHWAY.
(A) Flux through the serine pathway is shown as a function of light and Rubisco oxygenase flux, in presence of NADH/PQ dehydrogenase. Light-independent serine production pathway and photorespiration may both contribute to increasing serine production; (B) Flux through the serine pathway is shown as a function of light and RuBisCO oxygenase flux, in absence of NADH/PQ dehydrogenase. Increase in light-independent serine production pathway results in decrease in photorespiration, while maintaining constant serine production; (C) Reaction scheme explaining the futile cycle that exists between the two pathways which results in increasing serine production as flux through light-independent pathway is increases.

We also simulated possible scenarios that result into flux through both the pathways. We constrained the flux through light-independent pathway and calculated the flux through

116

photorespiration. Interestingly, we came up with two possible scenarios for serine production: (i) total serine production through both the pathways combined increases with increase in flux through light-independent pathway (Figure 4.5A); and (ii) serine production via photorespiratory pathway decreases with increased flux through light-independent pathway (Figure 4.5B). The former scenario is associated with increase in flux through NADP reduction (to facilitate hydroxypyruvate reductase, HPYRR1i_syn) and NADH oxidation (to facilitate phosphoglycerate dehydrogenase, 3PGDH, and glycolate dehydrogenase, GLYCTO_syn). This scenario creates an intracellular loop given by hydroxypyruvate reductase (NADH utilizing) and 3-phosphoglycerate dehydrogenase (NADH utilizing) (grey box, Figure 4.5C). Therefore, our simulations suggest that increase in pools of NAD (oxidized state) and NADPH (reduced state) may result in increased photorespiratory activity and light-independent serine production simultaneously without resulting in a trade-off between these two pathways. Existence of intracellular loops in metabolism have long been predicted. In the case experiments match our predictions, we would see an example of a diverging topology of metabolic network convert into a cyclic intracellular loop which may result in accumulation of serine.

4.5.   *MODEL PREDICTS THEORETICAL INCREASES IN METABOLIC LOADS AND CARBON FIXATION*

We evaluated the potential for the cyanobacterium to produce possible bio-products. We took a generalized approach by assuming that each metabolite within the metabolic network can produce a bio-product. We, then, implemented a simple reaction scheme for converting each metabolite (precursor) into a bio-product. The flux of the bio-product production was constrained to an arbitrary value (0.1 mmol/gDW/h) and other intracellular fluxes were calculated.

117

Specifically, we analyzed HCO_3 uptake, RuBisCO carboxylase, RuBisCO oxygenase and oxygen production.

Given this constant metabolic load, our simulations indicate that only 426 metabolic loads could be simulated with the same light uptake as the wild-type light uptake. Our results for all the cases considered, under constant light and light-independent serine production, indicate that it is
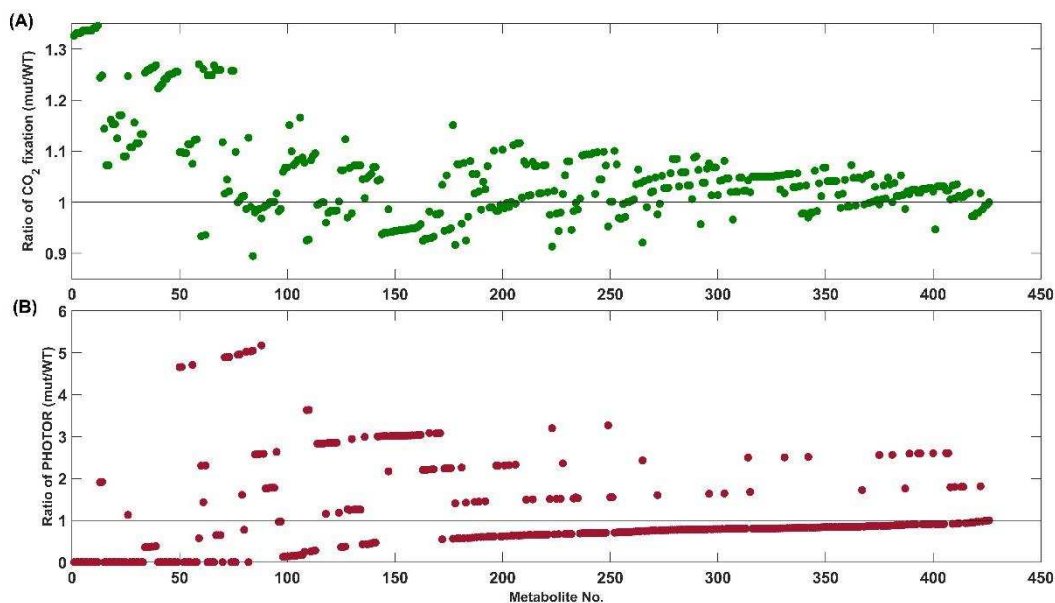
FIGURE 4.6: SECRETION OF VARIOUS METABOLITES MAY RESULT IN INCREASED (A) CO_2 FIXATION OR DECREASE IN (B) PHOTORESPIRATION.
The y-axis represents the ratio between a strain that secretes a given metabolite and the wild-type value of either CO_2 fixation (A) or PHOTOR (B). The x-axis represents the metabolite number being secreted.

theoretically possible to increase inorganic carbon (Ci) uptake (Figure 4.6A), increase CO_2 fixation, and reduce photorespiratory flux (Figure 4.6B). An important criteria to consider was also to quantify increases in photosynthetic rate. Here, we define photosynthetic rate as a ratio of light uptake per carbon fixed. We find that 76 metabolic loads resulted in a higher photosynthetic rate.

Another possible interpretation of these results concerns CO_2 remediation. CO_2 fixation reactions are faster than intracellular reactions contributing to growth, creating a bottleneck that

prevents greater CO_2 fixation (Knoop & Steuer, 2015). It has been argued that this bottleneck can be relieved by introducing a metabolic load, a high flux pathway, which transforms a metabolite into a molecule that does not contribute to growth and can be secreted form the cell (Oliver & Atsumi, 2015). By inserting a metabolic load (production of bio-product), our simulations of 426 metabolic loads show that it may be possible to process up to ~23% more inorganic carbon, ~12% more oxygen production, ~35% more CO_2 fixation (Figure 4.6A), and ~87% less photorespiration (Figure 4.6B). Though there are 55 cases of zero photorespiration, experimentally, it has been argued that photorespiration may be a necessary process (Bauwe et al., 2012; Eisenhut et al., 2008). Our simulations suggest that a two pronged approach which involves applying metabolic load (for industrially important molecules) and improving CO2 fixation is, indeed, possible.

4.6.   METABOLITE SECRETION

Our model currently has 929 metabolites. To calculate predicted theoretical yield, we removed the nutrient metabolites, protons, inorganic phosphates (pi, ppi, and pppi), and extracellular metabolites. After removing this metabolites, we were left with a list of 819 metabolites. Secretion of metabolites may lead to growth trade-off. Therefore, we performed simulations for 0%, 30%, 50%, 70% and 100% of the wild-type growth rate. 265 metabolites could not be secreted under any trade-off, leaving only 554 metabolites. For all growth shut-off conditions, the metabolite that achieves the maximum secretion (q_P) was molecular hydrogen (H_2). The plot (growth rate vs. q_P) (Figure D5) generates a decreasing trend with positive x-intercept and y-intercept.

Out of 819 metabolites, 265 metabolites could not secreted out. These also include the 136 blocked metabolites discussed in section 4.1. The remaining 129 metabolites cannot be secreted because of constraints placed by the network. One of the examples of metabolites which are not

part of the blocked metabolites but are present in the remaining 129 metabolites are those that are involved in intracellular loops, for e.g. individual S-states in oxygen evolving complex (S0, S1, S2, S3, and S4) which are modeled as individual metabolites. For the 554 metabolites remaining, we analyzed the flux distribution which results in the maximum product secretion under 50% growth rate reduction to obtain insight into the metabolic load on the network. Since high-dimensional flux vectors are difficult to analyze directly, we performed a principal component analysis and projected down to the space of the first three principal components that together capture 71% of variation. Principal component 1 (PC1) and principal component 2 (PC2) were effective in clustering metabolites which led to changes in PSI/PSII ratio (Figure 4.7A), while principal component 3 (PC3) was effective in clustering metabolites which led to changes in Pentose phosphate pathway and glycolysis (Figure 4.7B). PC1 and PC2 also captured metabolites which led to changes in flux of linear electron transport involving Ferredoxin-quinone reductase (FQR), one of the prime reasons associated to changes in PSI/PSII ratio. PC3 clustered metabolites in two distinct regions (Figure 4.7B). Region 1 related to metabolites which caused an increase in flux through pentose phosphate pathway (PPP) and led to fructose 6-phosphate production using transaldolase via sedoheptulose 7-phosphate. Metabolites which are part of this region include ribose phosphates, cofactors, nucleotides, terpenopids, phospholipids, oxaloacetate, succinate, etc. Region 2 related to metabolites which caused an increase in flux through glycolysis and led to fructose 6-phosphate (F6P) production using fructose bisphosphate phosphatase. Metabolites which are part of this region include nucleotide sugars, xylulose 5-phosphate, 3-phosphoglycerate, pyruvate, amino acids, lactate, ethanol, citrate, 2-oxoglutarate, acetyl-CoA, malate, sugars, glycogen, etc.

We also looked at the flux distribution of metabolites which may play a major role production commercially important chemicals. These include acetyl-CoA, shikimate, and ethanol. We chose these metabolites for their usefulness as commercially important molecules. Acetyl-CoA serves as a precursor for many important class of molecules such terpenoids, fatty acids, lipids, polyhydroxybutarate (PHB), tetracycline, and amino acids. Ethanol also has wide range of
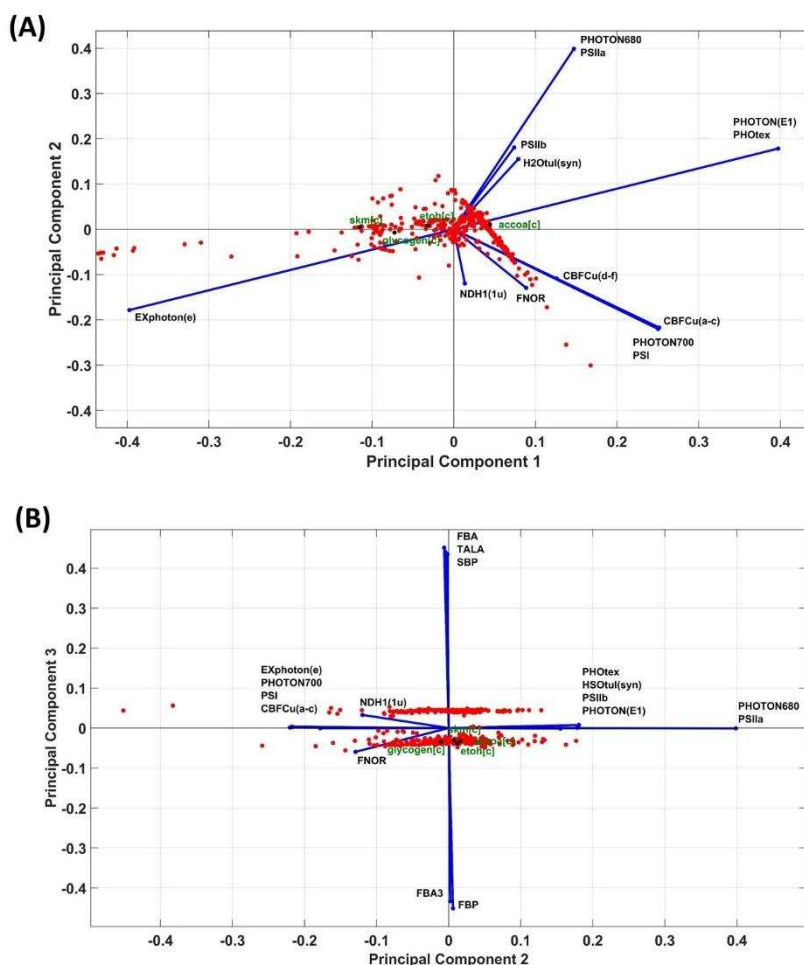
FIGURE 4.7: PRINCIPAL COMPONENT ANALYSIS ON FLUX DISTRIBUTION DATA OBTAINED FROM SIMULATION OF SECRETION OF A METABOLITE AT 50% GROWTH-RATE TRADE-OFF. (A) Comparison of PC1 and PC2 reveals the reactions which captured the most variation in PSI-to-PSII ratio, (B) Comparison of PC2 and PC3 reveals the reactions which captured the most variation in flux distribution to either glycolysis or TCA cycle, and pentose phosphate pathway. The flux distribution corresponding to each metabolite being secreted is represented by a red dot. The value of the dot represents the PC1 and PC2 of the flux distribution when transformed to the new space. Blue lines indicate the value of the coefficient of reactions that capture the most variation (>25% of maximum variation within a principal component) in either of the principal components.

industrial usefulness. Shikimate is an important precursor for antibiotic synthesis. Analysis of principal component 3 (Figure 4.7B, black dot, green text) indicates that, for all the four chemical species, fructose 6-phosphate is produced via gluconeogenesis, while their PSI/PII ratio reduced by ~20% of wild-type ratio. According to the Flux variability analysis for Acetyl-CoA production at a 50% growth reduction, there are 24 reactions for which flux varies significantly under these conditions. However, there were 38 more reactions for which flux varied more than 0.1 unit. These reactions belong to (i) pentose phosphate pathway such as transketolase 2, F6P and Xu5P phosphoketolases, and (ii) electron transfer machinery such as NADH dehydrogenase (active HCO3 transporters), NADPH dehydrogenase (4 protons utilizing). Overall, we find that oxidative reactions (cytochrome oxidase), decarboxylation reaction (malic enzyme), and cytoplasmic electron transport (cytochrome b6f) were turned on. As a result, water (leading to increased water splitting/PSII) and NADPH requirements were high. Similar trends were observed for lactate, ethanol, and glycogen secretion.

4.7.   FEATURES OF AUTOTROPHIC FLUX DISTRIBUTION

Autotrophic growth in the model can be simulated by applying constraints to light uptake and carbon dioxide (Figure D6). Other constraints on specific reactions can be found in the *Materials and Methods*. Given the flux constraint, an optimal solution was obtained using COBRA toolbox on MATLAB. Here, we leave the light uptake unconstrained such that maximum growth rate is obtained and then, we calculated minimum light uptake corresponding to maximum growth rate. RuBisCO oxygenase activity was left unconstrained but light-independent serine production was constrained such that ~3% of oxygenase was observed under low light conditions or optimal light conditions.

The simulations predicted some interesting flux distributions depending on whether the growth was simulated using bicarbonate (HCO_3^-) or carbon dioxide (CO_2) as carbon sources (Table D4). We used flux variability analyses at 100% growth to compare the flux distribution for these two conditions (CO_2 and HCO_3^-). Interestingly, we found that range of fluxes was wider for simulations with CO_2 compared to HCO_3^-. However, simulations with HCO_3^- resulted in overall range being closer to the experimentally determined solution. The three exceptions where experimentally observed flux value did not fall within calculated flux range for HCO_3^- growth simulation, but did for CO_2: (i) G3P dehydrogenase (GAPDi (nadp)), (ii) triose phosphate isomerase (TPI), and (iii) pyruvate kinase (PYK). It should also be noted that optimal light calculated for HCO_3^- simulations was higher than that of CO_2 simulations. Increased proton influx for HCO_3^- growth simulations suggests that ΔpH under these conditions is more positive than that for CO_2 growth simulations. The differences in growth condition seem to impact central carbon metabolism (CCM) flux distributions significantly. However, there is lack of experimental studies that compare these two autotrophic growth conditions.

Interestingly, we found that neither of the flux distribution gave an accurate description of entire flux distribution through CCM. In either growth simulations, we noticed discrepancy with experiments at three locations within CCM: (i) flux through oxidative pentose pentose phosphate pathway (~0.004% of carbon) was much lower than observed experimentally (~13.5% of carbon), (ii) flux calculated (~1% of carbon) through phosphoglucoisomerase also contradict experiments (~19% of carbon), and (iii) acetyl-CoA production from pyruvate was much lower (~0.02% of carbon) than that observed experimentally (35% of carbon). Interestingly, last two discrepancies arise due to a single pathway which consists of two reactions involved in production of acetyl-CoA (Figure D7): (i) acetyl phosphate production from fructose 6-phosphate via xylulose 5-

123

phosphate phosphoketolase (XU5PPK), and (ii) acetyl-CoA production from acetyl phosphate via phosphotransacetylase (PTAr). The first discrepancy only occurs when optimal light uptake is simulated. Interestingly, we find that by increasing the light uptake to ~7%, the flux through oxidative pentose phosphate pathway can increase to the experimentally observed value. Therefore, our analyses indicates that flux through this pathway is a mechanism to process excess electrons as a highly sensitive response to changes in light conditions.

As expected flux through TCA cycle is not cyclic, in either growth condition (CO_2 or HCO_3); only 2/3 of the pathway cycle is active. About 1% of total carbon flows through the reaction cascade consisting succinate to oxaloacetate conversion, and about 2% of total carbon flows through the reaction cascade from oxaloacetate to 2-oxoglutarate (AKG) production. We do not see any flux through GABA shunt or succinate semialdehyde. As discussed previously, oxaloacetate not directed towards citric acid cycle was involved in production of amino acids and nucleotides.

For photosynthesis to be functional, regeneration of ATP and electron carriers (NADPH) happen via photosynthetic light reactions. Previously published models have manually constrained the respiratory pathways to estimate flux distributions through photosynthetic system. Here, we refrain from doing this, to enable better understanding of the network properties and functional modes within photosynthesis and oxidative phosphorylation. Simulation with inorganic carbon, thylakoidal active CO_2 transport facilitator which hydrolyzes CO_2 to HCO_3 is an essential mechanism. As expected, the production of protons driving thylakoidal ATP synthase (ATPSu) is associated with water splitting at PSII by oxygen evolving complex (OEC), oxidation of PQH2 at cytochrome b6/f, reduction of NADPH via ferredoxin oxidoreductase (FNOR) supported by PSI.

Finally, our model correctly predicts the directionality among the flux branching points but differs in prediction of AcCoA which in our model is produced by F6PPK rather than PDH. It must be noted that identification of these pathways are novel in *Synechocystis sp.* PCC6803 and stoichiometric matrix of ^{13}C experimental flux measurements is devoid of these reactions, which are estimated by us to have significantly high activity during light-dependent growth.

4.8. HETEROTROPHIC FLUX DISTRIBUTION

Heterotrophic growth was simulated by allowing uptake of glucose in absence of light and maximizing growth rate [Supplementary Figure 8]. The glucose uptake was set to the experimentally determined value, 0.85 mmoles/gDW/h (C. Yang et al., 2002). Flux variability analysis simulations of reactions in central carbon metabolism were calculated and compared to experimentally determined fluxes (Table D5) (C. Yang et al., 2002). The main features of this comparison are discussed below.

Pentose phosphate pathway (PPP): Our simulations indicated that most of the glucose (~98% of total glucose uptake) taken up by the cell was routed through oxidative pentose phosphate pathway. However, experiments indicate that approximately 6% of glucose is funneled through glycolysis. Flux variability simulations (Table D5) does indicate a wide variability between these two pathways and also captures the experimental flux distribution (w.r.t. glucose funneled through glycolysis and PPP). Interconversions amongst pentose phosphate pathway, TCA cycle, and glycolysis produce precursors necessary for making other biomass components like amino acids, nucleic acids, fatty acids, and maintenance ATP. The main highlights of the simulation results of flux distribution through these three pathways are as follows. Production of R5P and XU5P occurs through ribose 5-phosphate isomerase (RPI, ~88% of total glucose uptake) and ribulose 5-phosphate 3-epimerase (RPE, ~160% of total glucose uptake), respectively. R5P production through RPI, results in favoring production of glyceraldehyde 3-phosphate (G3P) and

subsequently fructose 6-phosphate (F6P) via transketolase 1 (TKT1, ~82%) and transketolase 2 (TKT2, 79%), respectively. TKT2 is supported by transaldolase (TALA) via production of erythrose 4-phosphate (E4P). It should be noted that F6P and G3P produced here interface with glycolysis flux to produce substrates of TCA cycle, acetyl-CoA (AcCoA) and oxaloacetate (OAA). Consistent with experiments, we get a flux of about 1.4% of glucose uptake through sugar biosynthesis and metabolism. Hence, the glucose taken up by the cell gets partitioned towards sugar catabolism and PPP, F6P and G3P produced by PPP are utilized by to drive forward glycolysis to produce phosphoenolpyruvate (PEP) through to GAPD, PGK, PGM, and enolase (ENO, ~87% of glucose uptake). This PEP is divided towards production of Ac-CoA and OAA via PEP carboxylase (PPC) and pyruvate kinase (PYK). Inconsistent with experiments, the calculated ratio between pyruvate directed at amino acid synthesis and pyruvate directed towards TCA cycle was more than 1. Contrary to simulations of autotrophic growth, we found that TKT1, TKT2, and TALA reaction fluxes were more robust, as suggested by narrower flux range.

TCA cycle: As per recent state of knowledge (Zhang & Bryant, 2011), cyanobacterial TCA cycle has been found to be complete by identification of two reactions (EC 4.1.1.71, 2-oxoglutarate decarboxylase, *sll1981*; and EC 1.2.1.16, succinate semialdehyde dehydrogenase, *slr0370*) which convert, 2-oxoglutarate (AKG) to succinate via succinate semialdehyde as intermediate. This succinate interacts with the respiratory electron transport to produce fumarate via succinate dehydrogenase within the thylakoid and cytoplasmic membrane. We have also included the GABA shunt. Even, in presence of GABA shunt, we find that route through AKG decarboxylase is preferred. We also analyzed the changes that happen in the flux distribution when TCA cycle was forced to favor GABA shunt. However, no significant change in the growth rate or the fluxes of

126

TCA cycle was observed, but we do find additional flux modes are at play when the bypass is introduced. Flux through PGI becomes non-zero because flux through TALA increases 470% compared to that when newly identified bypass was allowed. Further, we find that fluxes of 44 reactions throughout the model changes by at least 10^{-4} mmoles/gDW/h. This could be because our simulations included loop law constraints. Loop law constraints choose the shortest possible path because of their basis being minimization of norm of fluxes. Further, a functional GABA shunt requires flux through two other reactions, one of which (ABTA, 4-aminobutanoate transaminase) siphons AKG from biosynthetic pathways to convert it to glutamate, forcing other reaction fluxes to readjust. Therefore, we argue that, in actuality, GABA shunt may be causing a weak metabolic overload making it an alternate optimal solution for cyclic flux through TCA cycle.

Serine production and RuBisCO: One major discrepancy that occurs with experiments is the flux estimated through RuBisCO oxygenase (RBCh). At the time ^{13}C heterotrophic flux measurement experiments (C. Yang et al., 2002) were conducted, the newly identified plant-like serine production pathway (Figure 4.5C) was still unknown in *Synechocystis sp.* PCC6803. For the same reason, the pathway has been missing from previously published models as well. In absence of light-independent plant-like pathway, RBCh flux was found to be ~9.5% of glucose uptake.

Respiratory electron transport: Succinate gets oxidized to produce fumarate via succinate dehydrogenase (SUCD) utilizing periplasmic membrane bound quinone, resulting from NADH dehydrogenase (NDH1_2p). In turn, quinol is oxidized by cytochrome oxidase bd (CYTBDpp) in periplasmic membrane. The oxygen reduction was associated to periplasmic membrane bound cytochrome oxidase (CYO1bpp). These set of reactions generate the necessary proton gradient

across, periplasmic space and cytosol, required to regenerate ATP within the cytosol. Our results improve upon previous published heterotrophic respiratory flux distribution by correctly predicting ATP production at the periplasmic membrane.

Overall, we find that our model qualitatively predicts heterotrophic flux distributions. However, we also find that in the light of recent advances made in identification of novel genes and reactions, such as plant-like serine production pathway and a functional cyclic TCA cycle, in *Synechocystis sp.* PCC6803, repeating 13C metabolic flux analysis to experimentally measure fluxes within glycolysis and pentose phosphate pathway under heterotrophic growth conditions may lead to better understanding of cyanobacterial metabolism under dark growth conditions.
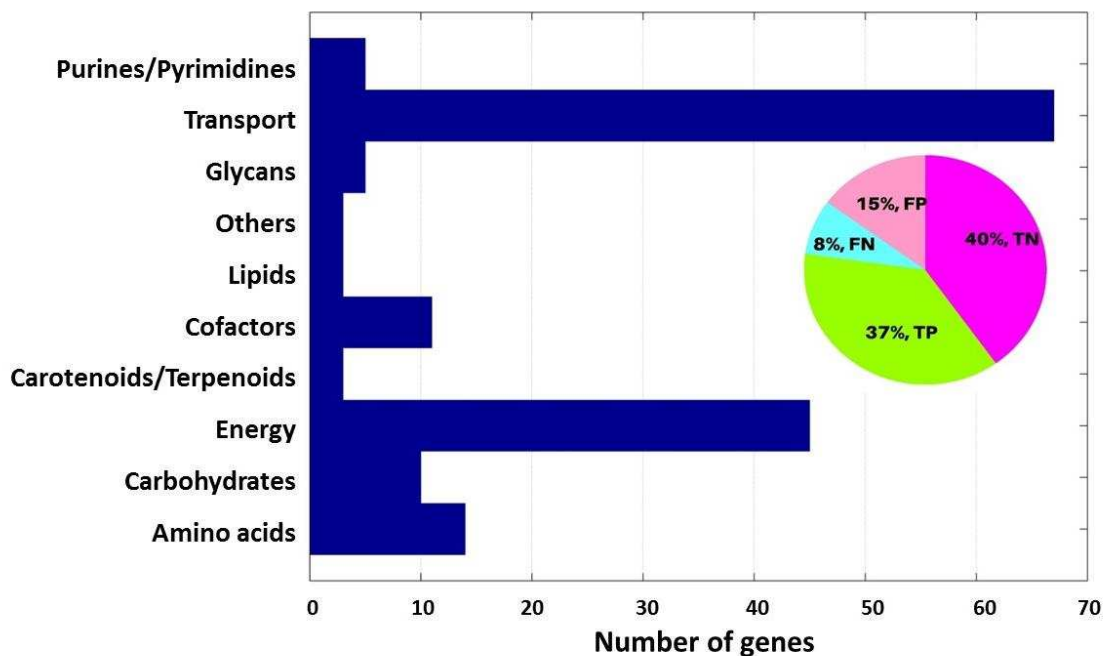

FIGURE 4.8: GENE DELETION ANALYSIS.
The distribution of genes, for which gene deletion was compared with experimental results, amongst various different subsystems. Inset shows true positive (TP), false positive (FP), true negative (TN), and false negative (FN) percentages.

4.9.   SINGLE GENE DELETION ANALYSIS (AUTOTROPHIC CONDITIONS)

Out of the 10 metabolic reconstructions published in the past, only two (J. Nogales et al., 2012; Saha et al., 2012) addressed performance of the model with respect to single gene deletions. Here, we compared single gene deletions with experimental studies. We manually performed a literature search to determine lethal gene deletions and compared it with model prediction in absence of reactions constrained by the gene. We also performed a global gene essentiality analysis. The numbers of essential genes predicted under autotrophic, heterotrophic, and mixotrophic conditions was 389, 307, and 306, respectively. A list of 305 (37.2%) essential genes appeared to be common to all the growth conditions. Comparison between literature search of 167 genes and their respective model predictions resulted in 37% true positive (growth predicted and observed), 40% true negative (growth neither predicted, nor observed), 15% false positive (growth predicted, but not observed), and 8% false negative (growth not predicted, but observed) (Figure 4.8). Therefore, out of 167 genes we looked at, 77% resulted in matches and 23% resulted in mismatches. Though, the false positive rate (FPR) of ~11% is low; the high proportion of mismatches is significant (p=0.002066; Fisher's exact test at significance level: 0.05).

The mismatches that were predicted as essential but are actually non-essential belong to Peptidoglycan biosynthesis, lipid biosynthesis, carotenoid biosynthesis, terpenoid biosynthesis, and ion transporters (zinc, calcium, and manganese). For example, slr0088, sll1653, and sll2010 which make transcripts of beta-carotene ketolase, demethylphylloquinone methyltransferase, and UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase, respectively are among this category of mismatches. The reason that deletion of these genes are predicted to be lethal in the model is that the products of the reaction catalyzed by these enzymes, which do not have an isozyme, are directly involved in the biomass growth objective. The mathematical formulation of the growth

rate objective is rigid and determines the optimal flux distribution. However, in reality, the organism can probably accommodate a different biomass composition; for example non-essential biomass components like terpenoids and peptidoglycans may be present in different amounts or may be substitutable by some other biomass components, and therefore show non-lethality. The rigidity of the growth rate objective function prevents growth being predicted in such instances by making all components of the biomass be equally essential. Hence, all the single deletions in non-isozymic or single pathway genes involved in synthesis of a non-essential biomass component will also lead up to lethality.

Consistent with previous studies (J. Nogales et al., 2012; Saha et al., 2012), our model continues to show a significantly low genetic robustness (47.4%) of *Synechocystis* under autotrophic conditions. Addition of newly added genes and pathways led to improvement in prediction of certain transporters, for e.g., sodium transport through cell, the single gene (slr1145) deletion which was predicted to be lethal in previous models, but is non-lethal in our model due to improvements in gene association (an isozymic multimer, sll1102 & sll1103 & sll1104) and addition of a literature-curated mechanism (Quintero, Montesinos, Herrero, & Flores, 2001).

5. CONCLUSION

Cyanobacteria have garnered much interest as a resource for harnessing naturally available sunlight and carbon dioxide to produce commercially important chemicals. Here, we presented a revised genome scale metabolic network model of a cyanobacterium, *Synechocystis sp.* PCC6803. This model makes improvements to the previously published models by including thermodynamics, incorporating a physiological mechanism for generating photorespiratory flux,

and increases the level of detail of the molecular mechanisms involved in photosynthesis and oxidative phosphorylation.

The model presented here was built on top of previous models. We also carried out free energy calculations for 510 reactions spanning central carbon metabolism (glycolysis, TCA cycle, and oxidative pentose phosphate pathway), amino acid biosynthesis and metabolism, purine metabolism, and pyrimidine metabolism in order to assess reaction directionality. We also identified thermodynamically unfavorable cycles between pyruvate kinase and nucleoside-diphosphate kinase (ATP/GTP utilizing) reactions. These were removed by applying reversibility constraints on these reactions. These were in addition to previously identified futile cycles in leucine transaminase (reversible and irreversible) reactions being catalyzed by two different enzymes.

Flux variability analysis also allowed us to identify the trade-offs between two primary serine biosynthesis mechanisms in *Synechocystis sp.* PCC6803: photorespiratory pathway and serine biosynthesis (recently discovered). Though, the model could be supported in absence of photorespiration, the energetic yields associated with these two pathways were found to be different. Our results also suggest that there could be additional metabolic roles of photorespiration which make this process indispensable under atmospheric CO_2 conditions. These metabolic roles maybe emerging from growth dependency on metabolites involved in 2-phosphoglycolate metabolism. Here we applied a constraint-based scheme, which simulates enzymatic bottlenecks in light-independent serine production pathway. However, other possible schemes could be designed such as growth dependence on 2-phosphoglycolate (by making part of biomass growth equation) or implementing a demand flux for 2-phosphoglycolate. Further, our results also allowed

us to make predictions and design constraints which capture some of the experimental observations of serine production from the two pathways discussed in this article.

We also analyzed the cyanobacterial phenotype (CO_2 fixation, Ci uptake, oxygen production, and photorespiration). Our analysis presents an example of how systems biology can be useful in designing strategies to improve CO_2 remediation, and producing bio-products, simultaneously. We also found difference in water uptake and light absorption when growth was simulated using CO_2 and carbonate from water. However, the differences in growth using CO_2 and HCO_3 have yet to be shown experimentally.

We also simulated gene deletions and found that, though growth under autotrophic conditions continues to exhibit low genetic robustness, the percentage of lethal gene deletions were lower than that observed by previous models. Therefore, this results suggest the metabolic models of *Synechocystis sp.* PCC6803 are still in its infancy, and as more data becomes available, genetic robustness of future models may increase. Further, we were able to compare 167 gene deletions to Cyanobase and literature surveys to predict a 77% accuracy of gene deletion simulations using iSyn816CJ.

Though, the expanded and updated model gives better validations for flux distributions and mutant growth rates under autotrophic growth, there still exist many gaps in understanding the experimentally determined flux distributions. These include balance between serine biosynthesis and photorespiration, as well as the high flux variability in PS and OXPHOS metabolic network. However, during changing light conditions, the metabolic composition could be changing which could result into high variability in internal fluxes. Therefore, additional 'omics data needs to be coupled with such stoichiometric models to explain and choose an alternate flux distribution which explains metabolic behavior coupled to photosynthesis.

CHAPTER 5. LEXICOGRAPHIC ANALYSIS OF DYNAMIC FLUX BALANCE MODEL OF

SYNECHOCYSTIS SP. PCC6803 METABOLIC NETWORK

1. SYNOPSIS

Cyanobacteria continue to gather interest due to their ability to assimilate atmospheric carbon dioxide and release oxygen, while making biomass for itself. This has generated significant attention towards using cyanobacteria for commercial production of novel substances using tools like genetic and metabolic engineering. However, to construct strains that not only grow optimally but also are efficient producers of a molecule of interest, it is important to understand intracellular metabolic regulation in these microorganisms in its full dynamic complexity. Photosynthetic organisms have an inherent dynamic complexity because in the natural habitat there are days and nights, as well as seasons, and the consequent changes in light intensity and composition. A variety of sustainable and green applications of metabolic engineering of cyanobacteria is ultimately possible only when translatable to utilization of the energy given out by the sun. Computational simulation methods for metabolic engineering have been largely based on Flux Balance Analysis (FBA) that cannot account for the natural cycles of sunlight. There is considerable value in developing dynamic methods for flux analysis that can overcome these limitations. Here, we apply a direct method of dynamic flux balance analysis that involves imbedding a Linear Programming problem within a set of kinetic equations, and using hierarchical or "lexicographic" optimization to study diurnal objective functions and lexicographic priority of substrate exchange, biomass growth, ATP synthase, and ATP maintenance in *Synechocystis sp.* PCC6803.

2. INTRODUCTION

Photosynthesis is a process which facilitates growth of an organisms using carbon dioxide, sunlight and water. Cyanobacteria are the only known prokaryotes that are capable of oxygenic photosynthesis and are believed to have had a significant role in the evolution of the oxygenic environment dating back to 3 billion years ago (Brocks et al., 1999). They have garnered much interest in the field of metabolic engineering, as they are great candidates for novel technologies including CO_2 remediation and production of chemicals such as biofuels (Nozzi et al., 2013), nutraceuticals (Gademann, 2011), and pharmaceuticals (Vijayakumar & Menakha, 2015). However, a huge chunk of information regarding diurnal metabolic rewiring, and proteomic and transcriptomic regulation has yet to be unraveled. Knowledge of such details about cyanobacteria can significantly expedite the design and implementation of commercial and technological applications.

Flux Balance analysis has been introduced and discussed in previous chapters. One of the limitations of FBA is that it is static, in other words it yields a distribution of fluxes that are frozen in time. As detailed previously, FBA begins with the dynamic metabolic equations and then assumes steady state to set all derivatives to zero. A more dynamic approach to the metabolic network would be to actually work with the full dynamic equations, called a kinetic model. However, kinetic models require a large amount of information that simply does not exist for the majority of in vivo reactions in the metabolic network.

As a consequence there has been significant work on developing a dynamical version of FBA, generically called Dynamic Flux Balance Analysis (DFBA). DFBA is a collection of mathematical frameworks to study and model metabolic rewiring of a microorganism as a consequence of its interaction with environment. DFBA models contain information on dynamics

134

of extracellular environment as well as the metabolic network of the microorganism. DFBA applications and implementations have increased, since the first formulation was released (Varma & Palsson, 1994) in 1994. DFBA has been applied to metabolic networks of *E. coli* (Radhakrishnan Mahadevan et al., 2002), *S. cerevisiae* (Hanly & Henson, 2013), *L. lactis* (Oddone et al., 2009), *C. reinhardtii* (Gomez et al., 2014), *H. sapiens* (red blood cells) (N Jamshidi et al., 2001), *S. stipites* (Hanly & Henson, 2013) etc. Some of the developments in DFBA in the past include integrated DFBA (Min Lee, Gianchandani, Eddy, & Papin, 2008), metabolic adjustment DFBA (R.-Y. Luo et al., 2006), dynamic multi-species metabolic models (Hanly & Henson, 2013), PROM-FBA (Chandrasekaran & Price, 2010), and DFBA-LQR (Uygun, Matthew, & Huang, 2006).

There are basically three types of DFBA methods that have been conceived of in previous work: (i) dynamic optimization approach (DOA) (Radhakrishnan Mahadevan et al., 2002), (ii) static optimization approach (SOA) (Radhakrishnan Mahadevan et al., 2002), and (iii) direct approach (DA). In SOA, the total batch time is divided into small time steps; an optimization problem is solved instantaneously at the beginning of each step; and then, the fluxes are integrated over the entire time step (Radhakrishnan Mahadevan et al., 2002). Thus SOA solves the dynamic problem by treating it as piecewise static, and then joining together all the static optimizations at different points. In DOA, an optimization problem is solved over the entire trajectory, which requires transforming the dynamic optimization to a non-linear programming problem, which is solved only once. The optimization here involves a terminal objective function and an instantaneous objective function. The solution is made possible by applying constraints of non-negative metabolite concentrations and fluxes, rates of exchange fluxes, and any non-linear constraints on transport fluxes (Radhakrishnan Mahadevan et al., 2002). Lastly, in DA, a system

of kinetic equations are solved with a linear program embedded (J L Hjersted & Henson, 2009; Jared L. Hjersted & Henson, 2006). The system of kinetic equations concern processes like metabolite uptake and secretion and are therefore experimentally accessible. More details about this method are presented in later sections.

The method of embedding a LP within a kinetic model, i.e. the DA, is the newest method and has some significant advantages over both the SOA and the DOA. It can be regarded as a method with more complexity that the SOA but less than the DOA (Gomez et al., 2014). In particular it requires less information than the DOA, and the additional information required when compared to static FBA is experimentally accessible. We therefore decided to use this method to study dynamic optimization in Synechocystis, using the iSynCJ816 model we developed.

Other groups have also developed cyanobacterial DFBA models using different approaches (Knoop et al., 2013). However, very little to no information on intracellular fluxes has been revealed computationally through these models. Recently, a DFBA-like approach was published which changed the biomass stoichiometry of the organism using transcriptomic information (Knoop et al., 2013). However, an issue with this approach is that it still lacks the information about metabolite formation rates. The study did give an insight into phosphoglycerate kinase, ribulose bisphosphate carboxylation, and phosphoglucomutase. This model was missing kinetic information and solved FBA at various light intensities to simulate diurnal cycles. A study on intracellular kinetics, which does not use a DFBA method, has also been published earlier (Zhu, Wang, Ort, & Long, 2013). No other study currently exists which simulated dynamics of cyanobacterial growth.

2.1. LEXICOGRAPHIC OPTIMIZATION

One of the challenges for extending FBA to tackle a dynamic diel cycle is the problem posed by non-unique intracellular fluxes. (Höffner et al., 2013). The problem is that FBA yields not just one optimal solution, but a family of optimal solutions. For dynamic optimization, the precise solution chosen determines the future trajectory of the simulation. This gets more complex if an organism has either more than one objective, or spends time in an environment where the objective could be completely different. To deal with this problem, a novel implementation of DA has recently been published called DFBAlab (Gomez et al., 2014) which implements what is called a lexicographic optimization to obtain a unique flux.

The idea of lexicographic optimization is quite simple. Instead of merely optimizing one objective function at every time step, we optimize a hierarchical list of objective functions that are user defined and chosen with reference to the physiology of the organism. We discuss this in greater detail below. DFBAlab can be used with the COBRA Toolbox within MATLAB environment.

3.   METHODS

3.1. STOICHIOMETRIC NETWORK

A stoichiometric network can be represented in the form of a matrix (S) which contains rows representative of metabolites (M), and columns representative of reactions (N). Therefore, the matrix, S of size M by N, itself denotes the chemical transformations with the metabolic network required to convert substrates to biomass. A negative entry in the matrix, S_{ij} corresponds to number of moles of reactant (M_i) consumed in the reaction (N_j). A positive entry in the matrix S_{ij} corresponds to number of moles of product (M_i) produced in the reaction (N_j). A null entry in

the matrix, $\mathbf{S}_{ij}$ corresponds to no involvement of a metabolite within the reaction. For our simulations, we used a newly developed stoichiometric network of *Synechocystis sp.* PCC6803 (Joshi, Peebles & Prasad, 2016, Submitted).

3.2. FLUX BALANCE ANALYSIS (FBA)

FBA is an extensively used constraint-based metabolic modeling approach for calculating flux through metabolic network (Orth et al., 2010). It has proved extremely successful in history; and has thus, become a standard approach in the field of metabolic modeling in systems biology (Gianchandani, Chavali, & Papin, 2010). Succinctly, FBA involves forming a stoichiometric matrix representing the metabolic network, as described in section 3.1, and forming a time-invariant linear programming problem to calculate fluxes using quasi steady state assumptions (Stephanopoulos, Aristidou, & Nielsen, 1998). This is achieved as follows.

Firstly, a stoichiometric network is formed, followed by writing out mass balance equations for each metabolite ($\mathbf{m}_i$) which results in equation (5.1).

$$\dot{\mathbf{z}}_1 = \sum_{j=1}^{N} \mathbf{S}_{ij} \cdot \mathbf{v}_j(z, t) \tag{5.1}$$

Here, dotted variable represents the time derivative of concentration ($\mathbf{z}_i$) of a metabolite ($\mathbf{M}_i$) and $\mathbf{v}_j(z,t)$ represents the flux of a reaction ($\mathbf{N}_j$) as a function of all metabolite concentrations ($\mathbf{z}$) and time ($\mathbf{t}$). Secondly, it is assumed that transients inside the metabolism are fast with respect to temporal changes in extracellular environment and quasi steady state assumptions are invoked, which leads to converting equation (5.1) to (5.1a) and (5.1b).

$$\sum_{j=1}^{N} \mathbf{S}_{ij} \cdot \mathbf{v}_j(z, t) = 0 \tag{5.1a}$$

$$\mathbf{S.v} = 0 \tag{5.1b}$$

Here, $\mathbf{v}$ represents the vector of reaction fluxes of size N by 1. It should be noted that equation (5.1b) is a set of linear algebraic equations. However, it is common for metabolic networks to have more reactions than there are metabolites; making (5.1b) an underdetermined problem, where there are more variables than there are equations. Such problems result in more than one solution. Therefore, to reduce the solution space, feasibility constraints can be applied in two ways: (i) by choosing a solution that optimizes some cellular/biological objective of the metabolic network (Feist & Palsson, 2010; Orth et al., 2010) given by equation (5.2a), and (ii) by applying biochemical considerations, topological or environmental constraints to the available flux space given by equation (5.2c).

$$\mu = \max_{\mathbf{v} \in R^N} \sum_{j=1}^{N} \mathbf{c}_j \mathbf{v}_j \ \ (\text{or}) \ \ \mu = \max_{\mathbf{v} \in R^N} \mathbf{c}^T \mathbf{v} \tag{5.2a}$$

$$\text{s.t. } \mathbf{S.v} = 0 \tag{5.2b}$$

$$\mathbf{v}^l \leq \mathbf{v} \leq \mathbf{v}^u \tag{5.2c}$$

Here, in equation (5.2), μ refers to the optimal value of the objective function; $\mathbf{c}_j$ refers to the weight of reaction ($\mathbf{N}_j$) in the objective function; $\mathbf{v}^u$ and $\mathbf{v}^l$ refer to the upper and lower limit of the allowable flux of reaction ($\mathbf{N}_j$); and $\mathbf{v}$ is the vector of N fluxes. These constraints together form a linear programming (LP) problem which defines the FBA model. Here, FBA was implemented using COBRA Toolbox with Gurobi 4.6.1 on MATLAB R2014b (D. Hyduke et al., 2011).

The choice of the cellular objective is still subject to discussion. Many different cellular objective have been tested in previous studies such as maximization of growth (Feist & Palsson, 2010; Ibarra et al., 2002), ATP maximization, ATP minimization, ATP per flux unit maximization, minimization of sum of fluxes (Robert Schuetz et al., 2007), minimization of substrate uptake

(Varma & Palsson, 1993), etc. However, an evolutionary argument can be made that a microorganism will maximize its growth if nutrients are not limiting (J S Edwards & Palsson, 2000b). It has also been shown experimentally that *E. coli* K-12 undergoes adaptive evolution to achieve the same growth rate as predicted by in silico simulations (Ibarra et al., 2002). We will mention wherever we used a different objective function, otherwise it should be assumed that growth rate objective function was used.

3.3.   DYNAMIC FLUX BALANCE ANALYSIS (DFBA)

DFBA is a collection of mathematical frameworks which allow simulation of time variant production of chemical species resulting from growth of a microorganism. As a general technique, DFBA couples FBA model with kinetics of growth, uptake of nutrients, and secretion of products. Here, we used DFBAlab with COBRA Toolbox on MATLAB R2014b (Gomez et al., 2014; Höffner et al., 2013), a toolbox that implements the DA for carrying out DFBA.

As described previously DFBA requires setting up uptake and secretion kinetics. This was done by writing out mass balance for extracellular/nutrient chemical species which can, generally, be described by equation (5.3).

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{f}\left(t, \mathbf{z}(t), \mathbf{v}\big(t, \mathbf{z}(t)\big)\right)$$

$$\mathbf{z}(t_0) = \mathbf{z}_0$$

(5.3)

Here, $\mathbf{f}$ represents the uptake kinetics as a function of time (t), metabolite concentration ($\mathbf{z}$), and the LP flux solution ($\mathbf{v}$) obtained from equation (5.2); and $\mathbf{v}$ also contains the exchange fluxes. Our next step is to modify the metabolic network such that all reactions can only carry positive flux and define the LP problem in context of this metabolic network. Therefore, equation (5.2) takes a time variant form given by equation (5.4)

$$U\big(t, \mathbf{z}(t)\big) = \arg \min_{\mathbf{v} \in R^N} \mathbf{c}^T \mathbf{v} \tag{5.4a}$$

$$\text{s.t. } \mathbf{S}.\mathbf{v} = \mathbf{b}\big(t, \mathbf{z}(t)\big) \tag{5.4b}$$

$$\mathbf{v} \geq \mathbf{0} \tag{5.4c}$$

Here, U (t, $\mathbf{z}$ (t)) is a subset of R^N and is called *the solution set* for the right-hand side $\mathbf{b}$, an element of R^M; and $\mathbf{b}$ represents the accumulation or utilization of chemical species. U (t, $\mathbf{z}$ (t)) is called *the solution set* because it contains all values of flux vector, $\mathbf{v}$, where the objective function attains the optimum value. In equation (5.4a), ***arg min*** refers to all the points where the function, $\mathbf{c}^T\mathbf{v}$, is minimized. Therefore, it should be noted that U may contain alternate solutions, if any, which correspond to optimum value of the objective function. It should also be noted that equation (5.4) has been written in standard LP form and maximizing $\mathbf{c}^T\mathbf{v}$ is equivalent to minimizing $-\mathbf{c}^T\mathbf{v}$. We have modified the weights ($\mathbf{c}$) in equation (5.4) as negative of the weights ($\mathbf{c}$) in equation (5.2). We refer to these equation combine as a dynamic system (5.3) with a LP embedded (5.4).

3.4.   *LEXICOGRAPHIC OPTIMIZATION*

As mentioned earlier, there is possibility of alternate solutions. It may be possible to reduce or resolve alternate solutions by formulating other alternate optimization problem (Gomez et al., 2014; Harwood, Höffner, & Barton, 2016) as described by equation (5.5) and (5.6)

$$g_1(V) = \min_{\mathbf{v} \in R^N} \mathbf{c}^T \mathbf{v}$$

$$\text{s.t. } \mathbf{S}.\mathbf{v} = \mathbf{b} \tag{5.5}$$

$$\mathbf{v} \geq \mathbf{0}$$

$$g_2(V) = \min_{\mathbf{v} \in R^N} \mathbf{c}_2{}^T \mathbf{v}$$

$$\text{s.t. } \mathbf{S.v} = \mathbf{b}$$

$$\mathbf{c^T v} = \mathbf{c^T u}$$

$$\mathbf{v} \geq \mathbf{0}$$

(5.6)

Here, $\mathbf{g}$:S $\rightarrow$ R^2 and c2 is a secondary optimization describing a secretion flux, uptake flux, and/or reaction flux like that of ethanol, oxygen, and/or ATP synthase. Lexicographic optimization works in following way. First, a priority list of a number of objective functions is ordered. Then, the highest priority objective function is optimized first, followed by the next priority objective function with the optimum value of last objective function as a constraint. The priority list is discussed further in the results section.

Each optimization level can be seen as picking out the optimal vector with reference to the objective function of that level from the solution set of the previous level. Thus the number of solutions rapidly decline and a unique optimum is obtained with only a few levels. It can be seen that the final solution can depend upon the order of the different functions being optimized. The biological implication of the lexicographic method is clear. It represents a series of secondary optimizations which can be hypothesized to be secondary objectives of the organism. Thus DFBA by the DA involves a series of additional hypotheses regarding metabolic objectives. These additional hypotheses are subordinate to the main one, for example growth maximization. The validity of a particular hierarchy of lexicographic optimization can thus only be determined through comparisons of predicted fluxes as a function of time with experimentally observed fluxes over time. For the moment we present and discuss predictions of the model utilizing three different schemes of lexicographic optimization coupled with DFBA.

4. RESULTS AND DISCUSSION

Implementation of DFBAlab has various user defined parts such as kinetic parameters of substrate consumption, diurnal (light and dark phase ratio) time spans, initial conditions, and lexicographic scheme. Here, we have analyzed various lexicographic schemes used to analyze growth dynamics of *Synechocystis sp.* PCC6803. The model used in our simulation (Joshi, Peebles, & Prasad, 2016, Submitted) contains 816 genes, 1045 reactions and 929 metabolites spanning 7 cellular compartments. The results discussed here were generated using 2.93 GHz Intel $^\circledR$ Core ™ i3 CPU in MATLAB (R2014b), Windows 7 64-bit operating system using Gurobi LP solver, and ode15s was used as the numerical integrator in MATLAB environment.

4.1. MODEL SETUP

This is a basic scheme as suggested by the authors of DFBAlab implementation (Gomez et al., 2014). Simulations begin with 12 hour light phase, followed by 12 hour dark phase. Further, we simulated a culture time of 54 hours; 2 light phases (12 hours each), 2 dark phases (12 hours each), and one light phase (6 hours). Michaelis-Menton expressions were implemented to uptake of carbon dioxide using parameters listed in Table 2 and reactor kinetics given by equations (5.7), and (5.8).

$$\dot{y}(t) = \mu\big(\boldsymbol{x}(t)\big)y(t) - \frac{F_{out}y(t)}{V}, \tag{5.7}$$

$$\dot{s}(t) = \frac{F_{in}s_0 - F_{out}s(t)}{V} + MT_s\big(\mathbf{x}(t)\big) + \Big(v_{Sp}\big(\mathbf{x}(t)\big) - v_{Sc}\big(\mathbf{x}(t)\big)\Big)y(t), \tag{5.8}$$

$$\text{for } s = g, o, c$$

Here, *y, g, o,* and *c* correspond to the concentrations of biomass, glycogen, oxygen, and carbon dioxide, respectively; $\mathbf{x} = [\ y\ g\ o\ c\ ]$; μ is the growth rate; v_{Sc} and v_{Sp} are the consumption and

production rates of substrate s calculated from lexicographic optimization; s_0 is the initial concentration in the feed; F_{in} and F_{out} are the inlet and outlet flows; V is the reactor volume; and MT_s is the mass transfer rate of s given by equation (9)

$$MT_s\big(\mathbf{x}(t)\big) = \begin{cases} (k_{sL}\theta)\left(\dfrac{s^{(g)}}{K_{Hs}} - s(t)\right) & \text{for } s = o,\, c \\[2ex] 0 & \text{for } s = g \end{cases} \tag{5.9}$$

Here K_{Hs} refers to Henry's constant of substrate s at 25° C; $k_{sL}\theta$ is the mass transfer coefficient; and $s^{(g)}$ is the concentration of s in the atmosphere.

The generalized form of uptake kinetics for carbon dioxide during light and glycogen during dark are given by equation (5.10)

$$v_s^{UB}\big(s(t)\big) = v_{s,max}\frac{s(t)}{K_s + s(t)}, \tag{5.10}$$

Here, v_s^{UB} is the upper bound of s exchange; $v_{s,max}$ is the maximum uptake rate of s; and K_s is Michaelis-Menton constant of s exchange. In addition to these, we fixed the light availability and all the available light (in mmol/gDW/h) was taken up by setting both bounds of light uptake reaction to value calculated by equations (5.11) and (5.12)

$$I_0(t) = 28\,\frac{\max\left(\sin^2\left(\dfrac{2\pi(t+6)}{48}\right),\sin^2\left(\dfrac{12\pi}{48}\right)\right) - \sin^2\left(\dfrac{12\pi}{48}\right)}{1 - \sin^2\left(\dfrac{12\pi}{48}\right)} \tag{5.11}$$

It should be noted that this light uptake function simulates 12h day/12 h night cycle. The prefactor was obtained from an algal study (de Oliveira Dal'Molin, Quek, Palfreyman, & Nielsen, 2011). Then, Beer-Lambert law was used average light available to cells within the reactor (A. Yang, 2011) using equations (5.12) and (5.13).

144

$$I_a\big(t, \mathbf{x}(t)\big) = I_0(t) \frac{1 - \exp\left(-LK_e\big(\mathbf{x}(t)\big)\right)}{LK_e\big(\mathbf{x}(t)\big)} \tag{5.12}$$

Here $K_e\left(\mathbf{x}(t)\right)$ is the extinction coefficient as a linear function of biomass concentration given by equation (5.13)

$$K_e\big(\mathbf{x}(t)\big) = K_{e1} + K_{e2}\mathbf{x}(t) \tag{5.13}$$

TABLE 5.1: INITIAL CONCENTRATIONS AND PARAMETERS

Variable/Parameter	Name	Value	Units
$v_{CO2,max}$	Maximum specific CO_2 uptake rate	0.249	mmol/gDW/h
K_{CO2}	MM constant for CO_2 uptake	0.034	mmol/L
MT_{CO2}	Mass transfer coefficient for CO_2	0.58	h^{-1}
$K_{H,CO2}$	Henry's constant of CO_2	0.035	mol/L/atm
s^{CO2}	concentration of CO_2 in atmosphere	0.035%	
$v_{max,glycogen}$	Maximum specific glycogen uptake rate	0.105	mmol/gDW/h
$K_{m,glycogen}$	MM constant for glycogen uptake	0.027	mmol/L
$v_{O2,max}$	Maximum specific O_2 uptake rate	0.383	mmol/gDW/h
K_{O2}	MM constant for O_2 uptake	0.135	mmol/L
MT_{O2}	Mass transfer coefficient for O_2	0.6	h^{-1}
$K_{H,O2}$	Henry's constant of CO_2	0.0013	mol/L/atm
s^{O2}	concentration of O_2 in atmosphere	21%	
L	Depth of reactor	0.1	m
K_{e1}	Extinction coefficient constant 1	0.32	m^{-1}
K_{e2}	Extinction coefficient constant 2	0.03	m^{-1} (g/m^3)$^{-1}$
c_0	Initial CO_2 concentration	0.61	mmol/L
o_0	Initial O_2 concentration	0.125	mmol/L
g_o	Initial glycogen concentration	0.566	mmol/L
y_o	Initial biomass concentration	0.153	g/L
V	Volume of the reactor	140	L
F_{in} (or) F_{out}	Incoming (or) Outgoing flow rate	1.001	L/h

The parameter values of carbon dioxide ($v_{CO2,max}$, K_{CO2}) and oxygen ($v_{O2,max}$, K_{O2}) kinetics were taken from previously published studies (Benschop, Badger, & Price, 2003). Mass transfer data (K_H and $k_{La}\theta$) for both, carbon dioxide and oxygen, were taken from previous study on algal-bacterial wastewater treatment pond (Buhr & Miller, 1983). Due to the lack of data on initial concentrations of carbon dioxide, oxygen, and glycogen, we used values similar to the ones found in DFBA of *C. reinhardtii*. Further, due to lack of information about the initial concentration of ATP, we used light as a starting condition rather than using a light phase from 5:00-17:00, and initialized the ATP concentration to zero. This may not necessarily be true; however, all ATP needed by the cyanobacterium to grow will have to be generated through metabolic network, initially. Initial biomass concentration was arbitrarily set to 0.153 g/L. The data presented here considers a batch reactor or a CSTR and we did not find qualitative changes due to low flow rate (F_{in} and F_{out}) conditions. The details of all the parameters and initial conditions have been mentioned in Table 5.1.

4.2. SCHEME 1

The lexicographic priority scheme used for these simulations has been mentioned in Table 2. This scheme was chosen based on experimental observations that (i) growth was observed only during light phase (Cheah et al., 2015); (ii) organism may undergo adaptive evolution to achieve optimal growth (Ibarra et al., 2002); (iii) ATP maximization performed well in *E. coli* cultures compared with biomass maximization (Robert Schuetz et al., 2007); (iv) ATP and glycogen levels were known to rise during light phase and decrease during dark phase in Synechocystis sp. PCC6803 (Saha et al., 2016); and (v) maximization of substrate consumption or product secretion at priority level 5-6 did not really make any qualitative difference in our simulation except during

dark phase when carbon dioxide production is physiologically known to happen in photosynthetic organisms. Imposing the maximization of substrate consumption and product secretion made sure of directionality of the substrate and product exchange.

TABLE 5.2 PRIORITY LIST ORDER. THE LEXICOGRAPHIC LP SCHEMES USED IN OUR SIMULATIONS

SCHEME 1				
Priority	Light Phase	Weights	Dark Phase	Weights
1	Maximize biomass production	1	Maximize ATP maintenance	1
2	Maximize ATP synthase (ATPase)	1	Minimize ATP synthase (ATPase)	1
3	Minimize ATP maintenance	1	Maximize biomass production	1
4	Maximize glycogen production	1	Maximize glycogen consumption	1
5	Maximize oxygen production	1	Maximize oxygen consumption	1
6	Maximize CO_2 consumption	1	Maximize CO_2 production	1
SCHEME 2				
Priority	Light Phase	Weights	Dark Phase	Weights
1	Maximize glycogen production	1	Maximize glycogen utilization	1
2	Maximize biomass production	1	Maximize biomass production	1
3	Maximize ATP synthase	1	Minimize ATP synthase	1
4	Minimize ATP maintenance	1	Maximize ATP maintenance	1
5	Maximize oxygen production	1	Maximize oxygen consumption	1
6	Maximize CO_2 consumption	1	Maximize CO_2 production	1
SCHEME 3				
Priority	Light Phase	Weights	Dark Phase	Weights
1	Maximize glycogen production	0.96	Maximize ATP maintenance	1
1	Maximize biomass production	0.04	Minimize ATP synthase (ATPase)	1
2	Maximize ATP synthase	1	Maximize biomass production	1
3	Minimize ATP maintenance	1	Maximize glycogen consumption	1
4	Maximize oxygen production	1	Maximize oxygen consumption	1
5	Maximize CO_2 consumption	1	Maximize CO_2 production	1

As mentioned previously in Section 3.1, our simulations begin with light phase and spans over 54 hours. We find that there are oscillations occurring with increasing amplitudes for carbon dioxide levels (Figure 5.1A, blue line); suggesting that as the culture grows, the pH increases in a batch culture. However, we did not find a study involving batch culture of Synechocystis sp. PCC6803, where net carbon dioxide increases over time. The concentration of oxygen tappers off

because of the constraints on concentration of oxygen which cannot be more than the value of Henry's constant. These constraints are applied to oxygen and carbon dioxide concentrations. It should be noted that laboratory batch cultures will likely become denser and inhibit growth due to various reasons such as (i) carbon dioxide is limited in isolated growth conditions, (ii) denser cultures will not get sufficient light, (iii) changes in pH may also have adverse effect on growth conditions. Further, as should be expected, we also find that near the highest light availability (6, 30, 54 h), the levels of carbon dioxide decrease at a faster rate. Additional glycogen production was not supported during the light phase (Figure 5.1A, red line). However, it should be noted that glycogen is part of the autotrophic growth within out model. Therefore, increase in biomass concentration also leads to increase in glycogen. Therefore, the total pool of glycogen does, in fact, increases. Considering this, our simulation indicates that glycogen pool increases (due to growth) during light and reduces during dark phase, as observed in a previously published experimental study (Saha et al., 2016).

The lexicographic scheme utilized for dark phase, resulted in no growth (Figure 5.1B). However, we find that glycogen is directed towards maintenance, as suggested by increased ATP maintenance flux. There is no ATP maintenance flux during light; however, ATP synthase is operational (Figure 5.1C). This suggests that during light phase ATP is stored in the form of other molecules such as glycogen, while in the dark phase glycogen is consumed to sustain ATP levels needed for cellular maintenance. Therefore, we interpret these results as being suggestive of reducing ATP levels in dark phase while high ATP levels during light phase (Saha et al., 2016). Interestingly, we find that ATP maintenance flux exceeds ATP synthase flux, which defies the physico-chemical understanding that more ATP is consumed by the cell during dark phase than can be produced during light phase. However, it is to be expected that ATP synthase is only active

148

during light phase because of availability of photons, which in turn drive electron transfer machinery and produce a proton gradient to drive ATP synthase.

4.3. SCHEME 2

The lexicographic priority scheme used for these simulations has been mentioned in Table 2. This scheme was used based on experimental observations that (i) glycogen levels increased during light phase and decreased during dark phase, this could be because glycogen is stored so that cells can utilize it during dark; (ii) fluxes could be predicted with ATP maximization than with biomass maximization when cells undergo change in environment (R. Schuetz et al., 2012), and (iii) ATP production consistently increased during light phase suggesting that no additional ATP was being produced during dark phase. We continue our simulations in similar way as done in Scheme 1; simulations begin with light phase and span over 54 hours. Similar to lexicographic scheme 1, we find that carbon dioxide and oxygen levels undergo oscillations with increasing amplitude (Figure 2A). However, we find that there are some significant discrepancies with known observations. In our simulations, growth only occurs during dark phase while large amounts of glycogen accumulation is also observed during light phase (Figure 2A & Figure 2B). This glycogen is then utilized during dark phase making the culture grow. Further, we also find that

ATP maintenance flux is completely dispensable according to these simulations (Figure 2C). It

should be noted that lexicographic priorities of the exchange fluxes and ATP are similar in both
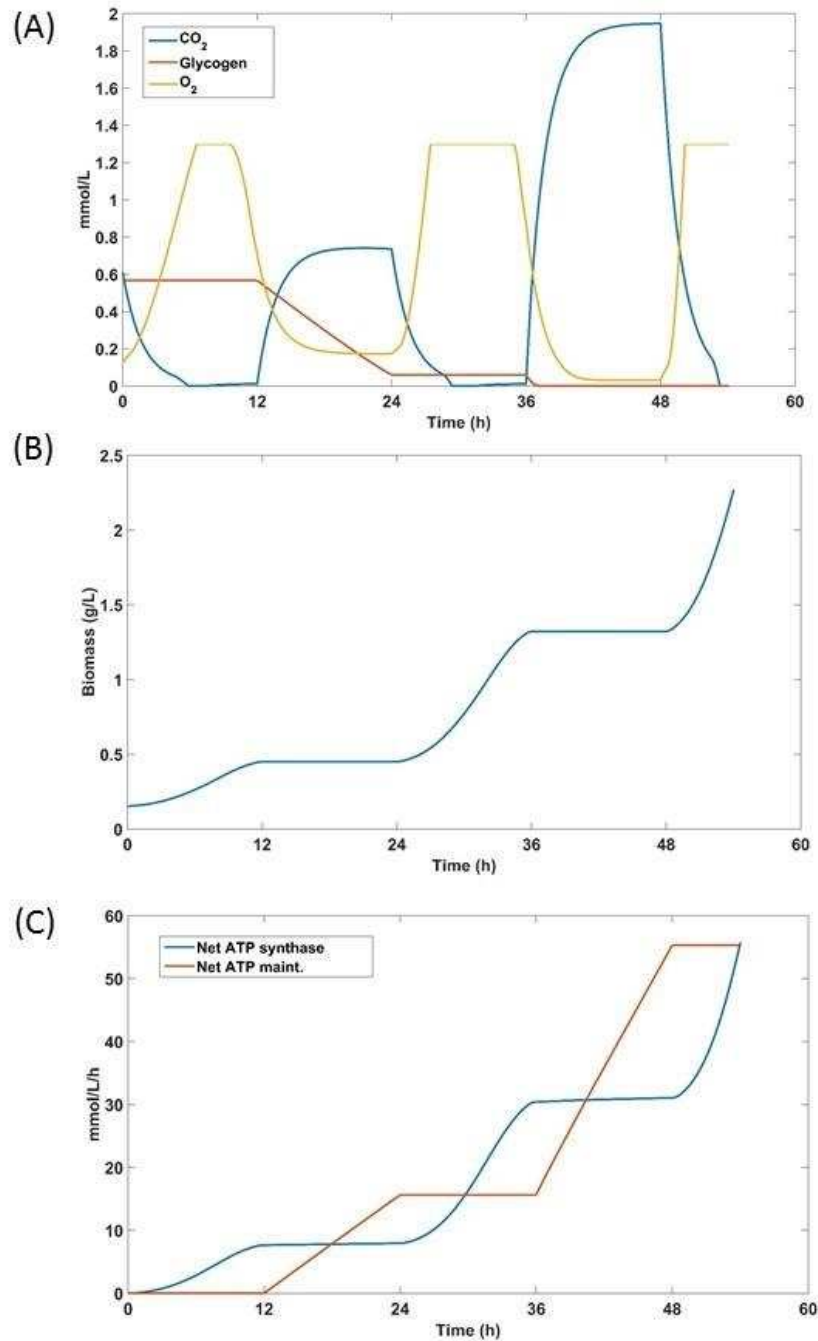
FIGURE 5.1: CONCENTRATION IN REACTOR, WHEN LEXICOGRAPHIC SCHEME 1 WAS USED for 54h
(LDLDL, 12h:12h).
(A) Plots for carbon dioxide (blue), oxygen (yellow) and glycogen (red) concentration (mmol/L). (B) biomass
concentration (g/L), and (C) Net ATP synthase (blue) and ATP maintenance flux (red).

cases, except in dark we order glycogen and oxygen utilization, and carbon dioxide production
which is known to be physiologically true.
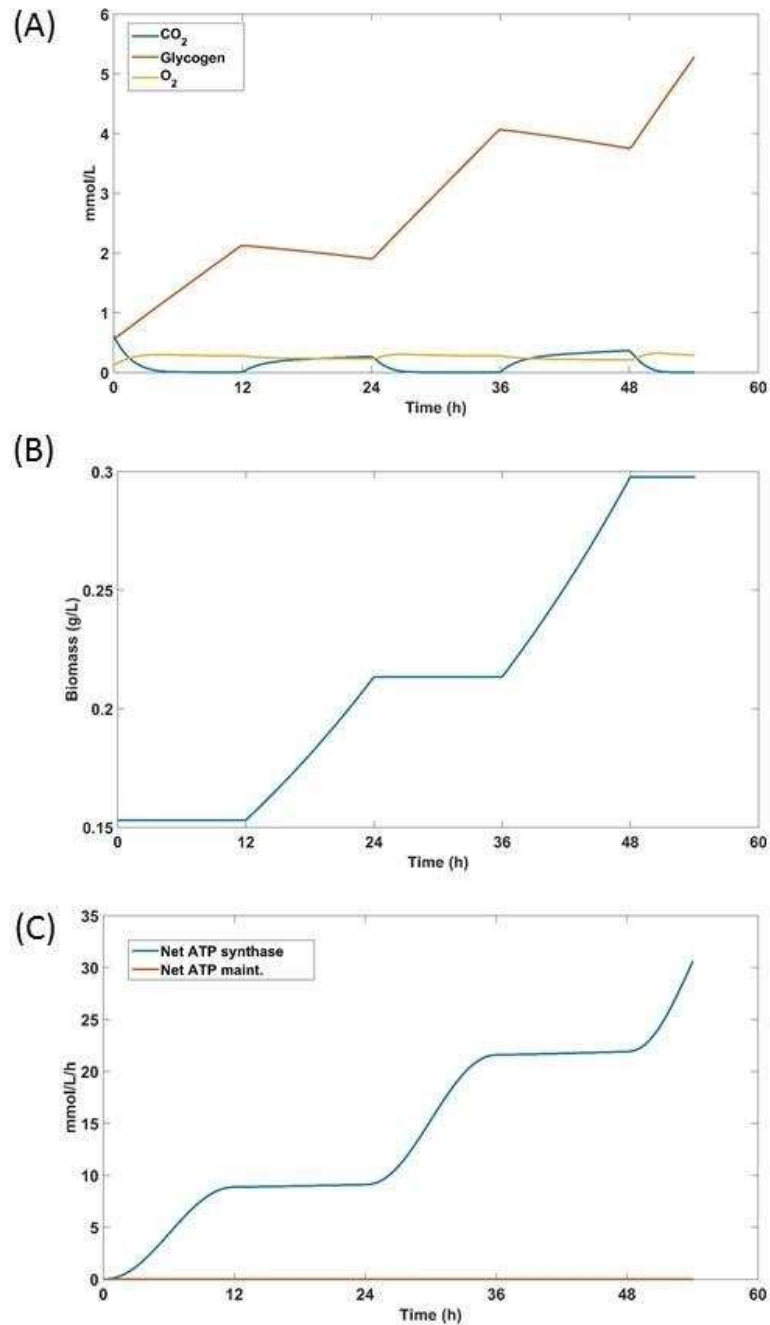
FIGURE 5.2: CONCENTRATION IN REACTOR, WHEN LEXICOGRAPHIC SCHEME 2 WAS USED for 54h
(LDLDL, 12h:12h).
(A) Plots for carbon dioxide (blue), oxygen (yellow) and glycogen (red) concentration (mmol/L). (B) biomass
concentration (g/L), and (C) Net ATP synthase (blue) and ATP maintenance flux (red).

151

4.4.   SCHEME 3

The lexicographic priority scheme used for these simulations has been mentioned in Table 2. The most important change we made in Scheme 3 was in the primary objective function, which here is a weighted combination of biomass growth and glycogen levels. The rationale behind this is the assumption that the organism "plans" for the dark phase by storing glycogen, therefore glycogen storage must be an objective for the photosynthetic organism. Therefore, we assigned a weight of 0.94 to glycogen sink and 0.06 to biomass growth; the higher weightage to glycogen sink is due to the fact that glycogen is a large molecule, and has a numerically small flux, so that the metabolic requirements of growth outcompetes the glycogen sink. The implementation of weighted objective function of biomass and glycogen can be justified because, as previously mentioned, organism does not grow during dark phase; and yet, oxygen is consumed and carbon dioxide is produced. We know that this is because of consumption of stored glycogen.

This scheme was used based on experimental observations that (i) glycogen levels increased during light phase and decreased during dark phase, but the decomposition of glycogen within the metabolic network does not contribute to growth and carbon dioxide evolution takes place; (ii) organism may have undergone adaptive evolution to synchronize with diurnal cycles; (iii) fluxes could be predicted with ATP maximization; and (iv) ATP levels increased during light phase and decreased during dark phase.

As in Scheme 1, the simulations using Scheme 3 showed no change in ATP synthase flux during the dark phase. However, unlike Scheme 1, ATP maintenance flux never exceeded ATP synthase (Figure 5.3C). The difference between ATP synthase and ATP maintenance was used a proxy for ATP levels. Our simulations also show that carbon dioxide and oxygen levels were in qualitative agreement between scheme 1 and scheme 3 (Figure 5.3A and 5.3B). Scheme 3 is based

on what we learned from Scheme 1&2 and overall does a better job in qualitatively capturing the

dynamics of carbon dioxide, oxygen, glycogen, ATP, and growth during the day-night cycle.
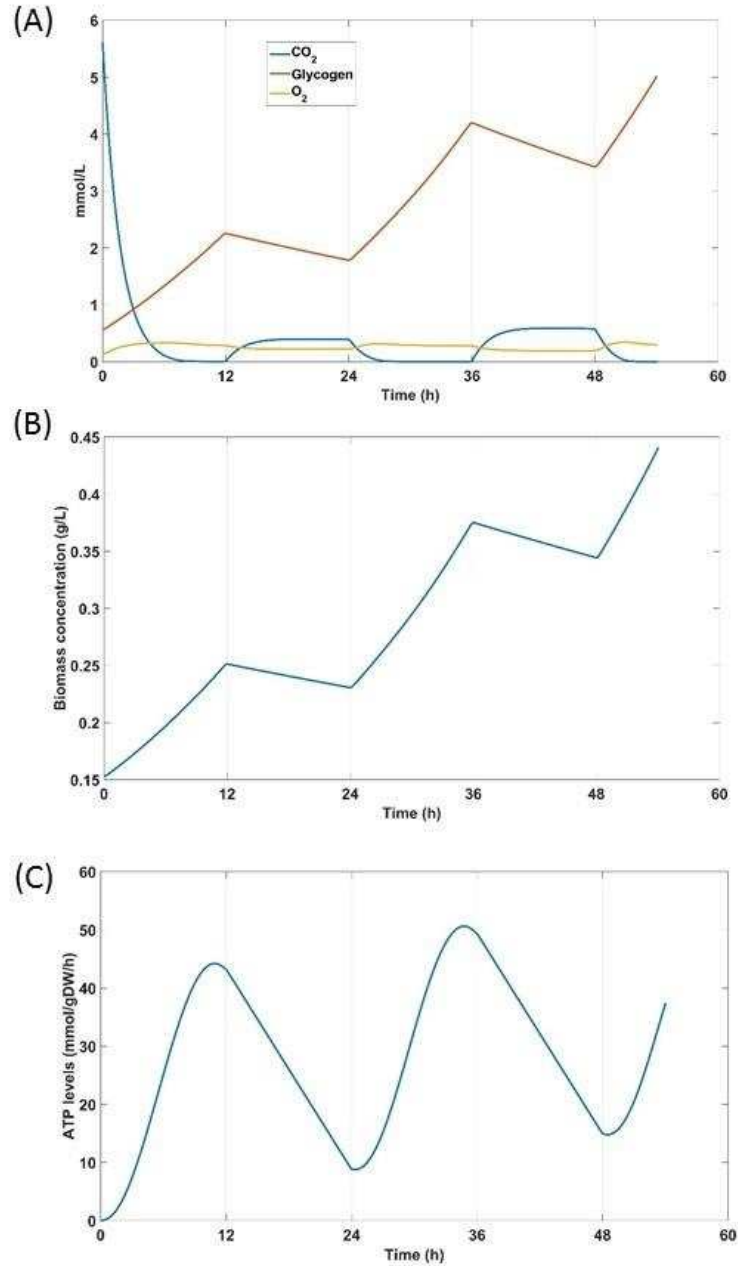
FIGURE 5.3: CONCENTRATION IN REACTOR, WHEN LEXICOGRAPHIC SCHEME 3 WAS USED for 54h (LDLDL, 12h:12h).
(A) Plots for carbon dioxide (blue), oxygen (yellow) and glycogen (red) concentration (mmol/L). (B) biomass concentration (g/L), and (C) difference of ATP synthase and ATP maintenance.

5.   CONCLUSION

Cyanobacteria continue to be of high interest because of its ability to assimilate naturally occurring sunlight and carbon dioxide around which useful technologies like CO2 remediation, and production of commercially valuable chemicals can be developed. Here, we presented dynamic flux balance simulations of extracellular metabolites of *Synechocystis sp.* PCC6803 under 12h light-to 12h dark growth conditions using a dynamic FBA method called the DA. We also showed the importance of lexicographic ordering using DFBAlab (Gomez et al., 2014).

We presented three priority schemes involving substrate uptake/production, biomass growth, ATP synthase and ATP maintenance flux. We found that a top priority maximization of glycogen utilization objective function performed poorly. This is because if glycogen accumulation was the top priority objective for the cell, growth would be observed during the dark phase rather than during light phase. However, previously published studies clearly indicate that there is little to no growth during the dark phase and ATP levels are higher during light phase than during dark (Saha et al., 2016). Temporally splitting the primary objective functions (maximization of growth during light phase and maximization of ATP maintenance during dark) resulted in better qualitative matches with experimental observations. Based on the results of the first two optimization schemes, we hypothesized that the organism maximizes a composite objective function, which is a convex combination of growth and glycogen storage. This objective function yielded a better match of the simulation results with the data.

It should be noted that we did not include any transcriptomic or proteomic constraints in the original LP problem embedded within the dynamic problem. However, inclusion of such constraints may result in more accurate lexicographic definition of the problem. Further, it should also be noted that some of the parameters, such as glycogen exchange rate, have been borrowed

from *C. reinhardtii* (Buhr & Miller, 1983; Gomez et al., 2014) or are arbitrarily used. WE have also not included any substrate or product inhibition which may change the growth and environmental dynamics. Therefore, given the correct kinetic parameters, a better dynamic problem may be formed which works with the embedded LP to give better solutions.

Our simulations involving change of primary objective function also suggests that there are likely two very different metabolic network objectives under two different environmental conditions (light and dark). However, previous studies in designing objective functions for flux balance analysis have speculated that though adaptive evolution reaches the in silico predicted growth maximizing flux distribution; it is likely that minimization of adjustment of metabolic fluxes maybe better suited during transitionary periods (Ibarra et al., 2002; Segrè et al., 2002). Minimization of metabolic adjustment (MoMA) is a method that involves minimizing the adjustment of metabolic fluxes from a user provided prior (to genetic, environmental, or temporal perturbation) flux distribution vector. Since each flux distribution is a point that lies in *n*-dimensional space, where *n* is the number of reaction fluxes; minimizing flux adjustment refers to finding a new point such that the distance from the old point prior to change or perturbation is minimized. Therefore, it is a quadratic programming problem. If this argument is true, an LP problem is not suited to address the dynamic changes from day to night, and a quadratic programming approach that minimizes adjustment may be needed. In particular we can postulate that MOMA could be an additional constraint that determines how the organism picks on a particular flux vector as the environmental conditions change. Incorporating some of these complexities in simulation methods may yield improved understanding and predictions of dynamic flux changes in diel cycles.

BIBLIOGRAPHY

Allahverdiyeva, Y., Ermakova, M., Eisenhut, M., Zhang, P., Richaud, P., Hagemann, M., … Aro, E. M. (2011). Interplay between flavodiiron proteins and photorespiration in Synechocystis sp. PCC 6803. *Journal of Biological Chemistry*, *286*(27), 24007–24014. Journal Article. http://doi.org/10.1074/jbc.M111.223289

Alper, H., Jin, Y. S., Moxley, J. F., & Stephanopoulos, G. (2005). Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. *Metabolic Engineering*, *7*(3), 155–164. Journal Article. http://doi.org/10.1016/j.ymben.2004.12.003

Arita, M. (2004). The metabolic world of Escherichia coli is not small. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(6), 1543–1547. Journal Article. http://doi.org/10.1073/pnas.0306458101

Asadollahi, M. A., Maury, J., Patil, K. R., Schalk, M., Clark, A., & Nielsen, J. (2009). Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering. *Metabolic Engineering*, *11*(6), 328–334. http://doi.org/10.1016/j.ymben.2009.07.001

Badger, M. R., von Caemmerer, S., Ruuska, S., & Nakano, H. (2000). Electron flow to oxygen in higher plants and algae: rates and control of direct photoreduction (Mehler reaction) and rubisco oxygenase. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *355*(1402), 1433–1446. http://doi.org/10.1098/rstb.2000.0704

Bailey, S., Melis, A., Mackey, K. R. M., Cardol, P., Finazzi, G., van Dijken, G., … Grossman, A. (2008). Alternative photosynthetic electron flow to oxygen in marine Synechococcus. *Biochimica et Biophysica Acta - Bioenergetics*, *1777*(3), 269–276. http://doi.org/10.1016/j.bbabio.2008.01.002

Balaram, P. (2003, January 18). Synthesizing life. *Current Science*, *85*(11), 1509–1510. http://doi.org/10.1038/35053176

Barabási, A.-L. (2009). Scale-free networks: a decade and beyond. *Science (New York, N.Y.)*, *325*(5939), 412–413. http://doi.org/10.1126/science.1173299

Barber, J. (2014). Photosystem II: its function, structure, and implications for artificial photosynthesis. *Biochemistry. Biokhimii͡a*, *79*(3), 185–96. Journal Article. http://doi.org/10.1134/S0006297914030031

Bateson, W. (1909). *Mendel's Principles of Heredity*. Cambridge: Cambridge University Press.

Battchikova, N., Eisenhut, M., & Aro, E. M. (2011). Cyanobacterial NDH-1 complexes: Novel insights and remaining puzzles. *Biochimica et Biophysica Acta - Bioenergetics*, *1807*(8), 935–944. Journal Article. http://doi.org/10.1016/j.bbabio.2010.10.017

Bauwe, H., Hagemann, M., Kern, R., & Timm, S. (2012). Photorespiration has a dual origin and manifold links to central metabolism. *Current Opinion in Plant Biology*, *15*(3), 269–275. Journal Article. http://doi.org/10.1016/j.pbi.2012.01.008

Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., van Dien, S. J., & Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature Chemical Biology*, *5*(8), 593–599. http://doi.org/10.1038/nchembio.186

Benschop, J. J., Badger, M. R., & Price, G. D. (2003). Characterisation of CO2 and HCO3-uptake in the cyanobacterium Synechocystis sp. PCC6803. *Photosynthesis Research*, *77*(2–3), 117–126. http://doi.org/10.1023/A:1025850230977

Bonven, B., & Gulløv, K. (1979). Peptide chain elongation rate and ribosomal activity in Saccharomyces cerevisiae as a function of the growth rate. *Molecular & General Genetics : MGG*, *170*(2), 225–30. http://doi.org/10.1007/BF00337800

Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C., & Kondrashov, F. a. (2012). Epistasis as the primary factor in molecular evolution. *Nature*, *490*(7421), 535–538. Journal Article. http://doi.org/10.1038/nature11510

Brocks, J. J., Logan, G. a, Buick, R., & Summons, R. E. (1999). Archean molecular fossils and the early rise of eukaryotes. *Science (New York, N.Y.)*, *285*(5430), 1033–1036. http://doi.org/10.1126/science.285.5430.1033

Brodie III, E. D. (2000). *Why evolutionary genetics does not always add up*. *Epistasis and the evolutionary process*. Book. Retrieved from http://books.google.fr/books?hl=fr&lr=&id=WS9BB3JeVxwC&oi=fnd&pg=PA3&ots=WUW9pp_7Y5&sig=efKRFMxQAMZR3sCty7NV8ntSaRY\nhttp://books.google.fr/books?hl=fr&lr=&id=WS9BB3JeVxwC&oi=fnd&pg=PA3&ots=WUW9pp_7Y5&sig=efKRFMxQAMZR3sCty7NV8ntSaRY#v=onepage&q&f=false

Buhr, H. O., & Miller, S. B. (1983). A dynamic model of the high-rate algal-bacterial wastewater treatment pond. *Water Research*, *17*(1), 29–37. http://doi.org/10.1016/0043-1354(83)90283-X

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., … ZILLIG, W. (1996). Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science (New York, N.Y.)*, *273*(5278), 1058–73. http://doi.org/10.1126/science.273.5278.1058

Burgard, A. P., Pharkya, P., & Maranas, C. D. (2003). OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnology and Bioengineering*, *84*(6), 647–657. Journal Article. http://doi.org/10.1002/bit.10803

Camas, F. M., & Poyatos, J. F. (2008). What determines the assembly of transcriptional network motifs in Escherichia coli? *PLoS ONE*, *3*(11), e3657. Journal Article. http://doi.org/10.1371/journal.pone.0003657

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. a., … Karp, P. D. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, *42*(D1), D459-71-D471. Journal Article. http://doi.org/10.1093/nar/gkt1103

Chandrasekaran, S., & Price, N. D. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences*, *107*(41), 17845–17850. http://doi.org/10.1073/pnas.1005139107

Chang, R. L., Ghamsari, L., Manichaikul, A., Hom, E. F. Y., Balaji, S., Fu, W., … Papin, J. a. (2011). Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. *Molecular Systems Biology*, *7*, 518. Journal Article. http://doi.org/10.1038/msb.2011.52

Cheah, Y. E., Zimont, A. J., Lunka, S. K., Albers, S. C., Park, S. J., Reardon, K. F., & Peebles, C. A. M. (2015). Diel light: Dark cycles significantly reduce FFA accumulation in FFA producing mutants of Synechocystis sp. PCC 6803 compared to continuous light. *Algal Research*, *12*, 487–496. http://doi.org/10.1016/j.algal.2015.10.014

Cheng, X. Y., Huang, W. J., Hu, S. C., Zhang, H. L., Wang, H., Zhang, J. X., … Ji, Z. L. (2012). A global characterization and identification of multifunctional enzymes. *PLoS ONE*, *7*(6). http://doi.org/10.1371/journal.pone.0038979

Chubukov, V., Uhr, M., Le Chat, L., Kleijn, R. J., Jules, M., Link, H., … Sauer, U. (2013). Transcriptional regulation is insufficient to explain substrate-induced flux changes in Bacillus subtilis. *Molecular Systems Biology*, *9*(709), 709. Journal Article. http://doi.org/10.1038/msb.2013.66

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703. Journal Article. Retrieved from http://epubs.siam.org/doi/abs/10.1137/070710111

Cooley, J. W., & Vermaas, W. F. J. (2001). Succinate dehydrogenase and other respiratory pathways in thylakoid membranes of Synechocystis sp. strain PCC 6803: Capacity comparisons and physiological function. *Journal of Bacteriology*, *183*(14), 4251–4258. Journal Article. http://doi.org/10.1128/JB.183.14.4251-4258.2001

Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J., & Palsson, B. O. (2004). Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, *429*(6987), 92–96. Journal Article. http://doi.org/10.1038/nature02456

D'Souza, G., Waschina, S., Kaleta, C., & Kost, C. (2015). Plasticity and epistasis strongly affect bacterial fitness after losing multiple metabolic genes. *Evolution*, *69*(5), n/a-n/a. Journal Article. http://doi.org/10.1111/evo.12640

Dawid, A., Kiviet, D. J., Kogenaru, M., de Vos, M., & Tans, S. J. (2010). Multiple peaks and reciprocal sign epistasis in an empirically determined genotype-phenotype landscape. *Chaos*, *20*(2), 26105. Journal Article. http://doi.org/10.1063/1.3453602

de Figueiredo, L. F., Gossmann, T. I., Ziegler, M., & Schuster, S. (2011). Pathways Analysis of NAD+ Metabolism. *Biochemical Journal*, *439*(2), 341–348. Journal Article. http://doi.org/10.1042/BJ20110320

de Oliveira Dal'Molin, C. G., Quek, L.-E., Palfreyman, R. W., & Nielsen, L. K. (2011). AlgaGEM – a genome-scale metabolic reconstruction of algae based on the

Chlamydomonas reinhardtii genome. *BMC Genomics*, *12*(Suppl 4), S5. http://doi.org/10.1186/1471-2164-12-S4-S5

de Visser, J. a. G. M., Cooper, T. F., & Elena, S. F. (2011). The causes of epistasis. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1725), 3617–3624. Journal Article. http://doi.org/10.1098/rspb.2011.1537

de Visser, J. A. G. M., & Elena, S. F. (2007). The evolution of sex: empirical insights into the roles of epistasis and drift. *Nature Reviews. Genetics*, *8*(2), 139–149. Journal Article. http://doi.org/10.1038/nrg1985

de Vos, M. G. J., Poelwijk, F. J., Battich, N., Ndika, J. D. T., & Tans, S. J. (2013). Environmental Dependence of Genetic Constraint. *PLoS Genetics*, *9*(6), e1003580. Journal Article. http://doi.org/10.1371/journal.pgen.1003580

Degtyarenko, K., De matos, P., Ennis, M., Hastings, J., Zbinden, M., Mcnaught, A., … Ashburner, M. (2008). ChEBI: A database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, *36*(SUPPL. 1). http://doi.org/10.1093/nar/gkm791

Demmig-Adams, B. (1992). Photoprotection and Other Responses of Plants to High Light Stress. *Annual Review of Plant Physiology and Plant Molecular Biology*, *43*(1), 599–626. Journal Article. http://doi.org/10.1146/annurev.arplant.43.1.599

Duarte, N. C., Becker, S. a, Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., … Palsson, B. Ø. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(6), 1777–1782. Journal Article. http://doi.org/10.1073/pnas.0610772104

Duarte, N. C., Herrgard, M. J., & Palsson, B. O. (2004). Reconstruction and Validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*, *14*, 198–1309. http://doi.org/10.1101/gr.2250904.

Edwards, J. S., Covert, M., & Palsson, B. (2002). Metabolic modelling of microbes: The flux-balance approach. *Environmental Microbiology*. http://doi.org/10.1046/j.1462-2920.2002.00282.x

Edwards, J. S., & Palsson, B. O. (2000a). Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics*, *1*, 1. Journal Article. http://doi.org/10.1186/1471-2105-1-1

Edwards, J. S., & Palsson, B. O. (2000b). The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(10), 5528–5533. Journal Article. http://doi.org/10.1073/pnas.97.10.5528

Eisenhut, M., Ruth, W., Haimovich, M., Bauwe, H., Kaplan, A., & Hagemann, M. (2008). The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(44), 17199–17204. Journal Article. http://doi.org/10.1073/pnas.0807043105

Esvelt, K. M., & Wang, H. H. (2013). Genome-scale engineering for systems and synthetic

Chlamydomonas reinhardtii genome. *BMC Genomics*, *12*(Suppl 4), S5. http://doi.org/10.1186/1471-2164-12-S4-S5

de Visser, J. a. G. M., Cooper, T. F., & Elena, S. F. (2011). The causes of epistasis. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1725), 3617–3624. Journal Article. http://doi.org/10.1098/rspb.2011.1537

de Visser, J. A. G. M., & Elena, S. F. (2007). The evolution of sex: empirical insights into the roles of epistasis and drift. *Nature Reviews. Genetics*, *8*(2), 139–149. Journal Article. http://doi.org/10.1038/nrg1985

de Vos, M. G. J., Poelwijk, F. J., Battich, N., Ndika, J. D. T., & Tans, S. J. (2013). Environmental Dependence of Genetic Constraint. *PLoS Genetics*, *9*(6), e1003580. Journal Article. http://doi.org/10.1371/journal.pgen.1003580

Degtyarenko, K., De matos, P., Ennis, M., Hastings, J., Zbinden, M., Mcnaught, A., … Ashburner, M. (2008). ChEBI: A database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, *36*(SUPPL. 1). http://doi.org/10.1093/nar/gkm791

Demmig-Adams, B. (1992). Photoprotection and Other Responses of Plants to High Light Stress. *Annual Review of Plant Physiology and Plant Molecular Biology*, *43*(1), 599–626. Journal Article. http://doi.org/10.1146/annurev.arplant.43.1.599

Duarte, N. C., Becker, S. a, Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., … Palsson, B. Ø. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(6), 1777–1782. Journal Article. http://doi.org/10.1073/pnas.0610772104

Duarte, N. C., Herrgard, M. J., & Palsson, B. O. (2004). Reconstruction and Validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*, *14*, 198–1309. http://doi.org/10.1101/gr.2250904.

Edwards, J. S., Covert, M., & Palsson, B. (2002). Metabolic modelling of microbes: The flux-balance approach. *Environmental Microbiology*. http://doi.org/10.1046/j.1462-2920.2002.00282.x

Edwards, J. S., & Palsson, B. O. (2000a). Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics*, *1*, 1. Journal Article. http://doi.org/10.1186/1471-2105-1-1

Edwards, J. S., & Palsson, B. O. (2000b). The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(10), 5528–5533. Journal Article. http://doi.org/10.1073/pnas.97.10.5528

Eisenhut, M., Ruth, W., Haimovich, M., Bauwe, H., Kaplan, A., & Hagemann, M. (2008). The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(44), 17199–17204. Journal Article. http://doi.org/10.1073/pnas.0807043105

Esvelt, K. M., & Wang, H. H. (2013). Genome-scale engineering for systems and synthetic

biology. *Molecular Systems Biology*, *9*(641), 641. http://doi.org/10.1038/msb.2012.66

Fani, R., & Fondi, M. (2009). Origin and evolution of metabolic pathways. *Physics of Life Reviews*, *6*(1), 23–52. Journal Article. http://doi.org/10.1016/j.plrev.2008.12.003

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., … Palsson, B. Ø. (2007). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, *3*, 121. Journal Article. http://doi.org/10.1038/msb4100155

Feist, A. M., & Palsson, B. O. (2010). The biomass objective function. *Current Opinion in Microbiology*, *13*(3), 344–349. http://doi.org/10.1016/j.mib.2010.03.003

Fischer, E., Zamboni, N., & Sauer, U. (2004). High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. *Anal Biochem*, *325*(2), 308–316. Journal Article. http://doi.org/S0003269703007528 [pii]

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., … al., et. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science (New York, N.Y.)*, *269*(5223), 496–512. http://doi.org/10.1126/science.7542800

Flynn, K. M., Cooper, T. F., Moore, F. B. G., & Cooper, V. S. (2013). The Environment Affects Epistatic Interactions to Alter the Topology of an Empirical Fitness Landscape. *PLoS Genetics*, *9*(4), e1003426. Journal Article. http://doi.org/10.1371/journal.pgen.1003426

Fong, S. S., Burgard, A. P., Herring, C. D., Knight, E. M., Blattner, F. R., Maranas, C. D., & Palsson, B. O. (2005). In silico design and adaptive evolution of Escherichia coli for production of lactic acid. *Biotechnology and Bioengineering*, *91*(5), 643–648. http://doi.org/10.1002/bit.20542

Fong, S. S., & Palsson, B. Ø. (2004). Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. *Nature Genetics*, *36*(10), 1056–1058. Journal Article. http://doi.org/10.1038/ng1432

Forchhammer, J., & Lindahl, L. (1971). Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of Escherichia coli 15. *Journal of Molecular Biology*, *55*(3), 563–568. http://doi.org/10.1016/0022-2836(71)90337-8

Förster, J., Famili, I., Fu, P., Palsson, B., & Nielsen, J. (2003). Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Research*, *13*(2), 244–253. Journal Article. http://doi.org/10.1101/gr.234503

Fu, P. (2009). Genome-scale modeling of Synechocystis sp. PCC 6803 and prediction of pathway insertion. *Journal of Chemical Technology and Biotechnology*, *84*(4), 473–483. Journal Article. http://doi.org/10.1002/jctb.2065

Fujisawa, T., Okamoto, S., Katayama, T., Nakao, M., Yoshimura, H., Kajiya-Kanegae, H., … Nakamura, Y. (2014). CyanoBase and RhizoBase: Databases of manually curated annotations for cyanobacterial and rhizobial genomes. *Nucleic Acids Research*, *42*(D1), D666-70-D670. Journal Article. http://doi.org/10.1093/nar/gkt1145

Gademann, K. (2011). Out in the Green: Biologically Active Metabolites Produced by Cyanobacteria. *Chimia*, *65*(6), 416–419. http://doi.org/10.2533/chimia.2011.416

Gianchandani, E. P., Chavali, A. K., & Papin, J. A. (2010). The application of flux balance analysis in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *2*(3), 372–382. http://doi.org/10.1002/wsbm.60

Gilbert, S. F., & Sarkar, S. (2000). Embracing complexity: Organicism for the 21st century. *Developmental Dynamics*, *219*(1), 1–9. http://doi.org/10.1002/1097-0177(2000)9999:9999<::AID-DVDY1036>3.0.CO;2-A

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., … Barrell, B. (1996). Life with 6000 genes. *Science (New York, N.Y.)*, *274*(5287), 546, 563–7. http://doi.org/10.1126/science.274.5287.546

Gomez, J. a, Höffner, K., & Barton, P. I. (2014). DFBAlab: a fast and reliable MATLAB code for dynamic flux balance analysis. *BMC Bioinformatics*, *15*(1), 409. http://doi.org/10.1186/s12859-014-0409-8

Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merkenschlager, M., Gisel, A., … Geschwind, D. (2014). Data integration in the era of omics: current and future challenges. *BMC Systems Biology*, *8*(Suppl 2), I1. http://doi.org/10.1186/1752-0509-8-S2-I1

Gonzalez, O., Gronau, S., Falb, M., Pfeiffer, F., Mendoza, E., Zimmer, R., & Oesterhelt, D. (2008). Reconstruction, modeling & analysis of Halobacterium salinarum R-1 metabolism. *Molecular bioSystems*, *4*(2), 148–59. http://doi.org/10.1039/b715203e

Griese, M., Lange, C., & Soppa, J. (2011). Ploidy in cyanobacteria. *FEMS Microbiology Letters*, *323*(2), 124–131. http://doi.org/10.1111/j.1574-6968.2011.02368.x

Hackenberg, C., Engelhardt, A., Matthijs, H. C. P., Wittink, F., Bauwe, H., Kaplan, A., & Hagemann, M. (2009). Photorespiratory 2-phosphoglycolate metabolism and photoreduction of O 2 cooperate in high-light acclimation of Synechocystis sp. strain PCC 6803. *Planta*, *230*(4), 625–637. Journal Article. http://doi.org/10.1007/s00425-009-0972-9

Hagemann, M., Weber, A. P., & Eisenhut, M. (2016). Photorespiration: origins and metabolic integration in interacting compartments. *Journal of Experimental Botany*, *67*(10), 2915–2918. http://doi.org/10.1093/jxb/erw178

Hamilton, J. J., Dwivedi, V., & Reed, J. L. (2013). Quantitative assessment of thermodynamic constraints on the solution space of genome-scale metabolic models. *Biophysical Journal*, *105*(2), 512–522. Journal Article. http://doi.org/10.1016/j.bpj.2013.06.011

Hanly, T. J., & Henson, M. A. (2013). Dynamic metabolic modeling of a microaerobic yeast co-culture: predicting and optimizing ethanol production from glucose/xylose mixtures. *Biotechnology for Biofuels*, *6*(1), 44. http://doi.org/10.1186/1754-6834-6-44

Harwood, S. M., Höffner, K., & Barton, P. I. (2016). Efficient solution of ordinary differential equations with a parametric lexicographic linear program embedded. *Numerische Mathematik*, *133*(4), 623–653. http://doi.org/10.1007/s00211-015-0760-3

He, X., Qian, W., Wang, Z., Li, Y., & Zhang, J. (2010). Prevalent positive epistasis in

Escherichia coli and Saccharomyces cerevisiae metabolic networks. *Nature Genetics*, *42*(3), 272–276. Journal Article. http://doi.org/10.1038/ng.524

Heldt, H.-W., & Piechulla, B. (2011). Plant Biochemistry. *Plant Biochemistry*, 163–191. http://doi.org/10.1016/B978-0-12-384986-1.00006-5

Hjersted, J. L., & Henson, M. a. (2009). Steady-state and dynamic flux balance analysis of ethanol production by Saccharomyces cerevisiae. *IET Systems Biology*, *3*(3), 167–179. Journal Article. http://doi.org/10.1049/iet-syb.2008.0103

Hjersted, J. L., & Henson, M. a. (2006). Optimization of fed-batch Saccharomyces cerevisiae fermentation using dynamic flux balance models. *Biotechnology Progress*, *22*(5), 1239–1248. Journal Article. http://doi.org/10.1021/bp060059v

Höffner, K., Harwood, S. M., & Barton, P. I. (2013). A reliable simulator for dynamic flux balance analysis. *Biotechnology and Bioengineering*, *110*(3), 792–802. Journal Article. http://doi.org/10.1002/bit.24748

Hong, S. J., & Lee, C. G. (2007). Evaluation of central metabolism based on a genomic database of Synechocystis PCC6803. *Biotechnology and Bioprocess Engineering*, *12*(2), 165–173. http://doi.org/10.1007/BF03028644

Hordijk, W., Hasenclever, L., Gao, J., Mincheva, D., & Hein, J. (2014). An investigation into irreducible autocatalytic sets and power law distributed catalysis. *Natural Computing*, *13*(3), 287–296. http://doi.org/10.1007/s11047-014-9429-6

Howitt, C. A., Udall, P. K., & Vermaas, W. F. (1999). Type 2 NADH dehydrogenases in the cyanobacterium Synechocystis sp. strain PCC 6803 are involved in regulation rather than respiration. *Journal of Bacteriology*, *181*(13), 3994–4003. http://doi.org/0021-9193/99/$04.00+0

Huege, J., Goetze, J., Schwarz, D., Bauwe, H., Hagemann, M., & Kopka, J. (2011). Modulation of the major paths of carbon in photorespiratory mutants of Synechocystis. *PLoS ONE*, *6*(1). http://doi.org/10.1371/journal.pone.0016278

Hutchison, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., … Renaudin, J. (2016). Design and synthesis of a minimal bacterial genome. *Science (New York, N.Y.)*, *351*(6280), aad6253. http://doi.org/10.1126/science.aad6253

Hyduke, D., Hyduke, D., Schellenberger, J., Que, R., Fleming, R., Thiele, I., … Palsson, B. (2011). COBRA Toolbox 2.0. *Protocol Exchange*, 1–35. http://doi.org/10.1038/protex.2011.234

Hyduke, D. R., Lewis, N. E., & Palsson, B. Ø. (2013). Analysis of omics data with genome-scale models of metabolism. *Molecular bioSystems*, *9*(2), 167–74. http://doi.org/10.1039/c2mb25453k

Ibarra, R. U., Edwards, J. S., & Palsson, B. O. (2002). Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, *420*(6912), 186–189. Journal Article. http://doi.org/10.1038/nature01149

Jain, S., & Krishna, S. (1998). Autocatalytic Sets and the Growth of Complexity in an

Evolutionary Model. http://doi.org/10.1103/PhysRevLett.81.5684

Jakubowska, A., & Korona, R. (2012). Epistasis for growth rate and total metabolic flux in yeast. *PLoS ONE*, *7*(3), e33132. Journal Article. http://doi.org/10.1371/journal.pone.0033132

Jamshidi, N., Edwards, J. S., Fahland, T., Church, G. M., & Palsson, B. O. (2001). Dynamic simulation of the human red blood cell metabolic network. *Bioinformatics (Oxford, England)*, *17*(3), 286–287. Journal Article. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=11294796&retmode=ref&cmd=prlinks

Jamshidi, N., & Palsson, B. Ø. (2007). Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC Systems Biology*, *1*(1), 26. Journal Article. http://doi.org/10.1186/1752-0509-1-26

Jankowski, M. D., Henry, C. S., Broadbelt, L. J., & Hatzimanikatis, V. (2008). Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophysical Journal*, *95*(3), 1487–1499. Journal Article. http://doi.org/10.1529/biophysj.107.124784

Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annual Review of Microbiology*, *30*(1), 409–425. http://doi.org/10.1146/annurev.mi.30.100176.002205

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, a L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*(6804), 651–654. Journal Article. http://doi.org/10.1038/35036627

Jin, Y. S., & Stephanopoulos, G. (2007). Multi-dimensional gene target search for improving lycopene biosynthesis in Escherichia coli. *Metabolic Engineering*, *9*(4), 337–347. Journal Article. http://doi.org/10.1016/j.ymben.2007.03.003

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, *36*(Web Server issue), W5–W9. Journal Article. http://doi.org/10.1093/nar/gkn201

Jones, P. R. (2008). Improving fermentative biomass-derived H2-production by engineering microbial metabolism. *International Journal of Hydrogen Energy*, *33*(19), 5122–5130. http://doi.org/10.1016/j.ijhydene.2008.05.004

Joshi, C. J., & Prasad, A. (2014). Epistatic interactions among metabolic genes depend upon environmental conditions. *Molecular bioSystems*, *10*(10), 2578–2589. Journal Article. http://doi.org/10.1039/c4mb00181h

Kanehisa, M. (2002). The KEGG database. *Novartis Foundation Symposium*, *247*, 91-101-103, 119–128, 244–252. Journal Article. http://doi.org/10.1038/nbt991

Kanehisa, M., & Goto, S. (2000). Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. Journal Article. http://doi.org/10.1093/nar/28.1.27

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, *40*(D1),

gkr988. Journal Article. http://doi.org/10.1093/nar/gkr988

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Research*, *42*(D1), D199–D205. Journal Article. http://doi.org/10.1093/nar/gkt1076

Kaneko, T., Sato, S., Kotani, H., Tanaka, a, Asamizu, E., Nakamura, Y., … Tabata, S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes*, *3*(3), 109–136. Journal Article. http://doi.org/10.1093/dnares/3.3.109

Kauffman, K. J., Prakash, P., & Edwards, J. S. (2003). Advances in flux balance analysis. *Current Opinion in Biotechnology*. http://doi.org/10.1016/j.copbio.2003.08.001

Kauffman, S. A. (1986). Autocatalytic sets of proteins. *Journal of Theoretical Biology*, *119*(1), 1–24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3713221

Kayser, A., Weber, J., Hecht, V., & Rinas, U. (2005). Metabolic flux analysis of Escherichia coli in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state. *Microbiology*, *151*(3), 693–706. Journal Article. http://doi.org/10.1099/mic.0.27481-0

Keating, S. M., Bornstein, B. J., Finney, A., & Hucka, M. (2006). SBMLToolbox: An SBML toolbox for MATLAB users. *Bioinformatics*, *22*(10), 1275–1277. http://doi.org/10.1093/bioinformatics/btl111

Keseler, I. M., Bonavides-Martínez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R. P., Johnson, D. A., … Karp, P. D. (2009). EcoCyc: A comprehensive view of Escherichia coli biology. *Nucleic Acids Research*, *37*(SUPPL. 1), D464–D470. Journal Article. http://doi.org/10.1093/nar/gkn751

Khanin, R., & Wit, E. (2006). How scale-free are biological networks. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *13*(3), 810–8. http://doi.org/10.1089/cmb.2006.13.810

Kibota, T. T., & Lynch, M. (1996). Estimate of the genomic mutation rate deleterious to overall fitness in E. coli. *Nature*, *381*(6584), 694–696. http://doi.org/10.1038/381694a0

Kim, M. K., & Lun, D. S. (2014). Methods for integration of transcriptomic data in genome-scale metabolic models. *Computational and Structural Biotechnology Journal*, *11*(18), 59–65. http://doi.org/10.1016/j.csbj.2014.08.009

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., … Lewis, N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, *44*(D1), D515-22. http://doi.org/10.1093/nar/gkv1049

Klamt, S., Saez-Rodriguez, J., & Gilles, E. (2007). Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biology*, *1*(1), 2. http://doi.org/doi:10.1186/1752-0509-1-2

Klamt, S., Stelling, J., Ginkel, M., & Gilles, E. D. (2003). FluxAnalyzer: Exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics*, *19*(2), 261–269. http://doi.org/10.1093/bioinformatics/19.2.261

Klemke, F., Baier, a., Knoop, H., Kern, R., Jablonsky, J., Beyer, G., … Hagemann, M. (2015). Identification of the light-independent phosphoserine pathway as additional source for serine in the cyanobacterium Synechocystis sp. PCC 6803. *Microbiology*. Journal Article. http://doi.org/10.1099/mic.0.000055

Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W., & Steuer, R. (2013). Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of Synechocystis sp. PCC 6803. *PLoS Computational Biology*, *9*(6), e1003081. Journal Article. http://doi.org/10.1371/journal.pcbi.1003081

Knoop, H., & Steuer, R. (2015). A Computational Analysis of Stoichiometric Constraints and Trade-Offs in Cyanobacterial Biofuel Production. *Frontiers in Bioengineering and Biotechnology*, *3*, 47. Journal Article. http://doi.org/10.3389/fbioe.2015.00047

Knoop, H., Zilliges, Y., Lockau, W., & Steuer, R. (2010). The metabolic network of Synechocystis sp. PCC 6803: systemic properties of autotrophic growth. *Plant Physiology*, *154*(1), 410–22. Journal Article. http://doi.org/10.1104/pp.110.157198

Knorr, A. L., Jain, R., & Srivastava, R. (2007). Bayesian-based selection of metabolic objective functions. *Bioinformatics*, *23*(3), 351–357. http://doi.org/10.1093/bioinformatics/btl619

Kochanowski, K., Sauer, U., & Chubukov, V. (2013). Somewhat in control-the role of transcription in regulating microbial metabolic fluxes. *Current Opinion in Biotechnology*, *24*(6), 987–993. http://doi.org/10.1016/j.copbio.2013.03.014

Kompala, D. S., Ramkrishna, D., Jansen, N. B., & Tsao, G. T. (1986). Investigation of bacterial growth on mixed substrates: experimental evaluation of cybernetic models. *Biotechnology and Bioengineering*, *28*(August 2015), 1044–1055. http://doi.org/10.1002/bit.260280715

Kompala, D. S., Ramkrishna, D., & Tsao, G. T. (1984). Cybernetic modeling of microbial growth on multiple substrates. *Biotechnology and Bioengineering*, *26*(11), 1272–1281. http://doi.org/10.1002/bit.260261103

Kopecná, J., Komenda, J., Bucinská, L., & Sobotka, R. (2012). Long-term acclimation of the cyanobacterium Synechocystis sp. PCC 6803 to high light is accompanied by an enhanced production of chlorophyll that is preferentially channeled to trimeric photosystem I. *Plant Physiology*, *160*(4), 2239–50. http://doi.org/10.1104/pp.112.207274

Kopf, M., Klähn, S., Scholz, I., Matthiessen, J. K. F., Hess, W. R., & Voß, B. (2014). Comparative analysis of the primary transcriptome of Synechocystis sp. PCC 6803. *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes*, *21*(5), 527–39. http://doi.org/10.1093/dnares/dsu018

Latifi, A., Ruiz, M., & Zhang, C. C. (2009). Oxidative stress in cyanobacteria. *FEMS Microbiology Reviews*, *33*(2), 258–278. Journal Article. http://doi.org/10.1111/j.1574-6976.2008.00134.x

Lea, P. J., & Leegood, R. C. (1999). *Plant biochemistry and molecular biology*. John Wiley.

Lee, H., Popodi, E., Tang, H., & Foster, P. L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(41), E2774-83. http://doi.org/10.1073/pnas.1210309109

Lee, J., Yun, H., Feist, A. M., Palsson, B., & Lee, S. Y. (2008). Genome-scale reconstruction and in silico analysis of the Clostridium acetobutylicum ATCC 824 metabolic network. *Applied Microbiology and Biotechnology*, *80*(5), 849–862. http://doi.org/10.1007/s00253-008-1654-4

Lew, V. L., & Bookchin, R. M. (1986). Volume, pH, and ion-content regulation in human red cells: analysis of transient behavior with an integrated model. *J Membr Biol*, *92*(1), 57–74.

Liao, Y. C., Huang, T. W., Chen, F. C., Charusanti, P., Hong, J. S. J., Chang, H. Y., … Hsiung, C. a. (2011). An experimentally validated genome-scale metabolic reconstruction of Klebsiella pneumoniae MGH 78578, iYL1228. *Journal of Bacteriology*, *193*(7), 1710–1717. Journal Article. http://doi.org/10.1128/JB.01218-10

Light, S., & Kraulis, P. (2004). Network analysis of metabolic enzyme evolution in Escherichia coli. *BMC Bioinformatics*, *5*, 15. Journal Article. http://doi.org/10.1186/1471-2105-5-15

Lima-Mendez, G., & van Helden, J. (2009). The powerful law of the power law and other myths in network biology. *Molecular BioSystems*, *5*(12), 1482. http://doi.org/10.1039/b908681a

Luo, R., Liao, S., Zeng, S., Li, Y., & Luo, Q. (2006). FluxExplorer: A general platform for modeling and analyses of metabolic networks based on stoichiometry. *Chinese Science Bulletin*, *51*(6), 689–696. http://doi.org/10.1007/s11434-006-0689-0

Luo, R.-Y., Liao, S., Tao, G.-Y., Li, Y.-Y., Zeng, S., Li, Y.-X., & Luo, Q. (2006). Dynamic analysis of optimality in myocardial energy metabolism under normal and ischemic conditions. *Molecular Systems Biology*, *2*, 2006.0031. http://doi.org/10.1038/msb4100071

Lv, Q., Ma, W., Liu, H., Li, J., Wang, H., Lu, F., … Mi, H. (2015). Genome-wide protein-protein interactions and protein function exploration in cyanobacteria. *Scientific Reports*, *5*, 15519. http://doi.org/10.1038/srep15519

Ma, W., Ogawa, T., Shen, Y., & Mi, H. (2007). Changes in cyclic and respiratory electron transport by the movement of phycobilisomes in the cyanobacterium Synechocystis sp. strain PCC 6803. *Biochimica et Biophysica Acta - Bioenergetics*, *1767*(6), 742–749. Journal Article. http://doi.org/10.1016/j.bbabio.2007.01.017

Maarleveld, T. R., Boele, J., Bruggeman, F. J., & Teusink, B. (2014). A data integration and visualization resource for the metabolic network of Synechocystis sp. PCC 6803. *Plant Physiology*, *164*(3), 1111–21. Journal Article. http://doi.org/10.1104/pp.113.224394

Mahadevan, R., Edwards, J. S., & Doyle, F. J. (2002). Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical Journal*, *83*(3), 1331–1340. Journal Article. http://doi.org/10.1016/S0006-3495(02)73903-9

Mahadevan, R., & Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, *5*(4), 264–276. Journal Article. http://doi.org/10.1016/j.ymben.2003.09.002

Martin, G., Elena, S. F., & Lenormand, T. (2007). Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature Genetics*, *39*(4), 555–560. Journal Article. http://doi.org/10.1038/ng1998

Maruyama, J., Yamaoka, S., Matsuo, I., Tsutsumi, N., & Kitamoto, K. (2012). A newly discovered function of peroxisomes: involvement in biotin biosynthesis. *Plant Signaling & Behavior*, *7*(12), 1589–93. http://doi.org/10.4161/psb.22405

Matsuo, M., Endo, T., & Asada, K. (1998). Properties of the respiratory NAD(P)H dehydrogenase isolated from the cyanobacterium Synechocystis PCC6803. *Plant & Cell Physiology*, *39*(3), 263–267.

Mattick, J. S., & Rinn, J. L. (2015). Discovery and annotation of long noncoding RNAs. *Nature Structural & Molecular Biology*, *22*(1), 5–7. http://doi.org/10.1038/nsmb.2942

Min Lee, J., Gianchandani, E. P., Eddy, J. A., & Papin, J. A. (2008). Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Computational Biology*, *4*(5). http://doi.org/10.1371/journal.pcbi.1000086

Monod, J. (1949). The Growth of Bacterial Cultures. *Annual Review of Microbiology*, *3*(1), 371–394. http://doi.org/10.1146/annurev.mi.03.100149.002103

Montagud, A., Navarro, E., Fernández de Córdoba, P., Urchueguía, J. F., & Patil, K. R. (2010). Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. *BMC Systems Biology*, *4*(1), 156. Journal Article. http://doi.org/10.1186/1752-0509-4-156

Montagud, A., Zelezniak, A., Navarro, E., de Córdoba, P. F., Urchueguía, J. F., & Patil, K. R. (2011). Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium Synechocystis sp. PCC6803. *Biotechnology Journal*, *6*(3), 330–342. Journal Article. http://doi.org/10.1002/biot.201000109

Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., Fujisawa, T., Sato, S., … Nakamura, Y. (2009). CyanoBase: The cyanobacteria genome database update 2010. *Nucleic Acids Research*, *38*(SUPPL.1), D379–D381. Journal Article. http://doi.org/10.1093/nar/gkp915

Nam, H., Lewis, N. E., Lerman, J. a., Lee, D.-H., Chang, R. L., Kim, D., & Palsson, B. O. (2012). Network Context and Selection in the Evolution to Enzyme Specificity. *Science*, *337*(6098), 1101–1104. Journal Article. http://doi.org/10.1126/science.1216861

Nikolaev, E. V. (2010). The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems. *Metabolic Engineering*, *12*(1), 26–38. Journal Article. http://doi.org/10.1016/j.ymben.2009.08.010

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O., & Thiele, I. (2012). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proceedings of the National Academy of Sciences*, *109*(7), 2678–2683. Journal Article. http://doi.org/10.1073/pnas.1117907109

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O., & Thiele, I. (2012). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(7),

2678–83. Journal Article. http://doi.org/10.1073/pnas.1117907109

Nozzi, N. E., Oliver, J. W. K., & Atsumi, S. (2013). Cyanobacteria as a Platform for Biofuel Production. *Frontiers in Bioengineering and Biotechnology*, *lipid*(September), 1–6. http://doi.org/10.3389/fbioe.2013.00007

Oberhardt, M. a, Palsson, B. Ø., & Papin, J. a. (2009). Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology*, *5*(320), 320. http://doi.org/10.1038/msb.2009.77

Oddone, G. M., Mills, D. A., & Block, D. E. (2009). A dynamic, genome-scale flux model of Lactococcus lactis to increase specific recombinant protein expression. *Metabolic Engineering*, *11*(6), 367–381. http://doi.org/10.1016/j.ymben.2009.07.007

Oh, Y. K., Palsson, B. O., Park, S. M., Schilling, C. H., & Mahadevan, R. (2007). Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *Journal of Biological Chemistry*, *282*(39), 28791–28799. Journal Article. http://doi.org/10.1074/jbc.M703759200

Oliveira, A. P., Nielsen, J., Förster, J., & Forster, J. (2005). Modeling Lactococcus lactis using a genome-scale flux model. *BMC Microbiology*, *5*(1), 39. Journal Article. http://doi.org/10.1186/1471-2180-5-39

Oliver, J. W. K., & Atsumi, S. (2015). A carbon sink pathway increases carbon productivity in cyanobacteria. *Metabolic Engineering*, *29*, 106–112. http://doi.org/10.1016/j.ymben.2015.03.006

Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., & Palsson, B. Ø. (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Molecular Systems Biology*, *7*, 535. http://doi.org/10.1038/msb.2011.65

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. Journal Article. http://doi.org/10.1038/nbt.1614

Overbeek, R., Larsen, N., Pusch, G. D., D'Souza, M., Selkov, E., Kyrpides, N., … Selkov, E. (2000). WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Research*, *28*(1), 123–125. http://doi.org/10.1093/nar/28.1.123

Paley, S. M., & Karp, P. D. (2006). The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Research*, *34*(13), 3771–3778. Journal Article. http://doi.org/10.1093/nar/gkl334

Palsson, B. (2006). Systems biology: Determining the capabilities of reconstructed networks. *Cambridge Univ Pr*. Journal Article. http://doi.org/10.1017/CBO9780511790515

Papadopoulos, F., Kitsak, M., Serrano, M. A., Boguna, M., & Krioukov, D. (2012). Popularity versus similarity in growing networks. *Nature*, *489*(7417), 537–540. http://doi.org/10.1038/nature11459

Papin, J. a., Stelling, J., Price, N. D., Klamt, S., Schuster, S., & Palsson, B. O. (2004). Comparison of network-based pathway analysis methods. *Trends in Biotechnology*, *22*(8),

400–405. Journal Article. http://doi.org/10.1016/j.tibtech.2004.06.010

Patil, K. R., Rocha, I., Förster, J., & Nielsen, J. (2005). Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics*, *6*(1), 308. Journal Article. http://doi.org/10.1186/1471-2105-6-308

Perfeito, L., Fernandes, L., Mota, C., & Gordo, I. (2007). Adaptive mutations in bacteria: high rate and small effects. *Science (New York, N.Y.)*, *317*(5839), 813–815. http://doi.org/10.1126/science.1142284

Pharkya, P., & Maranas, C. D. (2006). An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metabolic Engineering*, *8*(1), 1–13. http://doi.org/10.1016/j.ymben.2005.08.003

Phillips, P. C. (2008). Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews. Genetics*, *9*(11), 855–867. Journal Article. http://doi.org/10.1038/nrg2452

Poelwijk, F. J., Tănase-Nicola, S., Kiviet, D. J., & Tans, S. J. (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology*, *272*(1), 141–144. Journal Article. http://doi.org/10.1016/j.jtbi.2010.12.015

Poolman, M. G., Miguet, L., Sweetlove, L. J., & Fell, D. A. (2009). A genome-scale metabolic model of Arabidopsis thaliana and some of its properties. *Plant Physiol.*, *151*, pp.109.141267. http://doi.org/10.1104/pp.109.141267

Price, N. D., Reed, J. L., & Palsson, B. Ø. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews. Microbiology*, *2*(11), 886–897. Journal Article. http://doi.org/10.1038/nrmicro1023

Puchalka, J., Oberhardt, M. A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K. N., … Martins Dos Santos, V. A. P. (2008). Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology. *PLoS Computational Biology*, *4*(10). http://doi.org/10.1371/journal.pcbi.1000210

Quintero, M. J., Montesinos, M. L., Herrero, A., & Flores, E. (2001). Identification of genes encoding amino acid permeases by inactivation of selected ORFs from the Synechocystis genomic sequence. *Genome Research*, *11*(12), 2034–2040. http://doi.org/10.1101/gr.196301

Rae, B. D., Long, B. M., Whitehead, L. F., Förster, B., Badger, M. R., & Price, G. D. (2013). Cyanobacterial carboxysomes: Microcompartments that facilitate CO 2 fixation. *Journal of Molecular Microbiology and Biotechnology*, *23*(4–5), 300–307. http://doi.org/10.1159/000351342

Raman, K., & Chandra, N. (2009). Flux balance analysis of biological systems: Applications and challenges. *Briefings in Bioinformatics*, *10*(4), 435–449. Journal Article. http://doi.org/10.1093/bib/bbp011

Ravasz, E., Somera, a L., Mongru, D. a, Oltvai, Z. N., & Barabási, a L. (2002). Hierarchical organization of modularity in metabolic networks. *Science (New York, N.Y.)*, *297*(5586), 1551–1555. Journal Article. http://doi.org/10.1126/science.1073374

Reddy, T. B. K., Thomas, A. D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., … Kyrpides, N. C. (2015). The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Research*, *43*(Database issue), D1099-106. http://doi.org/10.1093/nar/gku950

Reed, J. L., & Palsson, B. (2004). Genome-scale in silico models of E. coli have multiple equivalent phenotypic states: Assessment of correlated reaction subsets that comprise network states. *Genome Research*, *14*(9), 1797–1805. Journal Article. http://doi.org/10.1101/gr.2546004

Remold, S. K., & Lenski, R. E. (2004). Pervasive joint influence of epistasis and plasticity on mutational effects in Escherichia coli. *Nature Genetics*, *36*(4), 423–426. Journal Article. http://doi.org/10.1038/ng1324

Roberts, S. B., Gowen, C. M., Brooks, J. P., & Fong, S. S. (2010). Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production. *BMC Systems Biology*, *4*(1), 31. Journal Article. http://doi.org/10.1186/1752-0509-4-31

Rocha, I., Förster, J., & Nielsen, J. (2008). *Microbial gene essentiality - protocols and bioinformatics*. *Microbial gene essentiality - protocols and bioinformatics* (Vol. 416). Book, Humana Press. http://doi.org/10.1007/978-1-59745-321-9

Roy, S. (1999). Multifunctional enzymes and evolution of biosynthetic pathways: Retro-evolution by jumps. *Proteins: Structure, Function and Genetics*, *37*(2), 303–309. Journal Article. http://doi.org/10.1002/(SICI)1097-0134(19991101)37:2<303::AID-PROT15>3.0.CO;2-6

Saha, R., Liu, D., Hoynes-O'Connor, A., Liberton, M., Yu, J., Bhattacharyya-Pakrasi, M., … Pakrasi, H. B. (2016). Diurnal Regulation of Cellular Processes in the Cyanobacterium *Synechocystis* sp. Strain PCC 6803: Insights from Transcriptomic, Fluxomic, and Physiological Analyses. *mBio*, *7*(3), e00464-16. http://doi.org/10.1128/mBio.00464-16

Saha, R., Verseput, A. T., Berla, B. M., Mueller, T. J., Pakrasi, H. B., & Maranas, C. D. (2012). Reconstruction and Comparison of the Metabolic Potential of Cyanobacteria Cyanothece sp. ATCC 51142 and Synechocystis sp. PCC 6803. *PLoS ONE*, *7*(10), e48285. Journal Article. http://doi.org/10.1371/journal.pone.0048285

Salimi, F., Mandal, R., Wishart, D., & Mahadevan, R. (2010). Understanding clostridium acetobutylicum ATCC 824 metabolism using genome-scale thermodynamics and metabolomics-based modeling. In *IFAC Proceedings Volumes (IFAC-PapersOnline)* (Vol. 11, pp. 126–131). http://doi.org/10.3182/20100707-3-BE-2012.0022

Satish Kumar, V., Dasika, M. S., & Maranas, C. D. (2007). Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics*, *8*, 212. http://doi.org/10.1186/1471-2105-8-212

Savinell, J. M., & Palsson, B. O. (1992a). Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *Journal of Theoretical Biology*, *154*(4), 421–454. Journal Article. http://doi.org/Doi 10.1016/S0022-5193(05)80161-4

Savinell, J. M., & Palsson, B. O. (1992b). Optimal selection of metabolic fluxes for in vivo measurement. II. Application to Escherichia coli and hybridoma cell metabolism. *Journal of Theoretical Biology*, *155*(2), 215–242. Journal Article. http://doi.org/10.1016/S0022-5193(05)80596-X

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., … Schomburg, D. (2011). BRENDA, the enzyme information system in 2011. *Nucleic Acids Research*, *39*(SUPPL. 1), D670-6. Journal Article. http://doi.org/10.1093/nar/gkq1089

Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., … Palsson, B. Ø. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, *6*(9), 1290–1307. http://doi.org/10.1038/nprot.2011.308

Schoenrock, A., Samanfar, B., Pitre, S., Hooshyar, M., Jin, K., Phillips, C. A., … Kerbosch, J. (2014). Efficient prediction of human protein-protein interactions at a global scale. *BMC Bioinformatics*, *15*(1), 383. http://doi.org/10.1186/s12859-014-0383-1

Schuetz, R., Kuepfer, L., & Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular Systems Biology*, *3*, 119. Journal Article. http://doi.org/10.1038/msb4100162

Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., & Sauer, U. (2012). Multidimensional Optimality of Microbial Metabolism. *Science*, *336*(6081), 601–604. Journal Article. http://doi.org/10.1126/science.1216882

Schwartz, J.-M., & Kanehisa, M. (2006). Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. *BMC Bioinformatics*, *7*(1), 186. Journal Article. http://doi.org/10.1186/1471-2105-7-186

Segrè, D., Deluna, A., Church, G. M., & Kishony, R. (2005). Modular epistasis in yeast metabolism. *Nature Genetics*, *37*(1), 77–83. Journal Article. http://doi.org/10.1038/ng1489

Segrè, D., & Marx, C. J. (2010). Introduction to focus issue: Genetic interactions. *Chaos*, *20*(2), 26101. Journal Article. http://doi.org/10.1063/1.3456057

Segrè, D., Vitkup, D., & Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(23), 15112–15117. Journal Article. http://doi.org/10.1073/pnas.232349399

Shastri, A. A., & Morgan, J. A. (2005). Flux balance analysis of photoautotrophic metabolism. *Biotechnology Progress*, *21*(6), 1617–1626. Journal Article. http://doi.org/10.1021/bp050246d

Sheikh, K., Forster, J., & Nielsen, L. K. (2005). Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of Mus musculus. *Biotechnology Progress*, *21*(1), 112–121. http://doi.org/10.1021/bp0498138

Shlomi, T., Berkman, O., & Ruppin, E. (2005). Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(21), 7695–7700. Journal Article. http://doi.org/10.1073/pnas.0406346102

Shlomi, T., Cabili, M. N., & Ruppin, E. (2009). Predicting metabolic biomarkers of human inborn errors of metabolism. *Molecular Systems Biology*, *5*(263), 263. http://doi.org/10.1038/msb.2009.22

Snitkin, E. S., & Segrè, D. (2011). Epistatic interaction Maps relative to multiple metabolic phenotypes. *PLoS Genetics*, *7*(2), e1001294. Journal Article. http://doi.org/10.1371/journal.pgen.1001294

Snoep, J. L., & Westerhoff, H. V. (2005). From isolation to integration, a systems biology approach for building the Silicon Cell. In *Systems Biology SE - 61* (Vol. 13, pp. 13–30). Berlin/Heidelberg: Springer-Verlag. http://doi.org/10.1007/b106456

Song, H. S., & Ramkrishna, D. (2009a). Reduction of a set of elementary modes using yield analysis. *Biotechnology and Bioengineering*, *102*(2), 554–568. http://doi.org/10.1002/bit.22062

Song, H. S., & Ramkrishna, D. (2009b). When is the quasi-steady-state approximation admissible in metabolic modeling? When admissible, what models are desirable? *Industrial and Engineering Chemistry Research*, *48*(17), 7976–7985. http://doi.org/10.1021/ie900075f

Song, H. S., & Ramkrishna, D. (2011). Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function. *Biotechnology and Bioengineering*, *108*(1), 127–140. Journal Article. http://doi.org/10.1002/bit.22922

Spirin, V., Gelfand, M. S., Mironov, A. a, & Mirny, L. a. (2006). A metabolic network in the evolutionary context: multiscale structure and modularity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8774–8779. Journal Article. http://doi.org/10.1073/pnas.0510258103

Stephanopoulos, G. N., Aristidou, A. a, & Nielsen, J. (1998). *Metabolic Engineering: Principles and Methodologies. Metabolic Engineering* (Vol. 54). Retrieved from http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0126662606

Stumpf, M. P. H., & Porter, M. A. (2012). Critical Truths About Power Laws. *Science*, *335*(6069), 665–666. http://doi.org/10.1126/science.1216142

Sun, J., Sayyar, B., Butler, J. E., Pharkya, P., Fahland, T. R., Famili, I., … Mahadevan, R. (2009). Genome-scale constraint-based modeling of Geobacter metallireducens. *BMC Systems Biology*, *3*, 15. http://doi.org/10.1186/1752-0509-3-15

Suthers, P. F., Zomorrodi, A., & Maranas, C. D. (2009). Genome-scale gene/reaction essentiality and synthetic lethality analysis. *Molecular Systems Biology*, *5*, 301. Journal Article. http://doi.org/10.1038/msb.2009.56

Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, *5*(1), 93–121. Journal Article. http://doi.org/10.1038/nprot.2009.203

Timm, S., & Bauwe, H. (2013). The variety of photorespiratory phenotypes - employing the current status for future research directions on photorespiration. *Plant Biology*. http://doi.org/10.1111/j.1438-8677.2012.00691.x

Tomo, T., Kusakabe, H., Nagao, R., Ito, H., Tanaka, A., Akimoto, S., … Okazaki, S. (2012). Luminescence of singlet oxygen in photosystem II complexes isolated from cyanobacterium Synechocystis sp. PCC6803 containing monovinyl or divinyl chlorophyll a. *Biochimica et Biophysica Acta - Bioenergetics*, *1817*(8), 1299–1305. Journal Article. http://doi.org/10.1016/j.bbabio.2012.02.018

Uygun, K., Matthew, H. W. T., & Huang, Y. (2006). DFBA-LQR: An Optimal Control Approach to Flux Balance Analysis. *Industrial & Engineering Chemistry Research*, *45*(25), 8554–8564. http://doi.org/10.1021/ie060218f

Varma, a., & Palsson, B. O. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Applied and Environmental Microbiology*, *60*(10), 3724–3731. Journal Article. http://doi.org/PMC201879

Varma, a, & Palsson, B. O. (1993). Metabolic capabilities of Escherichia coli: I. synthesis of biosynthetic precursors and cofactors. *Journal of Theoretical Biology*, *165*(4), 477–502. Journal Article. http://doi.org/10.1006/jtbi.1993.1202

Vermaas, W. F. J. (2001). *Photosynthesis and Respiration in Cyanobacteria. Life Sciences*. Book. http://doi.org/10.1038/npg.els.0001670

Vijayakumar, S., & Menakha, M. (2015). Pharmaceutical applications of cyanobacteria-A review. *Journal of Acute Medicine*. http://doi.org/10.1016/j.jacme.2015.02.004

Wang, X., Minasov, G., & Shoichet, B. K. (2002). Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *Journal of Molecular Biology*, *320*(1), 85–95. Journal Article. http://doi.org/10.1016/S0022-2836(02)00400-X

Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., & Bryant, S. H. (2009). PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkp456

Watanabe, M., Semchonok, D. a, Webber-Birungi, M. T., Ehira, S., Kondo, K., Narikawa, R., … Ikeuchi, M. (2014). Attachment of phycobilisomes in an antenna-photosystem I supercomplex of cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(7), 2512–7. Journal Article. http://doi.org/10.1073/pnas.1320599111

Weisz, P. B. (1973). Diffusion and Chemical Transformation. *Science*, *179*(4072), 433–440. http://doi.org/10.1126/science.179.4072.433

West, G. B., Brown, J. H., & Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science*, *276*(5309), 122–6. http://doi.org/10.1126/science.276.5309.122

Wuchty, S., Uetz, P., Hu, P., Butland, G., Yu, H., Stelzl, U., … Barabasi, A. L. (2014). Protein-protein Interaction Networks of E. coli and S. cerevisiae are similar. *Scientific Reports*, *4*, 7187. http://doi.org/10.1038/srep07187

Xavier, J. C., Patil, K. R., & Rocha, I. (2014). Systems biology perspectives on minimal and simpler cells. *Microbiology and Molecular Biology Reviews : MMBR*, *78*(3), 487–509.

http://doi.org/10.1128/MMBR.00050-13

Xu, L., Barker, B., & Gu, Z. (2012). Dynamic epistasis for different alleles of the same gene. *Proceedings of the National Academy of Sciences*, *109*(26), 10420–10425. http://doi.org/10.1073/pnas.1121507109

Xu, Z.-X. (2008). Constrain-based analysis of gene deletion on the metabolic flux redistribution of Saccharomyces Cerevisiae. *Journal of Biomedical Science and Engineering*, *1*(2), 121–126. Journal Article. http://doi.org/10.4236/jbise.2008.12020

Yang, A. (2011). Modeling and evaluation of CO 2 supply and utilization in algal ponds. *Industrial and Engineering Chemistry Research*, *50*(19), 11181–11192. http://doi.org/10.1021/ie200723w

Yang, C., Hua, Q., & Shimizu, K. (2002). Metabolic Flux Analysis in Synechocystis Using Isotope Distribution from 13C-Labeled Glucose. *Metabolic Engineering*, *4*(3), 202–216. Journal Article. http://doi.org/10.1006/mben.2002.0226

Yizhak, K., Le Dévédec, S. E., Rogkoti, V. M., Baenke, F., de Boer, V. C., Frezza, C., … Ruppin, E. (2014). A computational study of the Warburg effect identifies metabolic targets inhibiting cancer migration. *Molecular Systems Biology*, *10*(8), 744. Journal Article. http://doi.org/10.15252/msb.20134993

Yoshikawa, K., Kojima, Y., Nakajima, T., Furusawa, C., Hirasawa, T., & Shimizu, H. (2011). Reconstruction and verification of a genome-scale metabolic model for Synechocystis sp. PCC6803. *Applied Microbiology and Biotechnology*, *92*(2), 347–358. Journal Article. http://doi.org/10.1007/s00253-011-3559-x

Young, J. D., Shastri, A. a., Stephanopoulos, G., & Morgan, J. a. (2011). Mapping photoautotrophic metabolism with isotopically nonstationary 13C flux analysis. *Metabolic Engineering*, *13*(6), 656–665. Journal Article. http://doi.org/10.1016/j.ymben.2011.08.002

Young, J., Henne, K., Morgan, J., Konopka, A., & Ramkrishna, D. (2004). Cybernetic modeling of metabolism: Towards a framework for rational design of recombinant organisms. In *Chemical Engineering Science* (Vol. 59, pp. 5041–5049). http://doi.org/10.1016/j.ces.2004.09.037

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., … Brinkman, F. S. L. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, *26*(13), 1608–1615. http://doi.org/10.1093/bioinformatics/btq249

Yu, Y., You, L., Liu, D., Hollinshead, W., Tang, Y. J., & Zhang, F. (2013). Development of synechocystis sp. PCC 6803 as a phototrophic cell factory. *Marine Drugs*. http://doi.org/10.3390/md11082894

Zanghellini, J., Ruckerbauer, D. E., Hanscho, M., & Jungreuthmayer, C. (2013). Elementary flux modes in a nutshell: Properties, calculation and applications. *Biotechnology Journal*, *8*(9), 1009–1016. Journal Article. http://doi.org/10.1002/biot.201200269

Zhang, S., & Bryant, D. a. (2011). The Tricarboxylic Acid Cycle in Cyanobacteria. *Science*,

334(6062), 1551–1553. Journal Article. http://doi.org/10.1126/science.1210858

Zhu, X. G., Wang, Y., Ort, D. R., & Long, S. P. (2013). e-photosynthesis: A comprehensive dynamic mechanistic model of C3 photosynthesis: From light capture to sucrose synthesis. *Plant, Cell and Environment*, *36*(9), 1711–1727. Journal Article. http://doi.org/10.1111/pce.12025

APPENDIX A: SUPPLEMENTARY MATERIAL FOR CHAPTER 1

TABLE A1: LIST OF ALL THE ORGANISMAL METABOLIC MODELS PUBLISHED TILL DATE. THE DATA IS EXTRACTED FROM BIGG DATABASE. (sorted by organism)

BiGG ID	Organism	Metabolites	Reactions	Genes
iYO844	*Bacillus subtilis str.* 168	991	1250	844
iRC1080	*Chlamydomonas reinhardtii*	1706	2191	1086
iHN637	*Clostridium ljungdahlii* DSM 13528	698	785	637
iEC042_1314	*Escherichia coli* 042	1926	2715	1314
iECP_1309	*Escherichia coli* 536	1943	2740	1309
iEC55989_1330	*Escherichia coli* 55989	1953	2757	1330
iECABU_c1320	*Escherichia coli* ABU 83972	1944	2732	1320
iAPECO1_1312	*Escherichia coli* APEC O1	1944	2736	1313
iEcolC_1368	*Escherichia coli* ATCC 8739	1971	2769	1368
iECB_1328	*Escherichia coli* B *str.* REL606	1953	2749	1329
iB21_1397	*Escherichia coli* BL21(DE3)	1945	2742	1337
iECD_1391	*Escherichia coli* BL21(DE3)	1945	2742	1333
iECBD_1354	*Escherichia coli* 'BL21-Gold(DE3)pLysS AG'	1954	2749	1354
iBWG_1329	*Escherichia coli* BW2952	1949	2742	1328
ic_1306	*Escherichia coli* CFT073	1938	2727	1307
iEcDH1_1363	*Escherichia coli* DH1	1949	2751	1363
iECDH1ME8569_1439	*Escherichia coli* DH1	1950	2756	1439
iEcE24377_1341	*Escherichia coli* E24377A	1974	2764	1341
iECED1_1282	*Escherichia coli* ED1a	1929	2707	1279
iETEC_1333	*Escherichia coli* ETEC H10407	1964	2757	1333
iEcHS_1320	*Escherichia coli* HS	1965	2754	1321
iECIAI1_1343	*Escherichia coli* IAI1	1970	2766	1343
iECIAI39_1322	*Escherichia coli* IAI39	1957	2722	1321
iECOK1_1307	*Escherichia coli* IHE3034	1943	2730	1304
iEKO11_1354	*Escherichia coli* KO11FL	1974	2779	1354
iLF82_1304	*Escherichia coli* LF82	1940	2727	1302
iECNA114_1301	*Escherichia coli* NA114	1927	2719	1301
iECO103_1326	*Escherichia coli* O103:H2 *str.* 12009	1958	2759	1327
iECO111_1330	*Escherichia coli* O111:H- *str.* 11128	1959	2761	1328
iE2348C_1286	*Escherichia coli* O127:H6 *str.* E2348/69	1919	2704	1284
iECH74115_1262	*Escherichia coli* O157:H7 *str.* EC4115	1918	2695	1262
iZ_1308	*Escherichia coli* O157:H7 *str.* EDL933	1923	2722	1308
iECs_1301	*Escherichia coli* O157:H7 *str.* Sakai	1923	2721	1301
iECSP_1301	*Escherichia coli* O157:H7 *str.* TW14359	1920	2713	1299

iECO26_1355	*Escherichia coli* O26:H11 *str.* 11368	1965	2781	1355
iG2583_1286	*Escherichia coli* O55:H7 *str.* CB9615	1919	2705	1283
iNRG857_1313	*Escherichia coli* O83:H1 *str.* NRG 857C	1945	2736	1311
iECS88_1305	*Escherichia coli* S88	1944	2730	1305
iECSE_1348	*Escherichia coli* SE11	1957	2769	1348
iECSF_1327	*Escherichia coli* SE15	1951	2743	1327
iEcSMS35_1347	*Escherichia coli* SMS-3-5	1949	2747	1347
iECDH10B_1368	*Escherichia coli str.* K-12 *substr.* DH10B	1947	2743	1327
e_coli_core	*Escherichia coli* str. K-12 *substr.* MG1655	72	95	137
iAF1260	*Escherichia coli* str. K-12 *substr.* MG1655	1668	2382	1261
iAF1260b	*Escherichia coli* str. K-12 *substr.* MG1655	1668	2388	1261
iJO1366	*Escherichia coli* str. K-12 *substr.* MG1655	1805	2583	1367
iJR904	*Escherichia coli* str. K-12 *substr.* MG1655	761	1075	904
iY75_1357	*Escherichia coli* str. K-12 *substr.* W3110	1953	2760	1358
iUMN146_1321	*Escherichia coli* UM146	1944	2736	1319
iECUMN_1333	*Escherichia coli* UMN026	1935	2741	1332
iUMNK88_1353	*Escherichia coli* UMNK88	1971	2778	1353
iUTI89_1310	*Escherichia coli* UTI89	1942	2726	1310
iECW_1372	*Escherichia coli* W	1975	2783	1372
iWFL_1372	*Escherichia coli* W	1975	2783	1372
iAF987	*Geobacter metallireducens* GS-15	1109	1285	987
iIT341	*Helicobacter pylori* 26695	485	554	339
iAB_RBC_283	*Homo sapiens*	342	469	346
iAT_PLT_636	*Homo sapiens*	738	1008	636
RECON1	*Homo sapiens*	2766	3742	1905
iYL1228	*Klebsiella pneumoniae subsp.* pneumoniae MGH 78578	1658	2262	1229
iAF692	*Methanosarcina barkeri str.* Fusaro	628	690	692
iMM1415	*Mus musculus*	2775	3726	1375
iNJ661	*Mycobacterium tuberculosis* H37Rv	826	1025	661
iJN746	*Pseudomonas putida* KT2440	909	1056	746
iMM904	*Saccharomyces cerevisiae* S288c	1226	1577	905
iND750	*Saccharomyces cerevisiae* S288c	1059	1266	750
STM_v1_0	*Salmonella enterica subsp.* enterica serovar *Typhimurium str.* LT2	1802	2545	1271
iSbBS512_1146	*Shigella boydii* CDC 3083-94	1912	2592	1147
iSBO_1134	*Shigella boydii* Sb227	1910	2592	1134
iSDY_1059	*Shigella dysenteriae* Sd197	1890	2540	1059
iSFxv_1172	*Shigella flexneri* 2002017	1918	2639	1169
iS_1188	*Shigella flexneri* 2a *str.* 2457T	1914	2620	1188
iSF_1195	*Shigella flexneri* 2a *str.* 301	1917	2631	1195

iSFV_1184	*Shigella flexneri 5 str.* 8401	1917	2622	1184
iSSON_1240	*Shigella sonnei* Ss046	1938	2694	1240
iSB619	*Staphylococcus aureus subsp. aureus* N315	655	743	619
iJN678	*Synechocystis sp.* PCC 6803	795	863	622
iLJ478	*Thermotoga maritima* MSB8	570	652	482
iPC815	*Yersinia pestis* CO92	1552	1961	815

TABLE A2: LIST OF DFBA APPLICATIONS IN VARIOUS ORGANISMS

Reference	Organism(s)/Pathway(s)	Network size		Methods/Solvers
		Metabolites	**Reactions**	
Varma A, Palsson BØ. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol 60(10): 3724.	*Escherichia coli*	24	34	SOA/—
Mahadevan R, Edwards JS, Doyle FJ. 2002. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys J 83(3):1331–1340.	*Escherichia coli*	3	4	SOA/CPLEX DOA/fmincon
Sainz J, Pizarro F, Pe´rez-Correa JR, Agosin E. 2003. Modeling of yeast metabolism and process dynamics in batch fermentation. Biotechnol Bioeng 81(7):818–828.	*Saccharomyces cerevisiae*	43	38	SOA/—
Luo R, Liao S, Tao G, Li Y, Zeng S, Li Y, Luo Q. 2006. Dynamic analysis of optimality in myocardial energy metabolism under normal and ischemic conditions. Mol Syst Biol 2:0031.	**Mammalian:** Glycolysis, Fatty acid oxidation, Glycogen oxidation, Phosphocreatine synthesis and breakdown, and the TCA cycle.	7	8	DOA/fmincon
Hjersted JL, Henson MA. 2006. Optimization of fed-batch *Saccharomyces cerevisiae* fermentation using dynamic flux balance models. Biotechnol Prog 22(5):1239–1248. Hjersted JL, Henson MA. 2009. Steady-state and dynamic flux balance analysis of ethanol production by *Saccharomyces cerevisiae*. IET Syst Biol 3(3):167–179.	*Saccharomyces cerevisiae*	98	82	DA/CONOPT
Pizarro F, Varela C, Martabit C, Bruno C, Pe´rez-Correa JR, Agosin E. 2007. Coupling kinetic expressions and metabolic networks for predicting wine fermentations. Biotechnol Bioeng 98(5):986–998.	*Saccharomyces cerevisiae*	38	39	SOA/—
Hjersted JL, Henson MA, Mahadevan R. 2007. Genome-scale analysis of *Saccharomyces cerevisiae* metabolism and ethanol production in fed-batch culture. Biotechnol Bioeng 97(5):1190–1204.	*Saccharomyces cerevisiae*	1,059	1,265	DA/MOSEK
Anesiadis N, Cluett WR, Mahadevan R. 2008. Dynamic metabolic engineering for increasing bioprocess productivity. Metab Eng 10(5):255–266.	*Escherichia coli*	625	931	SOA/CPLEX

Lee J, Gianchandani E, Eddy J, Papin J. 2008. Dynamic analysis ofintegrated signaling, metabolic, and regulatory networks. PLoS Comput Biol 4(5):e1000086.	*Saccharomyces cerevisiae*	—	13	SOA/—
Luo R, Wei H, Ye L, Wang K, Chen F, Luo L, Liu L, Li Y, Crabbe MJC, Jin L, Li Y, Zhong Y. 2009. Photosynthetic metabolism of C3 plants shows highly cooperative regulation under changing environments: A systems biological analysis. Proc Natl Acad Sci 106(3):847–852.	Chloroplasts of C3 plants	8	5	DOA/fmincon
Oddone GM, Mills DA, Block DE. 2009. A dynamic, genome-scale flux model of *Lactococcus lactis* to increase specific recombinant protein expression. Metab Eng 11(6):367–381.	*Lactococcus lactis*	422	621	SOA/Mathematica
Lequeux G, Beauprez J, Maertens J, Horen EV, Soetaert W, Vandamme E, Vanrolleghem PA. 2010. Dynamic metabolic flux analysis demonstrated on cultures where the limiting substrate is changed from carbon to nitrogen and vice versa. J Biomed Biotechnol 2010: http://www. hindawi.com/journals/jbb/2010/621645/cta/	*Escherichia coli*	24	34	polynomial fitting
Salimi F, Zhuang K, Mahadevan R. 2010. Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. Biotechnol J 5(7):726–738.	*Clostridium acetobutylicum*	679	712	DA/—
	Clostridium cellulolyticum	603	621	
Zhuang K, Izallalen M, Mouser P, Richter H, Risso C, Mahadevan R, Lovley D. 2011. Genome-scale dynamic modeling of the competition between rhodoferax and geobacter in anoxic subsurface environments. ISME J 5:305–316.	*Geobacter sulfurreducens*	541	522	DA/LINDO
	Rhodoferax ferrireducens	790	762	
Meadows AL, Karnik R, Lam H, Forestell S, Snedecor B. 2010. Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. Metab Eng 12(2):150–160.	*Escherichia coli*	30	123	ODE15S/linprog
Vargas F, Pizarro F, Pe´rez-Correa J, Agosin E. 2011. Expanding a dynamic flux balance model of yeast fermentation to genome-scale. BMC Syst Biol 5(1):75.	*Saccharomyces cerevisiae*	590	1,181	SOA/LINDO
Nolan RP, Lee K. 2011. Dynamic model of CHO cell metabolism. Metab Eng 13(1):108–124.	Chinese hamster ovary (CHO) cells	150	136	SOA/—

Reference	Organism			
Hanly TJ, Henson MA. 2010. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. Biotechnol Bioeng 108(2):376–385.	*Escherichia coli* and *Saccharomyces cerevisiae*	625	931	DA/MOSEK
Hanly TJ, Urello M, Henson MA. 2011. Dynamic flux balance modeling of *S. cerevisiae* and *E. coli* co-cultures for efficient consumption of glucose/xylose mixtures. Appl Microbiol Biotechnol 93:2529–2541.	*Escherichia coli* and *Saccharomyces cerevisiae*	1,059	1,265	DA/MOSEK
Knoop H, Gründel M, Zilliges Y, Lehmann R, Hoffmann S, Lockau W, Steuer R. 2013. Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of *Synechocystis sp.* PCC 6803. PLoS Comput Biol 9(6):e1003081	*Synechocystis sp. PCC6803*	601	759	Temporal coordination

APPENDIX B: SUPPLEMENTARY MATERIAL FOR CHAPTER 2
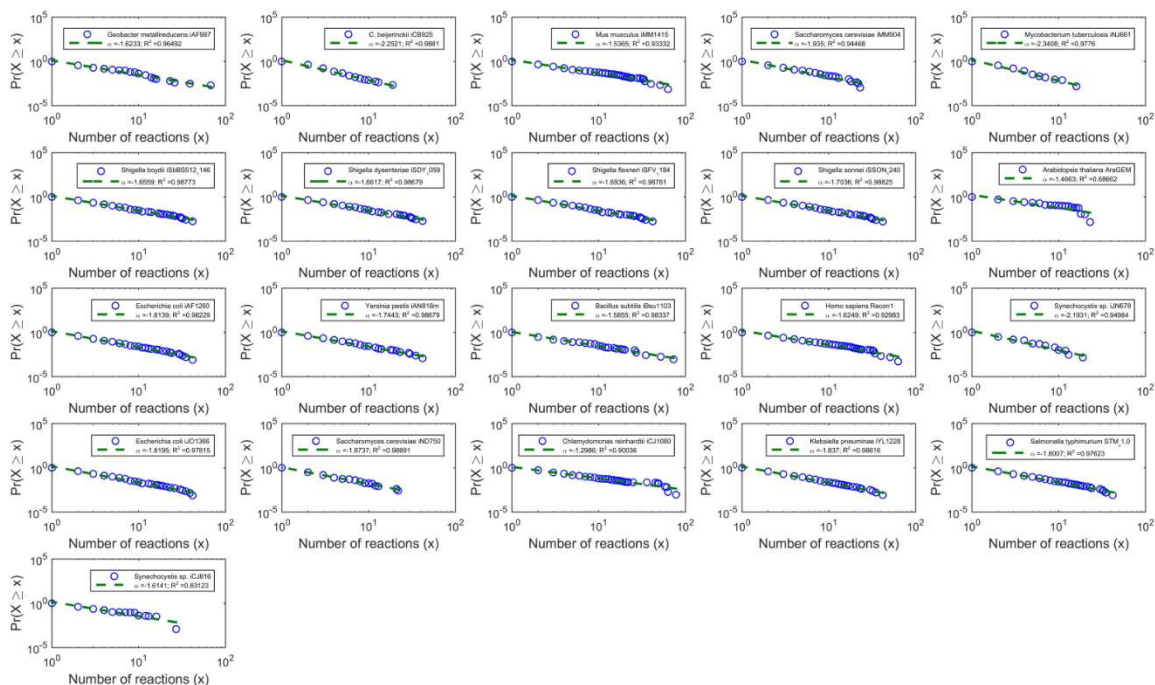
FIGURE B1: POWER-LAW FITS OF THE CUMULATIVE ENZYME-REACTION DISTRIBUTION OF ALL THE SPECIES AND ANOTHER E. COLI MODEL BY FITTING A LINEAR FUNCTION TO LOG-LOG DATA, USING MATLAB.

Blue circles indicate the model data and dark green dashed line represents the respective MATLAB fits. The legend reports the species, the exponent of the fit (alpha) and the R^2 of the fit. (Left to Right, then, Top to Bottom) *E. coli* (iAF1260), *Synechocystis sp.* PCC6803, *C. reinhardtii*, *S. cerevisiae*, *E. coli* (iJO1366), *K. pneumoniae*, *S. typhimurium*, *H. sapiens*, *A. thaliana*, *Y. pestis*, and *B. subtilis*. Links to the SBML files of each model, and references to the associated publications, are provided
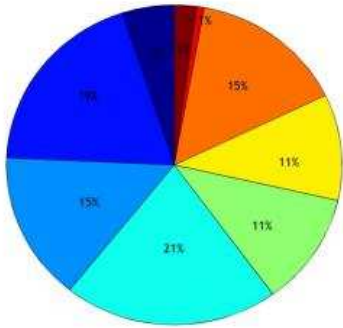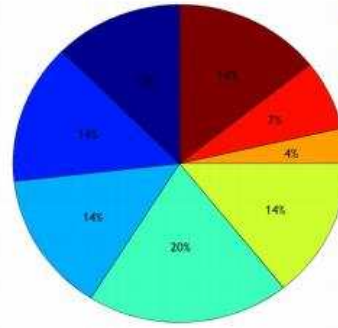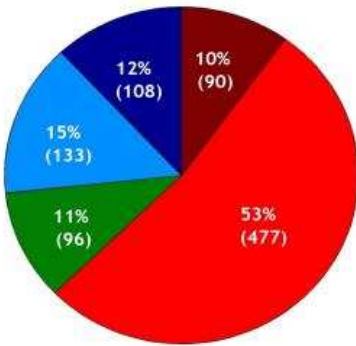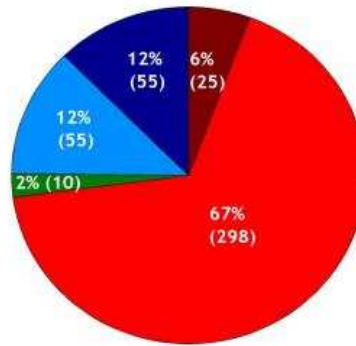
FIGURE B2: DISTRIBUTION AND CLASSIFICATION OF ENZYMES.
(A) The percentage distribution of all protein complexes among different subsystems in *E. coli* and *Synechocystis*, (B) Global classification of generalist and specialist enzymes in *E. coli* and *Synechocystis*.
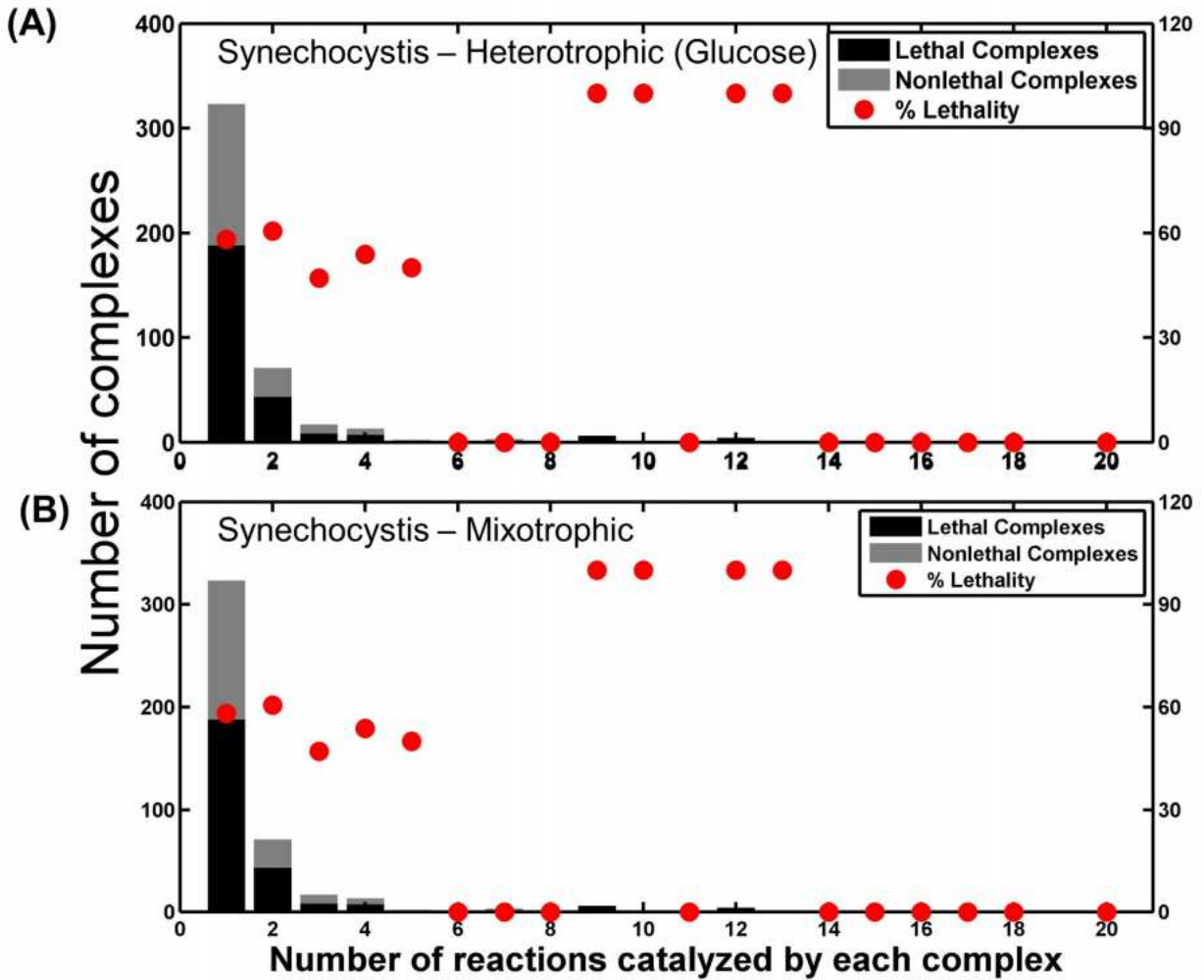
FIGURE B3: THE DISTRIBUTION OF LETHAL (ESSENTIAL) COMPLEXES AMONG COMPLEXES THAT CONTROL DIFFERENT NUMBERS OF REACTIONS IN SYNECHOCYSTIS UNDER DIFFERENT METABOLIC CONDITIONS.

(A) Heterotrophic growth and **(B)** Mixotrophic growth. As for autotrophic growth shown in the main text, the percentage of lethality (shown as grey dots) increases on average with the increase in the number of reactions constrained. This suggests that enzymes with a high degree of multifunctionality, ke, tend to be essential in Synechocystis even under non-autotrophic conditions.
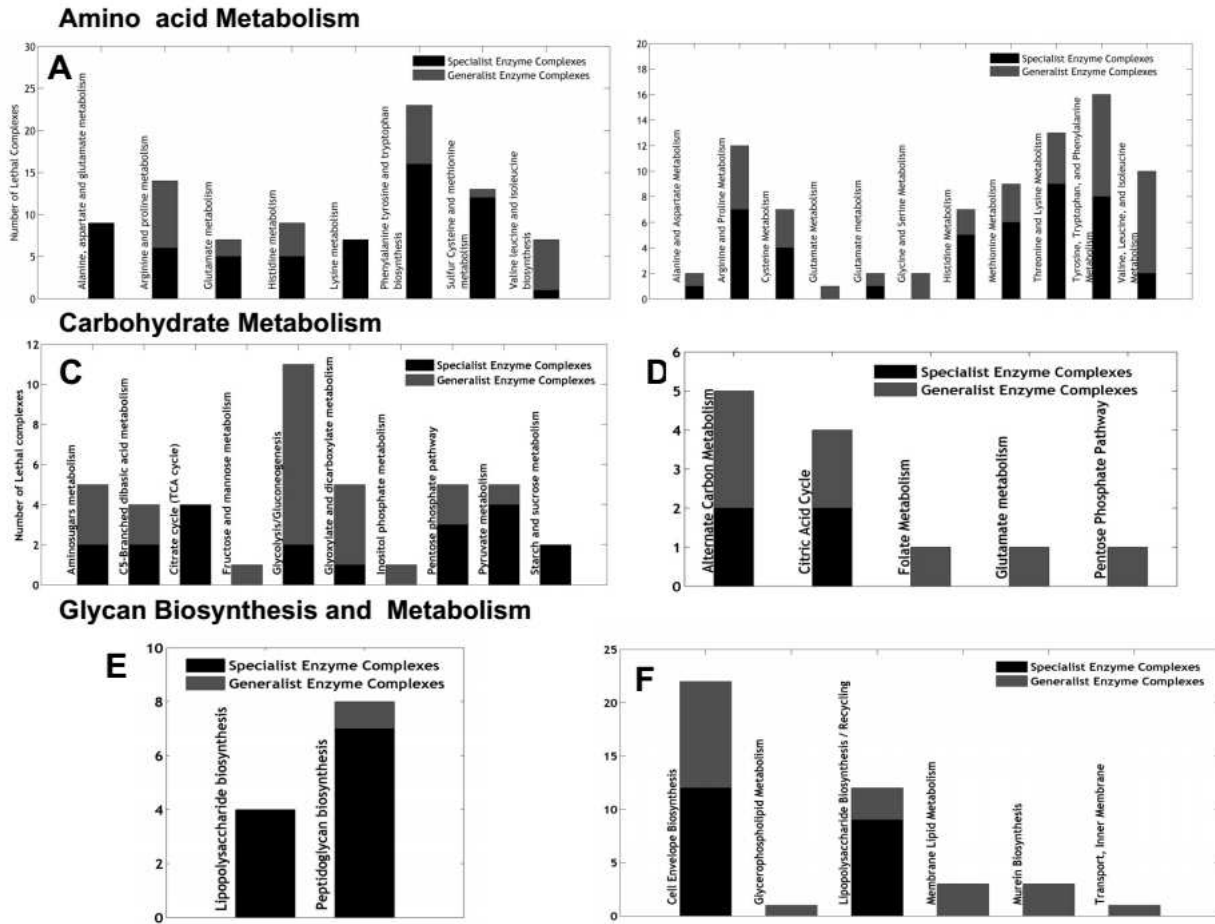
FIGURE B4: THE DISTRIBUTION OF LETHAL (ESSENTIAL) COMPLEXES IN AMINO ACID METABOLISM (A & B), CARBOHYDRATE METABOLISM (C & D), AND GLYCAN BIOSYNTHESIS (E & F) IN *SYNECHOCYSTIS* (A, C, & E) AND *E. COLI* (B, D, & F).

TABLE B1: INFORMATION ABOUT VARIOUS MODELS USED IN THIS STUDY.
The links to SBML, ZIP, or XLS file are provided and can be accessed by clicking in the model version column. The links to original article is provided and can be accessed by clicking in the reference column. The references are listed in Supplementary References at the end of this document.

Organism	Strain	Version	Genes	Metabolites	Reactions	Reference
BACTERIA						
Bacillus subtilis	168	iBsu1103	1103	1138	1437	Henry et al., 2009
Escherichia coli	K-12 MG1655	iJO1366	1366	1136	2251	Orth et al., 2011
Klebsiella pneumoniae	MGH 78578	iYL1228	1228	1658	1970	Liao et al., 2011
Salmonella typhimurium	LT2	STM_v1.0	1270	1119	2201	Thiele et al., 2011
Yersinia pestis	CO92	iPC815	815	963	1678	Charusanti et al., 2011
Escherichia coli	K-12 MG1655	iAF1260	1260	1039	2077	Feist et al.,2007
Synechocystis sp.	PCC6803	iJN678	678	795	863	Nogales et al., 2011
EUKARYOTES						
Saccharomyces cerevisiae	Sc288	iND750	750	646	1149	Duarte et al., 2004
Chlamydomonas reinhardtii		iRC1080	1080	1068	2190	Chang et al., 2011
Arabidopsis thaliana		AraGEM	1419	1748	1567	de Oliveira Dal'Molin et al., 2009
Homo sapiens		Recon 1	1,496	2,766	3,311	Duarte et al., 2007

TABLE B2: DESCRIPTION OF ALL THE COARSE-GRAINED SUBSYSTEMS USED IN THIS STUDY.
The subsystems in bold are the ones used in the study while the list that follows each bold entry are the KEGG subsystems that the coarse-grained definition includes.

Synechocystis sp. PCC6803	*E. coli*
Amino acid metabolism	**Amino acid metabolism**
Alanine, aspartate and glutamate metabolism	Alanine and Aspartate Metabolism
Arginine and proline metabolism	Arginine and Proline Metabolism
Glutamate metabolism	Cysteine Metabolism
Histidine metabolism	Glutamate Metabolism
Lysine metabolism	Glutamate metabolism
Phenylalanine tyrosine and tryptophan biosynthesis	Glycine and Serine Metabolism
Sulfur Cysteine and methionine metabolism	Histidine Metabolism
Valine leucine and isoleucine biosynthesis	Methionine Metabolism
Biomass	Threonine and Lysine Metabolism
Biomass	Tyrosine, Tryptophan, and Phenylalanine Metabolism
Carbohydrate metabolism	Valine, Leucine, and Isoleucine Metabolism
Aminosugars metabolism	**Carbohydrate metabolism**
C5-Branched dibasic acid metabolism	Alternate Carbon Metabolism
Citrate cycle (TCA cycle)	Anaplerotic Reactions
Fructose and mannose metabolism	Citric Acid Cycle
Glycolysis/Gluconeogenesis	Glycolysis/Gluconeogenesis
Glyoxylate and dicarboxylate metabolism	Glyoxylate Metabolism
Inositol phosphate metabolism	Methylglyoxal Metabolism
Nucleotide sugars metabolism	Pentose Phosphate Pathway
PHB byosynthesis	Pyruvate Metabolism
Pentose phosphate pathway	**Energy metabolism**
Pyruvate metabolism	Nitrogen Metabolism
Starch and sucrose metabolism	Oxidative Phosphorylation
Cyanophycin metabolism	**Glycan biosynthesis and metabolism**
Cyanophycin metabolism	Cell Envelope Biosynthesis
Energy metabolism	Lipopolysaccharide Biosynthesis / Recycling
Carbon fixation	Murein Biosynthesis
Hydrogen production	Murein Recycling
Nitrogen metabolism	**Lipid metabolism**
Oxidative phosphorylation	Glycerophospholipid Metabolism
Photosynthesis	Membrane Lipid Metabolism
Exchange reactions	**Metabolism of cofactors and vitamins**
Exchange reactions	Cofactor and Prosthetic Group Biosynthesis
Glycan biosynthesis and metabolism	Folate Metabolism
Lipopolysaccharide biosynthesis	**Nucleotide metabolism**
Peptidoglycan biosynthesis	Nucleotide Salvage Pathway

Peptidoglycan biosynthesis	Purine and Pyrimidine Biosynthesis
Lipid metabolism	**Unassigned**
Fatty acid biosynthesis	Unassigned
Galactolipids metabolism	**Translation**
Glycerolipid metabolism	Translation
Steroid biosynthesis	**Transport**
Sterol biosynthesis	Inorganic Ion Transport and Metabolism
Sulfolipid Biosynthesis	Transport, Inner Membrane
Metabolism of cofactors and vitamins	Transport, Outer Membrane
Biotin metabolism	Transport, Outer Membrane Porin
Folate biosynthesis	
Nicotinate and nicotinamide metabolism	
Pantothenate and CoA biosynthesis	
Porphyrin and chlorophyll metabolism	
Riboflavin metabolism	
Thiamine metabolism	
Ubiquinone and other pterpenoids biosynthesis	
Vitamin B6 metabolism	
Metabolism of other amino acids	
Glutathione metabolism	
Urea cycle and metabolism of amino groups	
Metabolism of terpenoids and polyketides	
Carotenoid Biosynthesis	
Terpenoid backbone biosynthesis	
Modeling	
Modeling	
Nucleotide metabolism	
Purine metabolism	
Pyrimidine metabolism	
Others	
Others	
Transport	
Transport	

TABLE B3: ANALYSIS OF ALL THE 39 SINGLE GENE DELETIONS THAT WAS LETHAL IN SYNECHOCYSTIS BUT NON-LETHAL IN *E. COLI*. The left column is the gene which is deleted, while the growth ratio is the ratio of the growth rate in E. coli after deletion to the growth rate of the wild type. The various reasons for escape are in color coded rows: Environment specific flux distribution (yellow), Distributed control (green), Existence of multiple/alternative pathways (red), Not required for growth (blue), and those that did not map back (white).

UNIQUE GENES	GROWTH RATIO	REASON FOR ESCAPE	JUSTIFICATION
b0118	1	DOES NOT MAP BACK!	In our analysis - it is required that gene be able to map back to the synechocystis gene it was derived from.
b0171	0.999	DOES NOT MAP BACK!	In our analysis - it is required that gene be able to map back to the synechocystis gene it was derived from.
b0469	1	Distributed Control	*E. coli* has a small flux through ADPT, when the b0469 deletion is made, the flux can be redirected through other pathway because it is associated to a different gene and is still active in the organism. However, when sll1430 is deleted, the possibility of rewiring no longer exists because both the reactions are deleted.
b0474	1	Distributed Control	It is also isozymic for another deletion b2518, which is involved in other reactions (NDPK1-8; nucleoside:diphosphate kinases), also provides and alternate route.
b0507	1	Existence of multiple pathways	There exists an alternate route to consumption of glyoxylate, such as Glycolate dehydrogenase (NAD or NADP dependent) and malate synthase.
b0728	0.993	Not required for growth	Succinyl CoA is not required for growth rate equation
b0910	0.998	Distributed Control	It is also isozymic for another deletion b0171, which is involved in other reactions (UMP kinase), also provides and alternate route.

b1064	1	DOES NOT MAP BACK!	In our analysis - it is required that gene be able to map back to the synechocystis gene it was derived from.
b1109	1	Distributed Control	This gene codes for reactions which are involved in reduction of menaquinol, quinone, and demethylmenaquinone; in 3 separate reactions (using NADH). This can be performed by other gene b3028, which catalyzes similar reactions (using NADPH).
b1207	1	Environment specific flux distribution	directionality, due to environment and growth conditions – Pentose phosphate pathway in Synechocystis
b1238	1	Distributed Control	A reversible reaction catalyzed by deoxyuridine phosphorylase makes this mutation a viable one.
b1270	1	DOES NOT MAP BACK!	In our analysis - it is required that gene be able to map back to the synechocystis gene it was derived from.
b1278	1	Not required for growth	Phosphatidylglycerols are not required for growth rate equation
b1300	1	Distributed Control	Aldehyde dehydrogenase (acetaldehyde, NAD utilizing) (ALDD2x) is responsible for producing acetate. This can be accomplished many other reactions which are not governed by this gene.
b1444	1	Distributed Control	4 amino butanoate can be made by other reactions which are not governed by this gene
b1748	1	DOES NOT MAP BACK!	In our analysis - it is required that gene be able to map back to the synechocystis gene it was derived from.
b1912	1	Not required for growth	Phosphatidylglycerophosphates are not required for growth rate equation

b2234	1	Existence of multiple pathways	deoxynucleotide phophates are produced using thioredoxin. These can also be made by a deoxynucleotide phosphate kinases
b2260	1	Existence of multiple pathways	A precursor step in making 2-demethylmenaquinol 8, it can be made/regenerated by other reactions such as fumarate reductase (demethylmenaquinone 8 dependent) and NADH dehydrogenase (demethylmenaquinone 8 dependent & 0 protons).
b2262	1	Existence of multiple pathways	A precursor step in making 2-demethylmenaquinol 8, it can be made/regenerated by other reactions such as fumarate reductase (demethylmenaquinone 8 dependent) and NADH dehydrogenase (demethylmenaquinone 8 dependent & 0 protons)
b2264	1	Existence of multiple pathways	A precursor step in making 2-demethylmenaquinol 8, it can be made/regenerated by other reactions such as fumarate reductase (demethylmenaquinone 8 dependent) and NADH dehydrogenase (demethylmenaquinone 8 dependent & 0 protons)
b2265	1	Existence of multiple pathways	An irreversible form of this enzyme exists, Isochorismate synthase (irreversible).
b2276	0.816	Distributed Control	This gene codes for reactions which are involved in reduction of menaquinol, quinone, and demethylmenaquinone; in 3 separate reactions (using NADH). This can be performed by other molecules which carry electrons
b2436	0.999	Existence of multiple pathways	Coproporphyrinogen III can be made by other reaction catalyzed by COproporphyrinogen III oxidase (oxygen independent).
b2501	0.988	Existence of multiple pathways	This is responsible for ATP to ADP conversions and production inorganic phosphates. This can be done in many different parts of the network like inorganic diphosphatase/triphosphatase.
b2508	0.999	Existence of multiple pathways	Xanthosine 5-phosphate can be made by many other reactions like xanthosine phosphotransferase, nucelotide pyrophosphate, etc.

b2551	0.986	Environment specific flux distribution	directionality, due to environment and growth conditions – photorespiration in Synechocystis.
b2661	1	Not required for growth	Succinic-semialdehyde can is not required for growth rate equation
b2688	1	Distributed Control	A precursor step in making glutathione. Glutathione can be made/regenerated by other reactions catalyzed by other enzymes such as glutathione oxidoreductases, and glutathionylsperimidine amidase.
b2744	1	Existence of multiple pathways	This happens to be a multifunctional enzyme in nucleotide metabolism. Some of these reactions are isozymic and do not completely delete the functionality. Presence of other genes cause the flux distribution through other reactions.
b2799	1	Existence of multiple pathways	L-lactaldehyde can be made by fuculose 1-phosphate, or rhamnulose 1-phosphate.
b2926	0.788	Environment specific flux distribution	directionality, due to environment and growth conditions – requirement of photosynthesis
b2947	1	Existence of multiple pathways	A precursor step in making glutathione. Glutathione can be made/regenerated by other reactions catalyzed by glutathione oxidoreductases, and glutathionylsperimidine amidase.
b3430	1	Not required for growth	ADP-glucose is not required for growth rate equation

b3731	0.322	Distributed Control	ATP synthase making ATP in cytosol utilizing protons from periplasmic membrane. This conversion is very essential and not all the required ATP can be generated through other cellular reactions. Therefore, the growth rate reduces drastically.
b3919	0.963	Environment specific flux distribution	directionality, due to environment and growth conditions – GLUCONEOGENESIS. The requirement that glucose 6-phosphate be produced from fructose 6-phosphate.
b3956	0.997	Existence of multiple pathways	Alternative routes like malate dehydrogenase, and aspartate transaminase facilitate flux redistribution and production of oxaloacetate.
b4025	0.994	Environment specific flux distribution	directionality, due to environment and growth conditions – GLUCONEOGENESIS. The requirement that glucose 6-phosphate be produced from fructose 6-phosphate.
b4383	1	Environment specific flux distributions	directionality, due to environment and growth conditions – Pentose phosphate pathway

TABLE B4: SUMMARY OF THE DATA PRESENTED IN TABLE B3.

SUMMARY	
Environment specific flux distribution	6
Distributed Control	11
Existence of multiple/alternative pathways	12
Not required for growth	5
DOES NOT MAP BACK!	5
TOTAL	39

TABLE B5: SUBSYSTEMS IN *E. COLI* TO WHICH THE REACTIONS, INVOLVED IN TABLE B3, BELONG.

SUBSYSTEMS	# OF REACTIONS
Alternate Carbon Metabolism	5
Nucleotide Salvage Pathway	19
Glyoxylate Metabolism	1
Citric Acid Cycle	1
Inorganic Ion Transport and Metabolism	1
Oxidative Phosphorylation	9
Histidine Metabolism	1
Cofactor and Prosthetic Group Metabolism	10
Glycerophospholipid Metabolism	21
Arginine and Proline Metabolism	3
Purine and Pyrimidine Biosynthesis	1
Glycine and Serine Metabolism	2
Glycolysis/Gluconeogenesis	4
Anaplerotic Reactions	1
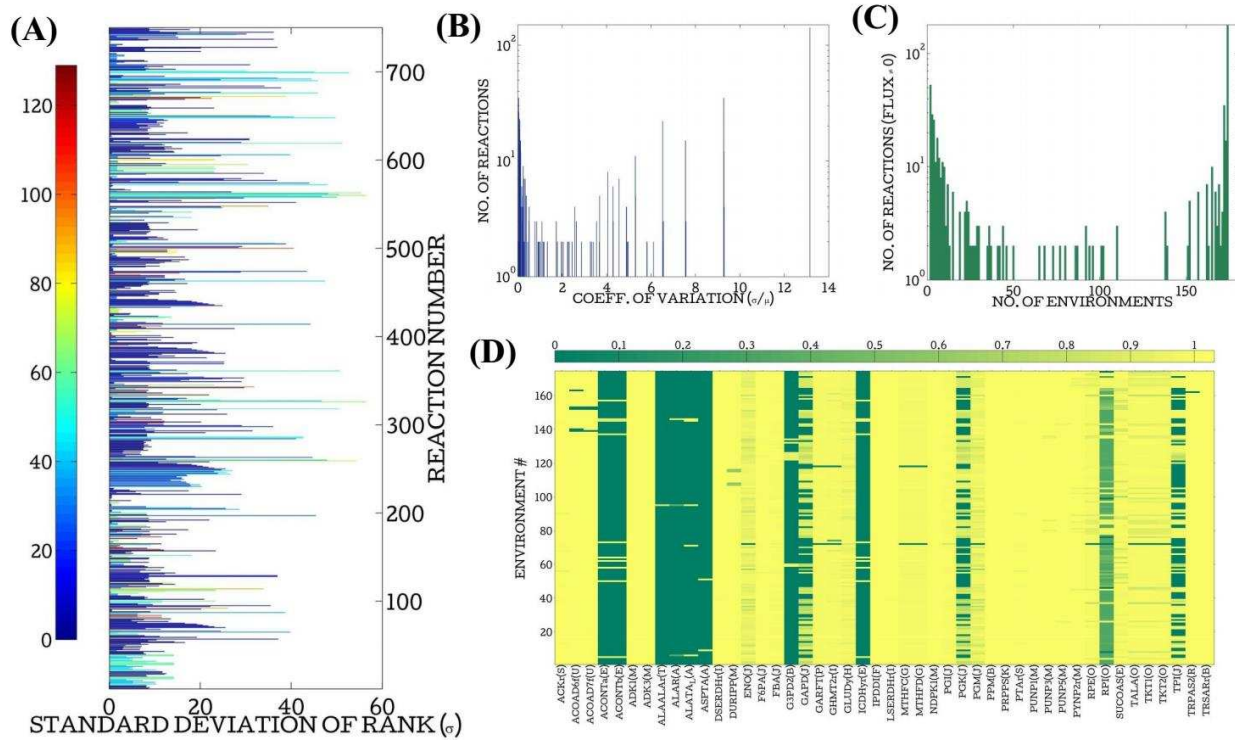
FIGURE C1: FLUXES FOR REACTIONS IN 174 DIFFERENT ENVIRONMENTAL CONDITIONS.
(A) Standard deviation of the rank calculated across 174 environmental conditions; color map represents the average of the rank across all 174 environmental conditions. **(B)** Distribution of coefficient of variation amongst reactions included in (A). **(C)** Distribution of all non-zero fluxes distributed across all 174 environmental conditions. **(D)** Fitness (growth rate ratio compared with Wild Type in a particular environmental condition) of single reaction deletions identified in Figure 1C; the letter in the parentheses represents the subsystem mentioned in Table 1.
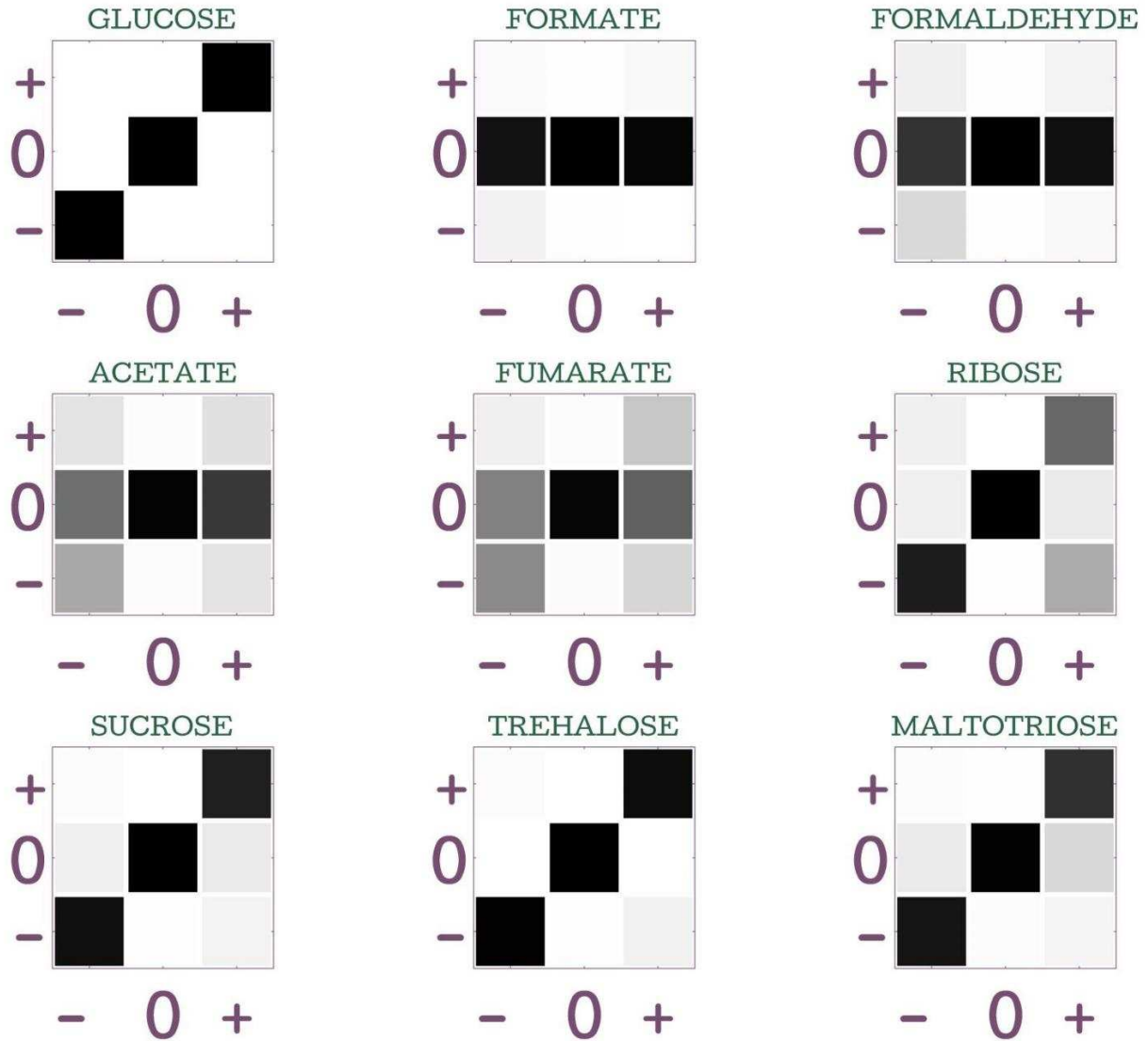
FIGURE C2: COMPARISON OF EPISTASIS IN 9 DIFFERENT GROWTH CONDITIONS, WITH RESPECT TO GLUCOSE.

The symbols -, 0 and + refer to negative, none and positive epistasis respectively. The x-axis is always glucose, while the y-axis is the other carbon source as stated. The number of interactions that fall in each category, i.e. matches or mismatches as shown are indicated by the shade of the corresponding box. The darker the box, more are the interactions in that category.

FIGURE C3: EPISTASIS INTERACTIONS CORRESPONDING TO INCREASING INCIDENT LIGHT.
Epistasis interactions calculated as incident light increases; red corresponds to negative interactions and green corresponds to positive interactions. The radius of the circle corresponds to number of interactions; the outer circle corresponds to total number of interactions (positive and negative).The letters on X-axis and Y-axis correspond to subsystems according to the legend shown in the figure.

TABLE C1: A LIST OF ALL THE 93 GENES INVOLVED IN MAKING PAIRS FOR COMPARISON ACROSS 9 DIFFERENT GROWTH CONDITIONS.

Column A represents the gene name and column B represents the reaction identifiers of the reactions constrained by the corresponding gene. The reaction identifiers are used from the model and are same as those used in the original published model (Feist et al. 2007).

GENES	REACTIONS
b0040	CRNt7pp, CRNt8pp
b0114	PDH
b0116	AKGDH, GLYCL, PDH
b0123	CU1Opp, FEROpp
b0171	URIDK2r
b0242	GLU5K
b0243	G5SD
b0335	ACCOAL
b0394	HEX7
b0429	CYTBO3_4pp
b0469	ADPT
b0474	ADK1, ADK3, ADK4, ADNK1, DADK
b0507	GLXCL
b0508	HPYRI
b0514	GLYCK
b0529	MTHFC, MTHFD
b0593	ICHORSi
b0721	SUCDi
b0726	AKGDH
b0728	SUCOAS
b0767	PGL
b0888	TRDR
b0910	CYTK1, CYTK2
b1006	URAt2rpp
b1015	PPAt4pp, PROt4pp
b1091	ACOATA, KAS15
b1198	DHAPT
b1207	PRPPS
b1216	CA2t3pp
b1602	THD2pp
b1761	GLUDy
b1779	GAPD
b1849	GART
b1852	G6PDH2r
b2029	GND
b2234	RNDR1, RNDR2, RNDR3, RNDR4

b2265	ICHORS
b2276	NADH16pp, NADH17pp, NADH18pp
b2297	PTA2
b2406	ADNt2rpp, CYTDt2rpp, INSt2rpp, THMDt2rpp, URIt2rpp, XTSNt2rpp
b2415	ACGAptspp, ACMANAptspp, ACMUMptspp, ASCBptspp, DHAPT, FRUpts2pp, FRUptspp, MANptspp, GALTptspp, GAMptspp, GLCptspp, MALTptspp, MANGLYCptspp, MNLptspp, SBTptspp, SUCptspp, TREptspp
b2417	ACMUMptspp, MALTptspp, SUCptspp, TREptspp
b2436	CPPPGO
b2463	ME2
b2497	URAt2pp
b2500	GARFT
b2501	PPK2r, PPKr
b2508	IMPD
b2551	GHMT2r
b2675	RNDR1b, RNDR2b, RNDR3b, RNDR4b
b2779	ENO
b2889	IPDDI
b2903	GLYCL
b2913	PGCD
b2920	PPCSCT
b2926	PGK
b2964	DGSNt2pp, DINSt2pp, GSNt2pp, INSt2pp
b2979	GLYCTO2, GLYCTO3, GLYCTO4
b3089	SERt4pp, THRt4pp
b3115	PPAKr
b3236	MDH
b3403	PPCK
b3417	MLTP1, MLTP2, MLTP3
b3431	GLDBRAN2
b3432	GLBRAN2
b3500	GTHOr
b3528	ASPt2_2pp, FUMt2_2pp, MALDt2_2pp, MALTt2_2pp, OROTt2_2pp, SUCCt2_2pp
b3565	XLYI1, XLYI2
b3572	VPAMT
b3616	THRD
b3617	GLYAT
b3653	GLUt4pp
b3708	TRPAS2
b3731	ATPS4rpp
b3744	ASNS2
b3835	OPHHX

b3919	TPI
b3924	FLDR
b3956	PPC
b4025	PGI
b4036	14GLUCANtexi, GLCtexi, MALTHXtexi, MALTPTtexi, MALTTRtexi, MALTTTRtexi, MALTtexi
b4067	ACt4pp, GLYCLTt4pp
b4077	ASPt2pp, GLUt2rpp
b4094	R15BPK
b4111	CRNDt2rpp, CRNt2rpp, CTBTt2rpp, PROt2rpp
b4139	ASPT
b4151	FRD2, FRD3
b4208	ALAt2pp, BALAt2pp, DALAt2pp, DSERt2pp
b4238	RNTR1c, RNTR2c, RNTR3c, RNTR4c
b4239	TRE6PH
b4240	TREptspp
b4384	PUNP1, PUNP2
b4388	PSP_L

TABLE C2: THE E. COLI SUBSYSTEMS CORRESPONDING TO LETTERS IN FIGURE 3.1, 3.3, AND 3.4.

Letter	Subsystem
A	Alanine and Aspartate Metabolism
B	Alternate Carbon Metabolism
C	Anaplerotic Reactions
D	Arginine and Proline Metabolism
E	Citric Acid Cycle
F	Cofactor and Prosthetic Group Biosynthesis
G	Folate Metabolism
H	Glutamate Metabolism
I	Glycine and Serine Metabolism
J	Glycolysis/Gluconeogenesis
K	Histidine Metabolism
L	Inorganic Ion Transport
M	Nucleotide Salvage Pathway
N	Oxidative Phosphorylation
O	Pentose Phosphate Pathway
P	Purine and Pyrimidine Biosynthesis
Q	Inner Membrane Transport
R	Tyrosine, Tryptophan, and Phenylalanine Metabolism
S	Pyruvate Metabolism
T	Cell Envelope Biosynthesis
U	Membrane Lipid Metabolism

TABLE C3: THE SYNECHOCYSTIS (HETEROTROPHIC GROWTH WITH GLUCOSE) SUBSYSTEMS CORRESPONDING TO LETTERS IN FIGURE 3.3

Letter	Subsystem
A	Carbon Fixation
B	TCA cycle
C	Glycerolipid Metabolism
D	Glycolysis
E	Glyoxylate and Dicarboxylate Metabolism
F	Nitrogen Metabolism
G	Others
H	Oxidative Phosphorylation
I	Pentose Phosphate Pathway
J	Photosynthesis
K	Pyrimidine Metabolism
L	Pyruvate Metabolism
M	Transport
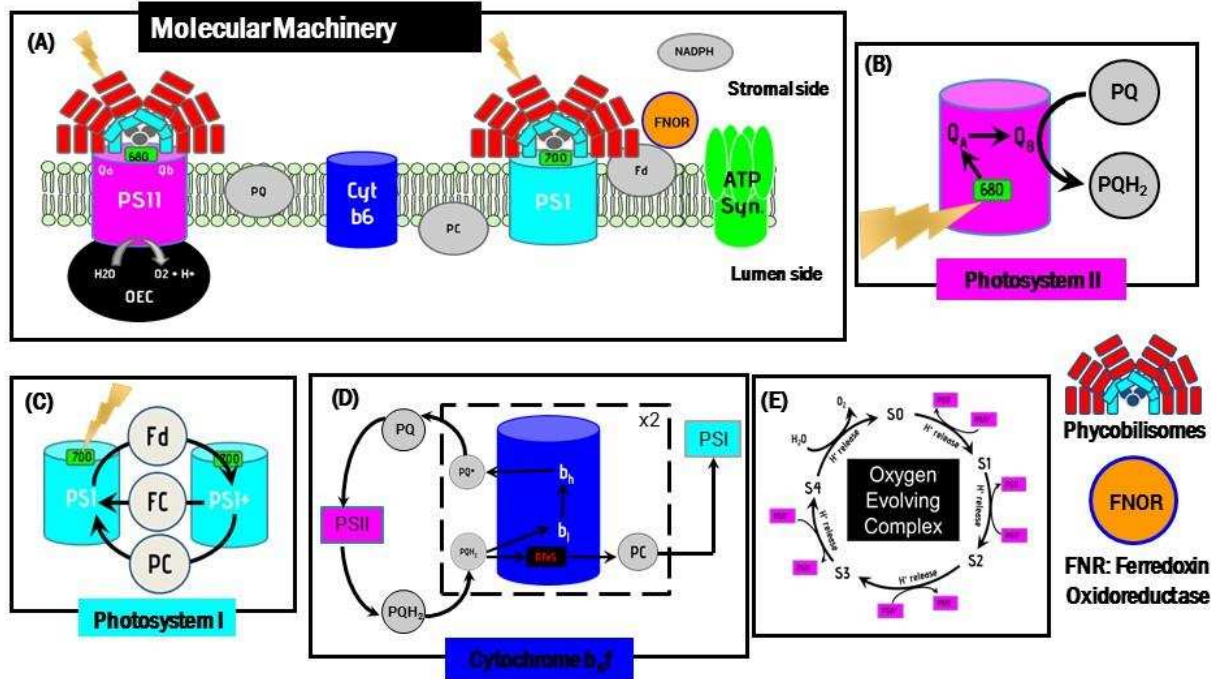
FIGURE D1: ORGANIZATION AND IMPLEMENTATION OF PHOTOSYNTHETIC MACHINERY WITHIN ISYNCJ816.
(A) Organization of photosynthetic machinery, (B) Organization of PSII, (C) Organization of PSI, (D) Organization of Cytochrome b6f, and (E) Organization of oxygen evolving complex. Fd, ferredoxin; FC, ferricytochrome c6; NADPH, Nicotinamide adenosine diphosphate; OEC, oxygen evolving complex; PC, plastocyanin; PSII; photosystem II; PSI, photosystem I; PQ, plastoquinone (oxidized); PQH2, plastoquinone (reduced); QA, quinone at site A; and QB, quione at site B.
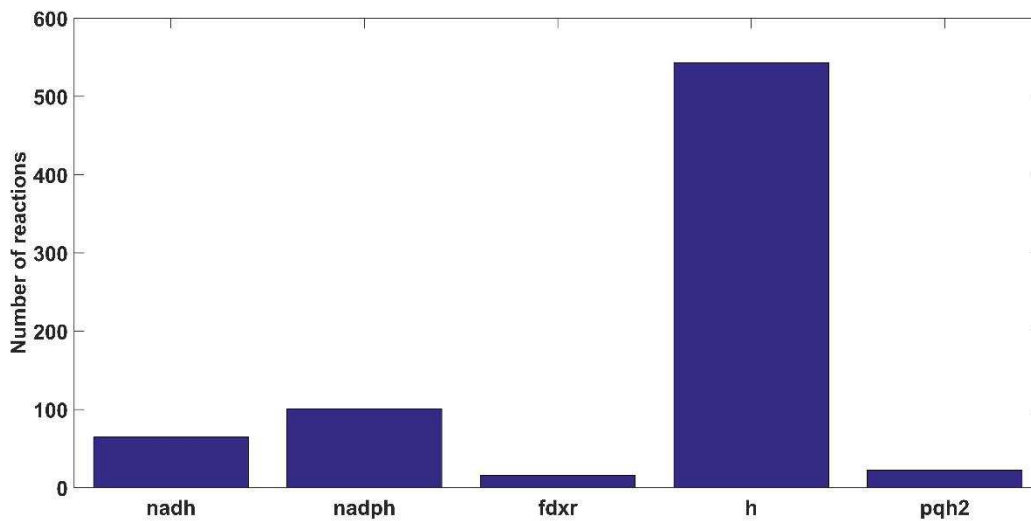
FIGURE D2: DISTRIBUTION OF MAIN ELECTRON CARRIERS INVOLVED WITHIN THE METABOLIC NETWORK. Abbreviation: nadh, nicotinamide adenine dinucleotide; nadph, nicotinamide adenine dinucleotide phosphate; fdxr, reduced form of ferredoxin; h, protons; pqh2, plastoquinol.
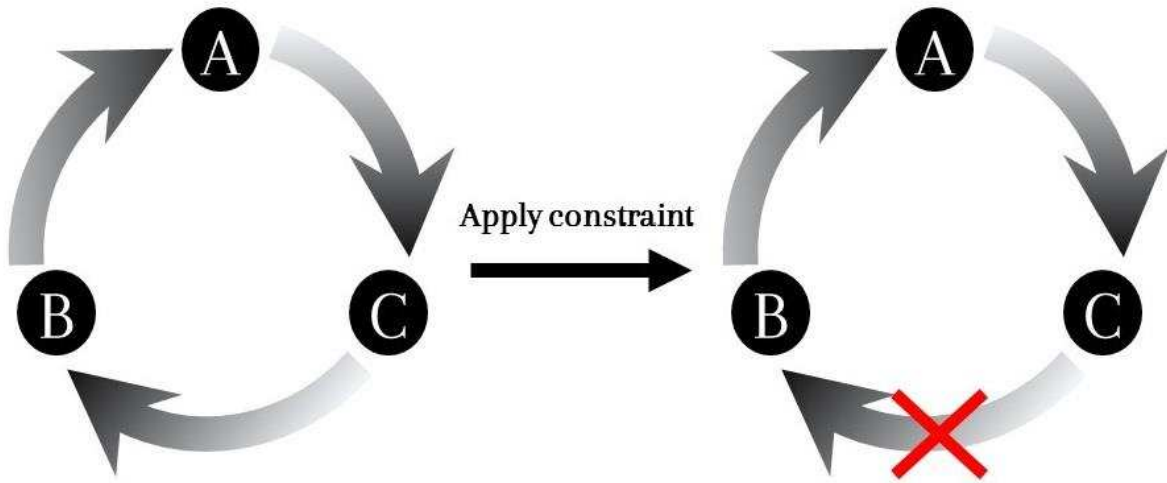
FIGURE D3: IDENTIFYING THERMODYNAMICALLY INFEASIBLE CYCLES OR FUTILE CYCLES.
The figure on the left describes a thermodynamically infeasible cycle, while the figure on the right describes resolving a thermodynamically infeasible cycle. Black circles represent metabolites and arrows indicate reactions.
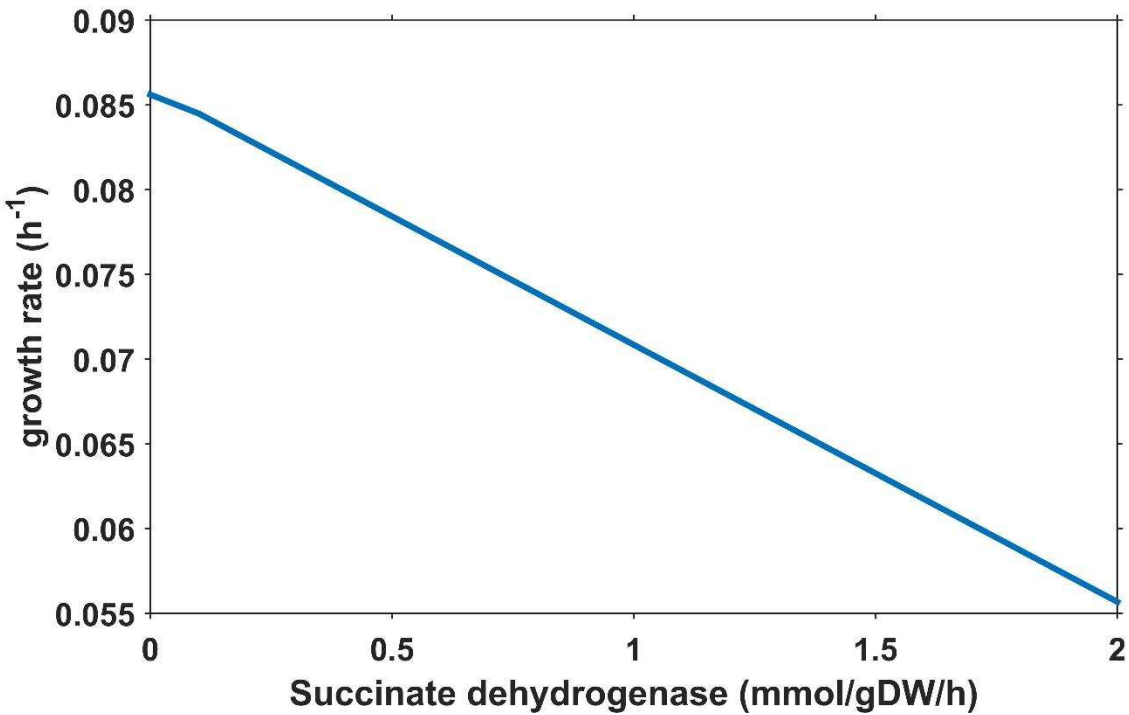
FIGURE D4: TRADE-OFF BETWEEN GROWTH RATE AND FLUX THROUGH SUCCINATE DEHYDROGENASE. Growth rate reduces as the flux through succinate dehydrogenase increases.
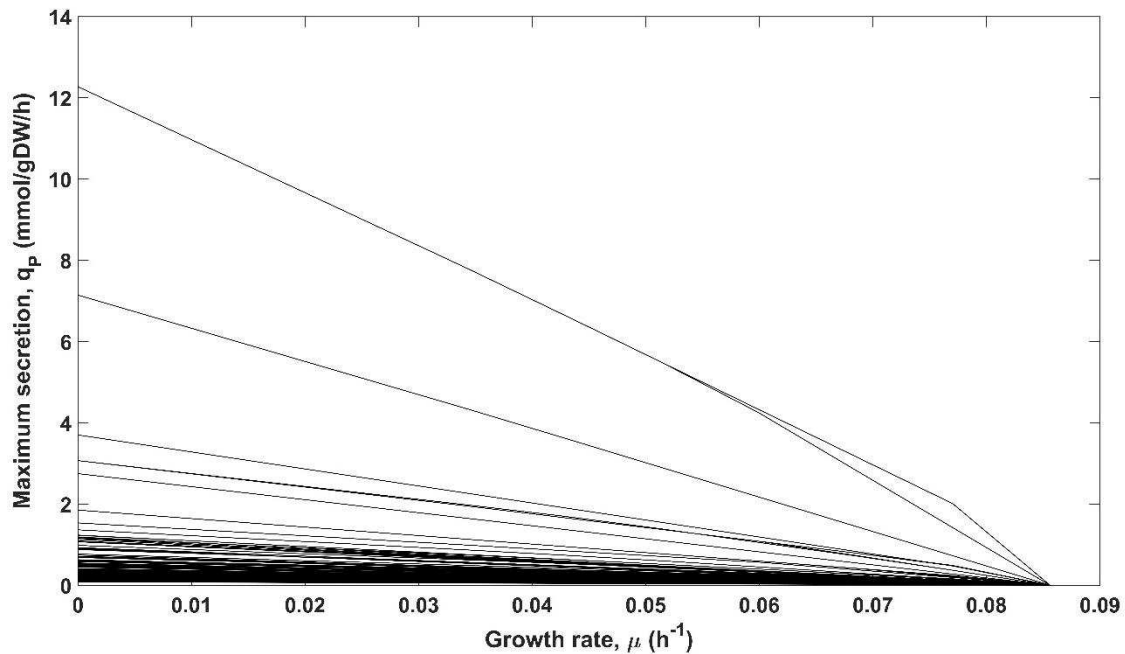
FIGURE D5: GROWTH TRADE-OFF AND SECRETION FLUX FOR ALL METABOLITES.
Growth rate decreases as the flux through secretion reaction increases. Each line represents growth trade-off for one metabolite.
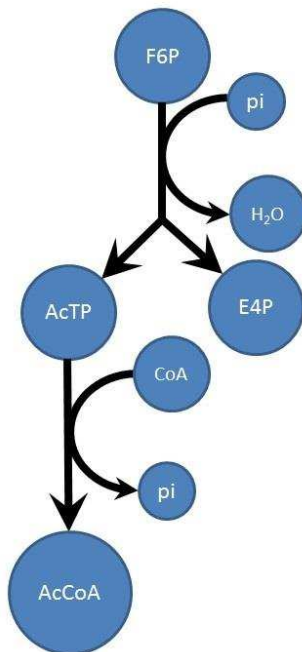
FIGURE D7: ADDITIONAL PATHWAY FOR ACETYL-COA PRODUCTION.
Production of Ac-CoA via pyruvate dehydrogenase or pyruvate oxidoreductase with Acetyl-CoA production from fructose 6-phosphate (F6P) via F6P phosphokinase and phosphotransacetylase. AcTP, acetyl triphosphate; AcCoA, acetyl-CoA; CoA, coenzyme A; E4P, erythrose 4-phosphate; F6P, fructose 6-phosphate; H2O, water; and pi, phosphate.
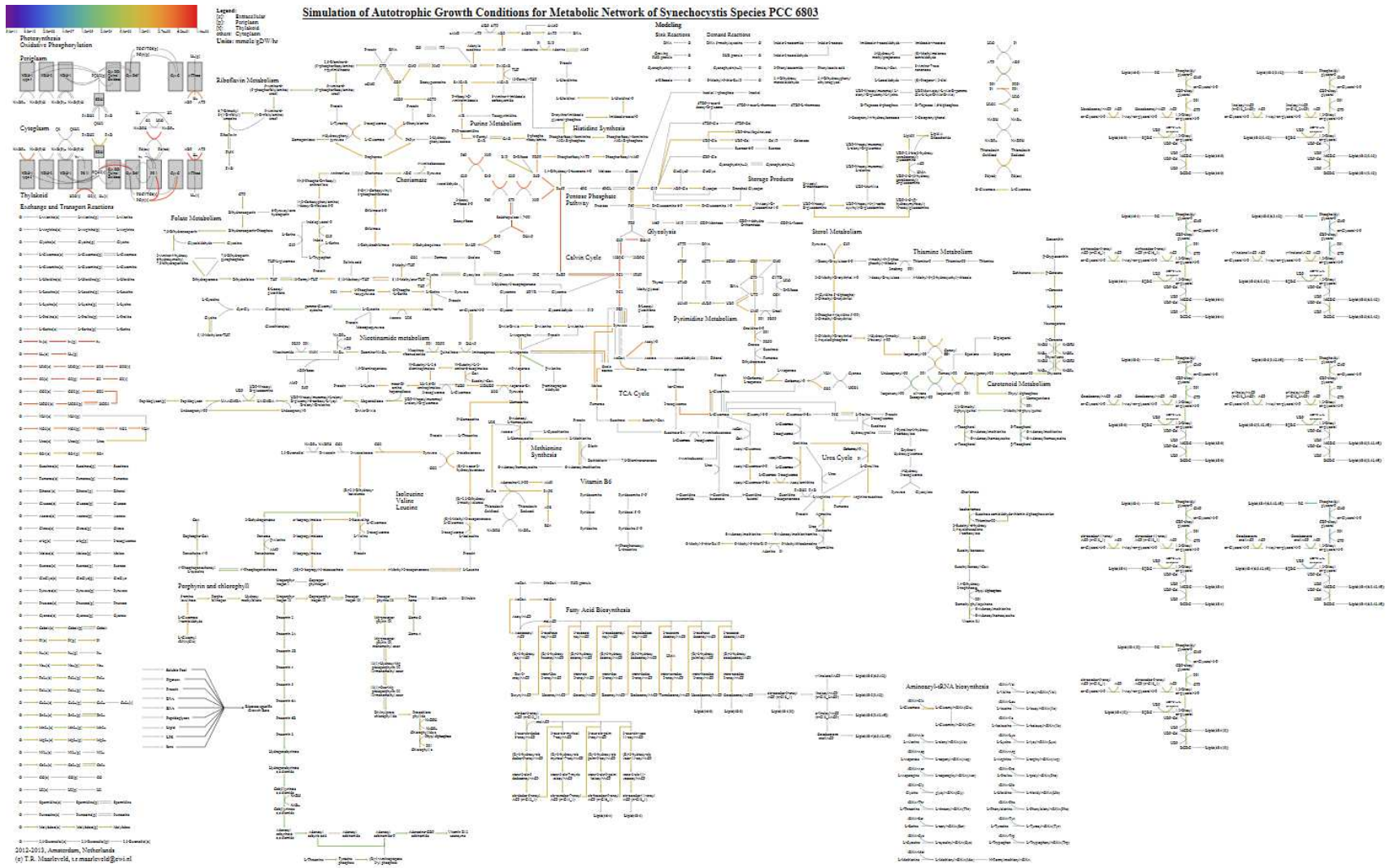
FIGURE D6: FLUX PREDICTIONS FOR AUTOTROPHIC GROWTH SIMULATIONS. HCO3 uptake was set at 100 mmol/gDW/h.
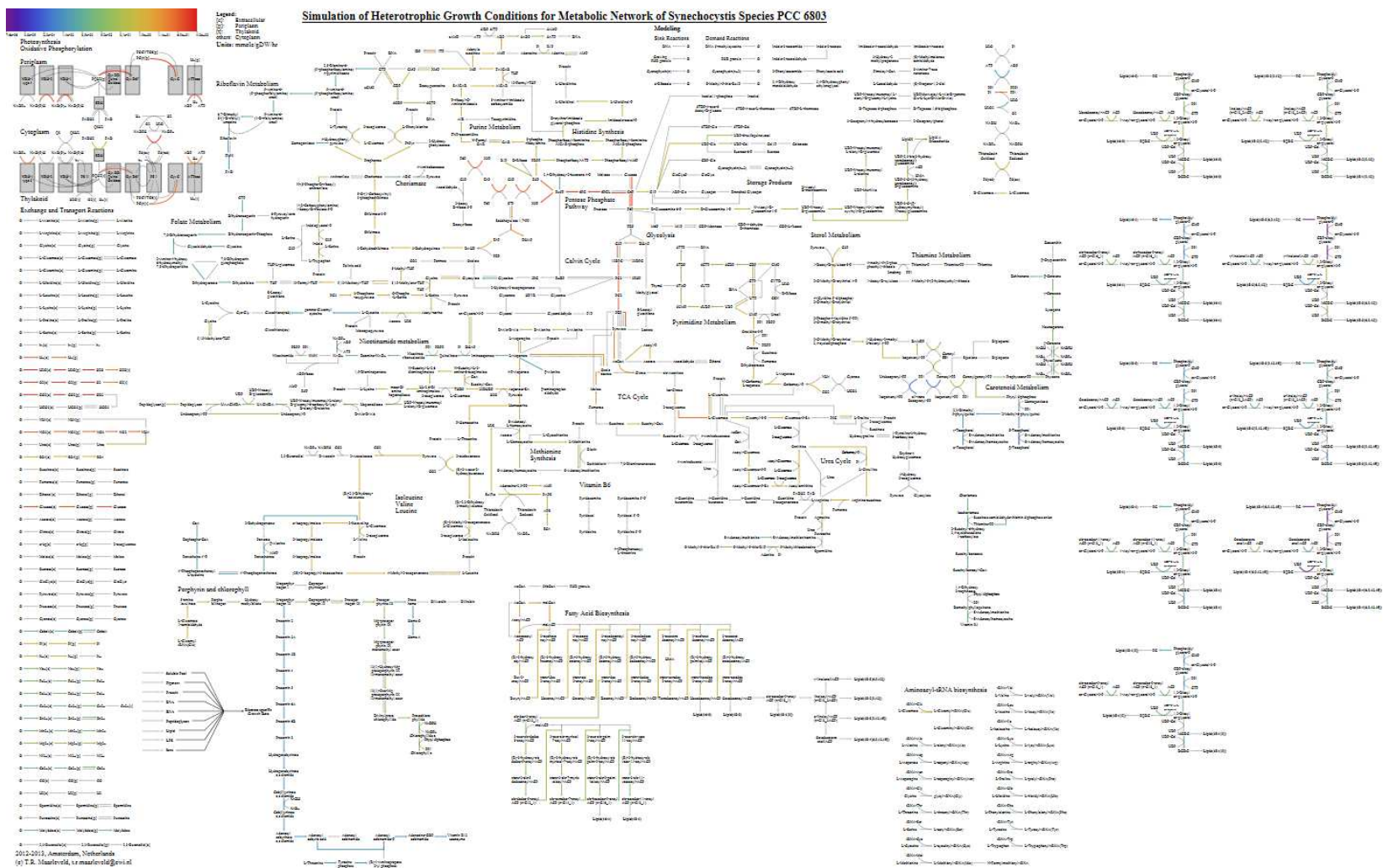
FIGURE D8: FLUX PREDICTIONS FOR HETEROTROPHIC GROWTH SIMULATIONS. Glucose uptake was set at 100 mmol/gDW/h.

TABLE D1a: THERMODYNAMIC INFORMATION OF REACTIONS.

Reaction ID	$\Delta_r G'^o$	$\Delta_r G'^m$	U_r	$\Delta_r G'^{max}$	$\Delta_r G'^{min}$	Reversible
PSTA	26.8	26.8	1.974842	35.80591	17.79409	1
ATPSu	20.1	24.8828	13.54991	46.57631	6.049693	0
ATPS4rpp	20.1	24.8828	13.54991	46.57631	6.049693	0
THRDR2	12.2	16.29232	7.106335	28.02645	5.511662	0
FMN2	13.3	13.3	6.418723	22.30591	4.294085	1
SADT	11.9	11.9	4.538722	20.90591	2.894085	0
THDPS	11.7	11.7	5.244044	20.70591	2.694085	0
MTHFD	11.6	11.6	6.496153	20.60591	2.594085	0
P5CRx	11.4	11.4	5.43139	20.40591	2.394085	0
UAGCVT	10	10	4.669047	19.00591	0.994085	0
GLGC	9.9	9.9	5.059644	18.90591	0.894085	0
PTPAT	9.9	9.9	4.949747	18.90591	0.894085	1
FMNAT	9.9	9.9	5.813777	18.90591	0.894085	0
NMNAT	9.9	9.9	5.291503	18.90591	0.894085	0
PSCVTi	9.9	9.9	3.03315	18.90591	0.894085	1
G1PTT	9.9	9.9	4.38178	18.90591	0.894085	0
GALU	9.9	9.9	4.460942	18.90591	0.894085	1
NNATr	9.9	9.9	5.263079	18.90591	0.894085	1
UAGDP	9.9	9.9	4.538722	18.90591	0.894085	0
OXGDC2	9.7	9.899334	4.135215	17.13051	-3.39824	0
NDPK1	9.5	9.5	5.761944	18.50591	0.494085	1
NDPK5	9.5	9.5	5.727128	18.50591	0.494085	1
NDPK3	9.5	9.5	5.329165	18.50591	0.494085	1
ICDHy	12.6	8.707014	5.310367	17.71293	-7.31877	0
GLYCTO_syn	7.6	7.6	5.069517	16.60591	-1.40591	0
G6PDH2	7.3	7.3	5.656854	16.30591	-1.70591	0
BKTIO_syn	7	7	7.771744	16.00591	-2.00591	0
ME1	10.8	6.907014	5.244044	15.91293	-9.11877	0
HISTP	10.5	6.40768	3.255764	12.68538	-0.82349	0
LDH_D	6.2	6.2	5.09902	15.20591	-2.80591	1
MDH	6.2	6.2	5.118594	15.20591	-2.80591	1
PDX5PSa	6.2	6.2	5.09902	15.20591	-2.80591	0
GND	10	6.107014	5.43139	15.11293	-9.91877	0
3PGDH	6.1	6.1	5.108816	15.10591	-2.90591	0
IPMD	6.1	6.1	5.234501	15.10591	-2.90591	0
MTHFR2	5.9	5.9	6.442049	14.90591	-3.10591	0
PRAIS	9.8	5.70768	5.51362	16.48834	-6.02645	0
PRAMPC	5.3	5.3	4.898979	9.802957	0.797043	0

P5CD	5.2	5.2	5.813777	14.20591	-3.80591	0
HSDx	5.1	5.1	5.118594	14.10591	-3.90591	1
PYK2	4.2	4.2	3.660601	13.20591	-4.80591	0
PYK	4.2	4.2	4.472136	13.20591	-4.80591	0
PYK5	4.1	4.1	4.27785	13.10591	-4.90591	0
SUCDu_syn	4.1	4.1	4.062019	13.10591	-4.90591	0
SUCDyy_syn	4.1	4.1	4.062019	13.10591	-4.90591	0
AGPR	-0.3	3.79232	5.477226	15.52645	-6.98834	1
ARGSL	7.8	3.70768	3.435113	9.985381	-3.52349	1
G3PD2	3.7	3.7	5.128353	12.70591	-5.30591	1
GLYCDx	3.7	3.7	5.157519	12.70591	-5.30591	1
GLCP2	-0.4	3.69232	4.764452	10.92349	-2.58538	0
GLCP	-0.4	3.69232	4.76313	10.92349	-2.58538	0
ADSL2r	7.7	3.60768	3.619392	9.885381	-3.62349	1
ADSL1r	7.7	3.60768	4.427189	9.885381	-3.62349	1
PPC1	3.2	3.2	2.387467	12.20591	-5.80591	0
ORPT	3.2	3.2	3.885872	12.20591	-5.80591	1
SBP	7.2	3.10768	2.828427	9.385381	-4.12349	0
HSTPT	2.6	2.6	2.915476	11.60591	-6.40591	1
DNMPPA	6.6	2.50768	4.242641	8.785381	-4.72349	0
PAPSP	6.6	2.50768	4.560702	8.785381	-4.72349	0
SPP	6.6	2.50768	4.301163	8.785381	-4.72349	0
MI1PP	6.6	2.50768	3.34664	8.785381	-4.72349	0
PGLYCP	6.6	2.50768	2.024846	8.785381	-4.72349	0
NTD7	6.6	2.50768	4.516636	8.785381	-4.72349	0
PGPP161	6.6	2.50768	4.195235	8.785381	-4.72349	0
PAPA160	6.6	2.50768	4.123106	8.785381	-4.72349	0
PAPA180	6.6	2.50768	4.312772	8.785381	-4.72349	0
ACP1(FMN)	6.6	2.50768	4.404543	8.785381	-4.72349	0
BPNT	6.6	2.50768	4.516636	8.785381	-4.72349	0
FBP	6.6	2.50768	3.224903	8.785381	-4.72349	0
NTD4	6.6	2.50768	3.741657	8.785381	-4.72349	0
PAPA161	6.6	2.50768	4.07431	8.785381	-4.72349	0
PAPA181	6.6	2.50768	4.242641	8.785381	-4.72349	0
PGPP180	6.6	2.50768	4.449719	8.785381	-4.72349	0
PGPP181	6.6	2.50768	4.370355	8.785381	-4.72349	0
PMDPHT	6.5	2.40768	3.687818	8.685381	-4.82349	1
PGPP160	6.5	2.40768	4.266146	8.685381	-4.82349	0
IMPD	2.4	2.4	6.457554	11.40591	-6.60591	0
MTHFD2i	2.1	2.1	6.503845	11.10591	-6.90591	0
P5CR	1.9	1.9	5.422177	10.90591	-7.10591	0

TKT1	1.9	1.9	2.932576	10.90591	-7.10591	1
AHCi	5.7	1.60768	4.593474	7.885381	-5.62349	0
TPI	1.5	1.5	1.341641	6.002957	-3.00296	1
PHACA	0.86	0.86	2.469818	5.362957	-3.64296	0
THRS	4.9	0.80768	2.258318	7.085381	-6.42349	0
IPDDI	0.8	0.8	1.788854	5.302957	-3.70296	1
PRAIi	0.5	0.5	2.828427	5.002957	-4.00296	0
MTRI	0.5	0.5	2.588436	5.002957	-4.00296	1
DASYN181	0.5	0.5	4.898979	9.505915	-8.50591	0
MAN1PT	0.4	0.4	4.909175	9.405915	-8.60591	0
PRMICIi	0.4	0.4	4.785394	4.902957	-4.10296	0
TMPPP	0.4	0.4	4.123106	9.405915	-8.60591	0
RPI	0.4	0.4	2.280351	4.902957	-4.10296	1
MEPCT	0.4	0.4	3.807887	9.405915	-8.60591	0
NMNS	0.4	0.4	3.847077	9.405915	-8.60591	0
DASYN160	0.4	0.4	4.806246	9.405915	-8.60591	0
DASYN161	0.4	0.4	4.764452	9.405915	-8.60591	0
DASYN180	0.4	0.4	4.959839	9.405915	-8.60591	0
FBA	4.2	0.10768	2.345208	6.385381	-7.12349	1
PAPPT3	0.1	0.1	7.576279	9.105915	-8.90591	0
IPPMIa	0.1	0.1	2.302173	4.602957	-4.40296	1
ORNTACr	0.1	0.1	2.569047	9.105915	-8.90591	1
PPA	4.1	0.00768	3	6.285381	-7.22349	0
FBA2	4.1	0.00768	2.366432	6.285381	-7.22349	1
NDPK2	1.14E-13	1.14E-13	5.310367	9.005915	-9.00591	1
ASPTA	5.68E-14	5.68E-14	2	9.005915	-9.00591	1
ADK_syn	5.68E-14	5.68E-14	5.882176	9.005915	-9.00591	1
CYTK1	5.68E-14	5.68E-14	5.329165	9.005915	-9.00591	1
DTMPK	5.68E-14	5.68E-14	5.25357	9.005915	-9.00591	1
VALTA	2.84E-14	2.84E-14	2.097618	9.005915	-9.00591	1
PHETA1	1.42E-14	1.42E-14	2.626785	9.005915	-9.00591	1
LEUTA	1.42E-14	1.42E-14	2.097618	9.005915	-9.00591	1
ILETA	1.42E-14	1.42E-14	2.097618	9.005915	-9.00591	1
DAPE	0	0	1.788854	4.502957	-4.50296	1
GHMT2r	0	0	4.549725	9.005915	-9.00591	1
PGM	0	0	1.264911	4.502957	-4.50296	1
PGMT	0	0	2.898275	4.502957	-4.50296	1
PMANM	0	0	2.898275	4.502957	-4.50296	1
UDPG4E	0	0	4.195235	4.502957	-4.50296	1
PGAMT	0	0	2.898275	4.502957	-4.50296	1
GLUR	0	0	1.414214	4.502957	-4.50296	1

ALAR	0	0	1.264911	4.502957	-4.50296	1
RPE	0	0	1.732051	4.502957	-4.50296	1
GK1	0	0	5.761944	9.005915	-9.00591	1
ADK1	0	0	7.224957	9.005915	-9.00591	1
NDPK7	0	0	5.282045	9.005915	-9.00591	1
URIDK2r	0	0	5.25357	9.005915	-9.00591	1
NDPK4	0	0	5.25357	9.005915	-9.00591	1
ICHORS	0	0	2.792848	4.502957	-4.50296	1
GLBRAN2	0	0	5.829872	4.502957	-4.50296	0
TMPKr	0	0	5.639149	9.005915	-9.00591	1
Htex	0	0	0	0	0	1
O2tex	0	0	1.732051	5.102254	-5.10225	1
CO2tex	0	0	1.732051	7.019868	-7.01987	1
CO2tpp	0	0	1.732051	7.019868	-7.01987	0
H2Otex	0	0	1.732051	0	0	1
HCO3tex	0	0	1.732051	4.502957	-4.50296	1
GLCtex	0	0	2.932576	4.502957	-4.50296	1
GLNtex	0	0	1.843909	4.502957	-4.50296	1
ARGtex	0	0	3	4.502957	-4.50296	1
GLUtex	0	0	1.414214	4.502957	-4.50296	1
HIStex	0	0	2.529822	4.502957	-4.50296	1
UREAtex	0	0	1.732051	4.502957	-4.50296	1
NO3tex	0	0	1.732051	4.502957	-4.50296	1
NH4tex	0	0	1.732051	4.502957	-4.50296	1
SO4tex	0	0	1.732051	4.502957	-4.50296	1
PItex	0	0	1.732051	4.502957	-4.50296	1
FE2tex	0	0	1.732051	4.502957	-4.50296	1
FE3tex	0	0	1.732051	4.502957	-4.50296	1
PTRCtex	0	0	1.67332	4.502957	-4.50296	1
SPMDtex	0	0	2.48998	4.502957	-4.50296	1
DRIBtex	0	0	2.720294	4.502957	-4.50296	0
DRU1Ptex	0	0	2.236068	4.502957	-4.50296	0
Ht2cax	0	0	0	0	0	1
O2t2cax	0	0	1.732051	5.102254	-5.10225	0
HCO3t2cax	0	0	1.732051	4.502957	-4.50296	0
RBPt2cax	0	0	2.720294	4.502957	-4.50296	1
3PGt2cax	0	0	1.264911	4.502957	-4.50296	1
2PGt2cax	0	0	1.095445	4.502957	-4.50296	0
H2Otul_syn	0	0	1.732051	0	0	1
O2tl	0	0	1.732051	5.102254	-5.10225	1
PQ9tu	0	0	3.898718	4.502957	-4.50296	1

H2Otpp	0	0	1.732051	0	0	1
O2tc	0	0	1.732051	5.102254	-5.10225	1
PQ9tm	0	0	3.898718	4.502957	-4.50296	1
H2tex	0	0	1.732051	4.817314	-4.81731	1
H2tpp	0	0	1.732051	4.817314	-4.81731	1
SUCRtex	0	0	3.949684	4.502957	-4.50296	1
ACtex	0	0	1.732051	4.502957	-4.50296	1
AKGtex	0	0	1.48324	4.502957	-4.50296	1
ALAtex	0	0	1.264911	4.502957	-4.50296	1
CITtex	0	0	1.67332	4.502957	-4.50296	1
CYNTtex	0	0	5.477226	4.502957	-4.50296	1
FRUtex	0	0	2.75681	4.502957	-4.50296	1
FUMtex	0	0	2.04939	4.502957	-4.50296	1
GLYtex	0	0	1.183216	4.502957	-4.50296	1
LEUtex	0	0	1.549193	4.502957	-4.50296	1
LYStex	0	0	1.732051	4.502957	-4.50296	1
MALtex	0	0	1.264911	4.502957	-4.50296	1
PROtex	0	0	2.32379	4.502957	-4.50296	1
PYRtex	0	0	1.341641	4.502957	-4.50296	1
SERtex	0	0	1.341641	4.502957	-4.50296	1
SUCCtex	0	0	1.183216	4.502957	-4.50296	1
G3PD	0	0	5.648008	9.005915	-9.00591	0
NDPK10	0	0	5.727128	9.005915	-9.00591	1
PPM	0	0	2.720294	4.502957	-4.50296	1
UGLT	0	0	5.09902	9.005915	-9.00591	1
LALDO	-4.1	-0.00768	5.882176	11.72645	-10.7883	1
NDPK6	-0.1	-0.1	5.263079	8.905915	-9.10591	1
NDPK9	-0.1	-0.1	5.75326	8.905915	-9.10591	1
CYSTA	-0.1	-0.1	2.428992	8.905915	-9.10591	1
SDPTA	-0.1	-0.1	2.683282	8.905915	-9.10591	1
TYRTA	-0.1	-0.1	2.701851	8.905915	-9.10591	1
IPPMIb	-0.1	-0.1	2.302173	4.402957	-4.60296	1
NDPK8	-0.1	-0.1	5.882176	8.905915	-9.10591	1
UMPK	-0.1	-0.1	5.300943	8.905915	-9.10591	1
SSALyr	-0.4	-0.4	5.215362	8.605915	-9.40591	1
FTHFLi	-0.4	-0.4	6.196773	13.10887	-13.9089	0
MTHFC	-0.5	-0.5	4.449719	4.002957	-5.00296	1
FUM	-0.6	-0.6	2.097618	3.902957	-5.10296	1
GLCt2pp	0	-0.68129	2.932576	3.82167	-5.18425	1
NH4tpp	0	-0.68129	1.732051	3.82167	-5.18425	0
ACt2rpp	0	-0.68129	1.732051	3.82167	-5.18425	1

AKGtpp	0	-0.68129	1.48324	3.82167	-5.18425	1
CITtpp	0	-0.68129	1.67332	3.82167	-5.18425	1
CYNTt2pp	0	-0.68129	5.477226	3.82167	-5.18425	0
FRUt3	0	-0.68129	2.75681	3.82167	-5.18425	0
FUMtpp	0	-0.68129	2.04939	3.82167	-5.18425	1
GLUttrappp	0	-0.68129	1.414214	3.82167	-5.18425	0
MALtpp	0	-0.68129	1.264911	3.82167	-5.18425	1
PTRCt2pp	0	-0.68129	1.67332	3.82167	-5.18425	0
PYRtpp	0	-0.68129	1.341641	3.82167	-5.18425	1
SPMDt2pp	0	-0.68129	2.48998	3.82167	-5.18425	0
SUCCtpp	0	-0.68129	1.183216	3.82167	-5.18425	1
PGI	-0.8	-0.8	2.810694	3.702957	-5.30296	1
MAN6PI	-0.8	-0.8	2.810694	3.702957	-5.30296	1
SUCOAS	-0.9	-0.9	6.284903	12.60887	-14.4089	1
FBA3	3.1	-0.99232	2.167948	5.285381	-8.22349	1
ENO	-1	-1	2	3.502957	-5.50296	1
ACOTA	-1	-1	2.50998	8.005915	-10.0059	1
ACCOACr	-1	-1	6.371813	12.50887	-14.5089	1
ABTA	-1.1	-1.1	1.923538	7.905915	-10.1059	0
ORNTA	-1.1	-1.1	2.12132	7.905915	-10.1059	0
G1SATi	-1.4	-1.4	1.414214	3.102957	-5.90296	0
MOHMT	-1.4	-1.4	4.64758	7.605915	-10.4059	0
GLYAT	-1.4	-1.4	1.897367	7.605915	-10.4059	0
CTPS2	2.6	-1.49232	5.830952	13.7913	-17.7294	0
SPT_syn2	-1.5	-1.5	1.788854	7.505915	-10.5059	1
SPT_syn	-1.5	-1.5	1.81659	7.505915	-10.5059	1
DHNPA	2.4	-1.69232	3.937004	4.585381	-8.92349	0
AOXSr	2.2	-1.69299	4.888763	7.312929	-17.7188	1
TKT2	-1.7	-1.7	2.720294	7.305915	-10.7059	1
TALA	-1.7	-1.7	2.898275	7.305915	-10.7059	1
SUCDi	-2.2	-2.2	5.394442	6.805915	-11.2059	0
QULNS	-2.3	-2.3	4.147288	6.705915	-11.3059	0
IMPC	-2.8	-2.8	3.885872	1.702957	-7.30296	1
DRPA	1.2	-2.89232	2.387467	3.385381	-10.1235	1
MI1PS	-3	-3	2.863564	1.502957	-7.50296	0
DHPTS	-7.2	-3.10768	4.427189	4.123491	-9.38538	0
DDPA	-3.2	-3.2	2.810694	5.805915	-12.2059	0
UPPRT	-3.2	-3.2	3.860052	5.805915	-12.2059	0
ADPT	-3.3	-3.3	4.658326	5.705915	-12.3059	1
PHTHAT	-3.4	-3.4	1.907878	5.605915	-12.4059	1
ADNCYC	0.5	-3.59232	4.593474	2.685381	-10.8235	0

212

MTHFR1	-3.6	-3.6	6.434283	5.405915	-12.6059	0
G5SD	0.3	-3.79232	5.282045	6.988338	-15.5264	0
GAPDi(nadp)	0.3	-3.79232	5.263079	6.988338	-15.5264	0
SERATi	-3.8	-3.8	4.722288	5.205915	-12.8059	0
ASADi	0.2	-3.89232	5.272571	6.888338	-15.6264	0
OPHBDC	-0.1	-3.99299	5.01996	0.509971	-15.5158	0
RBFSa	-4	-4	4.615192	5.005915	-13.0059	0
GTHRDH_syn	-0.3	-4.39232	3.098387	1.885381	-11.6235	0
CYGLDP	-0.3	-4.39232	2.607681	1.885381	-11.6235	0
SDPDS	-0.3	-4.39232	2.50998	1.885381	-11.6235	0
AMPN	-0.8	-4.89232	4.658326	1.385381	-12.1235	0
DAD5N	-0.9	-4.99232	4.658326	1.285381	-12.2235	0
ACKr	-5	-5	4.438468	4.005915	-14.0059	1
CYTDH	-1	-5.09232	3.847077	1.185381	-12.3235	0
NNDPR	-1.2	-5.09299	4	3.912929	-21.1188	0
TRSARr	-5.1	-5.1	5.108816	3.905915	-14.1059	1
ALCD19	-5.1	-5.1	5.157519	3.905915	-14.1059	1
LCARS	-5.1	-5.1	5.108816	3.905915	-14.1059	1
PGL	-5.3	-5.3	2.683282	-0.79704	-9.80296	0
OMPDC	-1.6	-5.49299	3.49285	-0.99003	-17.0158	0
ACGS	-5.5	-5.5	4.795832	3.505915	-14.5059	0
G1PACT	-5.5	-5.5	5.347897	3.505915	-14.5059	0
DHORTS	-5.5	-5.5	2.915476	-0.99704	-10.003	1
PTAr	-5.7	-5.7	4.711688	3.305915	-14.7059	1
ALCD2x	-5.7	-5.7	5.234501	3.305915	-14.7059	1
GAPDi	-9.8	-5.70768	5.272571	6.026449	-16.4883	0
AORNAH	-1.8	-5.89232	2.54951	0.385381	-13.1235	1
UHGADA	-1.9	-5.99232	4.690416	0.285381	-13.2235	0
MGSA	-2	-6.09232	1.81659	0.185381	-13.3235	0
AICART	-6.1	-6.1	5.412947	2.905915	-15.1059	1
HPYRR2_syn	-6.1	-6.1	5.118594	2.905915	-15.1059	0
LGTHL	-10.3	-6.20768	3.209361	1.023491	-12.4854	1
UAMAGS	-6.3	-6.3	6.403124	7.208872	-19.8089	0
DHFS	-6.4	-6.4	6.049793	7.108872	-19.9089	0
GLUCYS	-6.4	-6.4	4.969909	7.108872	-19.9089	0
GTHS	-6.4	-6.4	5.186521	7.108872	-19.9089	0
PRAGS	-6.4	-6.4	5.234501	7.108872	-19.9089	0
UAAGDS	-6.4	-6.4	6.640783	7.108872	-19.9089	0
UGMDDS	-6.4	-6.4	7.007139	7.108872	-19.9089	0
ALAALAr	-6.5	-6.5	4.929503	7.008872	-20.0089	1
ASPKi	-6.5	-6.5	4.404543	2.505915	-15.5059	0

UAMAS	-6.5	-6.5	6.236986	7.008872	-20.0089	0
PGK	-6.6	-6.6	4.38178	2.405915	-15.6059	1
GLU5K	-6.6	-6.6	4.41588	2.405915	-15.6059	0
ACGK	-6.6	-6.6	4.64758	2.405915	-15.6059	0
GF6PTA	-6.7	-6.7	3.255764	2.305915	-15.7059	1
DHAD1	-6.9	-6.9	2	-2.39704	-11.403	0
ADMDC	-3.1	-6.99299	4.764452	-2.49003	-18.5158	0
ASP1DC	-3.1	-6.99299	1.788854	-2.49003	-18.5158	0
OMCDC	-3.1	-6.99299	2	-2.49003	-18.5158	0
LYSDC	-3.1	-6.99299	2.12132	-2.49003	-18.5158	0
PPCDC	-3.1	-6.99299	3.209361	-2.49003	-18.5158	0
GLUDC	-3.1	-6.99299	1.843909	-2.49003	-18.5158	0
DHAD2	-7	-7	2	-2.49704	-11.503	0
ARGDC	-3.2	-7.09299	3.24037	-2.59003	-18.6158	0
DAPDC	-3.2	-7.09299	2.144761	-2.59003	-18.6158	0
H2ASE_syn	-1	-7.09554	5.07937	-2.59258	-16.4158	1
RBPC	-6.9	-7.09933	3.146427	6.198235	-14.3305	0
G3PDau	-7.2	-7.2	3.949684	1.805915	-16.2059	0
G3PDap	-7.2	-7.2	3.949684	1.805915	-16.2059	0
MALT	-3.2	-7.29232	5.215362	-1.01462	-14.5235	0
GARFT	-7.3	-7.3	5.22494	1.705915	-16.3059	1
ADCS	-7.4	-7.4	3.24037	1.605915	-16.4059	0
PUNP1	-7.4	-7.4	4.658326	1.605915	-16.4059	1
MTAP	-7.5	-7.5	4.732864	1.505915	-16.5059	0
PYNPa	-7.5	-7.5	3.834058	1.505915	-16.5059	1
METAT	-3.5	-7.59232	5.09902	3.188338	-19.3264	0
IPPS	-7.6	-7.6	4.909175	1.405915	-16.6059	0
PURT	-7.7	-7.7	5.449771	5.808872	-21.2089	0
DXPS	-8.1	-7.90067	2.144761	-0.66949	-21.1982	0
PDH	-8.2	-8.00067	6.83374	3.733463	-25.8012	0
AKGCL	-4.2	-8.09299	1.81659	-3.59003	-19.6158	0
DHQTi	-8.4	-8.4	2.588436	-3.89704	-12.903	0
NADN	-4.5	-8.59232	5.357238	-2.31462	-15.8235	0
ARGSS	-8.7	-8.7	5.282045	4.808872	-22.2089	0
AGMT	-4.7	-8.79232	2.983287	-2.51462	-16.0235	0
ARGN	-4.7	-8.79232	2.983287	-2.51462	-16.0235	0
CS	-8.9	-8.9	4.878524	0.105915	-17.9059	0
CHORS	-4.9	-8.99232	2.966479	-2.71462	-16.2235	0
ATPPRT	-9.1	-9.1	4.91935	-0.09409	-18.1059	0
ASPCT	-9.2	-9.2	2.607681	-0.19409	-18.2059	1
GMPS	-9.2	-9.2	6.024948	4.308872	-22.7089	0

OCBT	-9.3	-9.3	2.720294	-0.29409	-18.3059	1
NADTRHD	-9.5	-9.5	6.978539	-0.49409	-18.5059	0
DPGM	-9.5	-9.5	1.414214	-4.99704	-14.003	1
DHFR	-9.7	-9.7	6.457554	-0.69409	-18.7059	1
GCALDD	-9.8	-9.8	5.196152	-0.79409	-18.8059	0
GLYALDDr	-9.8	-9.8	5.272571	-0.79409	-18.8059	1
GLYOX	-5.8	-9.89232	3.405877	-3.61462	-17.1235	0
ABUTD	-9.9	-9.9	5.272571	-0.89409	-18.9059	0
PUTA3	-9.9	-9.9	5.282045	-0.89409	-18.9059	0
CYSTL	-1.8	-9.98464	2.898275	-1.9322	-19.944	0
DHDPRy	-10	-10	5.630275	-0.99409	-19.0059	0
ACS	-10.1	-10.1	6.348228	3.408872	-23.6089	0
SPMSx	-10.2	-10.2	4.898979	-1.19409	-19.2059	0
SPMS	-10.2	-10.2	4.898979	-1.19409	-19.2059	0
PRASCS	-10.3	-10.3	5.59464	3.208872	-23.8089	1
DHORD2	-10.3	-10.3	3.391165	-1.29409	-19.3059	0
FTHFD	-6.3	-10.3923	4.582576	-4.11462	-17.6235	0
SHK3D	-10.4	-10.4	5.440588	-1.39409	-19.4059	1
GLXCL	-10.6	-10.4007	2.345208	-3.16949	-23.6982	0
UDCPDP	-6.6	-10.6923	5.882176	-4.41462	-17.9235	0
ATPM	-6.7	-10.7923	4.516636	-4.51462	-18.0235	0
PPK2r	-10.8	-10.8	4.516636	-1.79409	-19.8059	1
PPK1r	-10.8	-10.8	4.516636	-1.79409	-19.8059	0
IGPDH	-10.8	-10.8	2.828427	-6.29704	-15.303	0
LPADSS	-10.9	-10.9	6.082763	-1.89409	-19.9059	0
UAGPT3	-10.9	-10.9	7.648529	-1.89409	-19.9059	0
NDH2_syn	-10.9	-10.9	6.17252	-1.89409	-19.9059	0
NDH2_1p	-10.9	-10.9	6.17252	-1.89409	-19.9059	0
SPS	-10.9	-10.9	5.059644	-1.89409	-19.9059	0
FOLD3	-11.3	-11.3	4.427189	-2.29409	-20.3059	0
ANPRT	-11.4	-11.4	3.549648	-2.39409	-20.4059	0
BCT1_syn	-6.7	-11.4736	4.837355	-0.69295	-23.2077	0
GLNabcpp	-6.7	-11.4736	4.878524	-0.69295	-23.2077	0
ARGabcpp	-6.7	-11.4736	5.422177	-0.69295	-23.2077	0
HISabcpp	-6.7	-11.4736	5.176872	-0.69295	-23.2077	0
UREAabcpp	-6.7	-11.4736	4.837355	-0.69295	-23.2077	0
NO3abcpp	-6.7	-11.4736	4.837355	-0.69295	-23.2077	0
SULabcpp	-6.7	-11.4736	4.837355	-0.69295	-23.2077	0
PIuabcpp	-6.7	-11.4736	5.138093	-0.69295	-23.2077	0
FE2abcpp	-6.7	-11.4736	4.837355	-0.69295	-23.2077	0
FE3abc	-6.7	-11.4736	4.837355	-0.69295	-23.2077	0

PTRCabcpp	-6.7	-11.4736	4.816638	-0.69295	-23.2077	0
SPMDabcpp	-6.7	-11.4736	5.157519	-0.69295	-23.2077	0
SUCRabcpp_syn	-6.7	-11.4736	6	-0.69295	-23.2077	0
ALAabcpp	-6.7	-11.4736	4.690416	-0.69295	-23.2077	0
GLYabcpp	-6.7	-11.4736	4.669047	-0.69295	-23.2077	0
LEUabcpp	-6.7	-11.4736	4.774935	-0.69295	-23.2077	0
LYSabcpp	-6.7	-11.4736	4.837355	-0.69295	-23.2077	0
PROabcpp	-6.7	-11.4736	5.07937	-0.69295	-23.2077	0
SERabcpp	-6.7	-11.4736	4.711688	-0.69295	-23.2077	0
PPBNGS	-15.6	-11.5077	3.72827	-4.27651	-17.7854	0
SUCBZL	-11.6	-11.6	6.503845	1.908872	-25.1089	0
MMHL	-11.7	-11.7	1.545962	-7.19704	-16.203	1
GFUCS	-11.9	-11.9	6.708204	-2.89409	-20.9059	1
ALDD2xr	-12	-12	5.375872	-2.99409	-21.0059	1
ADSS	-12.7	-12.7	5.899152	0.808872	-26.2089	0
APRAUR	-12.7	-12.7	5.98331	-3.69409	-21.7059	0
CYSS	-12.8	-12.8	2.529822	-3.79409	-21.8059	0
KARA1i	-13.2	-13.2	5.196152	-4.19409	-22.2059	0
PGSA160	-13.2	-13.2	4.816638	-4.19409	-22.2059	0
RBK	-13.2	-13.2	4.98999	-4.19409	-22.2059	0
DHQS	-9.2	-13.2923	2.44949	-7.01462	-20.5235	0
PNTK	-13.3	-13.3	4.711688	-4.29409	-22.3059	0
SHKK	-13.3	-13.3	4.764452	-4.29409	-22.3059	0
HSK	-13.3	-13.3	4.393177	-4.29409	-22.3059	0
PFK	-13.3	-13.3	4.97996	-4.29409	-22.3059	0
HEX7	-13.3	-13.3	4.98999	-4.29409	-22.3059	1
NADK	-13.3	-13.3	6.473021	-4.29409	-22.3059	0
DPCOAK	-13.3	-13.3	6.090977	-4.29409	-22.3059	0
CDPMEK	-13.3	-13.3	5.403702	-4.29409	-22.3059	0
DMATT	-13.3	-13.3	2.75681	-4.29409	-22.3059	0
PFK_2	-13.3	-13.3	4.97996	-4.29409	-22.3059	0
TMDK1	-13.3	-13.3	5.25357	-4.29409	-22.3059	0
PGSA161	-13.3	-13.3	4.764452	-4.29409	-22.3059	0
PGSA180	-13.3	-13.3	4.969909	-4.29409	-22.3059	0
PGSA181	-13.3	-13.3	4.898979	-4.29409	-22.3059	0
RBFK	-13.3	-13.3	5.813777	-4.29409	-22.3059	0
TMK	-13.3	-13.3	5.630275	-4.29409	-22.3059	0
THRPS	-13.3	-13.3	4.438468	-4.29409	-22.3059	0
GLYCK_2	-13.3	-13.3	4.370355	-4.29409	-22.3059	0
HEX1	-13.3	-13.3	5.089204	-4.29409	-22.3059	0
GLYK	-13.3	-13.3	4.427189	-4.29409	-22.3059	0

GALK	-13.3	-13.3	5.089204	-4.29409	-22.3059	0
GLYCK	-13.3	-13.3	4.370355	-4.29409	-22.3059	0
ADSK	-13.3	-13.3	5.932959	-4.29409	-22.3059	0
GRTT	-13.4	-13.4	3.224903	-4.39409	-22.4059	0
PRPPS	-13.4	-13.4	4.97996	-4.39409	-22.4059	1
HPPK	-13.4	-13.4	5.674504	-4.39409	-22.4059	0
GLUSx	-13.7	-13.7	5.585696	-0.19113	-27.2089	0
GMANDi	-13.8	-13.8	4.711688	-9.29704	-18.303	0
PFK_3	-13.9	-13.9	4.732864	-4.89409	-22.9059	0
PRUK	-14.3	-14.3	4.753946	-5.29409	-23.3059	0
DXPRIi	-14.5	-14.5	5.215362	-5.49409	-23.5059	0
HSD2	-14.6	-14.6	5.108816	-5.59409	-23.6059	0
TRSAR2	-14.6	-14.6	5.09902	-5.59409	-23.6059	0
ALCD2ai	-14.6	-14.6	5.147815	-5.59409	-23.6059	0
GUACYC	-10.6	-14.6923	4.438468	-8.41462	-21.9235	0
PPS	-10.9	-14.9923	4.795832	-4.21166	-26.7264	0
CBPS	-11	-15.0923	8.854377	4.694253	-35.8324	0
G5SADi	-15.1	-15.1	2.213594	-10.597	-19.603	0
PROD2	-15.1	-15.1	5.924525	-6.09409	-24.1059	0
ALCD2yi	-15.2	-15.2	5.22494	-6.19409	-24.2059	0
HPYRR1i_syn	-15.6	-15.6	5.108816	-6.59409	-24.6059	0
DPRi	-15.6	-15.6	5.205766	-6.59409	-24.6059	0
GLUPRT	-11.6	-15.6923	3.619392	-4.91166	-27.4264	0
PPNCL	-15.9	-15.9	4.604346	-2.39113	-29.4089	0
METS	-16	-16	4.582576	-6.99409	-25.0059	0
TMDS	-16	-16	5.272571	-6.99409	-25.0059	0
MMS	-16.2	-16.2	4.752894	-7.19409	-25.2059	0
PRFGS	-12.2	-16.2923	5.907622	-1.0087	-32.5294	0
AHMMPSi	-8.7	-16.8846	3.449638	-8.8322	-26.844	0
PANTS	-17.2	-17.2	4.84768	-3.69113	-30.7089	0
PPNCL1	-17.2	-17.2	5.25357	-3.69113	-30.7089	0
NADH5	-17.2	-17.2	7.120393	-8.19409	-26.2059	0
DAPAT	-17.3	-17.3	2.949576	-8.29409	-26.3059	0
CHORMi	-17.4	-17.4	2.701851	-12.897	-21.903	0
GMPS2	-14.1	-18.1923	6.244998	-2.9087	-34.4294	0
PPND	-14.3	-18.193	5.648008	-9.18707	-34.2188	0
UAG2EMAi	-14.2	-18.2923	4.549725	-12.0146	-25.5235	0
IGPS	-15.2	-19.093	3.255764	-14.59	-30.6158	0
GTPH	-16.2	-20.2923	4.335897	-14.0146	-27.5235	0
F6PPK	-20.3	-20.3	2.966479	-11.2941	-29.3059	0
NDH1_1u	-20.4	-20.4	6.164414	-11.3941	-29.4059	0

MAN1PT2	-20.7	-20.7	4.909175	-11.6941	-29.7059	0
FE2abcpp2	-16.2	-20.9736	4.669047	-10.1929	-32.7077	0
PACL	-21.2	-21.2	6.426508	-7.69113	-34.7089	0
GTPCI	-17.3	-21.3923	4.289522	-15.1146	-28.6235	0
NDH1_3u	-21.2	-21.3993	6.519202	-3.59881	-33.1335	0
PRATPP	-17.4	-21.4923	4.857983	-15.2146	-28.7235	0
ASNS1	-17.4	-21.4923	5.069517	-6.2087	-37.7294	0
NTPP8	-17.4	-21.4923	3.701351	-15.2146	-28.7235	0
NTPP9	-17.4	-21.4923	4.32435	-15.2146	-28.7235	0
NADS2	-17.5	-21.5923	6.892024	-6.3087	-37.8294	0
XU5PPK	-22	-22	2.48998	-12.9941	-31.0059	0
GLYCTO1	-24.1	-22.3817	2.073644	-11.0018	-29.6129	0
PROD5u	-22.3	-22.9813	4.335897	-13.9754	-31.9872	0
PROD5p	-22.3	-22.9813	4.335897	-13.9754	-31.9872	0
NDH1_1p	-20.4	-23.1252	6.164414	-14.1192	-32.1311	0
SUCBZS	-23.7	-23.7	2.792848	-19.197	-28.203	0
NDH1_4pp	-21.2	-24.1245	6.519202	-6.32396	-35.8586	0
PSP	-20.1	-24.1923	2.19089	-17.9146	-31.4235	0
GLCS1	-20.3	-24.3923	6.073501	-18.1146	-31.6235	0
DBTSi	-24.5	-24.6993	5.138093	-6.89881	-36.4335	0
UAPGRi	-24.8	-24.8	6.549809	-15.7941	-33.8059	0
DB4PS	-20.9	-24.9923	1.974842	-18.7146	-32.2235	0
DMPPS_syn	-25.7	-25.7	5.385165	-16.6941	-34.7059	0
TMDS2	-25.7	-25.7	7.218033	-12.1911	-39.2089	0
PDX5POi	-28.1	-26.3817	3.549648	-15.0018	-33.6129	0
PYDXNOi	-28.1	-26.3817	3.549648	-15.0018	-33.6129	0
IPDPS_syn	-26.5	-26.5	5.385165	-17.4941	-35.5059	0
PPNCL3	-26.7	-26.7	4.604346	-13.1911	-40.2089	0
MPBQ	-27.7	-27.7	4.868265	-18.6941	-36.7059	1
DHDPS	-31.8	-27.7077	3.420526	-20.4765	-33.9854	0
PYAM5PO	-25.5	-27.874	3.949684	-14.7193	-37.8334	0
NTPTP1	-24.4	-28.4923	4.289522	-22.2146	-35.7235	0
IG3PS	-26.3	-30.3923	4.711688	-19.6117	-42.1264	0
DNTPPA	-26.9	-30.9923	4.242641	-24.7146	-38.2235	0
NTPP2	-27	-31.0923	4.335897	-24.8146	-38.3235	0
NTPP4	-27	-31.0923	3.741657	-24.8146	-38.3235	0
PPNDH	-27.2	-31.093	2.966479	-26.59	-42.6158	0
USHD	-27.4	-31.4923	4.949747	-25.2146	-38.7235	0
ADPRDP	-27.4	-31.4923	4.969909	-25.2146	-38.7235	0
GTPCII	-26.2	-34.3846	5.60357	-26.3322	-44.344	0
PQBS2	-35.3	-35.3	5.656854	-26.2941	-44.3059	0

RBFSb	-38.2	-38.2	6.625708	-29.1941	-47.2059	0
CHORPL	-34.5	-38.5923	2.720294	-32.3146	-45.8235	0
ADCL	-35.7	-39.7923	2.738613	-33.5146	-47.0235	0
CAT	-45.3	-42.926	3.674235	-37.4695	-51.5777	0
SHCHCS2	-39.5	-43.5923	4.753946	-32.8117	-55.3264	0
ANS	-43.1	-47.1923	3.193744	-36.4117	-58.9264	0
FE3R	-50.3	-50.3	6.033241	-36.7911	-63.8089	0
ASPO6	-52.4	-50.6817	2.966479	-39.3018	-57.9129	0
MEHLER	-63.85	-60.9447	5.116151	-53.8906	-65.4477	0
GTHP	-74.3	-66.1154	5.80517	-56.156	-74.1678	0
CYTBDu	-86.9	-81.0894	7.893035	-66.9812	-90.0953	0
CYTBDpp	-86.9	-83.8146	7.893035	-69.7064	-92.8205	0
PYDXOi	-93.45	-92.263	4.845101	-80.5289	-105.595	0
PPOR	-116.74	-114.822	2.774887	-105.217	-126.345	0
RBCh	-123.5	-121.782	2.569047	-110.402	-129.013	0

TABLE D1b: THERMODYNAMIC INFORMATION OF METABOLITES.

Names	Gf	Uf
(1R,6R)-6-Hydroxy-2-succinylcyclohexa-2,4-diene-1-carboxylate	-193.5	3.3
(2R,3S)-3-Isopropylmalate	-199.5	1.6
(2S)-2-Isopropyl-3-oxosuccinate	-189.1	1.4
(R)-2,3-Dihydroxy-3-methylpentanoate	-159.1	1.4
(R)-2-Methylmalate	-209.7	0.29
(R)-4-Phosphopantothenoyl-L-cysteine	-426.5	4.4
(R)-Lactate	-123.1	0.7
(R)-Pantoate	-159.1	1.4
(R)-S-Lactoylglutathione	-291.4	5
(S)-1-Pyrroline-5-carboxylate	-60.3	2.4
(S)-2,5-Diaminopentanoate	-85.7	1.4
(S)-2-Aceto-2-hydroxybutanoate	-151.1	1.3
(S)-3-Methyl-2-oxopentanoic acid	-109.4	1.1
(S)-4-Amino-5-oxopentanoate	-111.4	1
(S)-Dihydroorotate	-151.5	4.2
(S)-Malate	-201	0.8
(S)-Malate	-201	0.8
(S)-Malate	-201	0.8
(S)-Propane-1,2-diol	-79.7	0.9
1-(2-Carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate	-391.6	3.3
1-(5-Phospho-D-ribosyl)-5-amino-4-imidazolecarboxylate	-374.1	6
1-(5-Phospho-D-ribosyl)-ATP	-1014.6	10.3
1-(5-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole	-486.9	5.5
1-(5-Phosphoribosyl)-5-amino-4-imidazolecarboxamide	-335.5	5.5
1-(5-Phosphoribosyl)-5-formamido-4-imidazolecarboxamide	-372	5.8
1,2-Diacyl-sn-glycerol (dihexadec-9-enoyl, n-C16_1)	-100	6.8
1,2-Diacyl-sn-glycerol (dihexadecanoyl, n-C16_0)	-144.7	7
1,2-Diacyl-sn-glycerol (dioctadec-11-enoyl, n-C18_1)	-93.5	7.5
1,2-Diacyl-sn-glycerol (dioctadec-9-enoyl, n-C18_1)	-93.5	7.5
1,2-Diacyl-sn-glycerol (dioctadecanoyl, n-C18_0)	-138.1	7.8
1,2-dihexadec-9-enoyl-sn-glycerol 3-phosphate	-311.9	6.8
1,2-dihexadecanoyl-sn-glycerol 3-phosphate	-356.6	7
1,2-dioctadec-11-enoyl-sn-glycerol 3-phosphate	-305.4	7.5
1,2-dioctadecanoyl-sn-glycerol 3-phosphate	-350	7.8
1,4-Dihydroxy-2-naphthoate	-122.8	6.3
10-Formyltetrahydrofolate	-144.3	9.3
1-deoxy-D-xylulose	-144	3.3
1-Deoxy-D-xylulose 5-phosphate	-356.2	1.3

1D-myo-Inositol 3-phosphate	-432.6	4
1-hexadec-9-enoyl-sn-glycerol 3-phosphate	-319.9	3.6
1-hexadecanoyl-sn-glycerol 3-phosphate	-342.2	3.8
1-Hydroxy-2-methyl-2-butenyl 4-diphosphate	-474.5	1.6
1-octadec-11-enoyl-sn-glycerol 3-phosphate	-316.6	4
1-octadecanoyl-sn-glycerol 3-phosphate	-338.9	4.2
2-(Formamido)-N1-(5-phosphoribosyl)acetamidine	-379.6	6.6
2,3,4,5-Tetrahydrodipicolinate	-139.4	3.5
2,3-Dihydroxy-3-methylbutanoate	-160.8	1.4
2,3-Dimethyl-5-phytylquinol	-6.9	2.1
2,3-Disphospho-D-glycerate	-585.8	1
2,5-Diamino-6-(5-phospho-D-ribosylamino)pyrimidin-4(3H)-one	-326.2	7.0
2-Acetolactate	-152.8	1.3
2-Amino-3-oxo-4-phosphonooxybutyrate	-361.8	0.4
2-Amino-4-hydroxy-6-(D-erythro-1,2,3-trihydroxypropyl)-7,8-dihydropteridine	-75.6	7.5
2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine	-3	7.4
2-Amino-7,8-dihydro-4-hydroxy-6-(diphosphooxymethyl)pteridine	-423.1	7.4
2-C-Methyl-D-erythritol 4-phosphate	-365.5	1.6
2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate	-467.9	1.9
2-Dehydropantoate	-148.7	1.4
2-Deoxy-D-ribose 5-phosphate	-352.9	3.3
2-Hydroxy-3-oxopropanoate	-149.8	0.8
2-Hydroxyphenylacetate	-1.1	1.5
2-Isopropylmaleate	-142.7	2.2
2-Methyl-4-amino-5-hydroxymethylpyrimidine diphosphate	-412.4	4.9
2-Methyl-6-phytylquinol	1.7	2.0
2-Methyl-6-solanyl-1,4-benzoquinol	198.6	4.8
2-Methylmaleate	-141.3	0.6
2-Octaprenylphenol	214.7	11.8
2-Oxo-3-hydroxy-4-phosphobutanoate	-400.3	0.34
2-Oxobutanoate	-110.9	0.9
2-Oxoglutarate	-188.8	1.1
2-Oxoglutarate	-188.8	1.1
2-Oxoglutarate	-188.8	1.1
2-Phospho-4-(cytidine 5-diphospho)-2-C-methyl-D-erythritol	-857	5.9
2-Phospho-D-glycerate	-371.1	0.8
2-Phosphoglycolate	-334.8	0.6
2-Phosphoglycolate	-334.8	0.6
2-Succinylbenzoate	-160.5	3
2-Succinylbenzoyl-CoA	-858.3	10.6
3-(4-Hydroxyphenyl)pyruvate	-114.9	2.6

3-(Imidazol-4-yl)-2-oxopropyl phosphate	-242.1	3.2
3,5-Cyclic AMP	-192.1	10.9
3,5-Cyclic GMP	-243.4	10.3
3-Dehydroquinate	-215.1	2.6
3-Dehydroshikimate	-166.8	2.6
3-Methyl-2-oxobutanoic acid	-111	1.1
3-Octaprenyl-4-hydroxybenzoate	132	11.9
3-Phosphoadenylyl sulfate	-580.1	8.9
3-Phospho-D-glycerate	-371.1	0.8
3-Phospho-D-glycerate	-371.1	0.8
3-Phospho-D-glyceroyl phosphate	-585.8	1
3-Phosphonooxypyruvate	-360.7	0.9
4-(Cytidine 5-diphospho)-2-C-methyl-D-erythritol	-645.1	5.9
4-Amino-4-deoxychorismate	-135.1	3.9
4-Amino-5-hydroxymethyl-2-methylpyrimidine	7.6	4.9
4-Aminobenzoate	-48.7	2.7
4-Aminobutanal	-31.7	1
4-Aminobutanoate	-84.5	0.9
4-Hydroxybenzoate	-92.2	2.6
4-Hydroxy-benzyl alcohol	-47.3	2.6
4-Methyl-2-oxopentanoate	-109.4	1.1
4-Methyl-5-(2-hydroxyethyl)-thiazole	-5.4	3.4
4-Methyl-5-(2-phosphoethyl)thiazole	-217.3	3.4
4-Phospho-L-aspartate	-380.5	1.1
5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)-imidazole-4-carboxamide	-649.7	12.2
5,10-Methenyltetrahydrofolate	-96.6	9
5,10-Methylenetetrahydrofolate	-93.5	8.9
5-Amino-6-(1-D-ribitylamino)uracil	-189.2	5.3
5-Amino-6-(5-phospho-D-ribitylamino)uracil	-401	5.3
5-Amino-6-(5-phosphoribosylamino)uracil	-393.5	6.2
5-Aminolevulinate	-112.8	1
5-Deoxyadenosine	-9.2	8.7
5-Deoxy-D-ribose	-144.8	3.7
5-Deoxy-D-ribose	-144.8	3.7
5-Methyltetrahydrofolate	-101.4	8.2
5-Methylthioadenosine	1.4	8.9
5-O-(1-Carboxyvinyl)-3-phosphoshikimate	-427.4	3.4
5-Phospho-alpha-D-ribose 1-diphosphate	-812.9	3.7
5-Phosphoribosylamine	-357.7	3.7
5-Phosphoribosylglycinamide	-388.5	4.1
5-Phosphoribosyl-N-formylglycinamide	-416.7	5.2

6,7-Dimethyl-8-(D-ribityl)lumazine	-137.7	7.6
6-Carboxyhexanoyl-CoA	-856	10.3
6-Phospho-D-gluconate	-480	2.2
7,8-Diaminononanoate	-80.9	1.8
7,8-Dihydroneopterin 3-triphosphate	-694.3	7.5
8-Amino-7-oxononanoate	-108	1.5
Acetaldehyde	-33.4	1.5
Acetate	-88.3	1.5
Acetate	-88.3	1.5
Acetate	-88.3	1.5
Acetoacetyl-CoA	-811.3	10.2
Acetyl phosphate	-301.4	0.8
Acetyl-CoA	-784.6	10.1
Adenine	78	7.8
Adenosine	-45.4	8.7
Adenosine 3,5-bisphosphate	-469.2	8.7
Adenylyl sulfate	-368.2	8.9
ADP	-465.4	8.7
ADP-glucose	-621.8	11.2
ADP-ribose	-585	10.8
Agmatine	6.6	4.5
alpha,alpha-Trehalose 6-phosphate	-587.4	8.4
alpha-D-Galactose 1-phosphate	-429.6	4.2
alpha-D-Glucosamine 1-phosphate	-394.5	4.2
alpha-D-Ribose 1-phosphate	-392.8	3.7
alpha-Isopropylmalate	-199.5	1.6
alpha-Ribazole	-77.4	5.9
Aminoimidazole ribotide	-296.9	4.9
AMP	-257.3	8.7
Anthranilate	-48.7	2.7
ATP	-673.5	8.7
beta-Alanine	-86.2	0.8
Biotin	-109.8	6.4
Cadaverine	-4.4	1.5
Carbamoyl phosphate	-300.6	1.6
CDP	-552.8	5.5
CDP-1,2-dihexadec-9-enoylglycerol	-591.5	8.9
CDP-1,2-dihexadecanoylglycerol	-636.2	9.1
CDP-1,2-dioctadec-11-enoylglycerol	-584.9	9.5
CDP-1,2-dioctadecanoylglycerol	-629.6	9.8
Chorismate	-170.3	3.9

Citrate	-280.3	1.4
Citrate	-280.3	1.4
Citrate	-280.3	1.4
CMP	-344.7	5.5
CO2	-92.3	1.5
CO2	-92.3	1.5
CO2	-92.3	1.5
CO2	-92.3	1.5
CoA	-750.9	9.8
CTP	-751.4	5.5
Cyanate	-23.3	15
Cyanate	-23.3	15
Cyanate	-23.3	15
Cys-Gly	-114.1	2.7
Cytidine	-132.8	5.5
Cytosine	-9.5	4
D-4-Phosphopantothenate	-400.2	2.4
dADP	-425.4	8.6
D-Alanine	-87.9	0.8
D-Alanyl-D-alanine	-118.9	2.2
dAMP	-217.3	8.6
dATP	-633.6	8.6
dCDP	-512.9	5.2
dCTP	-721	5.3
Deamino-NAD+	-571.6	12.1
Deoxyguanosine	-55.1	7.7
Dephospho-CoA	-539	9.9
D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate	-288	3.3
D-Erythrose 4-phosphate	-354.6	1.3
Dethiobiotin	-115.3	4.2
D-Fructose	-218.5	3.8
D-Fructose	-218.5	3.8
D-Fructose	-218.5	3.8
D-Fructose 1,6-bisphosphate	-642.3	3.7
D-Fructose 1-phosphate	-430.4	3.7
D-Fructose 6-phosphate	-430.4	3.7
D-Galactose	-217.7	4.3
dGDP	-475.2	7.7
D-Glucono-1,5-lactone	-216	1.2
D-Glucono-1,5-lactone 6-phosphate	-427.5	3.5
D-Glucosamine 6-phosphate	-394.5	4.2

D-Glucose	-217.7	4.3
D-Glucose	-217.7	4.3
D-Glucose	-217.7	4.3
D-Glucose 1-phosphate	-429.6	4.2
D-Glucose 6-phosphate	-429.6	4.2
D-Glutamate	-164.2	1
D-Glyceraldehyde	-106.5	1
D-Glyceraldehyde 3-phosphate	-318.3	0.9
D-Glycerate	-159.2	0.9
dGTP	-673.8	7.7
dIDP	-480.7	7.7
Dihydrofolate	-109.4	8.7
Dihydroneopterin phosphate	-287.5	7.5
Dihydropteroate	-2.2	8
Dimethylallyl diphosphate	-438.3	1.6
Diphosphate	-480.9	1.5
dITP	-688.8	7.7
di-trans,poly-cis-Undecaprenyl diphosphate	-154	15.8
di-trans,poly-cis-Undecaprenyl phosphate	54.2	15.8
D-Lactaldehyde	-70.3	0.8
D-Mannose 1-phosphate	-429.6	4.2
D-Mannose 6-phosphate	-429.6	4.2
D-Ribose	-181	3.8
D-Ribose 5-phosphate	-392.8	3.7
D-Ribulose 1,5-bisphosphate	-605.3	3.7
D-Ribulose 1,5-bisphosphate	-605.3	3.7
D-Ribulose 5-phosphate	-392.4	1.5
D-Tagatose 1,6-biphosphate	-642.3	3.7
D-Tagatose 6-phosphate	-430.4	3.7
dTDP	-564.8	5.1
dTDP-glucose	-721.2	8.4
dTDP-L-rhamnose	-685	8.3
dTMP	-356.7	5.1
dTTP	-772.9	5.1
dUDP	-564.7	5.1
dUMP	-356.6	5.1
dUTP	-772.9	5.2
D-Xylulose 5-phosphate	-392.4	1.5
Ethanol	-43.4	1.5
FAD	-528.5	15.4
FADH2	-536.5	14.6

Fe2+	-18.9	1.5
Fe2+	-18.9	1.5
Fe2+	-18.9	1.5
Fe2+	-1.1	1.5
Fe3+	-1.1	1.5
Fe3+	-1.1	1.5
Formate	-83.9	1.5
Fumarate	-143.7	2.1
Fumarate	-143.7	2.1
Fumarate	-143.7	2.1
gamma-L-Glutamyl-L-cysteine	-190.5	2.9
GDP	-515.1	7.9
GDP-4-dehydro-6-deoxy-D-mannose	-628.6	10.2
GDP-L-fucose	-635.3	10.5
GDP-mannose	-671.5	10.5
Geranyl diphosphate	-409.8	2.9
Glutathione	-221.3	4.4
Glutathione disulfide	-435.6	8.6
Glycerol	-115.9	1.2
Glycerone	-107.9	1
Glycerone phosphate	-319.8	0.9
Glycine	-87.8	0.7
Glycine	-87.8	0.7
Glycine	-87.8	0.7
Glycogen	-167.2	16.98741
Glycogen molecules	-167.2	17
Glycolaldehyde	-70.2	0.6
Glycolate	-122.9	0.5
Glyoxylate	-111	0.8
GMP	-307	7.9
GTP	-713.7	7.9
H+	-9.5	0
H+	-9.5	0
H+	-9.5	0
H+	-9.5	0
H+	-9.5	0
H2O	-56.7	1.5
H2O	-56.7	1.5
H2O	-56.7	1.5
H2O	-56.7	1.5
H2O	-56.7	1.5

H2O	-56.7	1.5
HCO3	-140.3	1.5
HCO3	-140.3	1.5
HCO3	-140.3	1.5
HCO3	-140.3	1.5
Hydrogen	4.2	1.5
Hydrogen	4.2	1.5
Hydrogen	4.2	1.5
Hydrogen peroxide	-32.1	1.5
Hydrogen sulfide	-6.7	1.5
Hydroxymethylbilane	-478.7	17.9
Hydroxypyruvate	-148.8	0.9
IDP	-520.6	7.8
Iminoaspartate	-172.8	4.9
IMP	-312.5	7.8
Indoleglycerol phosphate	-267.3	4.3
Isochorismate	-170.3	3.9
Isocitrate	-279	1.3
Isopentenyl diphosphate	-439.1	1.6
ITP	-728.8	7.9
L-1-Pyrroline-3-hydroxy-5-carboxylate	-106.6	2.4
L-2,3-Dihydrodipicolinate	-134.6	3.9
L-3,4-Dihydroxybutan-2-one 4-phosphate	-319.9	0.9
L-Alanine	-87.9	0.8
L-Alanine	-87.9	0.8
L-Alanine	-87.9	0.8
L-Arginine	-73	4.5
L-Arginine	-73	4.5
L-Arginine	-73	4.5
L-Asparagine	-123.2	1.7
L-Aspartate	-165.9	0.9
L-Aspartate 4-semialdehyde	-113.1	0.9
L-Citrulline	-124.1	2.9
L-Cystathionine	-154.8	2.6
L-Cysteine	-83.3	1.9
L-Glutamate	-164.2	1
L-Glutamate	-164.2	1
L-Glutamate	-164.2	1
L-Glutamate 5-semialdehyde	-111.4	1
L-Glutamine	-121.6	1.7
L-Glutamine	-121.6	1.7

L-Glutamine	-121.6	1.7
L-Glutamyl 5-phosphate	-378.9	1.1
L-Histidine	-46.4	3.2
L-Histidine	-46.4	3.2
L-Histidine	-46.4	3.2
L-Histidinol	0.9	4.4
L-Histidinol phosphate	-214.9	3.2
L-Homocysteine	-81.7	1.9
L-Homoserine	-122.5	0.9
Lipid A disaccharide	-724.3	13.9
Lipid X	-498.1	7
L-Isoleucine	-84.8	1.2
LL-2,6-Diaminoheptanedioate	-163.7	1.6
L-Lactaldehyde	-70.3	0.8
L-Leucine	-84.8	1.2
L-Leucine	-84.8	1.2
L-Leucine	-84.8	1.2
L-Lysine	-84.1	1.5
L-Lysine	-84.1	1.5
L-Lysine	-84.1	1.5
L-Methionine	-75.7	2.2
L-Phenylalanine	-50.8	2.4
L-Proline	-62.7	2.7
L-Proline	-62.7	2.7
L-Proline	-62.7	2.7
L-Serine	-124.1	0.9
L-Serine	-124.1	0.9
L-Serine	-124.1	0.9
L-Threonine	-124.2	1.1
L-Threonine O-3-phosphate	-336.1	1.2
L-Tryptophan	-25.7	4.2
L-Tyrosine	-90.2	2.6
L-Valine	-86.4	1.2
Malonyl-CoA	-862.5	10.1
Maltose	-375.5	8.5
Mercaptopyruvate	-108	1.9
meso-2,6-Diaminoheptanedioate	-163.7	1.6
Methylglyoxal	-59.8	0.9
myo-Inositol	-220.7	4.2
N-(5-Phospho-D-1-ribulosylformimino)-5-amino-1-(5-phospho-D-ribosyl)-4-imidazolecarboxamide	-649.3	10.7
N-(5-Phospho-D-ribosyl)anthranilate	-392.1	4.7

N-(L-Arginino)succinate	-224.5	5.2
N6-(1,2-Dicarboxyethyl)-AMP	-408.7	8.8
N-Acetyl-alpha-D-glucosamine 1-phosphate	-424.2	4.5
N-Acetyl-L-glutamate	-193.9	2.1
N-Acetyl-L-glutamate 5-phosphate	-408.6	2.1
N-Acetyl-L-glutamate 5-semialdehyde	-141.1	2.1
N-Acetylornithine	-115.5	2.1
NAD+	-529	12.3
NADH	-523.8	12.1
NADP+	-740.9	12.2
NADPH	-726.2	12.1
N-Carbamoyl-L-aspartate	-204.2	2.8
NH4+	-19	1.5
NH4+	-19	1.5
NH4+	-19	1.5
Nicotinamide	4.3	4.1
Nicotinate D-ribonucleotide	-379.4	5.4
Nitrate	-26.6	1.5
Nitrate	-26.6	1.5
Nitrate	-26.6	1.5
Nitrite	-8.9	1.5
Nitrous oxide	18.3	4.6
NMN	-336.8	5.5
N-Succinyl-2-L-amino-6-oxoheptanedioate	-294.4	2.6
N-Succinyl-LL-2,6-diaminoheptanedioate	-269.7	2.5
O-Acetyl-L-serine	-161.6	1.5
O-Phospho-4-hydroxy-L-threonine	-372.3	1.2
O-Phospho-L-homoserine	-334.4	1
O-Phospho-L-serine	-328.3	0.9
Orotate	-142.5	4.5
Orotidine 5-phosphate	-477.7	5.4
Orthophosphate	-262	1.5
Orthophosphate	-262	1.5
Orthophosphate	-262	1.5
Oxaloacetate	-190.5	1
oxidized FMN	-336.3	8.2
Oxygen	3.9	1.5
Oxygen	3.9	1.5
Oxygen	3.9	1.5
Oxygen	3.9	1.5
Oxygen	3.9	1.5

Pantetheine 4-phosphate	-346.8	4.4
Pantothenate	-188.3	2.4
Phenyl acetate	-49.6	2.3
Phenylacetyl-CoA	-747.5	10.3
Phenylpyruvate	-75.4	2.4
Phosphatidylglycerol (dihexadec-9-enoyl, n-C16_1)	-366.5	7.3
Phosphatidylglycerol (dihexadecanoyl, n-C16_0)	-411.2	7.6
Phosphatidylglycerol (dioctadecanoyl, n-C18_0)	-404.6	8.4
Phosphatidylglycerol (dioctadec-enoyl, n-C18_1)	-359.9	8.1
Phosphatidylglycerophosphate (dihexadec-9-enoyl, n-C16_1)	-578.4	7.3
Phosphatidylglycerophosphate (dihexadecanoyl, n-C16_0)	-623	7.6
Phosphatidylglycerophosphate (dioctadec-11-enoyl, n-C18_1)	-571.8	8
Phosphatidylglycerophosphate (dioctadecanoyl, n-C18_0)	-616.5	8.4
Phosphoenolpyruvate	-315.4	1.7
Phosphoribosyl-AMP	-598.3	10.3
Plastoquinol-9	167.2	6.1
Plastoquinol-9	167.2	6.1
Plastoquinone-9	182.4	7.6
Plastoquinone-9	182.4	7.6
Plastoquinone-9	182.4	7.6
Plastoquinone-9	182.4	7.6
Porphobilinogen	-118.3	3.9
Prephenate	-187.7	3.4
Putrescine	-6.1	1.4
Putrescine	-6.1	1.4
Putrescine	-6.1	1.4
Pyridoxal	-59.8	4.8
Pyridoxal phosphate	-271.7	4.8
Pyridoxamine	-32.6	4.8
Pyridoxamine 5-phosphate	-244.5	4.8
Pyridoxine	-67.7	4.8
Pyridoxine phosphate	-279.6	4.8
Pyruvate	-112.6	0.9
Pyruvate	-112.6	0.9
Pyruvate	-112.6	0.9
Quinolinate	-119.5	3.9
Reduced FMN	-344.4	8.7
Riboflavin	-124.4	8.2
S-Adenosyl-L-homocysteine	-76.1	9
S-Adenosyl-L-methionine	-66.5	10.6
S-Adenosylmethioninamine	13.2	10.6

Sedoheptulose 1,7-bisphosphate	-558.6	2.5
Sedoheptulose 7-phosphate	-465	2.5
Shikimate	-172	2.7
Shikimate 3-phosphate	-383.9	2.6
S-Methyl-5-thio-D-ribose 1-phosphate	-346.1	4.2
S-Methyl-5-thio-D-ribulose 1-phosphate	-345.6	2.5
S-Methyl-5-thio-D-ribulose 1-phosphate	-345.6	2.5
sn-Glycerol 3-phosphate	-327.8	1
Spermidine	5	3.1
Spermidine	5	3.1
Spermidine	5	3.1
Succinate	-163	0.7
Succinate	-163	0.7
Succinate	-163	0.7
Succinate semialdehyde	-110.2	0.7
Succinate semialdehyde-thiamin diphosphate anion	-462.3	7.3
Succinyl-CoA	-860.9	10.1
Sucrose	-376.3	7.8
Sucrose	-376.3	7.8
Sucrose	-376.3	7.8
Sucrose 6-phosphate	-588.2	7.7
Sulfate	-178	1.5
Sulfate	-178	1.5
Sulfate	-178	1.5
Sulfite	-126.2	1.5
Tetrahydrofolate	-113.9	8.7
Thiamin	54	7.1
Thiamin diphosphate	-366	7.2
Thiamin monophosphate	-157.9	7.2
Thymidine	-144.8	5.1
trans,trans-Farnesyl diphosphate	-381.4	4.4
Triphosphate	-699.8	1.5
Ubiquinol-8	124.7	13.2
Ubiquinone-8	146.2	13.1
UDP	-604.7	5.4
UDP-2,3-bis(3-hydroxytetradecanoyl)glucosamine	-829.5	10.7
UDP-3-O-(3-hydroxytetradecanoyl)-D-glucosamine	-781.6	9.3
UDP-3-O-(3-hydroxytetradecanoyl)-N-acetylglucosamine	-811.3	9.7
UDP-alpha-D-galactose	-761.1	8.8
UDP-glucose	-761.1	8.8
UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine	-799.1	9.4

UDP-N-acetyl-alpha-D-glucosamine	-755.7	9.2
UDP-N-acetyl-D-mannosamine	-212.4	4.6
UDP-N-acetylmuramate	-818.7	9.2
UDP-N-acetylmuramoyl-L-alanine	-849.7	10
UDP-N-acetylmuramoyl-L-alanyl-D-gamma-glutamyl-meso-2,6-diaminopimelate	-1063.5	12.5
UDP-N-acetylmuramoyl-L-alanyl-D-glutamate	-956.8	11.1
UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-carboxy-L-lysyl-D-alanyl-D-alanine	-1125.4	15.5
UMP	-396.5	5.3
Undecaprenyl-diphospho-N-acetylmuramoyl-(N-acetylglucosamine)-L-alanyl-D-glutamyl-meso-2,6-diaminopimeloyl-D-alanyl-D-alanine	-827	23.1
Undecaprenyl-diphospho-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimeloyl-D-alanyl-D-alanine	-674.6	20.8
Uracil	-61.3	4.4
Urea	-48.7	1.5
Urea	-48.7	1.5
Urea	-48.7	1.5
UTP	-812.8	5.4
Xanthosine 5-phosphate	-362.5	8

TABLE D2a: DOWNSTREAM GAPS EXISTING WITHIN iSynCJ816.

Metabolite ID	Metabolite Name
15gl[c]	D-Glucono-1,5-lactone
glcglyc[p]	2-(beta-D-Glucosyl)-sn-glycerol
2mm[c]	2-Methylmaleate
4a-hthptn[c]	4a-Hydroxytetrahydrobiopterin
pmcoa[c]	6-Carboxyhexanoyl-CoA
78dhptn[c]	7,8-Dihydrobiopterin
8aonn[c]	8-Amino-7-oxononanoate
adocbi[c]	Adenosyl cobinamide
ala-L[p]	L-Alanine
alatrna[c]	L-Alanyl-tRNA
arg-L[p]	L-Arginine
argtrna[c]	L-Arginyl-tRNA(Arg)
arsn[c]	Arsenite
arsni[c]	Arsenate ion
asntrna[c]	L-Asparaginyl-tRNA(Asn)
asptrna[c]	L-Aspartyl-tRNA(Asp)
btn[c]	Biotin
cn[c]	Cyanide ion
cynt[c]	Cyanate
cynt[p]	Cyanate
cystrna[c]	L-Cysteinyl-tRNA(Cys)
dhptn[c]	Dihydrobiopterin
dna5mtc[c]	DNA 5-methylcytosine
s[c]	Sulfur donor
dtbt[c]	Dethiobiotin
flutox[c]	Glutaredoxin disulfide
fru[p]	D-Fructose
g1p-B[c]	beta-D-Glucose 1-phosphate
glc-B[c]	beta-D-Glucose
gln-L[p]	L-Glutamine
glntrna[c]	Glutaminyl-tRNA
glu-L[p]	L-Glutamate
glutrd[c]	Glutaredoxin
gly[p]	Glycine
glytrna[c]	Glycyl-tRNA(Gly)
hcn[c]	Hydrogen cyanide
hco3[p]	HCO3
his-L[p]	L-Histidine
histrna[c]	L-Histidyl-tRNA(His)

photon[u]	Light
ind3ac[c]	Indole-3-acetate
lys-L[p]	L-Lysine
lystrna[c]	L-Lysyl-tRNA
maltp[c]	Maltopentose
n2o[c]	Nitrous oxide
nh4[p]	NH4+
no[c]	Nitric oxide
o2-[c]	O2.
p680[u]	Neutral reaction centre of the Photosystem II
p680p[u]	Positive charged reaction centre of the Photosystem II
p700[u]	Uncharged reaction centre of the Photosystem I
p700p[u]	Positive charged reaction centre of the Photosystem I
palmcoa[c]	Palmitoyl-CoA
pqn[u]	Semiplastoquinone radical loosely bound to the Photosystem II
pro-L[p]	L-Proline
protrna[c]	L-Prolyl-tRNA(Pro)
provitd3[c]	7-Dehydrocholesterol; Provitamin D3
ptrc[p]	Putrescine
qa[u]	Internal bound plastoquinone of the Photosystem II
qan[u]	Internal bound semiquinone radical of the Photosystem II
s0[u]	Starting state of a cluster of probably four manganese atoms
s1[u]	First oxidation state of the cluster of manganese atoms in the Photosystem II
s2[u]	Second oxidation state of the cluster of manganese atoms in the Photosystem II
s3[u]	Third oxidation state of the cluster of manganese atoms in the Photosystem II
s4[u]	Fourth oxidation state of the cluster of manganese atoms in the Photosystem II
spmd[p]	Spermidine
sucr[p]	Sucrose
thioc[c]	Thiocyanate
thios[c]	Thiosulfate
thptn[c]	Tetrahydrobiopterin
thrtrna[c]	L-Threonyl-tRNA(Thr)
trptrna[c]	L-Tryptophanyl-tRNA(Trp)
tyrtrna[c]	L-Tyrosyl-tRNA(Tyr)
valtrna[c]	L-Valyl-tRNA(Val)
mmtsa[c]	(S)-Methylmalonate semialdehyde
im4ac[c]	Imidazole-4-acetate
lald-L[c]	L-Lactaldehyde

12ppd-S[c]	(S)-Propane-1,2-diol
4aabutn[c]	4-Acetamidobutanoate
q8h2[c]	Ubiquinol-8
didp[c]	dIDP
ditp[c]	dITP
2oph[c]	2-Octaprenylphenol
tagdp-D[c]	D-Tagatose 1,6-biphosphate
pydx[c]	Pyridoxal
4hba[c]	4-Hydroxy-benzyl alcohol
uGgla[c]	UDP-N-acetylmuramoyl-L-alanyl-gamma-D-glutamyl-L-lysyl-D-alanyl-D-alanine
cpppg1[c]	Coproporphyrinogen I
34dhmald[c]	3,4-Dihydroxymandelaldehyde

TABLE D2b: ROOT GAPS EXISTING WITHIN iSynCJ816.

Metabolite ID	Metabolite Name
1p3h5c[c]	L-1-Pyrroline-3-hydroxy-5-carboxylate
4gudbd[c]	4-Guanidinobutanamide
appl[c]	D-1-Aminopropan-2-ol
asptrna(asn)[c]	L-Aspartyl-tRNA(Asn)
dann[c]	7,8-Diaminononanoate
dnacyt[c]	DNA cytosine
glutrna(gln)[c]	L-Glutamyl-tRNA(Gln)
h2o[u]	H2O
iad[c]	(Indol-3-yl)acetamide
ian[c]	3-Indoleacetonitrile
malt[c]	Maltose
malth[c]	Maltohexaose
trnaala[c]	tRNA(Ala)
trnaarg[c]	tRNA(Arg)
trnaasn[c]	tRNA(Asn)
trnaasp[c]	tRNA(Asp)
trnacys[c]	tRNA(Cys)
trnagly[c]	tRNA(Gly)
trnahis[c]	tRNA(His)
trnalys[c]	tRNA(Lys)
trnapro[c]	tRNA(Pro)
trnathr[c]	tRNA(Thr)
trnatrp[c]	tRNA(Trp)
trnatyr[c]	tRNA(Tyr)
trnaval[c]	tRNA(Val)

4abutn[c]	4-Aminobutanal
id3acald[c]	Indole-3-acetaldehyde
bamppald[c]	beta-Aminopropion aldehyde
cyst-L[c]	L-Cystathionine
drib[c]	Deoxyribose
gal[c]	D-Galactose
3hmp[c]	3-Hydroxy-2-methylpropanoate
im4act[c]	Imidazole-4-acetaldehyde
n4abutn[c]	N4-Acetylaminobutanal
q8[c]	Ubiquinone-8
3ophb[c]	3-Octaprenyl-4-hydroxybenzoate
tag6p-D[c]	D-Tagatose 6-phosphate
pyam5p[c]	Pyridoxamine 5-phosphate
pydxn[c]	Pyridoxine
pydam[c]	Pyridoxamine
dxyl[c]	1-deoxy-D-xylulose
thymd[c]	Thymidine
thm[c]	Thiamin
uGgl[c]	UDP-N-acetylmuramoyl-L-alanyl-gamma-D-glutamyl-L-lysine
uppg1[c]	Uroporphyrinogen I
R-actn[c]	Acetoin
23btdl[c]	Butane-2,3-diol

TABLE D3: FLUX VARIABILITY ANALYSIS OF REACTIONS WHICH ARE PART OF ELECTRON TRANSFER MACHINERY. EXCEL FILE CONTAINS FLUX VARIABILITY ANALYSIS OF PHOTOSYNTHETIC AND OXIDATIVE PHOSPHORYLATION MACHINERY WHICH IS INVOLVED IN ELECTRON TRANSFER.

	Opt. Light		< Opt. Light (30)		> Opt. Light (70)	
	min. flux	max. flux	min. flux	max. flux	min. flux	max. flux
H2ASE_syn	0	0.000221	0.000131	0.000131	0	0
MEHLER	0.558471	0.558692	0.558601	0.558601	0.558471	0.558471
NDH1_1u	0	1.691836	0.481337	0.481337	0	0
NDH1_3u	0.600007	0.725506	0.387897	0.387897	0.620627	0.620627
NDH2_syn	0	0	0	0	0	0
CYTBDu	0	0.000277	0.000164	0.000164	0	0
CYO1b2_syn	0.558471	2.250307	1.039807	1.039807	0.558471	0.558471
FQRa	0	0.000554	0.000327	0.000327	0	0
SUCDu_syn	0.026667	0.026747	0.014983	0.014983	0.027535	0.027535
SOR	0.132159	0.132528	0.132377	0.132377	0.132159	0.132159
PSI	24.60127	27.98494	15.04848	15.04848	25.36764	25.36764
PSII	24.72913	28.1128	15.76466	15.76466	25.45255	25.45255
FNOR	9.136994	10.86917	5.77488	5.77488	0	0

TABLE D4: COMPARISON OF SIMULATION OF AUTOTROPHIC FLUX VARIABILITY WITH EXPERIMENTALLY DETERMINED FLUXES.

Excel file contains comparison of autotrophic flux distribution between experimental study (Young et al., 2011) and our simulations.

Reactions	Exp. Value	Flux_HCO3	LB_HCO3	UB_HCO3
PGI	-0.703	-0.0336	-0.0338	-0.03363
G6PDH2	0.592	0	0	0.000158
FBP	2.22	0	0	2.931658
FBA	-2.22	0	-2.9317	0.000475
TPI	-3.515	-2.9462	-2.9464	-2.94621
GAPDi(nadp)	8.436	6.9043	6.90424	6.90471
PGM	-0.8584	-0.54778	-0.5482	-0.54761
ENO	0.8732	0.54778	0.54761	0.548245
PYK	0.3515	0.235376	0	0.235851
RPE	-2.8083	-2.4453	-2.4457	-2.44524
RPI	1.3172	1.359	1.35895	1.359006
PRUK	4.699	3.8041	3.80408	3.80455
RBPC	4.699	3.69	3.68993	3.690404
TKT2	-1.4245	-1.0176	-1.4614	-1.01762
TKT1	-1.3801	-1.4276	-1.4277	-1.42762
TALA	-0.037	1.5036	-1.4281	1.504036
FBA3	-1.332	-2.9312	-2.9317	0.000475
SBP	1.332	2.9312	0	2.931658
PDH	0.4366	0	0	0.000237
POR_syn	0.4366	0	0	0.000475
CS	0.1184	0.10678	0.10678	0.106837
ACONTa	0.1184	0.10678	0.10678	0.106837
ICDHy	0.111	0.10678	0.10678	0.106837
SUCDu_syn	0.0074	0.026687	0.02667	0.026747
FUM	0.0074	0.072571	0.07255	0.072807
MDH	-0.1332	0.072571	-0.0529	0.072807
ME1	0.1961	0	0	0.125494
PPC1	0.4292	0.24286	0.24286	0.368355
RBCh	0.0148	0.11415	0.11415	0.114205
PGLYCP	0.0148	0.11415	0.11415	0.114205
GLYCTO_syn	0.0148	0.11422	0.11381	0.114282

TABLE D5: COMPARISON OF SIMULATION OF HETEROTROPHIC FLUX VARIABILITY WITH EXPERIMENTALLY DETERMINED FLUXES.

Excel file contains comparison of heterotrophic flux distribution between experimental study (Yang et al., 2002) and our simulations.

Reactions	Minimum Flux	Maximum Flux	Sim. Flux (Gluc = 100)	MFA (Gluc = 100)
HEX1	0.85	0.85	100	100
PGI	-1.280790416	0.83824948	98.61756772	5.7
PFK	0	0.083454645	0	58.9
FBA	-0.000285427	0.083454645	0	58.9
GAPDi	0.817690408	0.000332998	96.43519887	142.2
PGK	-0.819910395	-0.817690408	96.43519887	142.2
PGM	-0.745401297	-0.743221673	87.67405587	142.2
ENO	0.743221673	0.745401297	87.67405587	142.2
PYK	0	0.510692196	0	72.7
PYK2	0	0.510692196	0	72.7
PYK5	0	0.510692196	60.04864599	72.7
POR_syn	0.331486752	0.333666375	39.23401544	117.5
PDH	0	0.000332998	0	117.5
CS	0.080259015	0.08044065	9.442361425	42.5
ACONTa	0.080259015	0.08044065	9.442361425	42.5
ME1	0	0.001997988	0	68.5
PPC1	0.182547662	0.18454565	21.47647834	55.7
FUM	0.054525724	0.054729195	6.417444522	33.2
G6PDH2	0	2.119039896	0	90.2
PGL	2.116497002	2.119039896	249.0591325	90.2
GND	2.116497002	2.119039896	249.0591325	90.2
RPI	-0.748501621	-0.748319986	88.05754014	33.7
RPE	1.368090851	1.37045211	160.9914552	56.5
TKT2	0.671392072	0.673571695	79.0070781	26.4
TKT1	0.69669878	0.696880415	81.98437708	30.1
TALA	0.613347927	0.697165842	72.19831752	30.1