Genome-scale metabolic model for the green alga *Chlorella vulgaris* UTEX 395 accurately predicts phenotypes under autotrophic, heterotrophic, and mixotrophic growth conditions

Cristal Zuñiga¹, Chien-Ting Li², Tyler Huelsman¹, Jennifer Levering¹, Daniel C. Zielinski¹, Brian O. McConnell³, Christopher P. Long³, Eric P. Knoshaug⁴, Michael T. Guarnieri⁴, Maciek R. Antoniewicz³, Michael J. Betenbaugh², Karsten Zengler*¹

¹ Department of Bioengineering, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093-0412, USA
² Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA
³ Department of Chemical and Biomolecular Engineering, Metabolic Engineering and Systems Biology Laboratory, University of Delaware, 150 Academy St, Newark, DE 19716, USA
⁴ National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO 80401, USA

Short title: Metabolic model of *Chlorella vulgaris* accurately predicts increased growth rates

Corresponding author: Karsten Zengler

Department of Bioengineering, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093-0412, USA

Phone: (858) 822-1168, Fax: (858) 822-3120, E-mail: kzengler@ucsd.edu

SUMMARY: Genome-scale metabolic model for *Chlorella vulgaris* UTEX 395 accurately predicts phenotypes under different growth conditions
AUTHOR CONTRIBUTIONS AND FUNDING INFORMATION

CZ, MJB, and KZ conceived and designed the study. CZ, TH, and JL performed the reconstruction. CTL, MTG, EPK, BOM, CPL and MRA performed the experiments and analyzed the data. CZ and DZ carried out the simulations and analysis. CZ and KZ wrote the manuscript with assistance from all co-authors. All authors have read and approved the manuscript.

This material is based upon work supported by the National Science Foundation under Grant No. 1332344, the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences Energy Frontier Research Centers program under Award Number DE-SC0012658 and Cristal Zuniga by Mexican National Research Council, CONACYT, fellowship No. 237897.

ABSTRACT

The green microalgae *Chlorella vulgaris* has been widely recognized as a promising candidate for biofuel production due to its ability to store high lipid content and its natural metabolic versatility. Compartamentalized genome-scale metabolic models constructed from genome sequences enable quantitative insight into the transport and metabolism of compounds within a target organism. These metabolic models have long been utilized to generate optimized design strategies for an improved production process. Here, we describe the reconstruction, validation, and application of a genome-scale metabolic model for *C. vulgaris* UTEX 395, iCZ843. The reconstruction represents the most comprehensive model
for any eukaryotic photosynthetic organism to date, based on the genome size and number of genes in the reconstruction. The highly curated model accurately predicts phenotypes under photoautotrophic, heterotrophic, and mixotrophic conditions. The model was validated against experimental data and lays the foundation for model-driven strain design and medium alteration to improve yield. Calculated flux distributions under different trophic conditions show that a number of key pathways are affected by nitrogen starvation conditions, including central carbon metabolism and amino acid, nucleotide, and pigment biosynthetic pathways. Furthermore, model prediction of growth rates under various medium compositions and subsequent experimental validation showed an increased growth rate with the addition of tryptophan and methionine.

**Keywords:** Metabolic modeling, *Chlorella vulgaris* UTEX 395, genome-scale metabolic reconstruction, photoautotrophy, heterotrophy, mixotrophy.
BACKGROUND

Photosynthetic microorganisms have gained attention for their potential utility in various biotechnology applications due to their polytrophic metabolism, including photoautotrophy, heterotrophy, and mixotrophy, enabling them to take advantage of different energy and carbon sources for growth. These organisms are capable of fixing CO₂ with light as an energy source to produce biomass and O₂, thus having a significant impact on the global O₂ and CO₂ budget. Algae, as well as microalgae, play a profound role in the carbon cycle and are responsible for 50% of the atmospheric oxygen release and carbon fixation on the planet (Tabatabaei et al., 2011).

The unicellular microalgae *Chlorella vulgaris* of the phylum Chlorophyta has long served as a model organism. Chlorella strains exhibit substantial metabolic flexibility in response to environmental perturbations (Mitra et al., 2012; Liu et al., 2013). Their robust metabolic capabilities suggest the potential for industrial production of desired biomass components, most notably lipids, to serve as precursors for biofuel production, as well as other value-added products (Perez-Garcia et al., 2011). In addition, Chlorella strains are capable of using nutrients (i.e. organic carbon and minerals) directly from wastewater for growth, making them attractive cell factories for biosustainable production processes (Liu et al., 2013). Although algal biodiesel production is technically feasible and environmentally desirable, several challenges remain to be overcome for the process to successfully compete with fossil fuel-derived products (Miao and Wu, 2006; Xiong et al., 2008).

Species such as *Chlorella protothecoides*, *C. minutissima*, and *C. vulgaris* can accumulate lipids up to 50%, 56%, and 60% of their dry weight, respectively (Wang et al., 2008; Guarnieri et al., 2013; Espinosa-Gonzalez et al., 2014). Microalgae with an oil content of
around 50% could reach a productivity of 86,515 kg biodiesel (per hectare per year),
compared to 170 and 340 kg (per hectare per year) from soy and canola (Savage, 2011).
Among microalgae, Chlorella strains possess significant advantages, such as higher
photosynthetic efficiency over other photosynthetic organisms (Doucha and Livanský,
2006). Due to this potential, efforts have been made to characterize the biomass production
of Chlorella strains in large-scale bioreactors. It has been reported that the synthesis of
lipids in Chlorella is substantially increased by deficiencies of nitrogen and phosphorus in
the culture medium, with nitrogen having the most important effect on lipid accumulation.
Other studies have shown that lower nitrogen-to-carbon ratios result in higher lipid storage
by *C. vulgaris* (Levering et al., 2015).

Industrial microorganisms, including *C. vulgaris*, have become targets for engineering
strategies to improve their performance in biotechnological applications. Genome-scale
models have proven to be useful tools for the targeted engineering of such organisms
(Monk et al., 2014). These genome-scale network reconstructions can help to understand
the compartmental organization of metabolism within an organism, discover new metabolic
gene functions, guide adaptive evolution approaches, and optimize the production of value-
added compounds, e.g., pigments and lipids (Espinosa-Gonzalez et al., 2014). Several
genome-scale reconstructions for photosynthetic organisms have been previously
generated. The most comprehensive of these, in terms of coverage of the genome, are for
the plants *Zea mays*, *Brassica napus*, and *Arabidopsis thaliana*, for the microalgae
*Chlamydomonas reinhardtii*, *Ostreococcus* sp. and *Tisochrysis lutea*, and for the bacteria
*Synechocystis* sp. PCC6803 and *Cyanothece* sp (Kim et al., 2012; Baroukh et al., 2015).
Additionally, smaller networks have recently emerged for *Chlorella* sp. FC2 IITG,*C.
protothecoides, C. variabilis, and C. pyrenoidosa. These smaller networks are limited in scope and contain only core carbon and photosynthetic metabolism, with an average of 500 reactions and 390 metabolites (Yang et al., 2000; Muthuraj et al., 2013; Wu et al., 2015; Juneja et al., 2016). Here we report the reconstruction of a genome-scale metabolic network for C. vulgaris UTEX 395 (iCZ843) based on the recently assembled genome sequence.

RESULTS

Reconstruction of the Chlorella vulgaris UTEX 395 metabolic network

Using the genome annotation for Chlorella vulgaris UTEX 395, as well as an existing and manually-curated genome-scale model for Chlamydomonas reinhardtii (iRC1080 (Chang et al., 2011)), a preliminary draft reconstruction for C. vulgaris UTEX 395 based on protein homology was assembled.

This draft reconstruction was subsequently subjected to an iterative manual curation process as depicted in Figure S1. During the conversion of the curated reconstruction to a mathematical model, QC/QA tests were performed according to established standards (Thiele and Palsson, 2010). We identified and resolved thermodynamically-infeasible reactions in the model that were capable of producing energy in the form of ATP, NADPH, or NADH without allowing carbon or photon input into the system. To guide this step we visualized flux-carrying reactions using the web application Escher for building, sharing, and embedding data-rich visualizations of biological pathways (King et al., 2015). In most cases, the addition of protons or cofactors was required to prevent these thermodynamically-infeasible energy producing reactions.

Non-enzymatic reactions were included in the model following the workflow for high quality reconstructions previously described (Thiele and Palsson, 2010) and were
associated with KEGG identifiers. A mechanistic representation of photosynthesis in *C. reinhardtii* (Chang et al., 2011), a closely related species, was adapted for the model for *C. vulgaris* UTEX 395 (*iCZ843*).

Once all the reactions were introduced into the model, mass and charge balances were checked computationally. All metabolic and transport reactions in the final model are mass balanced, except for the poorly-studied reactions related to metarhodopsin hydrolase and rhodopsin retinyltransferase. In these cases, an unknown part of the metabolite formula is present, which causes an imbalance in the reaction. However, the presence of this unknown moiety does not appreciably affect simulations of *iCZ843*. Complete lists of reactions and metabolites annotation are given in Supplementary Tables S1 and S2.

Refinement and gap analysis

Model refinement and gap analysis were performed after the conversion to a mathematical model to preserve metabolic pathway connectivity. Reactions identified during the gap filling process were categorized into different classes (non-GPR associated, non-enzymatic or spontaneous, and multi-step incomplete reactions) (see Supplementary Table S3). A detailed protocol used for gap filling is described in the supplementary information. Most of the non-GPR associated reactions were added in subsystems related to central carbon metabolism. Bioinformatics tools were used to predict the cellular compartmental location of lipid metabolism reactions (Table I). Retinol metabolism was placed in the chloroplast based upon predicted subcellular localization of proteins and the lack of an eyespot in *C. vulgaris*. We identified gene associations for 35% of the approximately 350 transport reactions. Non-gene associated reactions were included in the model using standard mechanisms, such as passive transport, transport linked to energy consumption, and salt efflux and influx, that were associated with transport of particular compounds. The
transport reactions for monocarboxylates, dicarboxylates, tricarboxylates, amino acids, and nucleic acids were included based on previous reports (Hanson, 1985). For more details see Supplementary Text.

**Model properties**

The model for *C. vulgaris* UTEX 395 (*iCZ843*) consists of six compartments: the cytoplasm, mitochondrion, chloroplast, thylakoid, glyoxysome, and the extracellular space. The model was tested for the ability to grow under known physiological growth conditions of *C. vulgaris*.

The detailed features of *iCZ843* are given in Figure 1. The model contains 843 out of 7100 annotated genes (around 12%), delineating 1770 metabolites and 2294 reactions. The largest subsystems in the reconstruction are amino acid metabolism and lipid metabolism. Other pathways with annotated GPRs include autotrophic metabolism, such as thiamine and biotin metabolism, and amino acid degradation, as well as ubiquinone, terpenoids, α-linolenic acid, and brassinosteroid biosynthesis. As shown in Table II, core models of various complexities have been generated for three Chlorella species (*C. sp. FC2 IITG, C. protothecoides*, and *C. pyrenoidosa*) and a genomic scale model for *C. variabilis* (Table II). A detailed comparison by subsystem with other photosynthetic models is shown in Table S4; the corresponding metabolic reactions in *iRC1080* and *iCZ843* are in Figure 1C.

**Defining the biomass objective function**

The biomass reaction accounts for all known biomass constituents in terms of their fractional abundance per gram of biomass. The biomass objective function (BOF) in *iCZ843* contains the stoichiometric coefficients for 140 metabolites, 39 measured and 101 estimated from the fatty acids profile.
The lipids produced by *C. vulgaris* can be grouped in the esters: triacylglyceride (TAG), phosphatidylglycerol (PG), phosphatidylinositol (PI), phosphatidylethanolamine (PE), sulphoquinovosyl diglyceride (SQDG), monogalactosyl diglyceride (MGDG), digalactosyl diglyceride (DGDG), and phosphatidylcholine (PC), where TAG can accumulate to 20–60% of dry cell weight (Guarnieri et al., 2011; Guarnieri et al., 2013; Bellou et al., 2014). Fatty acids in these esters commonly fall into seven major groups - C16:0, C16:1, C16:2, C16:3, C18:1, C18:2 and C18:3 (Nakamura and Imamura, 1985; Bellou et al., 2014). We also measured 14:0 and 18:0 fatty acid content for *C. vulgaris* UTEX 395 under the experimental conditions used in this study. The metabolic model encompasses pathways to synthesize each of these lipids. Using the experimentally determined fatty acids values, a theoretical categorization of the groups of lipids was conducted according to a previous study (Nichols et al., 1967).

The abundance of the groups of metabolites defined in the BOF are shown in Figure 1D. The light and dark conditions produce a difference in requirements of biomass precursor metabolites. Although metabolites specified in the BOF are the same under photoautotrophic and heterotrophic growth, experimental data revealed differences between the stoichiometric coefficients under these conditions (see Figure 2A). A ratio for each metabolite between both conditions was calculated. While the amino acid content is almost identical in light or dark, carbohydrates are accumulated under light conditions and excess carbohydrates are stored as starch.

Another useful way to constrain a metabolic model is to block the flux through enzymes that are not expressed under a certain condition. RNA-seq data revealed high expression of genes encoding ribulose-bisphosphate carboxylase (EC 4.1.1.39), phosphoribulokinase (EC 2.7.1.19), and NADP-dependent phosphorylating glyceraldehyde-3-phosphate...
dehydrogenase (EC 1.2.1.13) under light but not under growth in the dark (see Supplementary Table S5). This information was used to modify the flux bounds of related reactions depending on the experimental condition.

**Network evaluation and robustness**

The uptake rates used to constraint the model were calculated using our experimental data (see Supplementary Text, Figure S2 and Table S6). iCZ843 is able to simulate photoauto-, hetero-, and mixotrophic conditions according to Tables S7-S10. These condition-specific models are provided in the Supplementary information (SBML format).

Model functionality was validated by simulating growth under different conditions using flux balance analysis (Orth et al., 2010). Growth is dependent on the carbon and nitrogen source (organic as well as inorganic) and on the presence or absence of light. The model accurately predicts the consumption and release of metabolites from or into the medium.

For example, simulated nitrate uptake rates under photoautotrophic and heterotrophic conditions were 0.068 and 0.044 mmol h⁻¹, agreeing well with experimental data (0.063 and 0.049 mmolₙitrate h⁻¹). The simulated growth rates for photoautotrophic (0.025 h⁻¹), heterotrophic (0.016 h⁻¹), and mixotrophic growth (0.0419 h⁻¹) are consistent with our experimental data, Table III shows a comparison between experimental and predicted data for Chlorella and Chlamydomas.

The model demonstrated good accuracy for growth/no-growth prediction on 74 carbon sources and 53 nitrogen sources. Matthews correlation coefficients (MCC) were estimated for each conditions (Figure 3). Across carbon sources, the highest prediction accuracy was observed for photoautotrophy (MCC = 0.51), followed by heterotrophy, the positive predicted values ranged 0.77 – 0.85 for all the conditions (see Supplementary Table S11
and S12). iCZ843 was able to accurately predict the presence or absence of growth across several nitrogen sources as well, with a MCC of 0.68. The false negatives appear primarily to occur within particular subsystems, such as amino acid metabolism in charge of the degradation of L-homoserine, agmatine and putrescine, where the GPR associations were uncertain. Additionally, some of the false positive values under mixotrophic conditions appear to be due to the CO₂ uptake rate allowed to the model, true and false negatives were not predicted under mixotrophy due to the supply of CO₂ and light into the model.

It was found that certain enzymes such as mitochondrial cytochrome C peroxidase (EC 1.11.1.5), ferredoxin-NADP⁺ reductase (EC 1.18.1.2), ethanolamine kinase (EC 2.7.1.82), and formate dehydrogenase (EC 1.2.1.2) have a high impact on the flux through the BOF and are essential for a functional model (Figure S3).

Robustness analysis for cytochrome C peroxidase shows an inverse relationship between the reaction flux increase and the growth rate. Here, a flux of 2 mmol g_{DW}^{-1} h^{-1} results in a lethal phenotype under heterotrophic growth. According to our analysis, the enzymes ethanolamine kinase, formate dehydrogenase and ferredoxin-NADP⁺ reductase show a similar sensitivity; in all the cases, lower or higher fluxes result in a lethal phenotype.

**Flux distributions**

*Carbon and amino acid metabolism*

A general overview of the reconstruction, the connections, and transport of compounds is shown in Figure 4. The heat map shows the percentage of total flux by compartment. Negative and positive values represent the total mass shuttled outside or inside of the compartments, respectively, cytoplasm is considered as reference in order to establish
connections between the extracellular space and the rest of the organelles. Fluxes showed different behavior in light and dark conditions, cytoplasm and extracellular space are highly active in light conditions (photoauto- and mixotrophy); while under heterotrophy mitochondria, cytoplasm, and chloroplast have similar activity. The chloroplast has significant metabolic activity even in the dark due to the presence of some lipid and starch synthesis pathways in this compartment.

Under photoautotrophic and mixotrophic growth, CO₂ is shuttled from the cytoplasm to the chloroplast, while the flux is reversed during heterotrophy. Mitochondria provide CO₂ to the cytoplasm under all the conditions but this flux doubles when glucose is present in the culture medium. When *Chlorella vulgaris* UTEX 395 grows on NO₃ as nitrogen source, the synthesis of NH₄ is carried out intracellularly through six reactions. We found that NH₄ is imported to the chloroplast in light conditions, where it is used in the amino acid and nucleotide biosynthesis, but under heterotrophy NH₄ is released to the cytoplasm allowing the use of NH₄ in other pathways and organelles.

The ¹³C-tracing metabolic flux data set previously reported for *C. protothecoides* was used to validate the *i*CZ843’s flux distributions (Wu et al., 2015). The Pearson correlation \( R^2 \) in photoautotrophy was 0.85 and 0.77 for heterotrophy (see Supplementary Figure S4 and S5), the reactions were grouped by subsystem, where the flux distribution related to energy, carbon fixation, exchange and transport reactions were predicted accurately, some reactions flux in glycolysis, fructose metabolism and TCA cycle were sub-estimated by the model, the highest complexity of *i*CZ843 does not allow the direct comparison with the experimental data due to either the presence of the same reaction in several compartments, or the curtailment of pathways in just one experimental value, the reactions correspondence...
between iCZ843 and the reactions in *C. protothecoides* model is in the Supplementary Tables S13 and S14.

One of the highest flux pathways in iCZ842 was the TCA cycle, which contains 22 reactions. While most enzymes are highly active under heterotrophic growth, only 8 reactions are active under photoautotrophic and mixotrophic growth. The cytoplasmic enzymes aconitate hydratase (EC 4.2.1.3), isocitrate dehydrogenase (EC 1.1.1.42), mitochondrial citrate synthase (EC 2.3.3.1) and fumarate hydratase (EC 4.2.1.2), and chloroplastic malate dehydrogenase (EC 1.1.1.37) are active under the three conditions. On the other hand, the mitochondrial isocitrate dehydrogenase (EC 1.1.1.41) and chloroplastic malate and isocitrate dehydrogenase (EC 1.1.1.37, 1.1.1.42) only carry flux during heterotrophic growth.

The flux distributions of the pentose phosphate pathway (PPP) are similar under photoautotrophic and mixotrophic conditions but with lower activity under heterotrophy. Next we examined a set of enzymes within pyruvate metabolism, namely pyruvate decarboxylase (EC 4.1.1.1) and pyruvate dehydrogenase (EC 1.2.4.1; 2.3.1.12), which connect glycolysis/glucanogenesis, carbon fixation and alanine and aspartate metabolism. iCZ843 predicts these reactions with higher activity under photoautotrophic and mixotrophic compared to heterotrophic conditions, these fact is linked to greater growth efficiency at these conditions.

Phenylalanine, tyrosine, and tryptophan biosynthesis is an active subsystem powered by the phosphoenolpyruvate and acetoacetyl-CoA, key precursors in central carbon metabolism and lipid synthesis. Under mixotrophic growth, fluxes in this subsystem increased by around 25% during photoautotrophy and 60% during heterotrophy. The aminotransferase enzymes (EC 2.6.1.1) are highly active during photoautotrophic, heterotrophic, and
mixotrophic growth. These enzymes catalyze the synthesis of phenylalanine and tyrosine through α-ketoglutarate/glutamate, or the assimilation into hydroxyphenyl pyruvate, which participates in terpenoid and secondary metabolite biosynthesis. The flux distribution within glutamate and other amino acids metabolism is laid out in Figure S6. The predicted fluxes reveal the connectivity of glutamate in the cytoplasm with the synthesis of several amino acids (proline, histidine, valine, and arginine) in both light and dark conditions. Chloroplastic glutamate is used by amino acid transferases (EC 2.6.1.42) for the synthesis of isoleucine and tyrosine in the chloroplast under mixotrophic and photoautotrophic conditions.

**Pigment metabolism**

Certain pigments are synthesized in large quantities during heterotrophic growth. Members of the class Chlorophyceae contain α- and β-carotenes as well as the xanthophylls lutein, zeaxanthin, violaxanthin, and neoxanthin, all of which were included in the BOF of iCZ843. The model indicates that all of these pigments can be synthesized under either light or dark conditions.

The flux distribution showed that enzymes such as cryptoxanthin hydroxylase (CXHY EC 1.14.99.45) and zeinoxanthin forming α-carotene hydroxylase (CHYA2 EC 1.14.13.-) carry flux in both light and dark. During growth conditions lutein, loraxantin, and zeaxanthin are produced through the enzyme CHYA2 (EC 1.14.13.-), which condenses α-carotene to cryptoxanthin using NADPH. Ultimately, zeaxanthin is converted into lutein by CXHY and eventually to loraxanthin. The flux through CHYA2 is responsible for the large concentration of lutein observed in *C. protothecoides* growing mixotrophically (Shi and Chen, 1999). The pigment biosynthesis pathways also revealed changes in flux activity depending on the simulated conditions (see Figure S7).
Shadow prices

Shadow price analysis predicts the effect a single modification to the culture medium has on the objective function at the conditions studied, in this case growth rate. We estimated shadow prices for all 20 amino acid substitutions to the medium, as well as for acetate and glycerol addition and confirmed the predictions experimentally.

The comparison of growth rates under heterotrophy and estimated shadow prices are shown in Figure 5. Assuming an uptake rate of 1 mmol gₘ⁻¹ h⁻¹ the model predicted that addition of tryptophan or methionine leads to an increase in growth rate (tryptophan 0.040 h⁻¹, methionine 0.032 h⁻¹).

These hypotheses were tested experimentally by adding equimolar concentrations (10 mM) of tryptophan and methionine to the medium. The growth rate improved for both tryptophan (0.038 h⁻¹) and methionine (0.024 h⁻¹) addition (see Figure 5C). Addition of acetate to the medium was used such as negative control, and it had no effect on the growth rate in heterotrophic conditions.

DISCUSSION

Reconstruction

Due to recently developed tools, the creation of draft genome-scale models for bacteria is a largely automated process (Overbeek et al., 2014). However, for eukaryotes, this automation procedure often proved to be difficult, because of the complex compartmentalization present in these organisms. One tool that enables the automated generation of draft reconstructions suitable for eukaryotes is the RAVEN Toolbox (Agren et al., 2013). In contrast to what has been reported for other Chlorella reconstructions (Wu et al., 2015), we were able to find gene evidence for enzymes and transporters associated with amino acid metabolism, as well as steroids, porphyrin and chlorophyll metabolism.
Vitamins, ubiquinone, terpenoids, and brassinosteroid biosynthesis have also been included into the model, which among all photosynthetic organisms appears only in the *Zea mays* reconstruction (Saha et al., 2011).

Considering the genome size and number of genes in the reconstruction *iCZ843* is the most comprehensive model for any eukaryotic photosynthetic organism to date, as well as the detail of lipid metabolism. Previously, a comprehensive reconstruction of lipid biosynthesis has only been available for *C. reinhardtii* and *Synechocystis* sp. PCC6803, both unsuitable candidates for biofuel production because of their inability to accumulate high concentrations of TAGs (Chang et al., 2011; Saha et al., 2011). *C. vulgaris* has the ability to produce high amounts of triacylglyceride (TAG), which makes it a prime model organism for elucidating biofuel production.

Furthermore, amino acid biosynthesis has been reconstructed in detail for *C. vulgaris*. Amino acid metabolism in *iCZ843* contains 282 reactions and the total related reaction in other subsystems account for 312, compared with 34 reactions in *C. protothecoides*, 256 in *C. reinhardtii*, and 143 reactions in *Synechocystis* sp. (see Table S4) (Chang et al., 2011; Saha et al., 2011; Overbeek et al., 2014).

It is noteworthy that models for phototrophic bacteria, such as the cyanobacterium *Synechocystis* sp. PCC6803, often contain a larger fraction of genes per genome (see Table II). This discrepancy between kingdoms is due to the fact that functional genome annotations for bacteria are in general more comprehensive. The size and metabolic scope of *iCZ843* allow more detailed and complex simulations. *C. vulgaris* UTEX 395 drastically adapts its metabolism in response to changes in the environment, such as growth in light or dark, or availability of a nitrogen source (e.g.,
nitrate). These metabolic changes affect the biomass composition, primarily consisting of carbohydrates, proteins, lipids, and nucleic acids. We therefore measured changes of total carbohydrates, proteins, lipids, and RNA during photoautotrophic and heterotrophic (with glucose) growth with nitrate as the nitrogen source (Figure S2).

In other Chlorella species, such as C. zofingiensis, starch has been found to be the main stored carbohydrate accounting for 66.7% of total biomass (Zhu et al., 2014). Other studies with C. vulgaris showed a considerable breakdown of starch even during photosynthesis (Nakamura and Imamura, 1985).

We found a slightly higher content of TAG, PG, PI and PE under heterotrophic growth conditions, while the MGDG content was high during photoautotrophic growth. This could be related to the production of pools of apoproteins, such as those in the photosystem II complex that require lipids for their function (Thompson, 1996).

**Simulations using iCZ843**

Simulations confirmed that the metabolism of C. vulgaris UTEX 395 is redirected towards the synthesis of storage compounds like lipids and carbohydrates under nitrogen starvation (Guarnieri et al., 2013). The uptake of nutrients from the medium is in agreement with previously reported values as well as our experimental data, showing good accuracy in growth predictions for more than 120 carbon and nitrogen sources. The predictions on glucose confirm that the specific growth rate for mixotrophy can be represented approximately by the sum of heterotrophic and photoautotrophic growth rates (Killam and Myers, 1956; Perez-Garcia et al., 2011).

The efficiency of a microbial production process can be evaluated based on high titer, high yield, high productivity, and process robustness (Liu et al., 2013). Robustness analysis
allows the evaluation of the essentiality of the function of enzymes within the network. It correlates the enzyme-pathway relationship and the overall growth rate and can identify targets for metabolic engineering, a set of reactions that have the most profound effect on growth were identified.

**Flux distribution**

Genome-scale metabolic models help explain how targeted compounds can be synthesized through all possible routes. The predictions of *i*CZ843 provide metabolic details that correspond to the physiology of *C. vulgaris* and green algae.

The PPP and glycolysis have been fairly well documented in *C. vulgaris* (Baalan et al., 1973). According to the model the oxidative phase of the PPP, responsible for production of NADPH, is inactive under heterotrophic conditions. Glycolysis and gluconeogenesis have been intensively studied in *C. vulgaris* (Perez-Garcia et al., 2011). The degradation of glucose to pyruvate occurs throughout the cytoplasm, the mitochondria, and the chloroplast. It is known that central carbon metabolism is active in different compartments depending on the growth conditions (Hanson, 1985). The model shows interchanges of products and substrates from the active reactions uniquely present in the chloroplast with those functioning in the cytoplasm and mitochondria.

The model contains the reaction ACCOAth (gene associated to maker_Scaffold_645-augustus-gene-0.44), which facilitates the transport of acetyl-CoA from the cytoplasm to the chloroplast. When this reaction is deleted (knocked-out) under heterotrophy, the predicted growth rate decreases by 80%. After delving into the effects of some transporters on metabolism as a whole, we found that the flux of acetate from the chloroplast to the
cytoplasm correlates with the acetyl-CoA transport, and is vitally important for a viable growth prediction.

Acetate transport under dark conditions plays a very important role in green algae metabolism. This metabolite can be incorporated into acetyl coenzyme A (acetyl-CoA) following two possible pathways that both require ATP: a direct conversion with acetyl-CoA synthetase (ACS) or a two-step reaction involving acetate kinase (ACK) and phosphate acetyltransferase (PTArh) (Johnson and Alric, 2013). Analyzing our predicted flux distribution under heterotrophy enabled us to conclude that in order to get enough acetate into the cytoplasm, the same amount of acetyl-CoA need to be transported into the chloroplast. Then, a cascade effect activates chloroplastic enzymes such as PTArh and pyruvate synthase (PYRShi), where ferredoxin is reduced. Eventually ferredoxin will be precursor in the synthesis of NH₄, which is one of the main precursors needed for the synthesis of amino acids. The cytoplasmic acetate is incorporated into TCA cycle and lipids biosynthesis, and the ATP necessary for these reactions is produced in the mitochondria by the ATP synthase complex (ATPSm).

Furthermore, calculations obtained by iCZ843 agree with experimental ¹³C data. We found highest flux correlation with experimental data in photoautotrophy followed by heterotrophy. The discrepancies in the flux distributions were attributed to: 1) the condensed nature of the experimental data, where explicit comparison of predictions to data was not possible due to the presence of the same reaction in several compartments in the model, 2) the presence of anaplerotic (re-filling) reactions, which replenishes TCA cycle intermediates consumed in amino acid, lipids and nucleotide synthesis, allowing many alternate flux solutions, although the predicted and experimental data follow similar
behaviors. For heterotrophy we found also prediction errors with reactions related to carbon
fixation and pentose and fructose metabolism.

Glutamine, glutamate, asparagine, and aspartate contribute as building blocks for the
synthesis of organic nitrogen compounds such as other amino acids, chlorophylls,
nucleotides, alkaloids, and polyamines. Alternately, we found that glutamate needs to be
converted to glutamine and exported from the chloroplast to the cytoplasm in the dark,
where it is subsequently used for nucleic acid synthesis or incorporated in the TCA cycle.

It is well known that the enzyme glutamate ammonium ligase (EC 6.3.1.2) uses glutamine
and NH$_3$ as substrates. The direct reduction of nitrate to NH$_4$ and the availability of
 glutamine have been observed in chloroplasts during heterotrophy (Perez-Garcia et al.,
2011). The role of the chloroplast in the dark has been traced and attributed to the qualities
of the common ancestor, but a detailed experimental description of its metabolism has not
been reported in either plants or green algae. Some reports have proposed a shuttle of
metabolites between mitochondria, peroxisome, and chloroplast to fuel the glycolate
pathway or central carbon metabolism, but the transport of amino acid between different
 compartments is still poorly understood (Baalan et al., 1973; Hanson, 1985; Ren and
Paulsen, 2007). iCZ843 provides novel insights into the amino acid metabolism in
Chlorella. Several enzymes involved in amino acid metabolism were only found active
when carbon was redirected mainly to lipids production. The model not only enables the
study of metabolism at the reaction level, but the detailed compartmentalization of the
model additionally facilitates the analysis of enzyme activity in different organelles of the
cell under various conditions.

The three available core models of Chlorella (see Table II) assume that the fluxes differ
between light or dark only by the layout of the core carbon network (TCA, glycolysis, and
PPP), and the rest of the metabolism (synthesis of amino acids, DNA, and RNA) does not vary significantly in terms of relative fluxes (Yang et al., 2000; Muthuraj et al., 2013; Wu et al., 2015). Despite previous reports that assume the lack of change in the flux distribution for nucleotide metabolism (Yang et al., 2000; Muthuraj et al., 2013; Wu et al., 2015), we found that fluxes vary widely depending on the growth condition. Our results suggest that the anabolic part of the metabolism is not independent of growth conditions, and iCZ843 provides an accurate qualitative and qualitative analysis of the metabolism.

**Guided medium alteration**

An increase in growth rate will deplete nitrogen faster, which in turn will get you to a lipid accumulation state faster. The predicted and experimentally-determined growth rates for *C. vulgaris* UTEX 395 containing amino acid additions are 2-3 times higher than in regular medium. The growth rate for tryptophan-supplemented cultures can be attributed to its degradation to formate and kynurenine, which is further metabolized to generate alanine and eventually acetyl-CoA. Tryptophan catabolic products also can fuel the synthesis of arginine and serine, nucleotides, as well as NAD and NADP. Methionine is a source of methyl groups for a number of cellular components (e.g., for S-adenosylmethionine, purine and pyrimidine intermediary) and is used in cell wall formation. Methionine had been intimately associated with cell division and some analogues of methionine prevent cells of *C. vulgaris* from dividing, while allow them to maintain other cellular activities that can lead “giant cells”. Analysis of these abnormally large cells has revealed that there are increases in the rate of respiration, dry weight, and protein content (Shrift, 1960).

**CONCLUSIONS**
An in-depth understanding of metabolism is necessary to improve the production of desired products, such as nutraceuticals or biofuels, by microalgae. Genome-scale network reconstructions combine detailed biochemical and physiological information of an organism and provide new insights into growth conditions and subsequent manipulation strategies to enhance productivity. The final *C. vulgaris* UTEX 395 reconstruction contains 843 genes, 2,294 reactions, and 1,770 metabolites. The reconstruction was constrained during the validation process using transcriptomics and other experimental data under photoautotrophic, mixotrophic, and heterotrophic growth conditions. iCZ843 can accurately simulate the growth rates under these conditions. The model was successfully deployed to guide strategies altering the culture medium for increased growth performance.

**METHODS AND MATERIALS**

**Draft generation**

The draft was generated using the RAVEN Toolbox (Agren et al., 2013). As input, we provided the translated genome sequence of *C. vulgaris* UTEX 395 and the manually curated reference network of the microalgae *C. reinhardtii* (iRC1080, (Chang et al., 2011)) which was used as reference network. The *C. vulgaris* UTEX 395 genome sequence was taken from the previously deposited sequence (GenBank LDKB00000000). Several resources were used during the manual curation phase, such as primary literature and the databases KEGG, EMBL-EBI, ExplorEnz, BIGG, BRENDA, MetaCyc, and SwissProt (McDonald et al., 2007; Schellenberger et al., 2010; Scheer et al., 2011; McWilliam et al., 2013; Caspi et al., 2014; Consortium, 2014; Kanehisa et al., 2014). All the resources used during the reconstruction process are highlighted in Table I. Information regarding transport proteins was obtained from TransportDB and TCDB (Ren and Paulsen, 2004).
Subcellular protein localization was predicted using SignalP, ChloroP, HECTAR and WolF PSORT (Emanuelsson et al., 1999; Horton et al., 2007; Gschloessl et al., 2008; Petersen et al., 2011) (Table I).

**Metabolic network reconstruction**

The draft reconstruction was manually curated, following previously published protocols (Thiele and Palsson, 2010). The reconstruction was arranged using a KEGG pathway structure (e.g. systems: amino acid metabolism; subsystems: lysine biosynthesis). Metabolites and reactions were identified using KEGG IDs as well as EC codes for enzymes. For every metabolite, the elemental formula and charge were included in the annotation. The protonation state of each metabolite changes according to the pH of the compartment. The cytosolic pH was determined to be 7.2, at extracellular pH 6.5 (Komor and Tanner, 1974). The chloroplast and its sub-compartment the thylakoid, were assumed to share the same pH determined for the chloroplast to be 8.0 in light conditions (Hogetsu and Miyachi, 1979; Goss and Garab, 2001). The pH of the mitochondrial matrix has been measured in green algae at 7.8 (Giordano et al., 2003). The glyoxysome pH was assumed to be 8.2 (Dansen et al., 2001). The extracellular pH was 7.0±0.2 based on growth medium used. Metabolite charges and protonation states were calculated using ChemAxon (Marvin, 2015) (Table I).

New pathways and their annotation were added. Names for reactions and metabolites were assigned according to the information in BIGG and Sympheny (Schilling et al., 2008; Schellenberger et al., 2010); when the reactions and metabolites were not found in any database, a new naming was provided. Each pathway was manually curated; mass and charge balance, directionality, and cofactors involved in the reactions were accounted for (Figure S1). Initially, reactions in the draft reconstruction were imported keeping the same
location as in iRC1080. All information for these reactions was double checked and in some cases the reactions were relocated to a different compartment. The manual gene/protein/reaction (GPR) associations were set using BLAST to compare the protein sequences of *C. vulgaris* UTEX 395 to the corresponding sequences of *C.* sp. NC64A, *C.* variabilis and *C.* sorokiniana, as well as (in some cases) with the green algae *Coccomyxa subellipsoidea* and *Volvox carteri* (Altschul et al., 1990). The phylogenetic tree shows the relationship between these photosynthetic organisms (Figure S8).

Once the manual refinement was finished and the annotation of the model was completed, gap filling and dead-end identification were executed using the available COBRA tools (Schellenberger et al., 2011). The connectivity of the pathways were ensured by the addition of enzymes without gene associations but backed by literature evidence. When the missing transporter was detected, passive diffusion was assumed if the exact transport mechanism was unknown. Finally, quality control and assessment (QC/QA) tests for ATP, NADPH, and NADH maintenance were performed. These tests ensured that the model cannot maximize for any of the energy cofactors (ATP, NADPH, and NADH) without any input; the objective function has an expected result of zero. Analysis of loops and plot maps were done using ESCHER metabolic pathway visualization tool (King et al., 2015). The reconstruction was converted to JSON model for ESCHER using COBRApy (Ebrahim et al., 2013).

**Biomass composition and experimental data**

*C. vulgaris* can grow under different trophic conditions (e.g. auto-, hetero-, and mixotrophic). Each of these was represented mathematically through different biomass objective functions (BOF) (named in the model: Biomass_Cvu_auto-, Biomass_Cvu_mixo- and Biomass_Cvu_hetero-). Every equation contains the stoichiometric coefficients...
expressed in mmol g_{DW}^{-1} and all metabolites that are part of the biomass should be
included. For the BOF under mixotrophy we assumed the same data as for heterotrophy.
ATP maintenance was established according to (Boyle and Morgan, 2009).
Lipid, protein, carbohydrates and ribose in RNA contents were measured experimentally
under photoautotrophic and heterotrophic conditions following the method described
previously (Antoniewicz et al., 2007; Antoniewicz et al., 2011; Long and Antoniewicz,
2014). *C. vulgaris* UTEX 395 was grown in a 250 mL bottle with 200 mL of Bold’s Basal
medium (BBM) at 24°C (Guarnieri et al., 2013), and cycling of 14:10 h light:dark, at10,000
lux and a gas flow of 1% CO₂ (12 mL min⁻¹). For heterotrophic condition, *C. vulgaris* UTEX 395 was grown in a 500 mL of Bold’s Basal medium (BBM)+20mM Tris and a gas flow of air (400 mL min⁻1) using 6, 10, 28, 110 mM of glucose as organic
carbon source at 24 h dark. The urea uptake rate was measured in heterotrophy and
mixotrophy, testing an initial concentration of 1 mM and glucose 1%.
The amino acids content was completed with literature data: arginine, cysteine, glutamine
and tryptophan under photoautotrophy were obtained from (Faheed and Fattah, 2008);
cysteine content in heterotrophic was taken from (Wu et al., 2015).
The quantified fatty acids were spread out according to previously reported lipid profiles
(Nichols et al., 1967). The relation of nucleotides composition (28 RNA/DNA) was taken
from (Muthuraj et al., 2013). Literature and experimental data were normalized assuming
an idealized size and weight for *C. vulgaris* (data shown below). The specific starch
production and degradation rates were calculated from the experimental data, the curves
were adjusted to the Gompertz model using Kaleidagraph Synergy Software Inc. (Nichols
The CO₂ consumption rates and the chlorophylls a and b were measured under photoautotrophic conditions; experiments were performed in a 1.25 L reactor with 500 mL of BBM supplemented with 20 mM Tris, a total gas flow of 400 mL min⁻¹ with either 0.04, 3.0, 5.0, 10.0 and 12.5 % of CO₂, at an agitation rate of 200 rpm at ~25°C (room temperature) at light:dark cycles of 12:12 h and a light intensity of 10,000 lux (~300 µE m⁻² s⁻¹). Additional compositions of pigment were taken from (Safi et al., 2014).

The addition of amino acids was tested for heterotrophic conditions, adding to the culture medium equimolar concentrations, 0.01 M of glucose and either tryptophan or methionine.

Constraints and growth simulations

RNA-seq data and literature data such as specific CO₂ and glucose consumption rates were used to constrain the model under growth conditions in the light and dark (Guarnieri et al., 2011). For each growth condition, the storage and consumption of starch was taken into account; we calculated the rates using experimental data (Table S6).

The constraints related with mineral media composition and the reactions that were set to zero are summarized in the Supplementary Tables S7-S10. Growth simulations were performed in the COBRA Toolbox for MATLAB (Schellenberger et al., 2011), using the Flux Balance Analysis (FBA) procedure (Orth et al., 2010). The stoichiometric coefficients in the BOFs were set according to our experimental data. The △¹³C data set was taken from (Wu et al., 2015). Model benchmarking on carbon and nitrogen sources was performed using Biolog plates PM1-3, following the previously reported protocol described by (Chaiboonchoe et al., 2014), with the following modifications. C. vulgaris was grown to mid-log phase in modified BBM media (Guarnieri, 2011), pelleted via centrifugation at 4,000 x g for five minutes, washed and resuspended in fresh media to a final OD=0.1 (nitrate was excluded for nitrogen phenotyping). 100uL aliquots were inoculated into
Biolog plates and examined for 96 h in the plate reader with readings every 15 minutes. Plates were housed in a plate reader with no light (heterotrophic growth). The plates for both the PM1 and PM2 plate (carbon sources) were run at 490nm to examine dye absorbance alterations and 750nm to assess optical density. A confusion matrix and various measures of quality such as accuracy, specificity, sensitivity Matthews correlation coefficient were estimated according with (Matthews, 1975) (see Tables S11 and S12).

**Robustness and shadow prices**

The robustness analysis was performed under photoauto-, hetero- and mixotrophic conditions. The respective analyses were executed using the same constraints. The CO₂ uptake rates were plotted using MATLAB; the experimental CO₂ uptake rates reported by (Nascimento et al., 2015) were used to validate the simulated results.

The sensitivity of the FBA solution can be indicated by shadow prices, which represent the change of the BOF with respect to the external exchange flux for all the metabolites, computing the theoretical addition of one mole of every metabolite present in the model (Palsson, 2011). Negative values describe metabolites that are demanded and positive values identify metabolites that would be excreted in order to improve the objective value. Shadow prices were calculated using the COBRA Toolbox (Schellenberger et al., 2011) in order to compare the effect of the addition of amino acids, glycerol and acetate compounds.

**Supplemental Materials**

Supplementary Text and Figures S1-S8 (pdf)

Supplementary Information, Table S1. iCZ843 model (xlsx)

Supplementary Information, Table S2. Metabolites annotation (xlsx)

Supplementary Information, Table S3. Reactions non-gene associated (xlsx)
Supplementary Information, Table S4. Comparison of *iCZ843* with the available green algae models (xlsx)

Supplementary Information, Table S5. GPR for enzymes sensitive to light (xlsx)

Supplementary Information, Table S6. Starch synthesis and degradation (xlsx)

Supplementary Information, Table S7. Bold’s mineral medium (xlsx)

Supplementary Information, Table S8. Photoautotrophic solution (xlsx)

Supplementary Information, Table S9. Heterotrophic solution (xlsx)

Supplementary Information, Table S10. Mixotrophic solution (xlsx)

Supplementary Information, Table S11. Experimental and predicted growth of *Chlorella vulgaris* on various carbon sources (xlsx)

Supplementary Information, Table S12. Experimental and predicted growth of *Chlorella vulgaris* on various nitrogen sources (xlsx)

Supplementary Information, Table S13. Reactions association for $^{13}$C-tracing analysis (xlsx)

Supplementary Information, Table S14. $^{13}$C-tracing analysis of *C. protothecoides* under photoautotrophic and heterotrophic conditions (xlsx)

Supplementary Information, Model in photoautotrophy (sbml)

Supplementary Information, Model in mixotrophy (sbml)

Supplementary Information, Model in heterotrophy (sbml)

ACKNOWLEDGEMENTS

We thank Jared Broddrick, Adam M. Feist for assisting with the reconstruction process and Bernhard Palsson (UCSD) for guidance. We are also thankful for the careful evaluation of the reconstruction by an anonymous reviewer which significantly improved of the model.
This material is based upon work supported by the National Science Foundation under Grant No. 1332344, the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences Energy Frontier Research Centers program under Award Number DE-SC0012658 and Cristal Zuniga by Mexican National Research Council, CONACYT, fellowship No. 237897.

LIST OF AUTHOR CONTRIBUTIONS

CZ, MJB, and KZ conceived and designed the study. CZ, TH, and JL performed the reconstruction. CTL, MTG, EPK, BOM, CPL and MRA performed the experiments and analyzed the data. CZ and DZ carried out the simulations and analysis. CZ and KZ wrote the manuscript with assistance from all co-authors. All authors have read and approved the manuscript.
Figures legends

**Figure 1. Overview of network properties.** (A) Metabolic reactions of the reconstruction by subsystems (KEGG metabolic pathway diagrams). (B) Distribution of metabolites and reactions in the reconstruction. Most metabolites (44%) and reactions (59%) are located in the cytoplasm, while the chloroplast contains 30% of all metabolites and 25% of the reactions. The reconstruction has 1,770 metabolites, 1,180 of which are unique. Metabolites present in all five organelles are H₂O₂, H₂O, H, NADP, NADPH, and O₂. (C) Unique and shared metabolic reactions between *C. vulgaris* (*i* CZ843) and *Chlamydomonas rheinhardtii* (*i*RC1080), the exchange, demand, sink and transport reactions were excluded. (D) Distribution of the biomass objective function (BOF) in *i* CZ843. The BOF (140 metabolites) contains 20 amino acids, 8 nucleotides, 5 carbohydrates, 6 pigments and 101 lipids: 35 Triacylglycerols (TAG), 8 Phosphatidylglycerol (PG), 4 Phosphatidylinositol (PI), 4 Phosphatidylethanolamines (PE), 10 Sulfoquinovosyldiacylglycerols (SQDG), 15 Monogalactosyldiacylglycerols (MGDG), 16 Digalactosyldiacylglycerols (DGDG), and 9 Phosphatidylcholines (PC).

**Figure 2. Network evaluation under different conditions.** (A) Experimental coefficients in the biomass objective function under photoautotrophic and heterotrophic growth, the ratio between the values in heterotrophy and photoautotrophy (H/PA) was calculated and displayed in the y axis. (B) Specific growth rate simulated for photoautotrophic growth at different CO₂ uptake rates shown in blue. The predicted photon input flux was 646 (µE m⁻² s⁻¹). Red circles indicate experimental values.

**Figure 3. Model benchmarking through different carbon and nitrogen sources.** (A) Predicted growth rate in photoautotrophy (light + compound), heterotrophy (compound) and mixotrophy (CO₂ + light + compound) for the 74 carbon sources, the glucose uptake
rate was used to constraint the model for all the metabolites; markers represent experimental data for each condition. (B) Comparison of predicted growth rates using 53 nitrogen sources, the model was constraint using the urea uptake rate and glucose 1% uptake rate. (C) Statistics of the predictions under heterotrophy, true positive (TP), true negative (TN), false positive (FP), false negative, and Matthews correlation coefficients (MCC). The full data set and statistics under the three conditions is shown in the Supplementary Tables S11 and S12.

**Figure 4. Model overview and metabolites exchange between different compartments.** Color-coded values (heat map) refer to the total flux by compartment (mmol g\textsubscript{DW}^{-1} h^{-1}), the color map shows a negative value when the input mass is higher than output and vice versa. The arrows show how the reactions are carried out, an arrow with two points mean reversible reaction, which can carry flux in both directions; the irreversible reactions have just one head pointing to the products. Every reaction and metabolite contain in their name an location identifier; e, extracellular space; c, cytoplasm; h, chloroplast; u, thylakoid; m, mitochondria, and x, glyoxysome. For the naming and abbreviation of reactions and metabolites see Table S1 and S2.

**Figure 5. Predicted growth rates and shadow prices analysis for heterotrophic growth.** (A) Growth rates were determined experimentally and simulated for four different glucose concentrations. (B) The heat map displays the theoretical growth rate after the addition of 20 individual amino acids as well as glycerol and acetate for heterotrophic, mixotrophic, and photoautotrophic condition. The highest growth rate was achieved while adding tryptophan. The data of heterotrophy are displayed with a conversion factor 10\textsuperscript{-1}. (C) Experimental and predicted growth rates for medium containing tryptophan or methionine.
Tables legends

Table I. Online resources used during the reconstruction and curation process.

Table II. Properties of the genome-scale model for *Chlorella vulgaris* (iCZ843) and other prokaryotic and eukaryotic organisms are shown for comparison.

Table III. Experimental and predicted growth rates reported for *Chlorella* and *Chlamydomonas*.
LITERATURE CITED


Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. BBA - Protein Struct 405: 442–451
Ren Q, Paulsen IT (2007) Large-scale comparative genomic analyses of cytoplasmic
866
867
production, processing and applications of Chlorella vulgaris: A review. Renew
868
Sustain Energy Rev 35: 265–278
869
Saha R, Suthers PF, Maranas CD (2011) Zea mays irs1563: A comprehensive genome-scale
870
metabolic reconstruction of maize metabolism. PLoS One. doi:
871
10.1371/journal.pone.0021784
872
873
874
875
Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Söhngen C, Stelzer M,
876
Thiele I, Schomburg D (2011) BRENDA, the enzyme information system in 2011.
877
Nucleic Acids Res 39: 670–676
878
879
and genomic knowledgebase of large scale metabolic reconstructions. BMC
880
Bioinformatics 11: 213
881
Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC,
882
883
metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc 6:
884
1290–1307
885
886
Computational Infrastructure for Systems Biology.
887
Shi XM, Chen F (1999) Production and rapid extraction of lutein and the other lipid-
888
soluble pigments from Chlorella protothecoides grown under heterotrophic and
889
mixotrophic conditions. Nahrung - Food 43: 109–113
890
891
510–515
892
893
894
production from genetically engineered microalgae: Future of bioenergy in Iran.
895
Renew Sustain Energy Rev 15: 1918–1927
896
Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale
897
metabolic reconstruction. Nat Protoc 5: 93–121
898
Thompson GA (1996) Lipids and membrane function in green algae. Biochim Biophys
899
Acta - Lipids Lipid Metab 1302: 17–45
900
901
Microbiol Biotechnol 79: 707–718
902
903
Gompertz model. Poult Sci 83: 847–852
904
905
analysis reveal systematic properties of an oleaginous microalga Chlorella
906
907
Xiong W, Li X, Xiang J, Wu Q (2008) High-density fermentation of microalga Chlorella
908
protothecoides in bioreactor for microbio-diesel production. Appl Microbiol
909
Biotechnol 78: 29–36
910
911
microagalcells under photoautotrophic, mixotrophic and cycliclight-autotrophic/dark-
heterotrophic conditions. Biochem Eng J 6: 87–102
### Table I. Online resources used during the reconstruction and curation process.

<table>
<thead>
<tr>
<th>Resource</th>
<th>Link</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Functional annotation, draft generation and subcellular localization</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAVEN Toolbox</td>
<td><a href="http://biomet-toolbox.org/">http://biomet-toolbox.org/</a></td>
<td>(Agren et al., 2013)</td>
</tr>
<tr>
<td>Phylogeny.fr</td>
<td><a href="http://phylogeny.lirmm.fr/">http://phylogeny.lirmm.fr/</a></td>
<td>(Dereeper et al., 2008)</td>
</tr>
<tr>
<td><strong>Manual curation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEGG</td>
<td><a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a></td>
<td>(Kanehisa et al., 2014)</td>
</tr>
<tr>
<td>EMBL-EBI</td>
<td><a href="http://www.ebi.ac.uk/intenz/">http://www.ebi.ac.uk/intenz/</a></td>
<td>(McWilliam et al., 2013)</td>
</tr>
<tr>
<td>ExplorEnz</td>
<td><a href="http://www.enzyme-database.org/">http://www.enzyme-database.org/</a></td>
<td>(McDonald et al., 2007)</td>
</tr>
<tr>
<td>BIGG</td>
<td><a href="http://bigg.ucsd.edu/">http://bigg.ucsd.edu/</a></td>
<td>(Schellenberger et al., 2010)</td>
</tr>
<tr>
<td>BRENDA</td>
<td><a href="http://www.brenda-enzymes.org/">http://www.brenda-enzymes.org/</a></td>
<td>(Scheer et al., 2011)</td>
</tr>
<tr>
<td>MetaCyc</td>
<td><a href="http://metacyc.org/">http://metacyc.org/</a></td>
<td>(Caspi et al., 2014)</td>
</tr>
<tr>
<td>TransportDB</td>
<td><a href="http://www.membranetransport.org/">http://www.membranetransport.org/</a></td>
<td>(Ren and Paulsen, 2007)</td>
</tr>
<tr>
<td>SignalP 4.1 Server</td>
<td><a href="http://www.cbs.dtu.dk/services/SignalP/">http://www.cbs.dtu.dk/services/SignalP/</a></td>
<td>(Petersen et al., 2011)</td>
</tr>
<tr>
<td>ChloroP 1.1 Server</td>
<td><a href="http://www.cbs.dtu.dk/services/ChloroP/">http://www.cbs.dtu.dk/services/ChloroP/</a></td>
<td>(Emanuelsson et al., 1999)</td>
</tr>
<tr>
<td>HECTAR</td>
<td><a href="http://www.sb-roscoff.fr/hectar/">http://www.sb-roscoff.fr/hectar/</a></td>
<td>(Gschloessl et al., 2008)</td>
</tr>
<tr>
<td>WoLF PSORT</td>
<td><a href="http://www.genscript.com/psort/wolf_psort">http://www.genscript.com/psort/wolf_psort</a></td>
<td>(Horton et al., 2007)</td>
</tr>
<tr>
<td>TCDB</td>
<td><a href="http://www.tcdb.org/browse.php">http://www.tcdb.org/browse.php</a></td>
<td>(Saier et al., 2014)</td>
</tr>
<tr>
<td><strong>Conversion to mathematical model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SimPheny, Version 3.3</td>
<td><a href="http://systemsbiology.ucsd.edu/Downloads/SimPhenyPerlApplication">http://systemsbiology.ucsd.edu/Downloads/SimPhenyPerlApplication</a></td>
<td>(Schilling et al., 2008)</td>
</tr>
<tr>
<td>COBRA Toolbox 2.0</td>
<td><a href="http://sourceforge.net/projects/opencobra/files/">http://sourceforge.net/projects/opencobra/files/</a></td>
<td>(Schellenberger et al., 2011)</td>
</tr>
<tr>
<td>ESCHER</td>
<td><a href="http://escher.github.io/">http://escher.github.io/</a></td>
<td>(King et al., 2015)</td>
</tr>
<tr>
<td>COBRApy</td>
<td><a href="https://github.com/opencobra/cobraPy/blob/master/INSTALL.md">https://github.com/opencobra/cobraPy/blob/master/INSTALL.md</a></td>
<td>(Ebrahim et al., 2013)</td>
</tr>
<tr>
<td>GUROBI optimization</td>
<td><a href="http://www.gurobi.com/">http://www.gurobi.com/</a></td>
<td>(Gurobi Optimization, 2015)</td>
</tr>
</tbody>
</table>
Table II. Properties of the genome-scale model for *Chlorella vulgaris* (iCZ842) and other prokaryotic and eukaryotic organisms are shown for comparison.

<table>
<thead>
<tr>
<th>Microorganism</th>
<th>Synechocystis sp. PCC6803</th>
<th>Zea mays</th>
<th>Chlamydomonas reinhardtii</th>
<th>Chlorella pyrenoidosa</th>
<th>C. sp. FC2 IITG</th>
<th>C. protothecoides s</th>
<th>C. variabilis</th>
<th>C. vulgaris UTEX 395</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconstruction</td>
<td>Genomic</td>
<td>Genomic</td>
<td>Genomic</td>
<td>Genomic</td>
<td>Biochemical</td>
<td>Genomic</td>
<td>Genomic</td>
<td>Genomic</td>
</tr>
<tr>
<td>Genome size</td>
<td>3.57 Mb</td>
<td>2.4 Gb</td>
<td>100 Mb</td>
<td>NA</td>
<td>46.2 Mb</td>
<td>22.9 Mb</td>
<td>46.2 Mb</td>
<td>63 Mb</td>
</tr>
<tr>
<td>Genes in the model (Total genes/proteins)</td>
<td>678 (3575)</td>
<td>1563 (32540)</td>
<td>1073 (14,354)</td>
<td>0</td>
<td>0</td>
<td>461 (7039)</td>
<td>526 (9791)</td>
<td>843 (7100)</td>
</tr>
<tr>
<td>Total Reactions</td>
<td>863</td>
<td>1985</td>
<td>2190</td>
<td>67</td>
<td>159</td>
<td>272</td>
<td>1455</td>
<td>2286</td>
</tr>
<tr>
<td>Metabolites</td>
<td>790</td>
<td>1825</td>
<td>1707</td>
<td>61</td>
<td>114</td>
<td>144</td>
<td>1236</td>
<td>1770</td>
</tr>
<tr>
<td>Compartments</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Reference</td>
<td>(Nogales et al., 2012)</td>
<td>(Saha et al., 2011)</td>
<td>(Chang et al., 2011)</td>
<td>(Yang et al., 2000)</td>
<td>(Muthuraj et al., 2013)</td>
<td>(Wu et al., 2015)</td>
<td>(Juneja et al., 2016)</td>
<td>This work</td>
</tr>
</tbody>
</table>

*Based on the genome but the gene association is not shown in the model.*
Table III. Experimental and predicted growth rates reported for Chlorella and Chlamydomonas.

<table>
<thead>
<tr>
<th>Model</th>
<th>iCZ843 Predicted growth rate (h⁻¹)</th>
<th>Experimental</th>
<th>iRC1080 Predicted growth rate (h⁻¹)</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photoautotrophy</td>
<td>0.0252</td>
<td>0.014-0.025ᵃ</td>
<td>0.1538</td>
<td>0.035-0.09ᵇ</td>
</tr>
<tr>
<td>Heterotrophy</td>
<td>0.0168</td>
<td>0.018-0.025ᵃ</td>
<td>0.0299ᶜ</td>
<td>0.059-0.084ᵇ</td>
</tr>
<tr>
<td>Mixotrophy</td>
<td>0.0407</td>
<td>0.02-0.03ᵃ</td>
<td>0.1817ᶜ</td>
<td>0.066ᵇ</td>
</tr>
</tbody>
</table>

ᵃ Data from this study and (Baalan et al., 1973)
ᵇ (Boyle et al., 2009)
ᶜ Acetate used as organic carbon source, the model was constrained according to Chang et al., 2011.
Figure 1. Overview of network properties. (A) Metabolic reactions of the reconstruction by subsystems (KEGG metabolic pathway diagrams). (B) Distribution of metabolites and reactions in the reconstruction. Most metabolites (44%) and reactions (59%) are located in the cytoplasm, while the chloroplast contains 30% of all metabolites and 25% of the reactions. The reconstruction has 1,770 metabolites, 1,180 of which are unique. Metabolites present in all five organelles are H₂O₂, H₂O, H, NADP, NADPH, and O₂. (C) Unique and shared metabolic reactions between C. vulgaris (iCZ843) and Chlamydomonas reinhardtii (iRC1080), the exchange, demand, biomass, and maintenance reactions were excluded. (D) Distribution of the biomass objective function (BOF) in iCZ843. The BOF (140 metabolites) contains 20 amino acids, 8 nucleotides, 5 carbohydrates, 6 pigments and 101 lipids: 35 Triacylglycerols (TAG), 8 Phosphatidylglycerol (PG), 4 Phosphatidylinositol (PI), 4 Phosphatidylethanolamines (PE), 10 Sulfoquinovosyldiacylglycerols (SQDG), 15 Monogalactosyldiacylglycerols (MGDG), 16 Digalactosyldiacylglycerols (DGDG), and 9 Phosphatidylycerolines (PC).
Figure 2. Network evaluation under different conditions. (A) Experimental coefficients in the biomass objective function under photoautotrophic and heterotrophic growth, the ratio between the values in heterotrophic and photoautotrophic growth are shown with blue. Red circles indicate experimental values. The simulation was performed with the assumption that photon input flux was 646 μE m⁻² s⁻¹ (PRISM_solar_litho), and 8.31 mmol g⁻¹ h⁻¹ (upper bound DM_o2D(u)) for oxygen evolution rate.
Figure 3. Model benchmarking through different carbon and nitrogen sources. (A) Predicted growth rate in photoautotrophy (light + compound), heterotrophy (compound) and mixotrophy (CO₂ + light + compound) for the 74 carbon sources, the glucose uptake rate was used to constrain the model for all the metabolites; markers represent experimental data for each condition. (B) Comparison of predicted growth rates using 53 nitrogen sources, the model was constraint using the urea uptake rate and glucose 1% uptake rate. (C) Statistics of the predictions under heterotrophy, true positive (TP), true negative (TN), false positive (FP), false negative, and Matthews correlation coefficients (MCC). The full data set and statistics under the three conditions is shown in the Supplementary Tables S11 and S12.
Figure 4. Model overview and metabolites exchange between different compartments. Color-coded values (heat map) refer to the total flux by compartment (nmol gDW⁻¹ h⁻¹), the color map shows a negative value when the input mass is higher than output and vice versa. The arrows show how the reactions are carried out, an arrow with two points mean reversible reaction, which can carry flux in both directions; the irreversible reactions have just one head pointing to the products. Every reaction and metabolite contain in their name a location identifier; e, extracellular space; c, cytoplasm; h, chloroplast; u, thylakoid; m, mitochondria, and x, glyoxysome. For the naming and abbreviation of reactions and metabolites see Table S1 and S2.
Figure 5. Predicted growth rates and shadow prices analysis for heterotrophic growth. (A) Growth rates were determined experimentally and simulated for four different glucose concentrations. (B) The heat map displays the theoretical growth rate after the addition of 20 individual amino acids as well as glycerol and acetate for heterotrophic, mixotrophic, and photoautotrophic condition. The highest growth rate was achieved while adding tryptophan. The data of heterotrophy are displayed with a conversion factor of 0.3. (C) Experimental and predicted growth rates for medium containing tryptophan or methionine.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title


Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title


Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. BBA - Protein Struct 405: 442-451


Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title


Shi XM, Chen F (1999) Production and rapid extraction of lutein and the other lipid-soluble pigments from Chlorella protothecoides grown under heterotrophic and mixotrophic conditions. Nahrung - Food 43: 109-113
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title


Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title
Pubmed: Author and Title
CrossRef: Author Only Title Only Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title