UNIVERSITY OF CALIFORNIA, SAN DIEGO

Genome-Scale Reconstruction of the *Chlorella vulgaris* UTEX 395 Metabolic Network

A thesis submitted in partial satisfaction of the
requirements for the degree Master of Science

in

Bioengineering

by

Tyler Paul Huelsman

Committee in charge:

Professor Nathan Lewis, Chair
Professor Xiaohua Huang
Professor Christian Metallo

2015

The thesis of Tyler Paul Huelsman is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

Chair

University of California, San Diego

2015

DEDICATION

I dedicate this thesis to my parents, Michael and Rebecca, and to my brothers, Chad and Derek, for their love and guidance throughout the years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Palsson for giving me the opportunity to join his research lab. I would also like to thank my advisors, Dr. Zengler and Dr. Lewis, for their guidance in my research and in the development of my thesis. Their support was instrumental in the success of this thesis.

Next, I would like to thanks to Dr. Levering for teaching me the fundamentals of metabolic network reconstruction and for her frequent guidance throughout the project. Additionally, I would like to thank Dr. Zuniga for being a wonderful collaborator on this project, and for working with me to complete the reconstruction and validate the model.

I would also like to express my appreciation for Alex Thomas and Mona Ar for taking the time to provide me with assistance in the sampling of my data.

Lastly, I would like to thank all of my friends and family, who have supported me throughout the pursuit of my graduate degree.

ABSTRACT OF THE THESIS

Genome-Scale Reconstruction of the *Chlorella vulgaris* UTEX 395 Metabolic Network

by

Tyler Paul Huelsman

Master of Science in Bioengineering

University of California, San Diego, 2015

Professor Nathan Lewis, Chair

Genome-scale metabolic network reconstructions are organized knowledge bases consisting of the genomic information and metabolic pathways of a species. They provide a library of gene-protein-reaction relationships and a mathematical means for metabolic analysis of the organism. *Chlorella vulgaris* is a species of photosynthetic, eukaryotic microalgae that has received keen interest as a potential feed source for the manufacture of biofuels. Using its genome annotation, and implementing a homology-based reconstruction strategy, the metabolic network of *C. vulgaris* UTEX 395 was

reconstructed. The reconstruction was then formatted into a mathematical model to emulate the metabolism of *C. vulgaris* in photoautotrophic, mixotrophic, and heterotrophic growth conditions. The flux distributions of the reactions in each growth condition were then compared to identify key reactions in central carbon metabolism and lipid metabolism. The reconstruction network provides insight into the potential of *C. vulgaris* for metabolic engineering and represents a promising resource for the study of the metabolism of photosynthetic and algal organisms.

I:

Introduction

There is considerable interest in the use of microalgae as a feedstock for the production of biofuels due to their fast growth, high lipid yield, and low space requirement (Li et al. 2008; Savage 2011; Shi et al. 2012). Algae are comprised of a large group of photosynthetic eukaryotic organisms that display a wide variety of characteristics and come in diverse forms. They can be unicellular microalgae, as is the case with *Chlorella vulgaris*, or multicellular macroalgae, as in giant kelp. Unicellular microalgae are the most interesting as a potential biofuel feedstock due to their fast growth rate, photosynthetic efficiency, and high oil accumulation (Savage 2011; Shi et al. 2012). Sometimes referred to as "oleaginous" algae, microalgae species including *Chlamydomonas reinhardtii* and *C. vulgaris* are promising because of their ability to produce large amounts of neutral lipids (20-50% dry cell weight), their ability to grow at rates as high as one to three doublings per day, their ability to grow in lands unsuitable for agriculture of conventional crops, their ability to use nitrogen and phosphorous nutrients from wastewater for bio-remediation, and their ability to sequester CO_2 from point sources such as coal power plants to reduce carbon emissions (Chisti 2007; Hu et al. 2008; Shi et al. 2012).

Concern over the future availability of fossil fuels and the environmental impact of procuring, processing, and burning these fuels has become an impetus for seeking alternative energy sources (Bentley and Boyle 2008). Biofuels, such as biodiesel, bioethanol, and biohydrogen represent a renewable, sustainable, and carbon-neutral alternative to non-renewable liquid fuels such as petroleum; they can be used for transportation, power generation or heat. However, their greatest potential lies in the

transportation sector for which petroleum encompassed 90% of all energy used in the United States in 2012 (U.S. Department of Energy 2012).

Currently, oil is expected to maintain its dominance as the primary energy source for transportation worldwide for at least the next few decades; consumption of oil is predicted to increase from 87 million barrels per day in 2010 to 119 million barrels per day in 2040 (U.S. Energy Information Administration 2014). Biofuels are unlikely to replace oil as the dominant worldwide transportation energy source soon, but they do likely represent a renewable energy source that will displace fossil fuel consumption in the near future. Current projections predict consumption more than tripling from 1.3 million barrels of oil equivalent per day in 2011 to 4.6 million barrels of oil equivalent a day in 2040 (International Energy Agency 2014). At this consumption level, biofuels would occupy 8% of total transportation fuel demand in 2040, up from about 3% in 2011. Improving the efficiency of biofuel production by improving existing technology and using more effective feedstocks may make biofuels an even more appealing option and trigger a much faster expansion of the biofuels industry.

Biofuels have the advantage over fossil fuels of being renewable and carbon-neutral. Extraction and consumption of fossil fuels effectively means removing hydrocarbons from the earth and releasing them into the atmosphere in the form of carbon dioxide (CO_2) and other gases. In the atmosphere, CO_2 acts as a greenhouse gas, and is a major contributor to rising global average surface temperatures. Between the onset of the industrial revolution and 2013, global atmospheric CO_2 levels have risen from an annual average of about 280 ppm to about 396 ppm, an increase of 41% (U.S. Environmental Protection Agency 2014).

In contrast to fossil fuels, the production of biofuels operates within the natural global carbon cycle, fixing atmospheric CO_2 and converting it into lipids and other biofuel precursors. Biofuels are produced by photosynthetic organisms that can use CO_2 as their sole carbon source. These photosynthetic organisms, called photoautotrophs, convert CO_2 into organic compounds for growth and for biosynthesis of fatty acids used in biofuel production. Burning biofuels still releases CO_2 into the atmosphere, but it can be no more than the amount of CO_2 captured from the atmosphere or sequestered from point sources during the synthesis of the biofuel, meaning either atmospheric carbon dioxide levels are reduced, or else no net change in atmospheric CO_2 results from the consumption of biofuels.

Factors that may offset the theoretical carbon-neutrality of biofuels include carbon dioxide emissions from change of land use and carbon dioxide emissions from processing and transportation of biofuels (Wise et al. 2014). Change of land use is problematic for current biofuel feedstocks, such as ethanol from sugarcane and corn and biodiesel from palm, canola, and soybeans. Far less land would be required to garner an equivalent amount of fuel from algae, since it is estimated that algae could produce as much as 61,000 liters per hectare of fuel compared to 200 to 450 liters per hectare for soy and canola biodiesel (Savage 2011) and 5,900 and 3,100 liters per hectare for sugarcane and corn ethanol (Goldemberg 2008; Budny 2007). The result of switching biofuel paradigms to algae from current feedstocks would be a sharp reduction in carbon emissions from change of land use and a massive increase in fuel-per-land production efficiency.

The idea of using microalgae as a feedstock for biofuel is not new. The United States Department of Energy's Office of Fuels Development funded an extensive program from 1978 to 1996 with the goal of developing renewable transportation fuels from algae (Sheehan et al. 1998). One of the major focuses of this project was the Aquatic Species Program, which produced biodiesel from high lipid-content algae grown in open ponds with waste CO_2 from coal power plants. The Aquatic Species Program ultimately concluded that the most optimistic prediction of this technology is that it could produce biodiesel at a price two times that of diesel at the time. Sheehan et al. also concluded that major limitations preventing algae biodiesel from being cost effective are less engineering-related and more biological; there is a need for a highly productive algae species that can efficiently use sunlight for biomass production. However, the project also concluded that land requirement would not be an issue, as only 200,000 hectares of land, less than 0.1% of climatically suitable land in the United States, would be needed to produce a "quad" (quadrillion BTU). One quad is approximately 1% of the total energy consumption of the United States. A more recent estimate suggests that only 1-3% of the total cropping area in the United States would be necessary for microalgae to provide 50% of the country's transportation fuel (Chisti 2007).

Although the Aquatic Species Program was unable to prove the economic feasibility of producing biofuels from algae, the private sector has since pursued the technology and has achieved promising results. One group was able to cultivate the microalgae *Haematococcus pluvialis* in a coupled photobioreactor-open pond system that required two hectares of land to produce an annual average rate of oil production

of 420 GJ hectare^{-1} year^{-1}, which exceeded all estimations for any land crop to biodiesel system (Huntley and Redalje 2006). On top of this, they achieved a maximum oil production rate of 1014 GJ hectare^{-1} year^{-1}, and claimed that a rate of 3200 hectare^{-1} yr^{-1} is possible if a high photosynthetic efficiency organism such as Chlorella is used instead of *H. pluvialis*. *C. vulgaris* has been demonstrated to have a photosynthetic efficiency of 20% in large-scale cultures exposed to natural sunlight (Huntley and Redalje 2006) At this efficiency, they claim biodiesel would be cost-competitive with regular diesel.

Although *C. vulgaris* is considered one of the most suitable organisms for sustainable large-scale production of biodiesel, the process is still not considered by most to be commercially viable. There are two main strategies for addressing this issue; one is to improve the rate of biomass production of the organism and the other is to improve the rate of lipid biosynthesis within the available biomass production (Anthony et al. 2015). Much of the problem, however, lies in the fact that these two strategies seem to be mutually exclusive; one strategy cannot be employed without inhibiting the other (Sheehan et al. 1998). The simplest way to increase oil content in microalgae is through nutrient starvation, but this sharply decreases overall biomass production, and the net result of nutrient starvation is a lower yield of oil production (Sheehan et al. 1998; Chisti 2007; Hu et al. 2008). What is needed to move forward is to overcome the paradoxical nature of the algal species and to discover a method of cultivating a fast-growing and high oil-content organism.

Genome-scale metabolic reconstruction has been useful in guiding metabolic engineering and hypothesis-driven exploration in a variety of organisms including

Escherichia coli, *Saccharomyces cerevisiae*, *Mus musculus*, and even *Homo sapiens* (Feist and Palsson 2008; Oberhardt, Palsson, and Papin 2009). To date, only a few genome-scale metabolic reconstructions of photosynthetic organisms have been created, including *Synechocystis sp.* (Montagud et al. 2010), *Arabidopsis thaliana* (de Oliveira Dal'Molin et al. 2010), *Zea mays* (Saha, Suthers, and Maranas 2011), *Brassica napus* (Hay et al. 2014), and *C. reinhardtii* (Chang et al. 2011). Of these, *C. reinhardtii* is the only algae with a genome-scale metabolic reconstruction.

A metabolic reconstruction provides organized genomic, biochemical, and metabolic information on an organism that can be readily implemented into a functional metabolic model (Thiele and Palsson 2010). A high-quality reconstruction includes a network of reactions with proper stoichiometry, balanced mass and charge, gene-protein-reaction (GPR) associations, and confidence scores which indicate the amount of available evidence for the reaction. With this level of information, reconstructions can be implemented for use in constraint-based analyses, specifically using the methodological framework outlined by COnstraint-Based Reconstruction and Analysis (COBRA) (Schellenberger et al. 2011). Using this framework, constraints based on mass conservation, thermodynamics, and experimental data can be input into a reconstruction to model organism behavior based on specified growth conditions, genetic modifications (knockouts, knockdowns, upregulations), and various other metabolic scenarios.

In this thesis I report the creation of *i*CZ842, a genome-scale metabolic reconstruction of *Chlorella vulgaris* UTEX 395 based on homology with *Chlamydomonas reinhardtii*. The nearness of *C. reinhardtii* to *C. vulgaris* in phylogeny

made it an ideal candidate for a homology-based reconstruction. I created a draft reconstruction for *C. vulgaris* by computationally comparing the organism's genes to the metabolic network reconstruction of *C. reinhardtii*, *i*RC1080 (Chang et al. 2011), and continued to use *i*RC1080 as a reference network for the duration of the manual curation phase of the project. I then proceeded to compile the results of published computational algorithms for the determination of subcellular localization of proteins. The consolidation of these algorithms allowed for the deduction of the compartmentalization of metabolic pathways, which is a key feature as well as a key challenge of eukaryotic network reconstructions. I then commenced a pathway-by-pathway, reaction-by-reaction manual curation of the draft reconstruction. This involved identifying and characterizing every reaction in carbohydrate metabolism, energy metabolism, lipid metabolism, glycan metabolism, metabolism of cofactors and vitamins, and the metabolism of secondary metabolites. This was done by consulting literature and systematically identifying GPR associations using bioinformatics tools such as NCBI BLAST (Altschul et al. 1990). Amino acid metabolism and nucleotide metabolism pathway reconstruction were completed by another member of the lab. Following manual curation, I ensured that the network reactions were charge and mass-balanced and further network validation tests including identification of metabolic dead-ends and gap-filling were performed using COBRA Toolbox 2.0 by my collaborator (Schellenberger et al. 2011). I then used the completed model to simulate photoautotrophic, mixotrophic, and heterotrophic growth conditions in *C. vulgaris*.

II:

Materials and Methods

The Chlorella model reconstruction was created with assistance from the online tools and resources listed in Table M1.

Draft Reconstruction

The generation of a draft reconstruction for *Chlorella vulgaris* UTEX 395 was performed computationally by running a bidirectional BLAST (Altschul et al. 1990) with the manually curated reference network of the microalgae *Chlamydomonas reinhardtii*, *i*RC1080 (Chang et al. 2011). The selection of *C. reinhardtii* is based on the nearness in phylogeny of the photosynthetic organism based on the homology of multiple sequence alignments of ribulose-1,5-bisphosphate carboxylase (RuBisCO, EC 4.1.1.39) (Figure M1).

The initial draft reconstruction of the model was generated automatically using RAVEN (Reconstruction, Analysis and Visualization of Metabolic Networks) Toolbox (Agren et al., 2013). The RAVEN Toolbox incorporates subcellular localization of reactions, making it useful for reconstructions of eukaryotic organisms. RAVEN is also capable of interpreting input reference network reconstructions and genome sequences in order to automatically add reactions and genes from pre-existing models of similar organisms based on protein homology. It is also able to ascertain the KEGG Orthology ID of the genes of the subject organism based on the closest matches.

Using the predicted *C. vulgaris* gene models (Guarnieri et al. 2013) and the reference network iRC1080, a draft reconstruction based on homology with *C. reinhardtii* was obtained. The resulting homology reconstruction accounts for 621 genes associated with 1,108 reactions and 1,249 metabolites distributed across six

compartments, namely cytoplasm, extracellular space, chloroplast, mitochondria, thylakoid and glyoxysome. This draft reconstruction forms the basis for the final reconstruction.

A second draft reconstruction was also generated based on KEGG orthology. RAVEN Toolbox's KEGG Orthology ID capability matches each gene of the organism of interest with a predicted homologous KEGG gene. A draft reconstruction based on KEGG was created, which accounts for 470 genes associated with 804 reactions and 1,048 metabolites within the cytoplasm. Thus, the reconstruction based on KEGG orthology lacks compartmentalization. This network was utilized as an additional resource for the manual curation step, but was not the foundation for the final reconstruction. A reconstruction based on KEGG allowed for easy comparison between different resources, such as primary literature, KEGG and SwissProt.

Subcellular Localization and Compartments

Various algorithms were implemented in tandem to help determine the subcellular localization of the proteins. The outputs of the following algorithms were combined in a single spreadsheet and considered in conjunction with information from literature references and metabolic context from surrounding reactions to determine the location of each reaction:

ChloroP was used to predict chloroplastic proteins; it determines the presence of chloroplast transit peptides (cTP) in amino acid sequences and predicts the location of cTP cleavage sites (Emanuelsson, Nielsen, and Heijne 1999).

TargetP was used to predict the subcellular location of eukaryotic proteins as either mitochondrial, chloroplastic, signaling, or other; it assigns a location to an amino acid sequence based on the predicted presence of one of the N-terminal presequences including chloroplast transit peptides (cTP), mitochondrial targeting peptides (mTP) or secretory pathway signal peptides (SP) (Emanuelsson et al. 2000).

SignalP was used to discriminate between signal peptides and transmembrane proteins; it outputs a signal peptide/non-signal peptide prediction and a signal peptide cleavage site location prediction in amino acid sequences from three different types of organisms (eukaryotes, Gram-positive prokaryotes, and Gram-negative prokaryotes) using artificial neural networks (Petersen et al. 2011).

WoLF PSORT was used to predict subcellular localization based on similar protein sequences with known localizations (Horton et al. 2007). It converts amino acid sequences into numerical features and employs a k-nearest neighbor algorithm to determine similarity between the features of the input sequence and a dataset of either plant, fungi, or animal proteins from UniProt and other resources. WoLF PSORT classifies amino acid sequences as being located in either the nucleus, mitochondria, cytosol, plasma membrane, extracellular matrix, or chloroplast, and also includes potential dual-localizations for proteins that shuttle between compartments.

CELLO II was used to predict subcellular localization for prokaryotic or eukaryotic proteins based on sequence homology and coding schemes involving amino acid physicochemical features; it outputs localization predictions for amino acid sequences based on composition, dipeptide composition, partitioned composition, and chemo-typy, and it also outputs an aggregate prediction that incorporates all

aforementioned methods (Yu et al. 2006). CELLO II's possible subcellular location predictions for eukaryotes include chloroplast, cytoplasmic, cytoskeleton, endoplasmic reticulum, extracellular, Golgi apparatus, lysosomal, mitochondrial, nuclear, peroxisomal, and plasma membrane.

HECTARSEC is the general eukaryotic version of HECTAR, which specializes in predicting subcellular targeting in heterekonts; it outputs predictions for amino acid sequences as either containing a signal peptide, type II signal anchor, cTP, mTP or a protein that doesn't contain a N-terminal target peptide (Gschloessl, Guermeur, and Cock 2008).

Compartment pH and Protonation States

In order to determine the protonation states of metabolites in their respective compartments, apposite pH values were assigned to each compartment based on literature evidence. When extracellular pH was 6.5, the pH of the cytosol was determined to be 7.2 (Komor and Tanner 1974; Kusel et al. 1990). The pH of the chloroplast was shown to be 8.0 in light conditions (Goss and Garab 2001; Hogetsu and Miyachi 1979). The thylakoid was also assigned a pH of 8.0 because it is a sub-compartment of the chloroplast. The matrix of the mitochondria was found to have a pH of 7.8 in green algae (Giordano et al. 2003; Parisi et al. 2004). A pH of 7.0±0.2 was assigned to the extracellular matrix based on the pH of Bold's minimal growth medium. The glyosxysome was assigned a pH of 8.2 (Dansen et al. 2001).

Manual Curation

The manual curation phase of the reconstruction was performed by adapting a standardized high-quality metabolic reconstruction protocol to the requirements of our organism (Thiele and Palsson 2010). Metabolic pathways (e.g. pentose phosphate pathway) were added to the reconstruction one-by-one, using the primary pathways of the *C. reinhardtii* reconstruction as a reference initially, but also including additional pathways suggested by literature and gene presence to be present in *C. vulgaris*.

To begin the manual curation process for each individual pathway, metabolic pathway databases including KEGG (Kanehisa and Goto 2000) and BioCyc, and MetaCyc (Caspi et al. 2014) were consulted so that a general layout of the pathway could be readily perceived.

Each pathway was manually curated using available literature evidence from *C. vulgaris* and related species to determine the presence of particular enzymes and associated reactions, reaction directionality, cofactors involved in particular reactions and the mass and charge balance of the reactions. Protonation states of all compounds based on compartment-specific pH were accounted for when assigning charges.

For every reaction in each pathway, the gene of a phylogenetically close organism or genes from multiple organisms (*Chlorella variabilis, Coccomyxa subellipsoidea, Chlamydomonas reinhartii, Volvox carteri f. nagariensis, etc.)* was/were selected to be compared with *C. vulgaris* genes in order to identify gene-protein-reaction (GPR) associations. To perform this comparison and find the appropriate *C. vulgaris* gene, a local protein-protein BLAST (Altschul et al. 1990) of each characterized protein sequence selected from KEGG was performed against an

assembled database of the protein sequences of *C. vulgaris* to identify the individual protein(s) most likely responsible for catalyzing the given reaction. The resulting *C. vulgaris* protein sequences with expectation values (E values) smaller than 10^{-50}, or else the protein with smallest E value, was/were presumed to be potential GPR associations for the reaction. To test these assumptions, each of the proposed protein sequences were then entered into a global protein-protein BLAST against NCBI's non-redundant (nr) protein sequence database using NCBI's blastp suite (Altschul et al. 1990). The *C. vulgaris* protein sequence or sequences that resulted in high-identity (40+%) matches with protein sequences of the expected gene association using this method was/were then assigned as the correct GPR association for the given reaction. Reactions with multiple protein sequences in their GPR association required a Boolean notation to indicate whether or not each protein sequence represented a subunit of a complex of proteins necessary for a reaction to take place ("and") or an isozyme that also catalyzes the reaction ("or"). The absence of a high-identity protein sequence match for a GPR association for any given reaction resulted in low confidence for the existence of that reaction, and thus, that reaction would be left out of the model unless literature sources provided evidence for its existence, or else if the reaction was determined to be necessary to fill a gap within the network.

Every reaction was characterized with a reaction name, reaction stoichiometry and directionality (reversible or irreversible), a gene-reaction rule providing organized GPR association information, an EC Number, a KEGG ID, literature references if available, a confidence score based on the literature and genetic evidence for the reaction, and reaction notes describing the unique and/or features of the reaction.

Similarly, every metabolite was characterized with a metabolite name, chemical formula, charge, compartmentalization, KEGG ID and InChI String if available. Once all the aforementioned reaction and metabolite information for the reactions was retrieved, it was added to the network using scripts from the COBRA toolbox (Schellenberger et al. 2011).

The localization of each pathway in the network model was followed by assignment of transporters needed for functional conversion of pathway intermediates. Literature evidence and publicly available databases (BRENDA) were used to assign family and stoichiometry of transporters. In the absence of other evidence, necessary transporters were inferred from other organisms or else assumed to take the form of passive diffusion.

Network Evaluation

Following thorough manual curation of each pathway, the network was evaluated using the COBRA toolbox (Schellenberger et al. 2011). Charge and mass balance for every reaction was checked. Additionally, gap-filling analysis was performed to account for reaction gaps in the network and dead-ends in conversion of included intermediates and cofactors.

The initial and final reactants and products of each pathway were investigated to identify potential dead-ends, and additional metabolic or transport reactions were incorporated as appropriate. In addition, manual quality control tests on ATPm, NADPH and NADH were done, ending in the elimination of free energy loops.

Modeling-based gap-filling was also performed in the framework of flux balance analysis, with the addition of reactions determined *in silico* to be needed for growth.

Biomass Objective Functions

Protein, nucleotide, lipid, carbohydrate and ribose in RNA profiles were measured experimentally under both photoautotrophic and heterotrophic conditions using the method outlined by Long and Antoniewicz 2014. *C. vulgaris* UTEX 395 cells were grown in a 250 mL bottle containing 200 mL of Bold's Basal medium at a temperature of 24℃ (Guarnieri et al. 2013). A 14:10 hour light:dark cycle at 10,000 lux and a 1 percent CO_2 flow of (12 mL min^{-1}) were used. The tests for heterotrophic growth took place in identical conditions except instead using 10 g L^{-1} of glucose as an organic carbon source in 24 hours of dark. The protein profiles of *C. vulgaris* were completed by pulling from data in literature. Arginine, cysteine, glutamine, and tryptophan levels in photoautotrophic growth were found in Khairy, Ali, and Dowidar 2011 and Faheed and Abd-El Fattah 2008. Cysteine levels for heterotrophic growth were obtained from Wu et al. 2015. The determination of fatty acids group quantification was based on the lipid profile of *C. vulgaris* shown in Nichols, James, and Breuer 1967. The ratio of nucleotides (28 RNA per DNA) was obtained from Muthuraj et al. 2013. Experimental and literature data values were normalized considering a *C. vulgaris* cell of idealized weight and size. Rates of starch production and degradation were determined using experimental data.

Flux Balance Analysis and Flux Variability Analysis

Flux balance analysis (FBA) and flux variability analysis (FVA) were performed using COBRA Toolbox 2.0 and Gurobi Optimization tools. Autotrophic, mixotrophic, and heterotrophic growth conditions (Table M2) were simulated using Flux Balance Analysis to optimize flux through their respective biomass functions. Then Flux Variability Analysis was then run on the optimized models at 90% of optimal flux through the biomass function to return a flux distribution.

Flux Distribution Sampling

Reactions in the flux distributions returned by FVA analysis that carried a flux of 1E-9 or lower were removed from the dataset. The solution spaces were then sampled using optGPSampler (Megchelenbrink, Huynen, and Marchiori 2014). 50,000 sample steps and a quantity of sampling points equal to double the number of reactions were used for each solution space being sampled. The three resulting sampled solution spaces representing the three growth modes were then compared for significance using a p-value cutoff of 0.05.

Table M1: Online resources and tools used for the construction of the *Chlorella vulgaris* mathematical model.

Database	Link	Reference
Functional annotation, draft generation and subcellular localization		
InterPRO	http://www.ebi.ac.uk/interpro/	(Jones et al. 2014)
PRIAM	http://priam.prabi.fr/	(Claudel-Renard et al. 2003)
Blast	http://blast.ncbi.nlm.nih.gov/	(Altschul et al. 1990)
RAVEN Toolbox	http://www.biomet-toolbox.org/	(Agren et al. 2013)
SignalP 4.1 Server	http://www.cbs.dtu.dk/services/SignalP/	(Petersen et al. 2011)
ChloroP 1.1 Server	http://www.cbs.dtu.dk/services/ChloroP/	(Emanuelsson et al. 1999)
WoLF PSORT	http://www.genscript.com/psort/wolf_psort	(Horton et al. 2007)
CELLO	http://cello.life.nctu.edu.tw/	(Yu et al. 2006)
TargetP 1.1 Server	http://www.cbs.dtu.dk/services/TargetP/	(Emanuelsson et al. 2000)
HECTAR	http://www.sb-roscoff.fr/hectar/	(Gschloessl et al. 2008)
SwissProt	http://web.expasy.org/blast/	(The UniProt Consortium 2015)
TransportDB	http://www.membranetransport.org/	(Ren et al. 2007)
TCDB	http://www.tcdb.org	(Saier et al. 2014)
Manual curation		
KEGG	http://www.genome.jp/kegg/	(Kanehisa et al. 2000)
MetaCyc	http://metacyc.org/	(Caspi et al. 2010)
BIGG	http://bigg.ucsd.edu/	(Schellenberger et al. 2010)
BRENDA	http://www.brenda-enzymes.org/	(Scheer et al. 2011)
SwissProt	http://enzyme.expasy.org/	(The UniProt Consortium 2015)
EMBL-EBI	http://www.ebi.ac.uk/intenz/	(McWilliam et al. 2013)
ExplorEnz	http://www.enzyme-database.org/	McDonald et al., 2007
Conversion to mathematical model		
COBRA Toolbox 2.0	http://sourceforge.net/projects/opencobra/files/	(Schellenberger et al. 2011)
ChemAxon	http:// http://www.chemaxon.com/	ChemAxon, Inc.
GUROBI optimization	http://www.gurobi.com/	Gurobi Optimization, Inc
ESCHER	http://escher.github.io/	(King et al. 2015)
optGpSampler	http://cs.ru.nl/~wmegchel/optGpSampler/	(Megchelenbrink et al. 2014)
SimPheny, Version 3.3		Genomatica, Inc.

Table M2: Growth conditions and constraints for autotrophic, mixotrophic, and heterotrophic growth modes. Flux values are given in mmol gDW^{-1} h^{-1}. The "default" bounds for metabolic reactions are a lower bound of -1000 and upper bound of 1000. PRISM_solar_litho is the flux of photons into the network.

Growth Condition/Constraint	Autotrophic Flux	Mixotrophic Flux	Heterotrophic Flux	Bound
PRISM_solar_litho	-646	-646	0	lower/lower/both
EX_co2(e)	-11	-11	0	lower/lower/both
EX_glc-A(e)	0	-1	-1	both/lower/lower
EX_starch(h)	0	0	-10	both/both/lower
EX_o2(e)	-10	-10	-10	lower
EX_h(e)	-10	-10	-10	lower
EX_h2o(e)	-10	-10	-10	lower
EX_pi(e)	-10	-10	-10	lower
EX_nh4(e)	-10	-10	-10	lower
EX_no3(e)	-10	-10	-10	lower
EX_so4(e)	-10	-10	-10	lower
EX_fe2(e)	-10	-10	-10	lower
EX_fe3(e)	-10	-10	-10	lower
EX_mg2(e)	-10	-10	-10	lower
EX_na1(e)	-10	-10	-10	lower
H2Oth	0	0	0	upper/upper/lower
PCHLDR	0	0	default	both/both
G6PADHh	0	0	default	both/both
G6PBDHh	0	0	default	both/both
FBAh	0	0	default	both/both
ATPSh	default	default	0	both
GAPDH(nadp)hi	default	default	0	both
MDH(nadp)hi	default	default	0	both
MDHC(nadp)hr	default	default	0	both
RBPCh	default	default	0	both
RBCh	default	default	0	both
SBP	default	default	0	both
STARCH300DEGRA	1.7521e-07	1.7521e-07	0	upper
STARCH300DEGR2A	0	0	6.5705e-07	upper
STARCH300DEGRB	1.7521e-07	1.7521e-07	0	upper
STARCH300DEGR2B	0	0	6.5705e-07	upper

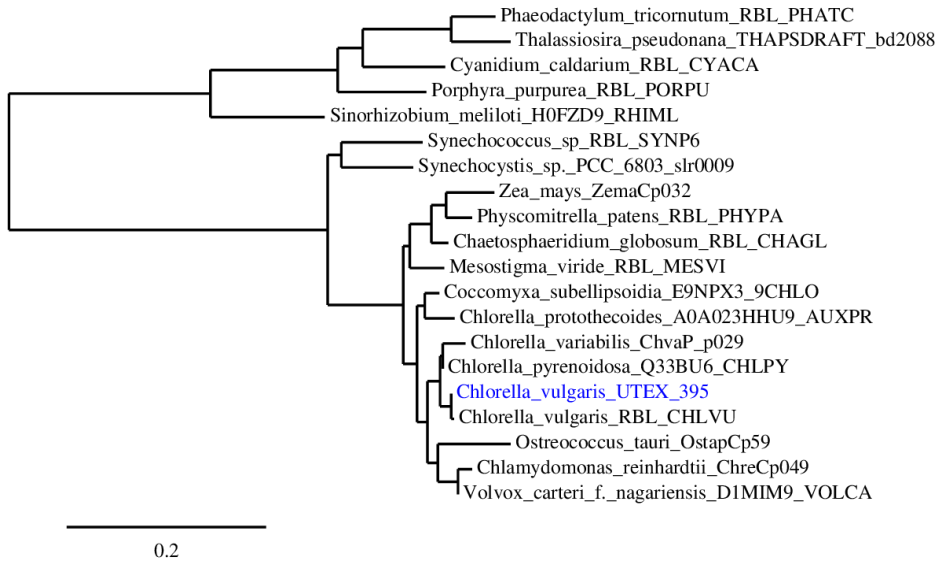

Figure M1: Arbitrarily rooted phylogenetic tree based on multiple sequence alignments of long-chain RuBisCO generated using Phylogeny.fr (Dereeper et al. 2008).

III:

Results

The genome-scale metabolic reconstruction of *Chlorella vulgaris* consisted of three major steps: 1) the generation of a draft reconstruction from homology with *C. reinhardtii*, 2) pathway-by-pathway manual curation of the draft reconstruction using literature and gene-reaction pathway databases and 3) network validation and simulation. The result of the reconstruction process can be compared to the reconstructions of other photosynthetic organisms in Table R1.

Draft Reconstruction Characteristics

The draft reconstruction consists of 621 genes, 1108 reactions, and 1249 metabolites that are distributed across six compartments. Namely, the cytosol, chloroplast, mitochondria, glyoxysome, thylakoid, and extracellular space. The aforementioned quantities of reactions and metabolites include reactions and metabolites that are present in multiple compartments. The number of compartments included in this reconstruction was reduced from the number of compartments in *i*RC1080, which includes also nucleus, flagellum, eyespot, and golgi. The reasoning for not including flagellum and eyespot compartments is obvious; Chlorella contains neither of these physiological structures. The golgi and nucleus were not included due to a general dearth of knowledge about reaction pathways in these bodies in Chlorella, and also, fewer subcellular localization methods were able to assign protein sequences into these compartments, making the potential certainty of such compartmentalization tenuous. Instead, reactions anticipated as occurring in compartments other than the six included were placed in the cytosol.

Reconstruction Subcellular Localization

Subcellular localization within *C. vulgaris* was assigned to reactions with the aid of information from literature, awareness of the localization of neighboring reactions, and various bioinformatics tools found online. Based on this information, each of the 1819 metabolic reactions were assigned to one of five compartments within the reconstruction (Figure R1). In this reconstruction, the thylakoid compartment is fully contained within the chloroplast just as it is located physiologically. The thylakoid is where the primary photosynthetic reactions, photosystems I and II occur, as well as chlorophyll and carotenoid synthesis. Meanwhile, the glyoxysome contains portions of the tricarboxylic acid (TCA) cycle, the glyoxylate cycle, fatty acid metabolism, and amino acid metabolism.

As mentioned previously, this reconstruction has four fewer compartments than the *i*RC1080 reconstruction. GPRs for retinol metabolism were found in *C. vulgaris* despite the absence of an eyespot. Since the eyespot is located in the chloroplast in *C. reinhardtii*, and because subcellular localization methods suggested the chloroplast as the most likely location for the reactions, retinol metabolism was placed in the chloroplast in *i*CZ842. The flagellar reactions in *i*RC1080 were not unique to the compartment, and simply were not included in *i*CZ842. A side-by-side comparison of the compartmentalization of *i*CZ842 and *i*RC1080 is shown in Figure R2.

Reconstruction Subsystems

The final reconstructed metabolic network of *i*CZ842 consists of 1812 metabolic reactions divided into 8 major groups of subsystems and 70 different subsystems (Figure

R3). Subsystems are simply a way of organizing the metabolic network into sets of pathways so that specific reaction pathways are easier to find and work with. These subsystems range in size between one reaction (Tryptophan metabolism) and 359 reactions (Glycerolipid metabolism).

*i*CZ842 exhibits more similarities to *i*RC1080 than differences in terms of the contents of the metabolic network. Of the 1812 reaction in the *i*CZ842 metabolic network, 209 of these reactions are not in *i*RC1080 (Table R2). Conversely, 109 reactions in the *i*RC1080 metabolic network are not in *i*CZ842. These differences span all major groups of subsystems (Figure R4). The reactions omitted from *i*CZ842 primarily include O-Glycan biosynthesis pathway reactions that were not found in *C. vulgaris* and pathways that were located in compartments not included in *i*CZ842 such as glycolysis in the flagellum, as well as a handful of reactions here and there for which no gene or insufficient evidence was found. The reaction pathways that are in just *i*CZ842 include thiamine metabolism, biotin biosynthesis, lipoic acid metabolism, ubiquinone and other terpenoid-quinone biosynthesis, brassinosteroid biosynthesis, alpha-linolenic acid metabolism, and the biosynthesis of phosphatidylcholines.

C. vulgaris can produce monogalactosyl diglycerides, digalactosyl diglycerides, phosphatidylglycerols, sulphoquinovosyl diglycerides, phosphatidylethanolamines, and phosphatidylcholines. For all of these, it produces significant amounts of C18:1, C18:2, and C18:3 fatty acid chains. It also produces C16:0, C16:1, C16:2, and C16:3 monogalactosyl and digalactosyl diglycerides, and C16:0 phosphatidylglycerols (Nichols et al. 1967). All of the pathways necessary to produce these lipids were included in the lipid metabolism section of the reconstruction.

Biomass Objective Functions

C. vulgaris can grow under three different conditions: photoautotrophic, mixotrophic, and heterotrophic. Each growth condition can be represented quantitatively by using different biomass objective functions (BOF). Each BOF is an equation that contains stoichiometric coefficients expressed in mmol gDW^{-1} of all metabolites that contribute to the biomass of the organism. The BOFs of the *C. vulgaris* model contain values for 20 amino acids, 4 nucleosides, 4 nuclotides, 5 carbohydrates (including starch), 101 lipid compounds, and 11 pigments. The Chlorella model includes three biomass objective functions corresponding to three different growth modes: autotrophic, mixotrophic, and heterotrophic.

Growth Mode Analysis

Following the computational determination of optimal growth using flux balance analysis (FBA), flux variability for each reaction in the Chlorella network was determined. By sampling the solution space, we were able to predict the average and range expected values for fluxes through each reaction in each growth mode. Statistically significant differences in sampled flux distributions for specific reactions between different growth modes provided insight into how Chlorella's metabolism responds to different growth scenarios. Photoautotrophic growth was modeled using inputs of light (i.e. photons) and carbon dioxide; mixotrophic growth was modeled using inputs of light, carbon dioxide, and glucose; heterotrophic growth was modeled using no light and an input of glucose as the lone carbon source.

Five reactions in the pentose phosphate pathway (PPP) were determined to have statistically significant differences in flux between autotrophic and heterotrophic growth modes and also between mixotrophic and heterotrophic growth modes (Table R3). *Ribose-phosphate diphosphokinase* (RPDPK), *D-ribulose-5-phosphate 3-epimerase* (RPEh), *ribose-5-phosphate isomerase* (RPIh), *transketolase 1* (TKT1h) and *transketolase 2* (TKT2h) play a crucial role in the central carbon metabolism as they act as intermediaries between glycolysis, carbon fixation, and nucleotide biosynthesis. A metabolic network map of the relevant portion of the PPP was created to show how the different reaction fluxes affected the flow of metabolites in each growth mode (Figures R5-R7). Immediately noticeable is that *phosphoribulokinase* (PRUK) and *ribulose-bisphosphate carboxylase* (RBPCh) carry a significant amount of flux away from the D-ribulose 5-phosphate node in the autotrophic and mixotrophic modes, but carries no flux in the heterotrophic mode. This is also expected, as a key characteristic of heterotrophic growth is the inability to fix carbon, and the light-activated RBPCh is the primary reaction in carbon fixation (Pedersen, Kirk, and Bassham 1966). Thus, RBPCh has been switched off in the heterotrophic growth mode. It is also apparent that fewer metabolites are channeled through the PPP in the heterotrophic growth mode (Figure R7). Yang et al. show that while the cytosolic portion of the PPP seems to carry more flux in heterotrophic growth than in photoautotrophic or mixotrophic growth in *Chlorella sp.*, the chloroplastic portion of the PPP is not significantly active in heterotrophic growth, and overall, the photoautotrophic and mixotrophic growth modes carry more flux through the PPP (Yang, Hua, and Shimizu 2000). This claim is substantiated in our simulation of *Chlorella vulgaris*, where PPP is much more active

in photoautotrophic and mixotrophic growth, as in these modes, glucose 6-phosphate metabolites are channeled through the PPP towards D-ribulose 5-phosphate to be used as a substrate for carbon fixation by RuBisCO (RBPCh). However, despite its low utilization of the PPP, the heterotrophic growth mode sees the most metabolites channeled towards nucleic acid production via *ribose-phosphate diphosphokinase* (RPDPK).

Sampling the solution space of growth modes also provided insight into the production of fatty acids and lipids in Chlorella. For instance, four fundamental reactions in the fatty acid biosynthesis pathway were determined to have statistically different sampled flux distributions between the three growth modes (Table R4). ACOATA and MCOATA produce acetyl-acyl-carrier-protein (acetyl-ACP) and malonyl-acyl-carrier-protein (malonyl-ACP) respectively. These are the two substrates required by KAS14 to initiate the entire process of building fatty acids, which are in turn used as substrates to produce all the lipids created by the organism. Intriguingly, the mixotrophic mode provides the largest flux through these three reactions, whereas the heterotrophic growth mode provides the least. However, *acyl-ACP delta9-desaturase ((9Z)-n-C16:1)* (ACP1619ZD9DS), which produces C16:1 monounsaturated fatty acids and leads to the production of C18:1 monounsaturated fatty acids, is slightly higher in heterotrophic mode than mixotrophic or autotrophic modes. Variations in activity for certain enzymes at these key junctures of lipid production could significantly alter the outcome of what types of lipids are produced, and in what quantity.

Table R1: Characteristics of metabolic network reconstruction of *Chlorella vulgaris* and other photosynthetic species. All species are eukaryotic, except *Synechocystis*, which is a cyanobacterium.

Organism	***Synechocystis sp.* PCC6803**	***Zea mays***	***Ostreococcus tauri***	***Chlamydomonas reinhardtii***	***Chlorella protothecoides***	***Chlorella vulgaris* UTEX 395**
Genome size	3.57 Mb	2.4 Gb	12.6 Mb	100 Mb	22.9 Mb	63 Mb
Model genes (Total genes)	678 (3,575)	1563 (32540)	0 (7699)*	1073(14,354)	461 (7039)	842 (7100)
Total Reactions	863	1985	801	2190	272	2261
Metabolites	790	1825	1014	1707	144	1760
Compartments	3	5	1	10	4	6
Reference	Nogales et al. 2012	Saha, Suthers, and Maranas 2011	Krumholz et al. 2012	Chang et al. 2011	Wu et al. 2015	This thesis

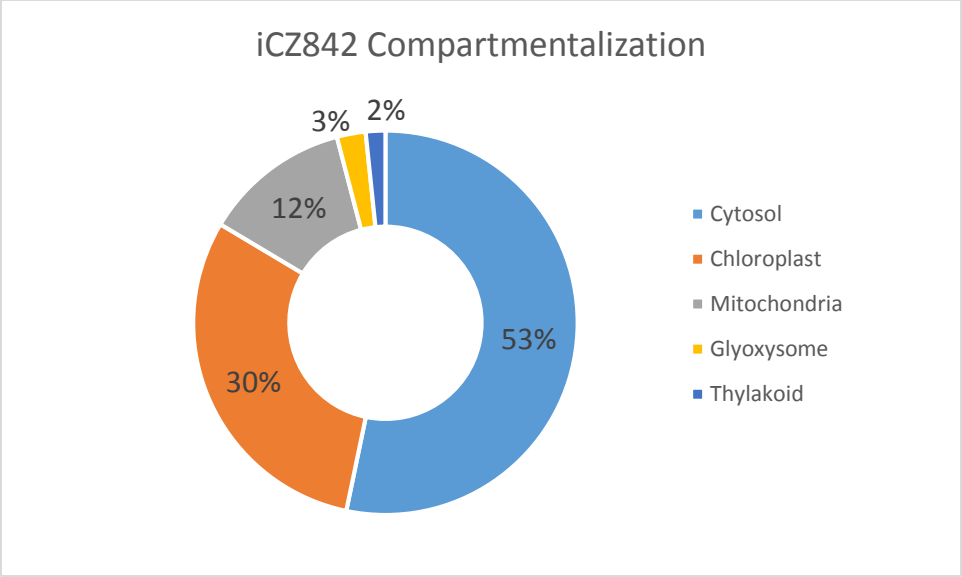

Figure R1: Distribution of the metabolic reactions of *i*CZ842 based on compartment.

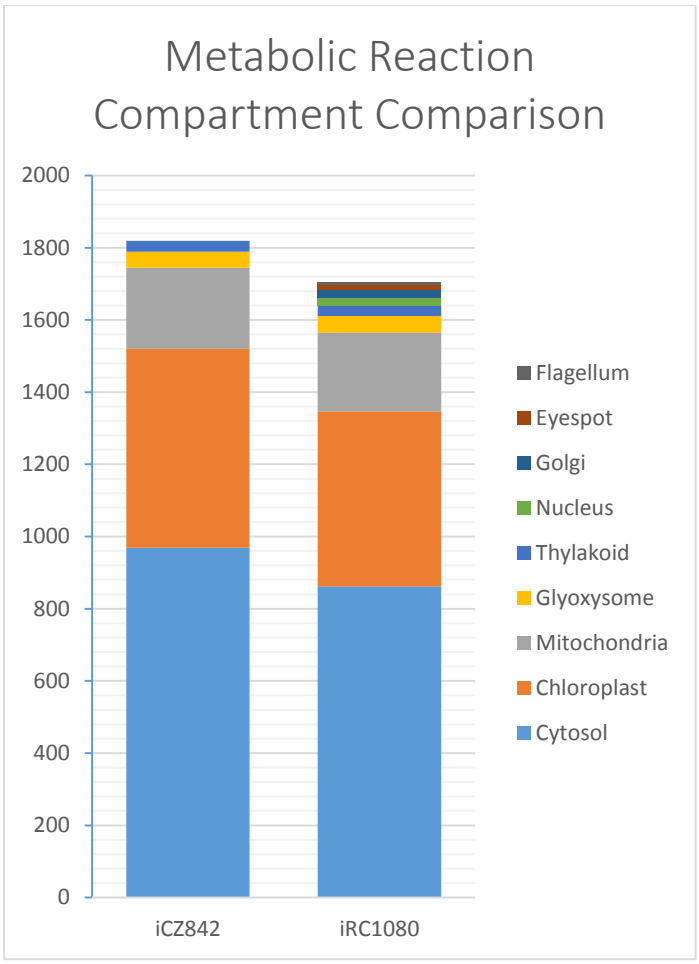

Figure R2: Quantity of metabolic reactions in the different compartments of *i*CZ842 and *i*RC1080.

Table R2: Quantity of metabolic reactions in each the major groups of subsystems in just *i*CZ842, *i*CZ842 and *i*RC1080, or just *i*RC1080

	***i*CZ842**	***i*CZ842 and *i*RC1080**	***i*RC1080**
Carbohydrate Metabolism	26	238	44
Energy Metabolism	0	52	1
Lipid Metabolism	68	623	4
Nucleotide Metabolism	7	175	9
Amino Acid Metabolism	29	283	3
Glycan Metabolism	0	28	23
Metabolism of Cofactors and Vitamins	72	143	24
Secondary Metabolites and Other Reactions	7	61	1
Total Metabolic Reactions	209	1603	109

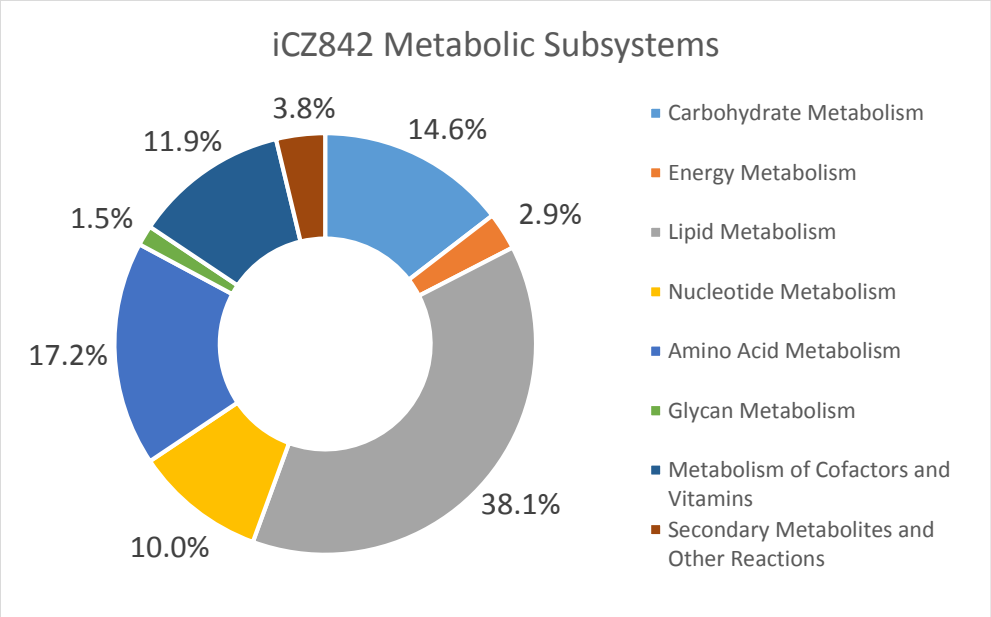

Figure R3: Percentage of metabolic reactions contained by each of eight major groups of subsystems in *i*CZ842.

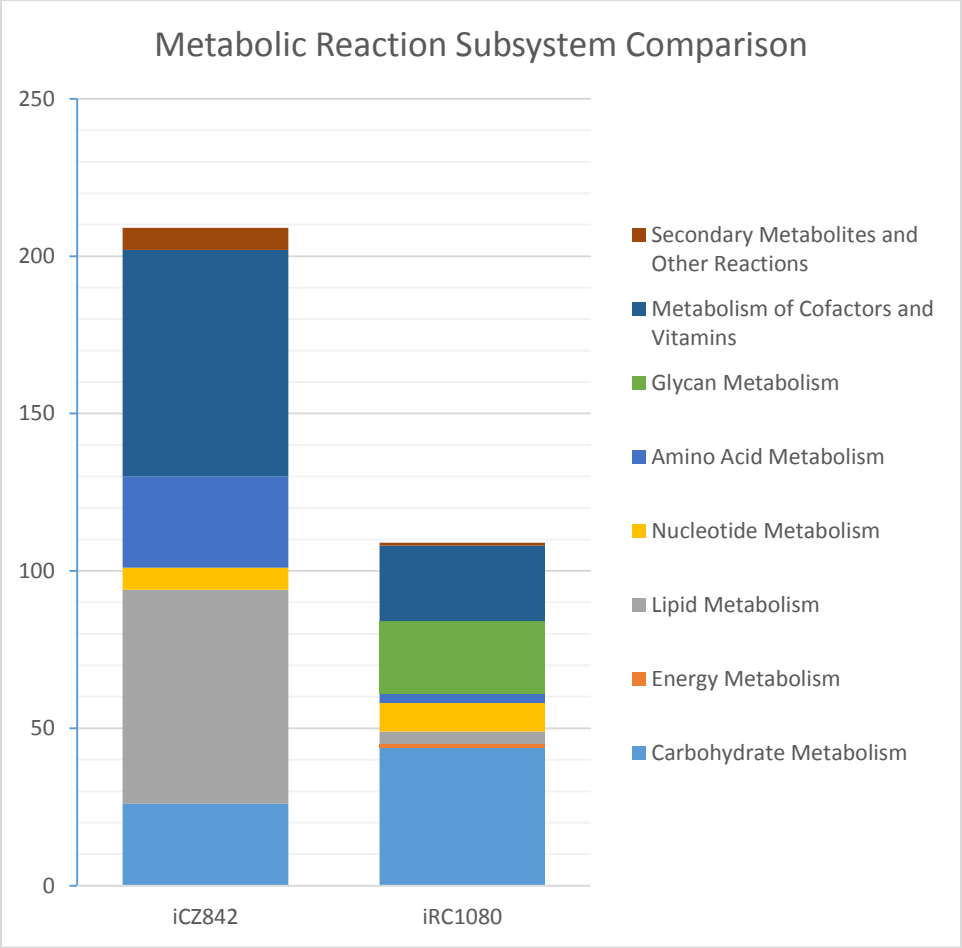

Figure R4: Quantity of metabolic reactions that are not in both *i*CZ842 and *i*RC1080 by subsystem. *i*CZ842 contains several pathways in *Metabolism of Cofactors and Vitamins* that are not in *i*RC1080 including thiamine metabolism, biotin biosynthesis, lipoic acid metabolism, ubiquinone and other terpenoid-quinone biosynthesis, brassinosteroid biosynthesis, alpha-linolenic acid metabolism. *i*CZ842 also contains a phosphatidylcholine biosynthesis pathway in *Lipid Metabolism* that is not in *i*RC1080. *i*CZ842 does not contain the O-glycan biosynthesis pathway that is present in *i*RC1080. iCZ842 also does not contain some duplicate pathways (pathways that occur in more than one compartment) in *Carbohydrate Metabolism* and *Nucleotide Metabolism* that are located in the flagellum and nucleus respectively in *i*RC1080, compartments which are not in *i*CZ842.

Table R3: The mean flux value of sampled points for five key reactions in the pentose phosphate pathway of *C. vulgaris* in the autotrophic, mixotrophic, and heterotrophic growth modes. For all five reactions, the difference in flux between autotrophic and heterotrophic as well as mixotrophic and heterotrophic modes were determined to be statistically significant ($p < 0.05$). Additionally, there is a significant difference in RPDPK flux between autotrophic and mixotrophic growth modes (RPDPK = *ribose-phosphate diphosphokinase*, RPEh = *D-ribulose-5-phosphate 3-epimerase*, RPIh = *ribose-5-phosphate isomerase*, TKT1h = *transketolase 1*, TKT2h = *transketolase 2*).

	Mean sampled flux value (mmol gDW^{-1} h^{-1})		
	Autotrophic	Mixotrophic	Heterotrophic
RPDPK	0.00029	0.00004	0.00041
RPEh	-7.24040	-9.80200	0.02999
RPIh	3.52250	4.35360	-0.07049
TKT1h	-3.61790	-4.88380	0.01823
TKT2h	-3.62250	-4.91820	0.01175

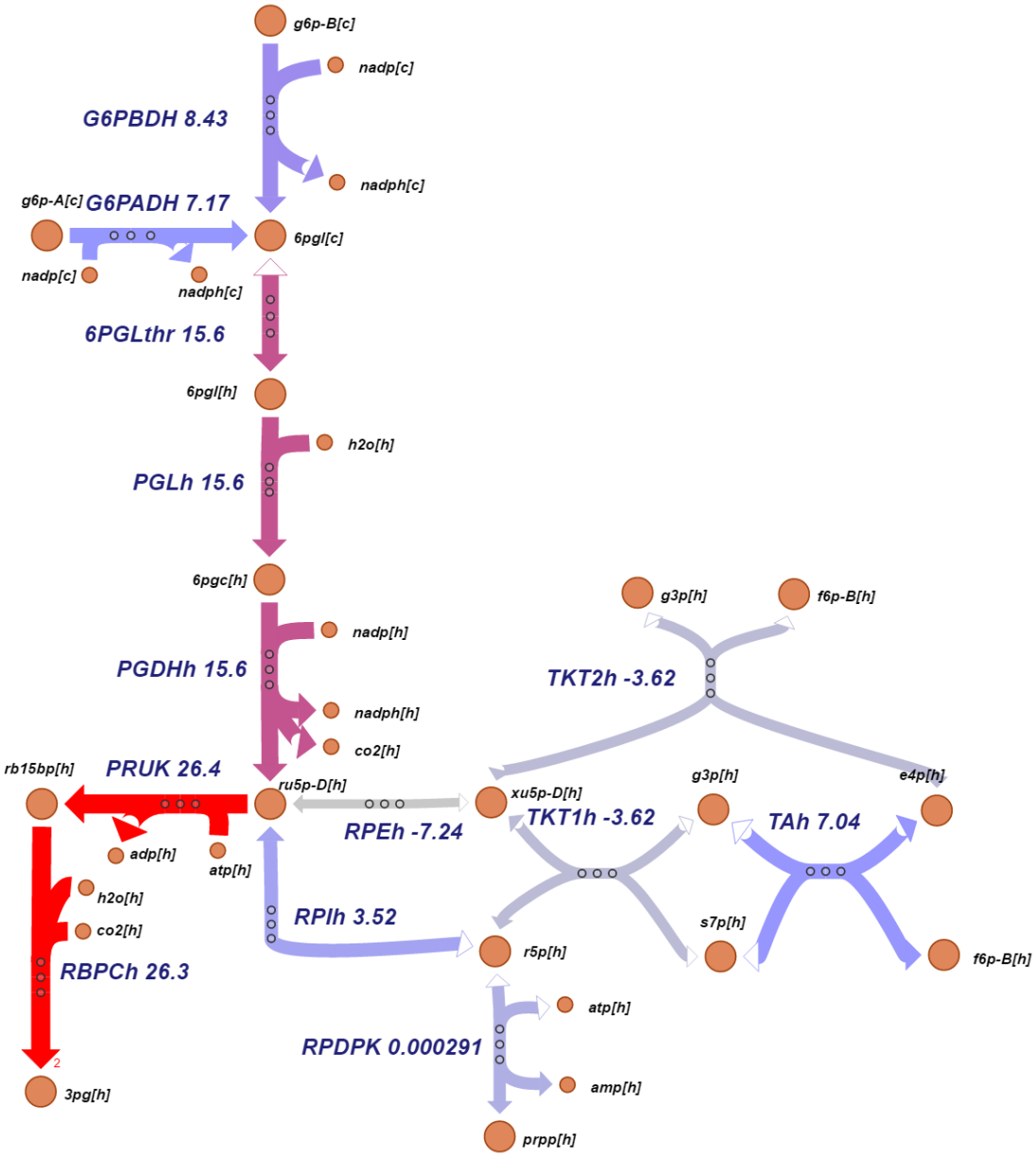

Figure R5: Pentose phosphate pathway map of the autotrophic growth mode. Flux values given in mmol gDW^{-1} h^{-1} are mean values received from sampling (G6PBDH = *glucose-6-phosphate 1-dehydrogenase* (g6p-B), G6PADH = *glucose-6-phosphate 1-dehydrogenase* (g6p-A), 6PGLthr = *6-phospho-D-glucono-1,5-lactone transport* (chloroplast) , PGLh = *6-phosphogluconolactonase*, PGDHh = *6-phosphogluconate dehydrogenase*, TAh = *transaldolase*, PRUK = *phosphoribulokinase*, RBPCh = *ribulose-bisphosphate carboxylase*, RPDPK = *ribose-phosphate diphosphokinase*, RPEh = *D-ribulose-5-phosphate 3-epimerase*, RPIh = *ribose-5-phosphate isomerase*, TKT1h = *transketolase 1*, TKT2h = *transketolase 2*).

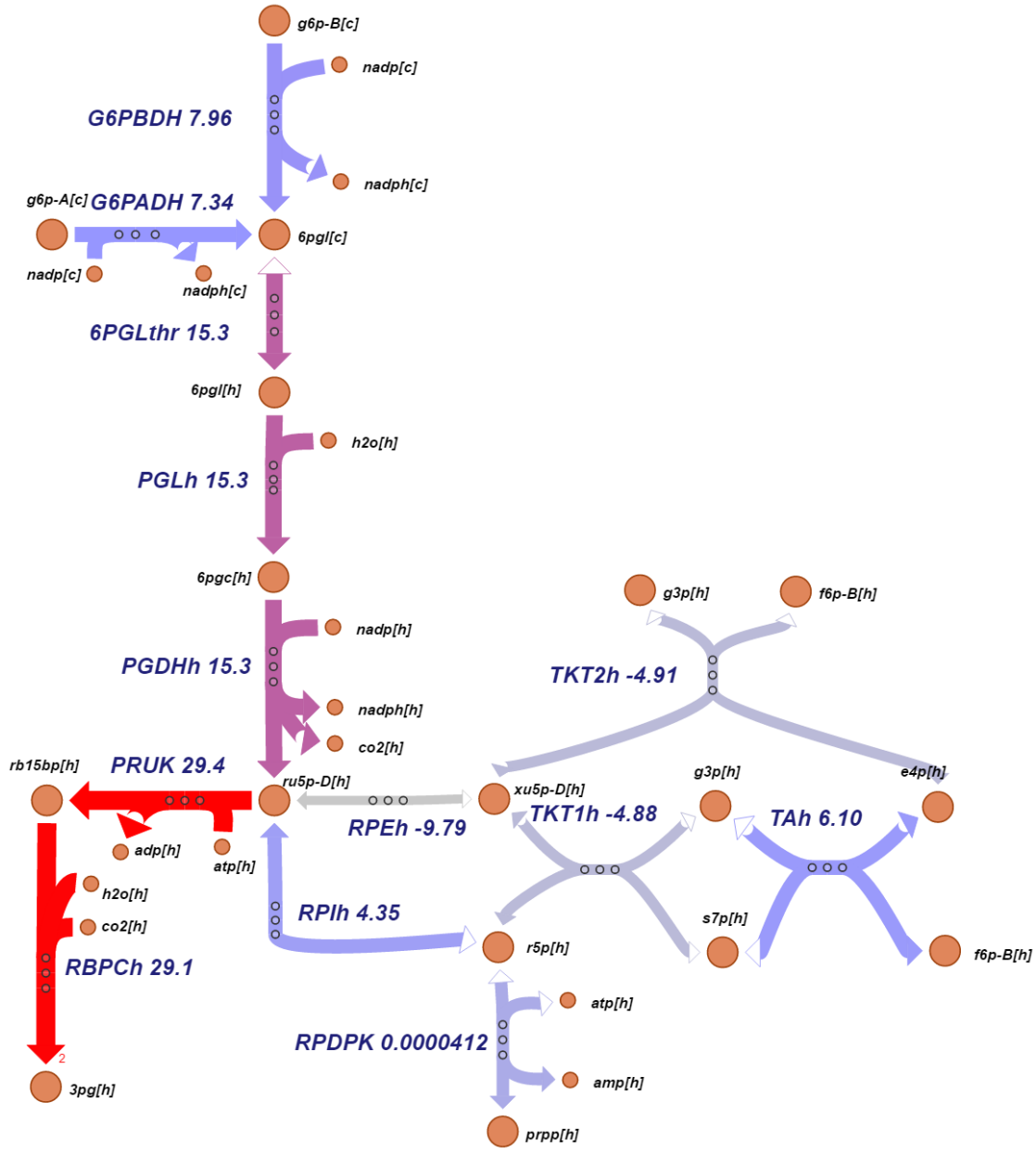

Figure R6: Pentose phosphate pathway map of the mixotrophic growth mode. Flux values given in mmol gDW^{-1} h^{-1} are mean values received from sampling (G6PBDH = *glucose-6-phosphate 1-dehydrogenase* (g6p-B), G6PADH = *glucose-6-phosphate 1-dehydrogenase* (g6p-A), 6PGLthr = *6-phospho-D-glucono-1,5-lactone transport* (chloroplast) , PGLh = *6-phosphogluconolactonase*, PGDHh = *6-phosphogluconate dehydrogenase*, TAh = *transaldolase*, PRUK = *phosphoribulokinase*, RBPCh = *ribulose-bisphosphate carboxylase*, RPDPK = *ribose-phosphate diphosphokinase*, RPEh = *D-ribulose-5-phosphate 3-epimerase*, RPIh = *ribose-5-phosphate isomerase*, TKT1h = *transketolase 1*, TKT2h = *transketolase 2*).

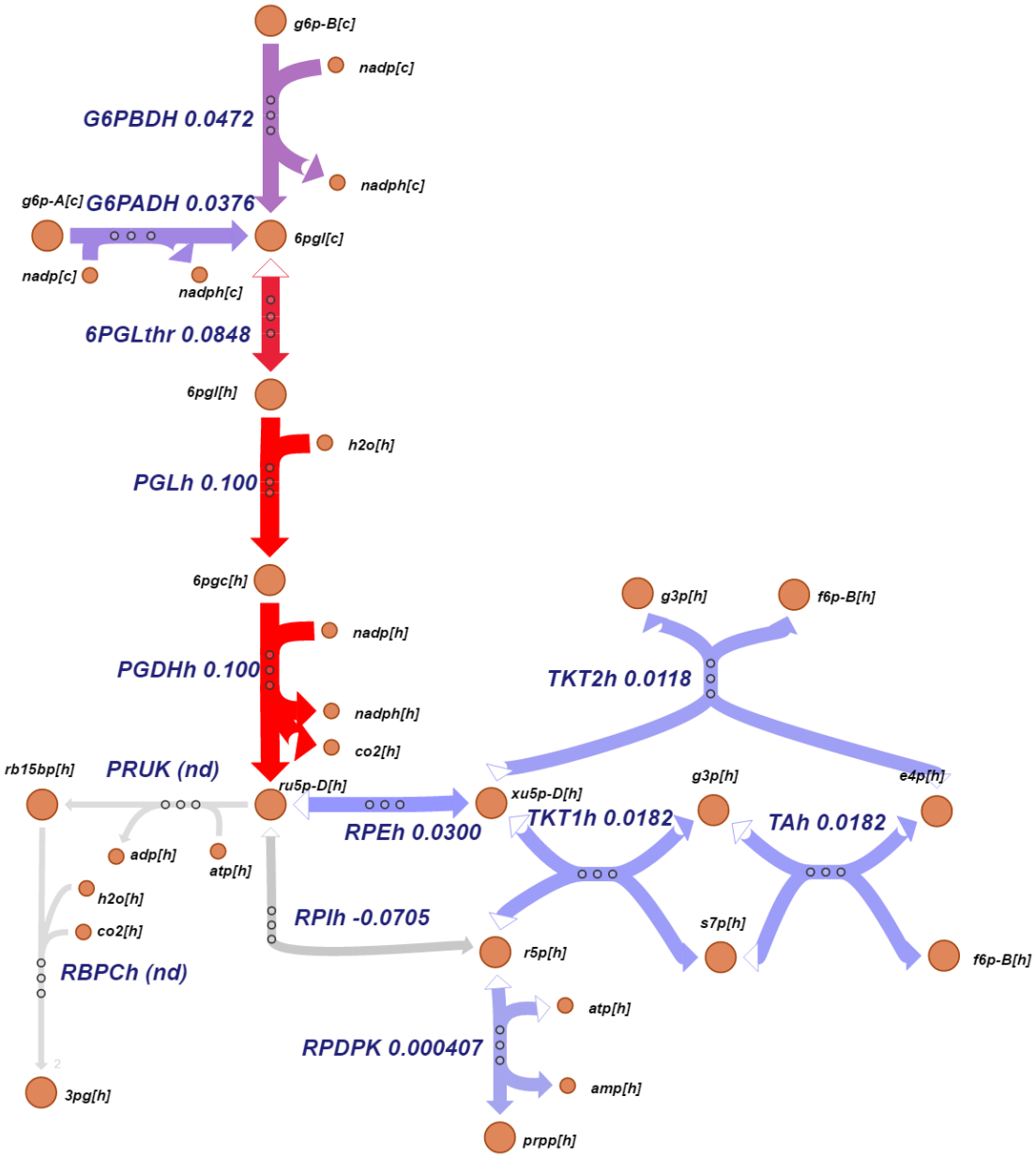

Figure R7: Pentose phosphate pathway map of the heterotrophic growth mode. Flux values given in mmol gDW^{-1} h^{-1} are mean values received from sampling (G6PBDH = *glucose-6-phosphate 1-dehydrogenase* (g6p-B), G6PADH = *glucose-6-phosphate 1-dehydrogenase* (g6p-A), 6PGLthr = *6-phospho-D-glucono-1,5-lactone transport* (chloroplast) , PGLh = *6-phosphogluconolactonase*, PGDHh = *6-phosphogluconate dehydrogenase*, TAh = *transaldolase*, PRUK = *phosphoribulokinase*, RBPCh = *ribulose-bisphosphate carboxylase*, RPDPK = *ribose-phosphate diphosphokinase*, RPEh = *D-ribulose-5-phosphate 3-epimerase*, RPIh = *ribose-5-phosphate isomerase*, TKT1h = *transketolase 1*, TKT2h = *transketolase 2*).

Table R4: The mean flux value of sampled points for four reactions in the fatty acid biosynthesis pathway of *C. vulgaris* in the autotrophic, mixotrophic, and heterotrophic growth modes. For all four reactions, the differences in flux between each pairing of growth modes were determined to be statistically significant ($p < 0.05$), except there was no significant difference between autotrophic and mixotrophic growth modes for ACP1619ZD9DS (ACP1619ZD9DS = *acyl-ACP delta9-desaturase ((9Z)-n-C16:1))*, KAS14 = *beta-ketoacyl-ACP synthase*, ACOATA = *beta-ketoacyl-acyl-carrier-protein synthase III*, MCOATA = *malonyl-CoA ACP S-malonyltransferase*).

	Mean sampled flux value (mmol gDW^{-1} h^{-1})		
	Autotrophic	Mixotrophic	Heterotrophic
ACP1619ZD9DS	0.0040	0.0036	0.0056
KAS14	0.0406	0.2235	0.0105
ACOATA	0.0406	0.2235	0.0105
MCOATA	0.2953	1.5719	0.0809

IV:

Discussion

The genome-scale reconstruction of the metabolic network of *C. vulgaris* UTEX 395 was reconstructed using a methodical procedure incorporating genomics, reaction databases, and literature searches. The reconstruction includes organized genomic, biochemical, and metabolic information of *C. vulgaris* and can be utilized to map its metabolism. The reconstruction has also been successfully implemented as a mathematical model to predict metabolic behavior in various growth conditions. The model used was able to predict a decrease in total flux through the pentose phosphate pathway when *C. vulgaris* is grown in heterotrophic conditions versus when it is grown in photoautotrophic or mixotrophic conditions.

Several vitamin and cofactor pathways of interest have been included in this reconstruction based on genomic and literature evidence. In particular, the pathways leading to the production of brassinosteroids, plant hormones recently isolated and characterized in *C. vulgaris* (Bajguz 2009). Brassinosteroids are important regulators of plant growth and factor significantly into plant response and resistance to environmental stresses (Bajguz and Hayat 2009) Similarly, pathways leading to jasmonic acid production were also included due to being identified in Chlorella (Ueda et al. 1991). Jasmonic acid is another plant hormone that is known to play a role in biotic and abiotic stress response (Bajguz and Hayat 2009). Inclusion of these pathways is potentially useful to researchers attempting to improve growth rates of *C. vulgaris*, as well as bolster its ability to resist environmental stresses. Additionally further investigation of vitamin biosynthesis pathways in *C. vulgaris* may be useful in investigating its potential to produce pharmaceuticals and other high value products (Skjånes, Rebours, and Lindblad 2012).

Because there is particular interest in *C. vulgaris* as a source of lipids for the production of biofuels, special attention was given to including all lipid metabolism pathways that are evidenced by literature and the genome functional annotation. The reconstruction includes pathways leading to the production of triacylglycerides (TAGs) with C16 and C18 fatty saturated, monounsaturated, and polyunsaturated fatty acid groups. These TAGs can be split via transesterification into C16 and C18 fatty acid esters to produce biodiesel. Much is known about the relative lipid production of *Chlorella vulgaris* in photoautotrophic, mixotrophic, and heterotrophic growth conditions with both glucose and acetate as a carbon source (Nichols 1965; Nichols et al. 1967; Chisti 2007; Liang, Sarkany, and Cui 2009). Namely, that Chlorella grown heterotrophically on glucose or acetate generally produces a larger quantity of lipids than photoautotrophically grown Chlorella. However, the various metabolic routes used (and unused) by Chlorella that determine which types of lipids are produced, and that ultimately result in higher lipid yield for heterotrophic conditions, is still unclear on a network-wide level. It is clear from both prior experimental results, as well as from our simulations, that the PPP and Embden-Meyerhof glycolytic pathway differ significantly between growth modes, and are fundamental in both the general growth of the organism, and in producing the building block metabolites that become lipids. The *i*CZ842 reconstruction allows for simple quantitative modeling of these central carbon metabolism pathways and others on a system-wide scale in specific growth conditions, and will help determine the optimal condition for growth and lipid production of *C. vulgaris*.

The *i*CZ842 reconstruction also provides a path forward towards understanding the metabolic pathways that lead to the production of the TAGs that are ideal for the manufacture of biodiesel; it can be used to model growth conditions as well as potential gene insertions and deletions that could optimize TAG production. The model was able to identify *acyl-ACP delta9-desaturase* (ACP1619ZD9DS), *beta-ketoacyl-ACP synthase* (KAS14), *beta-ketoacyl-acyl-carrier-protein synthase III* (ACOATA), and *malonyl-CoA ACP S-malonyltransferase* (MCOATA) as four fundamental fatty acid biosynthesis reactions whose fluxes varied significantly between autotrophic, mixotrophic, and heterotrophic growth conditions. These reactions and the gene-protein-reaction connections that govern them could be candidates for further study regarding their potential for increasing lipid production via metabolic engineering.

Beyond the inherent value of the *i*CZ842 reconstruction to investigating *C. vulgaris* for its scientific and industrial potential, the reconstruction can also be used as a template to study further algal and plant specimens. *C. vulgaris* is now just one of a few photosynthetic species with a genome-scale metabolic reconstruction (Table R1). Much as *i*RC1080 was an invaluable reference network reconstruction to the creation of *i*CZ842, it is evident that *i*CZ842 could be used in a similar way. Therefore, in addition to serving as a mathematical model and a genomic, metabolic, and biochemical knowledgebase for *C. vulgaris* UTEX 395, *i*CZ842 can be used to aid the construction of further high-quality metabolic network reconstructions.

References

Agren, R., L. Liu, S. Shoaie, W. Vongsangnak, I. Nookaew, and J. Nielsen. 2013. "The RAVEN Toolbox and Its Use for Generating a Genome-Scale Metabolic Model for Penicillium Chrysogenum." *PLoS Computational Biology* 9 (3): e1002980.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10.

Anthony, J., V. R. Rangamaran, D. Gopal, K. T. Shivasankarasubbiah, M. L. J. Thilagam, M. P. Dhassiah, D. S. M. Padinjattayil, V. N. Valsalan, V. Manambrakat, S. Dakshinamurthy, S. Thirunavukkarasu, and K. Ramalingam. 2015. "Ultraviolet and 5'fluorodeoxyuridine Induced Random Mutagenesis in Chlorella Vulgaris and Its Impact on Fatty Acid Profile: A New Insight on Lipid-Metabolizing Genes and Structural Characterization of Related Proteins." *Marine Biotechnology* 17 (1): 66–80.

Bajguz, A. 2009. "Isolation and Characterization of Brassinosteroids from Algal Cultures of Chlorella Vulgaris Beijerinck (Trebouxiophyceae)." *Journal of Plant Physiology* 166 (17). Elsevier: 1946–49.

Bajguz, A., and S. Hayat. 2009. "Effects of Brassinosteroids on the Plant Responses to Environmental Stresses." *Plant Physiology and Biochemistry* 47 (1). Elsevier Masson SAS: 1–8.

Bentley, R., and G. Boyle. 2008. "Global Oil Production: Forecasts and Methodologies." *Environment and Planning B: Planning and Design* 35 (4): 609–26.

Budny, D. 2007. "The Global Dynamics of Biofuels." *Brazil Institute Special Report*, no. 3: 1–8.

Caspi, R., T. Altman, R. Billington, K. Dreher, H. Foerster, C. A. Fulcher, T. A. Holland, I. M. Keseler, A. Kothari, A. Kubo, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, D. Weerasinghe, P. Zhang, et al. 2014. "The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases." *Nucleic Acids Research* 42 (Database Issue): D459–71.

Caspi, R., T. Altman, J. M. Dale, K. Dreher, C. A. Fulcher, F. Gilham, P. Kaipa, A. S. Karthikeyan, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, S. Paley, L. Popescu, A. Pujar, A. G. Shearer, P. Zhang, and P. D. Karp. 2010. "The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/genome Databases." *Nucleic Acids Research* 38 (Database): D473–79.

Chang, R. L., L. Ghamsari, A. Manichaikul, E. F. Y. Hom, S. Balaji, W. Fu, Y. Shen, T. Hao, B. Ø. Palsson, K. Salehi-Ashtiani, and J. A. Papin. 2011. "Metabolic Network Reconstruction of Chlamydomonas Offers Insight into Light-Driven Algal Metabolism." *Molecular Systems Biology* 7 (518): 1–13.

Chisti, Y. 2007. "Biodiesel from Microalgae." *Biotechnology Advances* 25 (3). Elsevier Inc.: 294–306.

Claudel-Renard, C., C. Chevalet, T. Faraut, and D. Kahn. 2003. "Enzyme-Specific Profiles for Genome Annotation: PRIAM." *Nucleic Acids Research* 31 (22): 6633–39.

Dansen, T. B., E. H. W. Pap, R. J. A. Wanders, and K. W. A. Wirtz. 2001. "Targeted Fluorescent Probes in Peroxisome Function." *The Histochemical Journal* 33 (2): 65–69.

de Oliveira Dal'Molin, C. G., L.-E. Quek, R. W. Palfreyman, S. M. Brumbley, and L. K. Nielsen. 2010. "AraGEM, a Genome-Scale Reconstruction of the Primary Metabolic Network in Arabidopsis." *Plant Physiology* 152 (2): 579–89.

Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J.-F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J.-M. Claverie, and O. Gascuel. 2008. "Phylogeny.fr: Robust Phylogenetic Analysis for the Non-Specialist." *Nucleic Acids Research* 36 (Web Server issue): W465–69.

Emanuelsson, O., H. Nielsen, S. Brunak, and G. von Heijne. 2000. "Predicting Subcellular Localization of Proteins Based on Their N-Terminal Amino Acid Sequence." *Journal of Molecular Biology* 300 (4): 1005–16.

Emanuelsson, O., H. Nielsen, and G. V. O. N. Heijne. 1999. "ChloroP, a Neural Network-Based Method for Predicting Chloroplast Transit Peptides and Their Cleavage Sites." *Protein Science* 8 (5): 978–84.

Faheed, F. A., and Z. Abd-El Fattah. 2008. "Effect of Chlorella Vulgaris as Bio-Fertilizer on Growth Parameters and Metabolic Aspects of Lettuce Plant." *Journal of Agricultural & Social Sciences* 4 (4): 165–69.

Feist, A. M., and B. Ø. Palsson. 2008. “The Growing Scope of Applications of Genome-Scale Metabolic Reconstructions Using Escherichia Coli.” *Nature Biotechnology* 26 (6): 659–67.

Giordano, M., A. Norici, M. Forssen, M. Eriksson, and J. A. Raven. 2003. “An Anaplerotic Role for Mitochondrial Carbonic Anhydrase in Chlamydomonas Reinhardtii.” *Plant Physiology* 132 (4): 2126–34.

Goldemberg, J. 2008. “The Brazilian Biofuels Industry.” *Biotechnology for Biofuels* 1 (1): 6.

Goss, R., and G. Garab. 2001. “Non-Photochemical Chlorophyll Fluorescence Quenching and Structural Rearrangements Induced by Low pH in Intact Cells of Chlorella Fusca (Chlorophyceae) and Mantoniella Squamata (Prasinophyceae).” *Photosynthesis Research* 67 (3): 185–97.

Gschloessl, B., Y. Guermeur, and J. M. Cock. 2008. “HECTAR: A Method to Predict Subcellular Targeting in Heterokonts.” *BMC Bioinformatics* 9 (January): 393.

Guarnieri, M. T., A. Nag, S. Yang, and P. T. Pienkos. 2013. “Proteomic Analysis of Chlorella Vulgaris: Potential Targets for Enhanced Lipid Accumulation.” *Journal of Proteomics* 93. Elsevier B.V.: 245–53.

Hay, J. O., H. Shi, N. Heinzel, I. Hebbelmann, H. Rolletschek, and J. Schwender. 2014. “Integration of a Constraint-Based Metabolic Model of Brassica Napus Developing Seeds with (13)C-Metabolic Flux Analysis.” *Frontiers in Plant Science* 5 (December): 724.

Hogetsu, D., and S. Miyachi. 1979. “Role of Carbonic Anhydrase in Photosynthetic CO2 Fixation in Chlorella.” *Plant & Cell Physiology* 20 (4): 747–56.

Horton, P., K.-J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier, and K. Nakai. 2007. “WoLF PSORT: Protein Localization Predictor.” *Nucleic Acids Research* 35 (Web Server issue): W585–87.

Hu, Q., M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, and A. Darzins. 2008. “Microalgal Triacylglycerols as Feedstocks for Biofuel Production: Perspectives and Advances.” *The Plant Journal* 54 (4): 621–39.

Huntley, M. E., and D. G. Redalje. 2006. “CO2 Mitigation and Renewable Oil from Photosynthetic Microbes: A New Appraisal.” *Mitigation and Adaptation Strategies for Global Change* 12 (4): 573–608.

International Energy Agency. 2014. “World Energy Outlook 2014.”

Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez, and S. Hunter. 2014. “InterProScan 5: Genome-Scale Protein Function Classification.” *Bioinformatics* 30 (9): 1236–40.

Kanehisa, M., and S. Goto. 2000. “KEGG : Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research* 28 (1): 27–30.

Khairy, H. M., E. M. Ali, and S. M. Dowidar. 2011. “Comparative Effects of Autotrophic and Heterotrophic Growth on Some Vitamins, 2, 2-Diphenyl-1 Picrylhydrazyl (DPPH) Free Radical Scavenging Activity, Amino Acids and Protein Profile of Chlorella Vulgaris Beijerinck.” *African Journal of Biotechnology* 10 (62): 13514–19.

King, Z. a., A. Dräger, A. Ebrahim, N. Sonnenschein, N. E. Lewis, and B. O. Palsson. 2015. “Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways.” *PLOS Computational Biology* 11 (8): e1004321.

Komor, E., and W. Tanner. 1974. “The Hexose-Proton Cotransport System of Chlorella pH-Dependent Change in Km Values and Translocation Constants of the Uptake System.” *Journal of General Physiology* 64 (5): 568–81.

Krumholz, E. W., H. Yang, P. Weisenhorn, C. S. Henry, and I. G. L. Libourel. 2012. “Genome-Wide Metabolic Network Reconstruction of the Picoalga Ostreococcus.” *Journal of Experimental Botany* 63 (6): 2353–62.

Kusel, A. C., J. Sianoudis, D. Leibfritz, L. H. Grimme, and A. Mayer. 1990. “The Dependence of the Cytoplasmic pH in Aerobic and Anaerobic Cells of the Green Algae Chlorella Fusca and Chlorella Vulgaris on the pH of the Medium as Determined by 31P in Vivo NMR Spectroscopy.” *Archiv Für Mikrobiologie* 153 (3): 254–58.

Li, Y., M. Horsman, N. Wu, C. Q. Lan, and N. Dubois-calero. 2008. “Biofuels from Microalgae.” *Biotechnology Progress* 24 (4): 815–20.

Liang, Y., N. Sarkany, and Y. Cui. 2009. “Biomass and Lipid Productivities of Chlorella Vulgaris under Autotrophic, Heterotrophic and Mixotrophic Growth Conditions.” *Biotechnology Letters* 31 (7): 1043–49.

Long, C. P., and M. R. Antoniewicz. 2014. “Quantifying Biomass Composition by Gas Chromatography/Mass Spectrometry.” *Analytical Chemistry* 86 (19): 9423–27.

McWilliam, H., W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley, and R. Lopez. 2013. “Analysis Tool Web Services from the EMBL-EBI.” *Nucleic Acids Research* 41 (Web Server issue): W597–600.

Megchelenbrink, W., M. Huynen, and E. Marchiori. 2014. “optGpSampler: An Improved Tool for Uniformly Sampling the Solution-Space of Genome-Scale Metabolic Networks.” *PLoS ONE* 9 (2).

Montagud, A., E. Navarro, P. Fernández de Córdoba, J. F. Urchueguía, and K. R. Patil. 2010. “Reconstruction and Analysis of Genome-Scale Metabolic Model of a Photosynthetic Bacterium.” *BMC Systems Biology* 4 (1). BioMed Central Ltd: 156.

Muthuraj, M., B. Palabhanvi, S. Misra, V. Kumar, K. Sivalingavasu, and D. Das. 2013. “Flux Balance Analysis of Chlorella Sp . FC2 IITG under Photoautotrophic and Heterotrophic Growth Conditions.” *Photosynthesis Research* 118 (1): 167–79.

Nichols, B. W. 1965. “Light Induced Changes in the Lipids of Chlorella Vulgaris.” *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 106: 274–79.

Nichols, B. W., A. T. James, and J. Breuer. 1967. “Interrelationships between Fatty Acid Biosynthesis and Acyl-Lipid Synthesis in Chlorella Vulgaris.” *The Biochemical Journal* 104: 486–96.

Nogales, J., S. Gudmundsson, E. M. Knight, B. O. Palsson, and I. Thiele. 2012. “Detailing the Optimality of Photosynthesis in Cyanobacteria through Systems Biology Analysis.” *Proceedings of the National Academy of Sciences* 109 (7): 2678–83.

Oberhardt, M. A., B. Ø. Palsson, and J. A. Papin. 2009. “Applications of Genome-Scale Metabolic Reconstructions.” *Molecular Systems Biology* 5 (320): 320.

Parisi, G., M. Perales, M. Fornasari, A. Colaneri, N. Gonzalez-Schain, D. Gomez-Casati, S. Zimmermann, A. Brennicke, E. Zabaleta, A. Araya, J. G. Ferry, J. Echave, and E. Zabaleta. 2004. “Gamma Carbonic Anhydrases in Plant Mitochondria.” *Plant Molecular Biology* 55 (2): 193–207.

Pedersen, T. A., M. Kirk, and J. A. Bassham. 1966. “Light-Dark Transients in Levels of Intermediate Compounds during Photosynthesis in Air-Adapted Chlorella.” *Physiologia Plantarum* 19 (1): 219–31.

Petersen, T. N., S. Brunak, G. von Heijne, and H. Nielsen. 2011. “SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions.” *Nature Methods* 8 (10). Nature Publishing Group: 785–86.

Ren, Q., K. Chen, and I. T. Paulsen. 2007. “TransportDB: A Comprehensive Database Resource for Cytoplasmic Membrane Transport Systems and Outer Membrane Channels.” *Nucleic Acids Research* 35 (SUPPL. 1): 274–79.

Saha, R., P. F. Suthers, and C. D. Maranas. 2011. “Zea Mays iRS1563: A Comprehensive Genome-Scale Metabolic Reconstruction of Maize Metabolism.” *PloS One* 6 (7): e21784.

Saier, M. H., V. S. Reddy, D. G. Tamang, and Å. Västermark. 2014. “The Transporter Classification Database.” *Nucleic Acids Research* 42 (D1): 251–58.

Savage, N. 2011. “Algae: The Scum Solution.” *Nature* 474 (7352): S15–16.

Scheer, M., A. Grote, A. Chang, I. Schomburg, C. Munaretto, M. Rother, C. Sohngen, M. Stelzer, J. Thiele, and D. Schomburg. 2011. “BRENDA, the Enzyme Information System in 2011.” *Nucleic Acids Research* 39 (Database): D670–76.

Schellenberger, J., J. O. Park, T. M. Conrad, and B. Ø. Palsson. 2010. “BiGG: A Biochemical Genetic and Genomic Knowledgebase of Large Scale Metabolic Reconstructions.” *BMC Bioinformatics* 11: 213.

Schellenberger, J., R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, and B. Ø. Palsson. 2011. “Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox v2.0.” *Nature Protocols* 6 (9): 1290–1307.

Sheehan, J., T. Dunahay, J. Benemann, and P. Roessler. 1998. “A Look Back at the U. S. Department of Energy’s Aquatic Species Program — Biodiesel from Algae Office of Fuels Development.”

Shi, F., P. Wang, Y. Duan, D. Link, and B. Morreale. 2012. “Recent Developments in the Production of Liquid Fuels via Catalytic Conversion of Microalgae: Experiments and Simulations.” *RSC Advances* 2 (26): 9727–47.

Skjånes, K., C. Rebours, and P. Lindblad. 2012. “Potential for Green Microalgae to Produce Hydrogen, Pharmaceuticals and Other High Value Products in a Combined Process.” *Critical Reviews in Biotechnology* 33 (January 2012): 1–44.

The UniProt Consortium. 2015. “UniProt: A Hub for Protein Information.” *Nucleic Acids Research* 43 (D1): D204–12.

Thiele, I., and B. Ø. Palsson. 2010. “A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction.” *Nature Protocols* 5 (1): 93–121.

U.S. Department of Energy. 2012. “Biomass Multi-Year Program Plan.”

U.S. Energy Information Administration. 2014. “International Energy Outlook 2014.”

U.S. Environmental Protection Agency. 2014. “Climate Change Indicators in the United States, 2014. Third Edition. EPA 430-R-14-004.”

Ueda, J., K. Miyamoto, M. Aoki, T. Hirata, T. Sato, and Y. Momotani. 1991. “Identification of Jasmonic Acid in Chlorella and Spirulina.” *Bulletin of the University of Osaka Prefecture. Ser. B, Agriculture and Biology* 43: 103–8.

Wise, M., J. Dooley, P. Luckow, K. Calvin, and P. Kyle. 2014. “Agriculture, Land Use, Energy and Carbon Emission Impacts of Global Biofuel Mandates to Mid-Century.” *Applied Energy* 114 (February). Elsevier Ltd: 763–73.

Wu, C., W. Xiong, J. Dai, and Q. Wu. 2015. “Genome-Based Metabolic Mapping and 13C Flux Analysis Reveal Systematic Properties of an Oleaginous Microalga Chlorella Protothecoides.” *Plant Physiology* 167 (2): 586–99.

Yang, C., Q. Hua, and K. Shimizu. 2000. “Energetics and Carbon Metabolism during Growth of Microalgalcells under Photoautotrophic, Mixotrophic and Cycliclight-Autotrophic/dark-Heterotrophic Conditions.” *Biochemical Engineering Journal* 6: 87–102.

Yu, C., Y. Chen, C. Lu, and J. Hwang. 2006. “Prediction of Protein Subcellular Localization.” *Proteins: Structure, Function, Bioinformatics* 64 (3): 643–51.