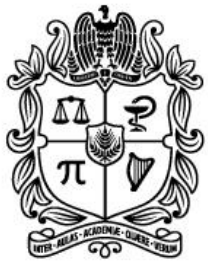

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Predicción del comportamiento dinámico de crecimiento fotoautotrófico de un alga oleaginosa utilizando un modelo metabólico de escala genómica espaciotemporal

Juan David Tibocha Bonilla, Ch. Eng.

Universidad Nacional de Colombia

Facultad de Ingeniería

Departamento de Ingeniería Química y Ambiental

Bogotá, Colombia

2019

Predicción del comportamiento dinámico de crecimiento fotoautotrófico de un alga oleaginosa utilizando un modelo metabólico de escala genómica espaciotemporal

En inglés: Prediction of the dynamic behavior of photoautotrophic growth of an oleaginous alga using a multiscale metabolic model

Juan David Tibocha Bonilla, Ing. Qco.

Tesis de investigación presentada como requisito parcial para optar al título de:

Magister en Ingeniería Química

Director:

Rubén Darío Godoy Silva, Ph.D.

Co-director:

Karsten Zengler, Ph.D.

Línea de investigación:

Bioprocesos – Bioingeniería de microalgas

Grupo de investigación:

Procesos Químicos y Bioquímicos

Universidad Nacional de Colombia

Facultad de Ingeniería

Departamento de Ingeniería Química y Ambiental

Bogotá, Colombia

2019

A mis padres, autores
últimos de lo que soy y seré

Acknowledgements

A mis padres, porque no hay mejores padres en este mundo. Es por ustedes que sé a dónde quiero llegar, y quién quiero llegar a ser. A mi mamá, por formarme y guiarme durante toda mi vida, con su incomparable altruismo hacia mí y en general nuestra familia. A mi papá, por dar siempre lo máximo de sí para que nunca le falte nada a nadie y por darme un ejemplo del ingeniero que quiero ser.

Al profe Rubén y a Anita, por ser mi familia en la universidad. Al profe, quien ha sido mi mentor como ingeniero y como investigador, y por haber sido quien me impulsó a abrir las alas; gracias por cada discusión, cada almuerzo y cada chiste. A Anita, mi mentora, amiga y guía tanto en los momentos difíciles, como en las oportunidades.

A mi hermano Pablito, por ser mi amigo y apoyo en la casa, y ahora en la distancia.

A Pipe y Yuyi, mis primos con quienes crecí y quienes me alegran la vida.

A Degly, mi segundo hermano, quien me ha acompañado como el amigo más leal que jamás pensé que tendría, casi desde el día en que lo conocí.

A Leslie, por completar mi corazón, haciéndome sentir que no me falta nada, ya sea de cerca o en la distancia, ya sea con su consejo o su sonrisa.

A Forero, mi amigo y compañero, por todas las charlas, los parciales, las tareas, las trasnochadas, los viajes y los asados. Sin usted no habría podido formarme como el ingeniero que soy.

A mi tío Henry, por ser un apoyo incondicional para mi familia, sin importar el día o la hora. Gracias, además, porque sin una mano no hubiera podido programar tan rápido.

A Cristal, quien ha sido un apoyo y una guía inmensos fuera de mi país, y quien me ha ayudado a aprovechar esta nueva etapa de mi vida al máximo.

To Karsten, for giving me the opportunity not only to boost my career at such a great lab, but also for being an amazing group leader and friend.

A la Facultad de Ingeniería y la Dirección de Investigación de la Sede Bogotá, por apoyar económicamente mi intercambio académico internacional, el cual fue crucial para mi desarrollo profesional.

A la Universidad Nacional de Colombia, mi alma mater, donde me formé como profesional y adulto, y la cual será siempre mi universidad.

A Dieguito, Juan, Juli, Johan y Javier, mis amigazos.

Contents

List of figures

List of tables

Chapter 1. Introduction

More than two decades ago, the first genome-scale metabolic (GSM) model of all history was generated, with the work of Fleischmann et al. [1] on the bacterium *Haemophilus influenzae*, and the most representative metabolic modeling strategy was initiated. During the first stage of GSM modeling, central carbon metabolism was the main object of study. After some time, a more engineering-specific orientation took shape when trying to adjust the model enough to make predictions of culture conditions as well as the best genetic modifications for specific purposes.

Today, the horizon of GSM modeling is mainly composed of its application to prediction in bioprocesses, more specifically, in bioreactors and other equipment where conditions vary both spatially and temporally. Although the most widely-employed strategy of utilizing GSM models, the Flux Balance Analysis (FBA), is based on uniform and constant conditions (steady state and homogeneity of flow and fluid properties), in recent years several authors have called attention to its future in "spatiotemporal" [2] and “modular” models [3].

Westermark et al. [2] devised a strategy in 2016 in which several types of models served as complementary modules to the FBA. Following this line of thought, Henson [1] proposed the spatiotemporal FBA (sFBA) as a holistic approach that allows not only to predict cellular phenotypic states, but also global culture conditions, by restricting the system of equations with kinetic information and spatial distributions of properties. At the moment,

only one model of this kind has been developed, which was based on the microalga *Chlamydomonas reinhardtii* [3]; however, in their approach they do not take into account the phenomena associated with the incidence of light.

In the cultivation of microalgae, the culture conditions contain spatial distributions of light intensity and phenomena of photoinhibition, which is framed in an entirely transient environment. Therefore, a photobioreactor model must, in addition to light distribution, include kinetics of nutrient consumption that allow it to reproduce the stationary phase. Under these conditions, the model would be able to predict conditions of cultivation time, photoperiod and light intensity, which will maximize the overall lipid productivity in the reactor. In the future, such model could be readily used for photobioreactor design, by helping the engineer decide upon the arrangement and types of light source, as well as for predicting the behavior of a mutant by blocking genes in the photobioreactor.

The reason why the cultivation of microalgae, and specifically its lipid productivity, are widely studied since these microorganisms are the core to biomass-based energy sources, namely third and fourth generation biofuels. These strategies are crucial as a result of the need to supply energy for the global exponentially-growing population, in a world with finite fossil fuel reserves, not to mention the long-term environmental and health effects that their use has caused.

At present, with a population 350% greater than a century ago, the planet has surpassed 400 ppm of carbon dioxide (CO_2) in the atmosphere [4]. It should be noted that this increase, which began around 285 ppm almost two centuries after the industrial revolution, has but intensified in the last 50 years [5]. It is not a coincidence that during this period

CO_2 emissions increased by approximately 90%, with a fossil fuel contribution of around 78%. Carbon dioxide is far from being the only undesirable product of fossil fuel combustion. For example, these fuels contribute to a 50% of the global emissions of nitrogen oxides (NO_x) [6], causing well-known effects including, but not limited to, the formation of *smog* (particulate matter) and acid rain.

Colombia, despite having a significant contribution of clean energy such as hydro power, is not excluded from the problem. According to the EPA, about a quarter of PM2.5 particulate matter is composed of nitrogen oxides, whose levels are high enough in Bogota to be listed as the fifth worst city in the Americas with a concentration of 24 $g \cdot m^{-3}$ [7]. Even more so, just this year several public health measures have been undertaken in the capital city for the alarming intensification of said pollution.

Numerous alternatives have been proposed to mitigate the anthropogenic impact on the ecosystem in terms of energy production. Biofuels are part of this group and have gained importance in recent years. Initially, these were proposed using food crops to produce oils (biodiesel) or bioethanol. These, called first-generation biofuels, have been widely criticized for the deviation of food crops toward the energy industry, causing an increase in their cost.

In response to this, alternatives were created in which the method is maintained, but the raw material is replaced by waste, by-products and inedible parts of crops. However, the need to use large areas of land persisted. In order to solve this, a novel strategy using microalgae was proposed, with which the soil requirement was much lower and did not require soil fertility for the production of biofuels. However, it is necessary to use lipid-

accumulation species that achieve contents of more than 20%, such as *Chlorella vulgaris* [8], called oleaginous microalgae [9-12].

It is then necessary to find those conditions under which an oleaginous microalga, in this case *Chlorella vulgaris*, produces the greatest amount of lipids. Until now, most studies have focused on studying alternatives in a purely experimental manner. However, an optimization of the kind proposed in this study would take decades of experimental work and would be extremely expensive. Consequently, modeling emerges as a useful tool in this kind of studies, although it requires a correct and well-curated definition. Specifically, GSM models are capable of predicting a greater amount of information than traditional experimental models, given their theoretical nature and their basis in species-specific omics data. As stated above, the GSM model alone is not able to predict macroscopic conditions of a culture in large-scale photobioreactors. In this work, the model *iCZ947* was modified with the influence of spatial distributions of photon flux, phenomena of photoinhibition and recovery of photosystem II, as well as kinetics of nutrient consumption as restrictions. By using this model, light- and nutrient-dependent growth and concentrations were predicted, in a way that allowed to accurately predict and theorize about the intracellular transient behavior of the microalga under several growth conditions.

References

1. Henson MA. Genome-Scale Modeling of Microbial Metabolism with Temporal and Spatial Resolution. 2017; 43: 1164-71.

2. Westermark S, Steuer R. Toward Multiscale Models of Cyanobacterial Growth: A Modular Approach. Front Bioeng Biotechnol. 2016; 4: 1-24.

3. Jeong DH, Lee JM. Integrating flux balance analysis into microalgae growth kinetics for dynamic simulation. IFAC Proceedings Volumes (IFAC-PapersOnline). IFAC; 2013.

4. Marland G, Boden TA, Andres RJ, Brenkert AL, CAJ Global, regional, and national fossil fuel CO_2 emissions. A compendium of data on global change. http://cdiac.ess-dive.lbl.gov/trends/emis/overview.html. Accessed December 6 2017.

5. Etheridge DM, Steele LP, Langenfelds RJRL, France and RJ, Barnola JM, Morgan VI. Historical CO2 records from the Law Dome DE08, DE08-2, and DSS ice cores. In trends: A compendium of data on global change. http://cdiac.ess-dive.lbl.gov/trends/co2/lawdome.html. Accessed December 6 2017.

6. Lee DS, Köhler I, Grobler E, Rohrer F, Sausen R, Gallardo-Klenner L, et al. Estimations of global NO_x emissions and their uncertainties. Atmos Environ. 1997; 31: 1735-49.

7. US EPA. The Particle Pollution Report Current Understanding of Air Quality and Emissions through 2003. Res Triangle Park NC US Environ Prot Agency, Off Air Qual Plan Stand Emiss Monit Anal Div. 2004; 28.

8. Chisti Y. Biodiesel from microalgae. Trends Biotechnol. 2008; 26: 126-31.

9. Sandeep K. Sub- and supercritical water technology for biofuels. In: Lee JW, editor. Advanced Biofuels and Bioproducts. New York: Springer; 2013. p. 147-83.

10. Samorì C, Torri C, Samorì G, Fabbri D, Galletti P, Guerrini F, et al. Extraction of hydrocarbons from microalga *Botryococcus braunii* with switchable solvents. Bioresour Technol. 2010; 101: 3274-9.

11. Huang L, Xue Z, Zhu Q. Method for the production of reservatrol in a recombinant oleaginous microorganism. 2010.

12. Yoshida M, Tanabe Y, Yonezawa N, Watanabe MM. Energy innovation potential of oleaginous microalgae. Biofuels. 2012; 3: 761-81.

Chapter 2. Advances in metabolic modeling of oleaginous microalgae

The content of this chapter was published in Biotechnology for Biofuels: Tibocha-Bonilla JD, Zuñiga C, Godoy-Silva RD, Zengler K. Advances in metabolic modeling of oleaginous microalgae. Biotechnol Biofuels. 2018;11:241.

2.1. Abstract

Production of biofuels and bioenergy precursors by phototrophic microorganisms, such as microalgae and cyanobacteria, are a promising alternative to conventional fuels obtained from non-renewable resources. Several species of microalgae have been investigated as potential candidates for the production of biofuels, for most part due to their exceptional metabolic capability to accumulate large quantities of lipids. Constraint-based modeling, a systems biology approach that accurately predicts the metabolic phenotype of phototrophs, has been deployed to identify suitable culture conditions as well as to explore genetic enhancement strategies for bioproduction. Core metabolic models were employed to gain insight into the central carbon metabolism in photosynthetic microorganisms. More recently, comprehensive genome-scale models, including organelle-specific information at high resolution, have been developed to gain new insight into the metabolism of phototrophic cell factories. Here, we review the current state-of-the-art of constraint-based modeling and computational method development and discuss how

advanced models led to increased prediction accuracy and thus improved lipid production in microalgae.

Key words – Oleaginous phototrophs, lipid production, constraint-based metabolic modeling, central carbon metabolism.

2.2. Background

Photosynthetic microorganisms have been recognized as one of the oldest life-forms on Earth [1]. These organisms, including microalgae such as *Chlamydomonas sp., Synechocystis sp.,* and *Chlorella sp.*, have attracted significant attention from the biotechnology industry because of their ability to efficiently transform renewable resources (CO_2, light, and water) into biomass and fuel precursors [2]. The photosynthetically produced biomass along with accumulated and secreted metabolites can be employed for the downstream synthesis of fuels (e.g. ethanol, biodiesel, and biocrude) and fine chemicals (e.g. pigments and organic acids) [3].

The world's ever-expanding requirement for cheap energy and fuel requires constant improvement of production platforms to meet the demand. The increased fuel consumption has led to an increase in global greenhouse gas emissions [4], exemplified by a sharp increase in CO_2 levels from 280 ppm before the industrial revolution to today's 407 ppm [5][6]. Over 75% of these CO_2 emissions have been attributed to the burning of fossil fuels [7,8], rendering the reduction of humanity's carbon footprint a major global technological challenge. One alternative to address this challenge is increased utilization

of biofuels from renewable resources and thus significant efforts have been under way to improve the efficiency of production of various biofuels [9].

Biofuels are categorized in first, second, and third generation biofuels depending on the type of raw material that is used for their production [10]. First generation biofuels are produced from agricultural crops; one example being bioethanol production from sugar cane. These biofuels have been widely criticized as they pose extra demands on food production, which consequently raises food prices. Additionally, intensive agricultural processes to satisfy cost-effective production of crops for biofuels can lead to eutrophication and contamination of environmental resources [8,11,12]. As an alternative second generation biofuels generated from woody waste and inedible food parts, such as biofuels from lignocellulosic biomass, have been proposed as a substitute for first generation biofuels generated from food sources [10]. Secondary biofuels still require fertile land and often substantial amount of water for irrigation, limiting their areas of production. Third generation biofuels, such as biosustainable production by microalgae, have thus been investigated to complement first and second-generation biofuels. Third generation biofuels also face several drawbacks which need to overcome before turning into an economically viable alternative [13]. One of the largest challenges for third generation biofuels from photosynthetic microorganisms lies in the harvesting process and downstream refinement of compounds of interest. For example, the costly recovery process of lipids from microalgal biomass, which in the case of biodiesel can account for up to 50% of the final cost [14], often prevent algae biofuel operations to be economically viable [14]. Higher lipid content would offset these staggering costs and would widely benefit profitability and applicability of a

third-generation biofuel technology. An early study by the U.S. Department of Energy from 1978, reported that a lipid content of 60% would be necessary for third generation biofuels to become economical feasible [15]. This number has now being revised to 20-40%, depending on strain and cultivation conditions [16]. Increasing the lipid content of phototrophs has thus been a major focus for the biofuel industry. Major efforts to improve lipid content have been focused on optimizing culture conditions and on advanced strain engineering designs. Both strategies greatly benefit from the use of metabolic modeling [98]. In this review we compare various computational methods used for the rational design of strains and culture media, including flux balance analysis (FBA), dynamic Flux Balance Analysis (dFBA), ^{13}C Metabolic Flux Analysis (^{13}C MFA), and Elementary Modes (EM) analysis. We focus in particular on the latest insights into central carbon metabolism (tricarboxylic acid cycle, the Calvin cycle, the glyoxylate shunt, glycolysis/gluconeogenesis, and the pentose-phosphate pathway) of oleaginous microalgae obtained by computational modeling as it is most relevant for production of biofuels and fuel precursors. Furthermore, we discuss the impact of time course modeling as well as the importance of incorporating compartmentalization into genome-scale models for microalgae and highlight the complexity of modeling lipid metabolism to increase biofuel productivity.

2.2.1. Oleaginous Photosynthetic Microorganisms

Microalgae have historically been classified in two classes: bacterial microalgae (C*yanophyta*) and eukaryotic microalgae, the latter including green algae (*Chlorophyta),* red algae (*Rhodophyta*), and diatoms (*Bacillariophyta*). Characteristic for all microalgae is

their ability to grow photoautotrophically with CO_2 and light as only carbon and energy source. Several microalgae are also able to grow heterotrophically in the absence of light using various organic substrates, or grow mixotrophically, which refers to the uptake of organic carbon, e.g. glucose, sucrose, or acetate during growth in the light [17]. Oleaginous microalgae are attractive cell factories for the production of third generation biofuels due to their ability to achieve an outstanding accumulation of lipids, in some cases surpassing 20% of total biomass in dry weight [13], reaching economic feasibility [16]. Some studies have reported microalgae lipid productivities around 136,900 L ha-1 year-1 [12], which are several times higher than those achieved by oil palm plantations (22,780 L ha-1 year-1) [12,18]. Microalgae have also been explored for the production of non-lipid-based biofuels [12]. Several genera of microalgae have been used for biofuel production and metabolic models now exist for organisms such as *Chlamydomonas* [19–30], *Chlorella* [31–35], *Nannochloropsis* [36–38], *Synechocystis* [39–46], *Tetraselmis* [47], *Monoraphidium* [48], *Ostreococcus* [49], *Tisochrysis* [50], and [51–54]. Several of these microalgae are also genetically tractable (*Chlamydomonas*, *Synechocystis*, *Phaeodactylum*) [55]. Key information regarding central carbon metabolism, nutrient dependence, and reaction distribution through different compartments in these organisms has been obtained.

2.2.2. Metabolic Modeling

Various modeling approaches have been deployed to improve applicability of microorganisms for industrial applications. Modeling efforts can be categorized into isotope labeling-based, kinetic-based, and constraint-based approaches [56]. Isotope

labeling studies and kinetic-based approaches are restricted to core metabolic networks or whole cell analysis and none of those methods is yet available on a genome-scale and neither of these approaches considers organelle-specific compartmentalization. Constraint-based modeling approaches are currently the most widely used methods in metabolic modeling of oleaginous microalgae. These models enable in-depth understanding of microorganisms and their metabolism by simulating intracellular fluxes throughout a metabolic network, often at genome-scale [57].

Genome-Scale Metabolic models (GSMs) are a mathematical representation of all the available biochemical and genomic information about a specific organism. GSMs have extensively been used to guide strain engineering designs by optimizing biochemical processes within an organism [33]. The reconstruction of a metabolic network can start *de novo* by identifying and adding reactions one by one, or it can be initiated by the creation of a draft reconstruction based on sequence homology to another related organism [33]. As of May 2018, 44 metabolic models for oleaginous microorganisms have been reported. Details about characteristics of available models are summarized in Table 1. Highlights of milestones in metabolic modeling of oleaginous microalgae are shown in Figure 2.1. While the first models for oleaginous microorganisms contained only core reaction, reaction size and complexity increased significantly over time (Figure 2.1).

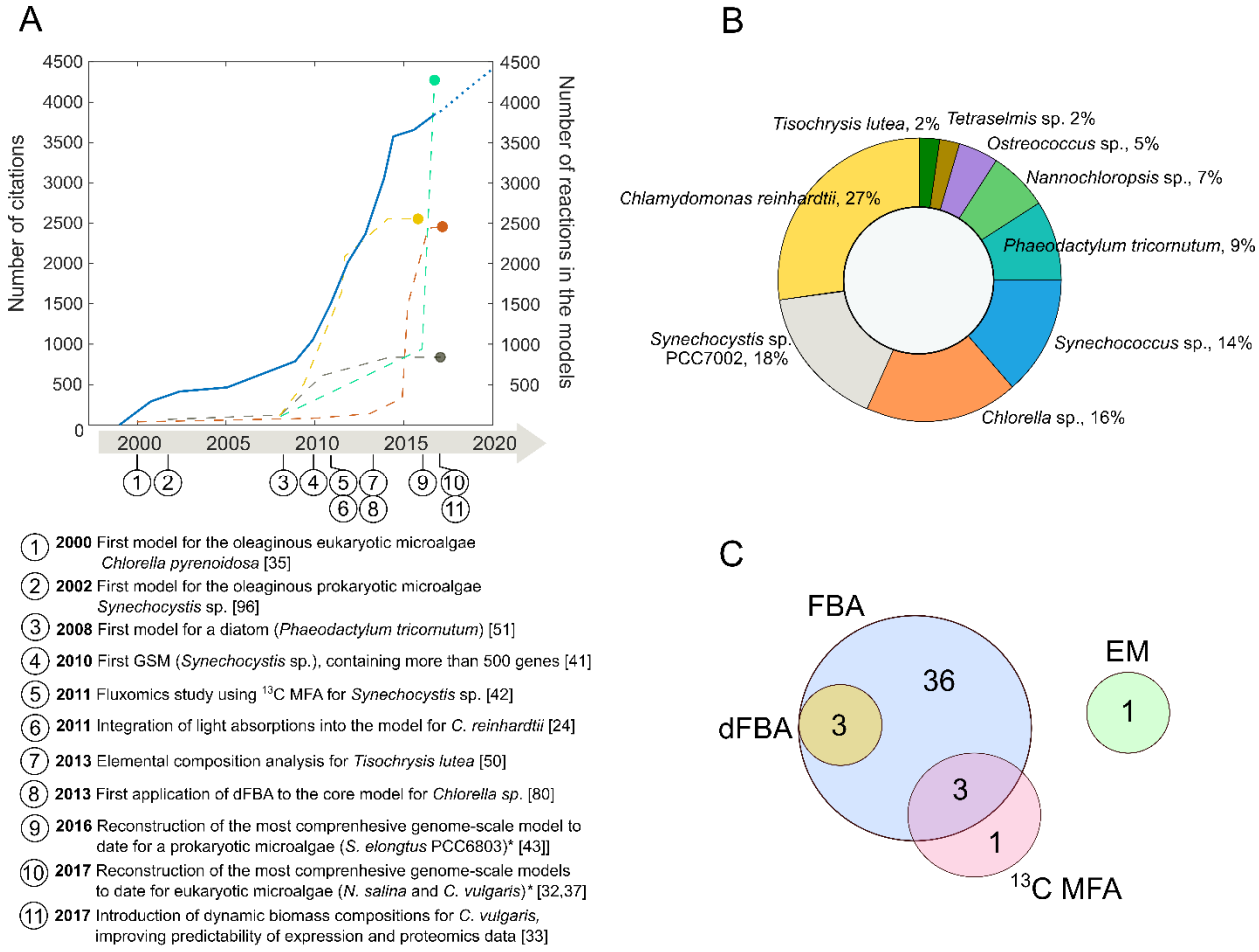

Figure 2.1. Key developments in the constraint-based metabolic modeling of oleaginous microalgae. A) Cumulative number of citations for all 44 publications related to "Metabolic Modeling of Oleaginous Microalgae and Cyanobacteria" (blue line) and estimated future citations is (blue dotted line). Dashed lines represent the number of reactions per model for *Chlamydomonas* (yellow), *Synechocystis* and *Synechococcus* (grey), *Chlorella* (orange), *Phaeodactylum* (green). B) Breakdown of total number of publications by microorganism (percentage) highlights the importance of model organisms such as *Synechocystis, Synechococcus, Chlorella, Chlamydomonas, and Chlorella*. C) Frequency of metabolic modeling approaches used to solve models: Flux Balance Analysis (FBA), followed by ^{13}C Metabolic Flux Analysis, dynamic Flux Balance Analysis (dFBA), and Elementary Modes (EM).

The first GSMs for oleaginous microalgae were reconstructed for *Chlamydomonas reinhardtii* [19] and *Synechocystis sp.* [41]. Reconstructing a GSM model, which has been and the reviewed in detail previously [58–60], requires a genome sequence and high quality information about gene function and metabolism. Manual curation is required to improve

accuracy of the model. This curation process is very time and labor intensive, often spanning weeks to months before completion. To facilitate rapid model generation, automated pipelines, such as ModelSEED [61] and PATRIC [62], have been made publically available. ModelSEED and PATRIC are reconstruction tools based on subsystems annotation, in which metabolic networks are decomposed into subsystems and analyzed individually. Both tools are based on RAST (Rapid Annotations using Subsystems Technology) that compares the genome sequence with existing information from phylogenetic neighbors [63]. However, it has to be noted that reconstructions created by automated tools are prone to errors and special attention must be directed towards quality control and quality assurance (QC/QA) tests, in particular in regards to mass balance and energy production without input [57,64]. Automatically and semi-automatically reconstructed models thus require intensive manual curation before detailed and accurate predictions can be made. Figure 2.2A compiles the number of core and genome-scale models created for oleaginous photosynthetic microorganisms reported to date.

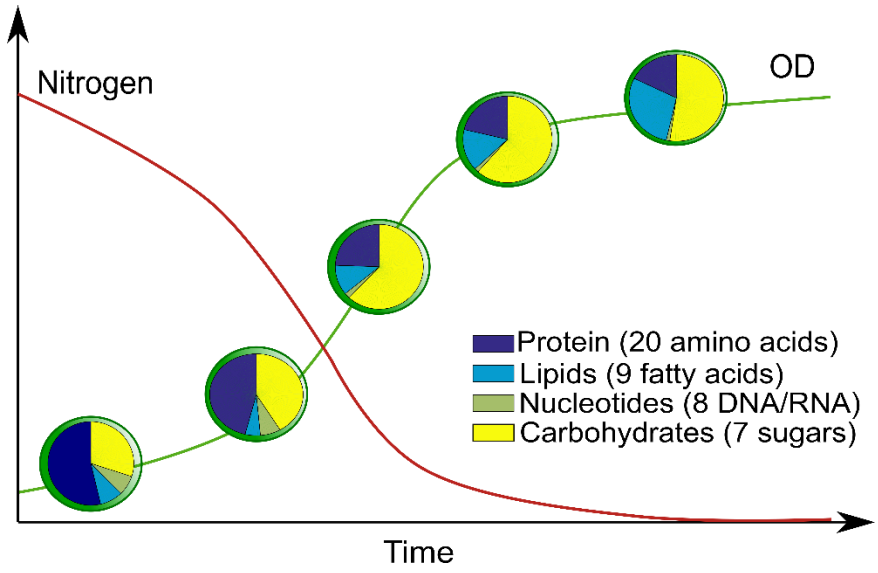

Figure 2.2. Changing biomass composition of *Chlorella vulgaris* in response to nitrogen depletion determined over time. While optical density (OD) increases of a growth course, the microalga accumulates storage compounds, such as lipids and carbohydrates, while reducing the total protein fraction of the biomass. Data collected from [32].

All GSMs can be expressed as a general mass balance, which includes every metabolite being produced or consumed within the network in its respective reaction. This mass balance takes the form shown in (1).

$$\frac{\mathrm{d}}{\mathrm{dt}}C = [\boldsymbol{S}]v \tag{1}$$

The vector C represents the instantaneous concentration of metabolites inside the cell, the vector contains all reaction rates and the matrix represents the stoichiometric information about reactions and participant metabolites. The stoichiometric matrix is a shared requirement among all constraint-based flux analysis approaches. Each column of this matrix contains the stoichiometric coefficients of a compound for all included reactions. In a similar fashion, each row represents the coefficients of all metabolites that take part in a single reaction [65]. A m number of metabolites would render the S matrix of mxn dimensions, with n always greater than m.

The rectangular nature of the S matrix is one of the most important obstacles to overcome when working with metabolic networks and is easily seen when taking into account that for m number of metabolites, there are m change rates inside vector C, m transport rates, and p intracellular rates that are unknown. The system of equations is then comprised of only m mass balances and as many as $n=2m+p$ variables [66]. This system indetermination is what has given birth to several different approaches to metabolic modeling, which are discussed below. For system determination to be achieved, the measurement of a total of

m-n variables would be required. Large metabolic networks contain degrees of freedom that can amount to several hundreds. Therefore, so-called core models, focusing on central metabolism have been developed. These core models are used in Metabolic Flux Analysis, such as the ^{13}C-MFA, i.e. fluxomics. However, it is currently computationally infeasible to use large and compartmentalized metabolic networks for fluxomics analysis. Due to this, metabolic engineers have simplified the problem by transforming reaction (1) into an optimization problem using an objective function and a defined set of constraints [65]. The definition of constraints results in a solution space, which delimits all possible functional states of a reconstructed network and a set of permitted phenotypes [67]. Metabolic models account for three types of constraints [65,67]: a) physico-chemical, which are based on conservation laws of mass and energy, dependency of reaction rates on biochemical loops and thermodynamics; b) environmental, such as availability of nutrients, electron acceptors, and other external conditions (e.g. photon uptake); and c) regulatory, including enzyme composition and performance, which helps to contextualize gene-related information, such as expression data and accurate gene-protein-reaction associations [68].

In phototrophic organisms some physicochemical constraints are decided upon by following thermodynamic limits, regarding direction, reversibility or non-reversibility of reactions, which can be determined by calculating the Gibbs free energy. Environmental constraints are usually based on measured experimental values of light quality, and nutrient and substrate uptake rates. Some regulatory constraints are those used in a study by Levering et al., in which the GSM of the diatom *Phaeodactylum tricornutum* was employed to capture the response to varying environmental conditions due to a

transcriptional regulatory network [69]. Despite this, there are still too many variables to account for in the dynamic system. Various approaches to analyze the metabolic network of oleaginous microalgae are discussed below.

2.2.3. Flux Balance Analysis (FBA)

Most metabolic modeling studies involving oleaginous microalgae have been using FBA for simulation. A few other approaches have been used as an alternative or complement, such as ^{13}C-MFA [22,31,34,42] or EM [50]. Figure 2.1B and C highlights existing models and methods used to interrogate these models. Currently, large–scale metabolic networks are analyzed mainly *in silico* using FBA. Analysis of dynamic data obtained by experimentally–intensive strategies like ^{13}C-MFA rely on simplified metabolic models, e.g. representing only central metabolism [22,31,34,42].

FBA refers to the application of linear programming to analyze fluxes under balanced metabolite conditions [65]. This statement is based on two assumptions: First, the cells are in steady state and second, all cells have a general objective while growing. The first assumption simplifies the system significantly by neglecting all transient behavior of the metabolite concentrations, thus yielding (2). The elimination of all the unknown concentration change rates inside is mathematically convenient, but forces the system, i.e. a culture flask or bioreactor, to theoretically exist in a steady state.

$$[\boldsymbol{S}]v = 0 \tag{2}$$

The second assumption of an objective function in the model implies that all cells grow with a specific objective, which is the same for every cell during the calculation time. The most widely used objective function for FBA is the maximization of biomass production, which implies that the organism has evolved sufficiently to have the optimal arrangement of fluxes so that its growth will be maximized. While this assumption is likely correct for certain microorganisms, it is not universally applicable [70]. For example, under nutrient-sufficient conditions the objective of a cell might not be biomass production but rather the optimization of the production rate of storage compounds for later use. In a similar way, we know that phenotypic states vary in accordance with the growth phase or environmental conditions (Figure 2.2), especially those that exhibit a dynamic biomass composition, such as phototrophs [71–73] and yeast [74]. Thus, time-specific biomass compositions are needed for light-dark cycles, considering degradation of storage pools during dark periods. This is of particular interest for production of biofuel precursors. Furthermore, maximization of carbon uptake rate as CO_2 has been proposed as a suitable objective function for autotrophic modeling during the light period [32]. FBA has proven to be useful and to reproduce overall experimental behavior *in silico*, although a true steady state is hardly encountered in experimental settings [58]. Its versatility and the accurate reproducibility of experimental results under several culture conditions make FBA one of the most widely used methods for metabolic modeling [75].

2.2.4. Biomass Objective Function

The biomass objective function (BOF) is a broadly used modeling reaction, which drives the supplemented resources across the metabolic network in order to produce all known cellular components in the model (such as amino acids, nucleotides, fatty acids, carbohydrates, vitamins, ions, and cofactors). Maximizing for the BOF allows simulating growth rate and biomass yields. The BOF can be determined from the genome sequence [59] or through experimentation. Both approaches have been successfully applied, especially for prokaryotic microorganisms. However, when microorganisms have been subjected to non-optimal conditions, such as extreme temperatures, pH, or limited nutrient concentrations, a single BOF is often not suitable to predict experimental data successfully [70,76]. For these cases, auxiliary objective functions have been proven necessary, such as minimization of ATP production, substrate uptake rate, or redox potential production rate [70].

There are several levels of refinement of the BOF [77], but it generally consists in the definition of a set of metabolites which compose the biomass. The set can be composed of just one reaction yielding a hypothetical compound called “biomass” or could otherwise be refined up to building blocks or biomass components (carbohydrates, lipids, proteins, DNA, RNA, pigments, etc.) [78]. The BOF of manually curated metabolic models of oleaginous microorganisms often account for hundreds of metabolites as part of the lipid metabolism, because lipids being the primary target for biofuel production in these organisms. Lipid chain fatty acids (14:0, 16:1, 18:1, 16:2) are usually summarized as triacylglycerols (TAG), monogalactosyldiacylglycerols (MGDG), etc. representing the

entirety off all lipids in the organism. Accurate BOF composition has enabled the improved prediction of phenotypic states. It has been claimed that constrained BOF furthers the predictability of experimental nutrient- and light-limited conditions [33]. In some cases, the BOF has been complemented by a two-step optimization approach with minimization of uptake rates. In autotrophic growth conditions, minimization of light uptake (photons) has been employed but no significant improvement of the growth rate prediction has been obtained [23,39]. In the same way, minimization of carbon source substrate uptake rate has been utilized for heterotrophic growth [25,47]. As alternatives, minimization of flux magnitudes across the network was used for *P. tricornutum* [51,54], maximization of ATP yield [28], and minimization of ATP demand [24] for *C. reinhardtii*, and maximization of hydrogen production rate for both *C. reinhardtii* [25] and *Synechocystis sp.* [40].

2.2.5. Dynamic FBA

Overcoming the steady-state assumption of standard FBA is vital for the modeling of highly dynamic systems, which are characteristic of photosynthetic microorganisms [33,37,79]. These organisms have evolved under cyclic light/dark conditions, which require switching between different phenotypic states. During light periods, inorganic carbon is fixed into storage carbon compounds, such as carbohydrates and lipids, which are consumed in the dark period to accommodate vital cell functions. The storing-for-later behavior results in a dynamic biomass composition that can change during the light period (hours) or along the course of growth (days). In the case of *C. vulgaris* and other phototrophs it has been shown that the biomass composition is also dependent on nitrogen availability (Figure 2.2). Since

FBA is used under a steady–state assumption, it is virtually disqualified for its use in the aforementioned cases. On the other hand, not including this assumption would add a set of ordinary differential equations to the problem, yielding a differential-algebraic system. To solve this, a dynamic FBA approach was proposed using either a Dynamic Optimization Approach (DOA) or a Static Optimization Approach (SOA) [80].

The DOA calculates time profiles of fluxes and metabolite concentrations by solving the optimization problem over the entire time span of interest, running the calculation only once. The dynamic system is transformed into a non-linear programming problem (NLP) by parameterizing the differential equations through the method of orthogonal collocation on finite elements, described by Cuthrell and Biegler [81]. The BOF is then rewritten as a weighted average of the instantaneous and the terminal objective functions, and is subjected to the system of differential equations along with the constraints. The SOA approach, on the other hand, solves the optimization problem multiple times, once for each time interval. At the end, an integration of the set of instantaneous rates of change over the interval is carried out for the calculation of metabolite concentrations.

Experiment-based BOF constraints are an alternative method to simulate dynamic metabolic behavior [33]. Changes in the BOF influence the state of the metabolic network, thus directly affecting predictions. This approach improved the accuracy of flux prediction by considering measurements over the course of growth under autotrophic and heterotrophic conditions in *Chlorella vulgaris*. The time series flux distributions accurately simulate 75% of expression and proteomics data collected over the course of growth, including allosteric reactions and multi-subunit enzymes. This approach also enabled the

determination of the net content of nitrogen pools at each condition [33]. When an experimental determination of metabolites constituting the BOF is not feasible, unsteady-state methods, such as unsteady-state FBA (uFBA). These unsteady-state methods operate with a limited number of measured metabolites. uFBA was recently developed and applied to study heterotrophic microorganism [86] but uFBA would be a promising approach for the analysis of photosynthetic microorganisms.

2.2.6. Unsteady-state FBA

The aim of uFBA is to calculate internal flux distributions from existing time-course data, e.g. target metabolomics data. These data sets typically contain information about several (five to ten) metabolites such as glycerol, ethanol, acetate, etc. It is necessary to determine the rate of change of these metabolites from the experimental data and to include these rates in the system of equations [82]. Ideally, all rates of change would be known and the uFBA could be run as a series of standard FBA methods. Since this is often not feasible, all immeasurable variables are assumed to be, initially, under steady-state conditions as well as under a closed system assumption, i.e. with no possibility of transport inside or outside the cell. Elimination of this amount of transport reactions can often over-determine the system and requires further conditioning. A “metabolite node relaxation” algorithm has been deployed that assigns sink reactions to unmeasured variables to allow for their accumulation or depletion. The algorithm is based on optimizations that find the minimum number of sink reactions that are necessary while keeping the model computable [86].

2.2.7. Metabolic Flux Analysis (MFA)

MFA is an alternative to FBA which also assumes a steady-state mass balance [83]. When working with small enough metabolic networks, it is possible to measure or define enough numbers of internal or external fluxes to determine the algebraic equation system. For this strategy, (2) is rewritten by decomposing the matrix and the vector into the measurable (known) and the immeasurable (unknown) fluxes, as shown in (3).

$$[\boldsymbol{S}]_u v_u + [\boldsymbol{S}]_m v_m = 0 \quad (3)$$

The larger the metabolic network, the more fluxes are necessary to measure for system determination. Therefore, metabolic networks of several hundred reactions require measurements of internal fluxes for most of the fluxes, e.g. by ^{13}C labeling [22,31,42].

2.2.8. Elementary Modes (EM)

EM is based on the calculation of all the solutions of the system in (2) in the allowable flux space, restricting the solution with a thermodynamic constraint and a non-decomposability constraint [84]. The latter renders each solution an elementary flux mode, which means it is a unique and minimal set of reactions. These sets can be rewritten into macroscopic reactions, thus reducing the degrees of freedom exhibited formerly by (2). Often EM is combined with core genome-scale metabolic models in order to provide energetic efficiencies and optimal flux distributions [84,85] . The use of EM analysis (Figure 2.1C) has been declined over the last years, in part due to the rapid development of omic tools applied to sequencing, which enables to generate genome-scale metabolic network reconstructions based on complete genome sequences.

2.3. Lessons learned from metabolic modeling of oleaginous phototrophs

Advances in modeling of microalgae are in part due to the availability of full genome sequences [23]. Having full genome sequences available was crucial for generating the initial genome-scale metabolic models for the microalgae *Chlamydomonas* [26] and opened the possibilities to create additional algae models based on homology [32]. In this section we describe different approaches to reconstruct and simulate with metabolic models for oleaginous microalgae to increase growth and lipid content and improve bioproduction.

2.3.1 Growth conditions

Several microalgae are able to grow as autotrophs, heterotrophs, or mixotrophs. Some metabolic pathways are only active under certain growth modes, including the oxidative/reductive pentose phosphate pathway [22,23,27,39,40], Calvin cycle, and presumably the glyoxylate shunt [39]. Hence, differential mathematical models are necessary for correct prediction for each growth condition, requiring unique stoichiometric matrixes and biomass formation equations. The study and prediction of phenotypes dependent on growth conditions is perhaps the most studied aspect regarding in oleaginous microalgae.

Experimentally, highest biomass yields have been reported for autotrophic conditions, while lowest were obtained under heterotrophic growth [23,39,53,79]. Mixotrophic growth, as a kind of hybrid condition, has shown biomass yields falling between ones observed for

autotrophic and heterotrophic. However, an exception is the study of Navarro et al. [40], in which a mixotrophic biomass yield (92%) higher than the autotrophic one (60%) was predicted as reported for *Synechocystis sp.* The constraints regarding growth conditions directly affect the way carbon is distributed across the metabolic network, and thus the biomass yield and production rate. So, it is crucial to determine and adjust these constrains if needed for each growth condition. For example, under autotrophic growth the biomass yields have been reported to be close to 100%, since no carbon is lost in the process [23,39,44,53,79], thus experimental measurements can be used directly. On the other hand, under heterotrophic growth conditions a significant carbon loss as CO_2 in oleaginous microalgae has been reported to vary between 37% [39] to 40% for *Synechocystis sp.* [40], 50% for *C. reinhardtii* [23] and 50.5% for *Chlorella sp.* [79] as a result of the carbon input flux is lost as CO_2 due to energy production through the TCA cycle and the oxidative pentose phosphate pathway (PPP) [23,39,44,53,79]. Mixotrophic biomass yields tend to be higher than under heterotrophy since part of the released CO_2 is fixed once again [39]. Reported net biomass yields are therefore around 92% (*Synechocystis sp.*), 100% (*C. reinhardtii*) and 80% (*Chlorella sp.*) assuming a closed system.

2.3.2. Light conditions

Since light directly impacts microalgae growth and behavior, efforts have been made to define the quality and quantity of light constraints in metabolic models [29, 37]. Models can be significantly improved by considering a more realistic light uptake mechanism. Chang et al. [26] proposed to divide the total light spectrum into effective spectral

bandwidths, each of which had an associated effective bandwidth coefficient. These coefficients, along with the activity spectra of light-driven reactions, allowed for the correct calculation of flux distribution along these reactions, taking into account that phototrophic organisms are strongly affected by the nature of the incoming light. It has been shown that the bandwidth coefficient varies from microorganism to microorganism. In general, microorganism-specific refining for light uptake modeling in the GSM models will be needed for further improvement [39].

2.3.3. Intracellular pools

Although metabolic modeling has focused on simulating the intracellular environment of a cell under steady state, allowing the accumulation of certain metabolites in pools has proven necessary for the correct prediction of phenotypic states [60,86]. Metabolite pools can play an important role in regulation of reactions, since different pathways find themselves interconnected by common metabolite collections. Target metabolomics data has been used successfully to constrain the metabolic model of *Chlorella* and determined the pool size of nitrogen [33]. Another target examples are energy-dependent and energy-replenishing processes are coordinated by the ATP, ADP and AMP pools [86] as well as nitrogen and chrysolaminarin pools in *P. tricornutum* [54].

2.3.4. Compartmentalization

Eukaryotic microalgae contain different organelles (e.g. cytosol, mitochondria, nucleus, endoplasmic reticulum, glyoxysome, chloroplast, Golgi apparatus, vacuole, thylakoid, eye spot, and the cell wall). The exact compartmentalization is species dependent. Accurate

annotation of proteins and compartmentalization in the model is necessary for maximizing information content and gaining detailed knowledge about microalgae metabolism. Flux distributions highly depend on the model's capability for metabolic exchange prediction between organelles. Careful manual curation of these models and delimitation of capabilities while adding reactions and reconstructing eukaryotic models in an automatic matter is thus crucial to achieve maximal predictability [63].

The example of nicotinamide adenine dinucleotide phosphate (NADPH) production in eukaryotic microalgae highlights the importance of compartmentalization. The PPP plays the role of producing NADPH in the cytosol, while the electron transport chain (ETC) is in charge of producing it in the chloroplast. Non-compartmentalized models can predict that the entire NADPH demand is supplied by the ETC, rather than PPP supplying NADPH demand outside the chloroplast. This issue was encountered in the first metabolic model of an oleaginous microalgae *C. pyrenoidosa* [35]. While the model can simulate central carbon metabolism in general, it cannot predict detailed engineering targets since information about where fluxes take place is not available.

Early metabolic models were focused on the reconstruction of core algae models, which were later expanded to include genome-scale information (Table 1) [19,20,35,49,50]. The least compartmentalized model included only the chloroplast and cytosol, in order to uncouple the NADPH consumption/production of the Calvin cycle and the PPP [28]. More refined models now account for the mitochondria, thylakoid lumen, glyoxysome (peroxisome), extracellular environment, nucleus, Golgi apparatus, endoplasmic reticulum, vacuoles and the cell wall [21,24,26,30,36,37].

2.3.5. Modeling lipid production

Phototrophs produce several different kinds of lipids, including tri- and diglycerides, phospho- and glycolipids and hydrocarbons [14]. More specifically, lipid production in oleaginous microalgae includes triacylglycerol (TAG), phosphatidylglycerol, phosphatidylinositol, phosphatidylethanolamine, sulfoquinovosyldiglycerol, MGDG, digalactosyldiglycerol, and phosphatidylcholine. TAG alone can accumulate from 20% to 60% of dry cell weight in some species like C. vulgaris [32]. Modeling and gaining insight into the increased lipid content of microalgae has been the object of several studies. Most of these studies have investigated the effect of nitrogen depletion [21,33,34,37,38,42,48,54], while others have studied the influence of low CO_2 and low light [54] on increasing overall lipid content. A study of *Nannochloropsis gaditana* reported increased lipid productivity under mixotrophic growth conditions [38].

When microalgae are subjected to nitrogen depletion conditions, carbon flux is shifted away from photosynthetic pathways as cells shift into the stationary phase and begin to store carbon as starch and lipids. This phenomenon and its effect on biomass composition is displayed in Figure 2.2, in which a general trend of biomass composition evolution as a function of time and nitrogen availability is presented. Under these non-optimal culture conditions, microalgae shift the central carbon flux from biomass production to the production of storage compounds. As a result the growth rate is decreased because carbohydrates and/or lipid are accumulated under stress conditions [40]. *C. protothecoides* was reported to redirect 81% of the input carbon flux towards fatty acid synthesis, but as a consequence decreased its growth rate by 30% [34]. In a similar way, *C. reinhardtii* was

found to accumulate TAG when faced with nutrient limitation, but its growth halted after 8 h of cultivation [22]. Interestingly, Lim et al. [47] reported for *Tetraselmis* sp. that after 24 h of culture, TAG biosynthesis was downregulated, though the accumulation of TAG continued. The authors explained their finding to be a result of decreased lipid degradation rather than lipid production.

2.4. New insights into the central carbon metabolism of microalgae

Most studies on oleaginous microalgae have focused on the central carbon metabolism and revealed new findings about glycolysis, PPP, TCA cycle, and the Calvin cycle. Figure 2.3 shows the most important metabolic pathways in microalgae and how they are linked to lipid metabolism. FBA has been used to study genome-wide fluxes through the metabolic network under mixotrophy, heterotrophy, and autotrophy. While most studies coincide in their assessment of functionality and fluxes in central carbon pathways, other pathways such as the glyoxylate and ornitine shunt are still not well understood, and modeling results are often not consistent between different studies and organisms [22,23,25,31,34,39,40,42,79]. As a general rule, significant carbon flow through the TCA cycle has been reported under heterotrophic conditions, which demand catabolism of external organic compounds, contributing to the reduction of flux through the electron transport chain (ETC) and the Calvin cycle [23,32,35,38]. During heterotrophic growth, most microalgage prefer glucose as carbon and energy source (Figure 2.3). Other microalgae, such as C. reinhardtii, are only capable of assimilating two-carbon organic compounds, like acetate [22]. When glucose enters the cytosol, its fate can either be oxidation via glycolysis

to pyruvate, oxidation via PPP to ribose 5-phosphate or transformation into storage compounds (lipids, glycogen and starch) [87]. In microalgae, acetate coming from the extracellular environment can be converted in the glyoxysome to succinate through the glyoxylate shunt, which can be considered as a variation of the TCA cycle. Succinate, an important biosynthetic precursor that can be converted into oxaloacetate, from which phosphoenolpyruvate (PEP) can be synthetized by the enzyme PEP carboxykinase, and enter gluconeogenesis for carbohydrate or lipid synthesis [17]. Under autotrophic growth, the carbon source is inorganic (CO_2) and the energy source is light. In the thylakoid lumen of eukaryotic microalgae, the ETC takes advantage of protons from light to store its energy in the form of ATP and NADPH, which are subsequently used to reduce CO_2 molecules into triose phosphates (G3P) in the Calvin cycle. G3P can then be converted into glucose 6-phosphate (G6P), sucrose, starch and other sugars in the chloroplast.

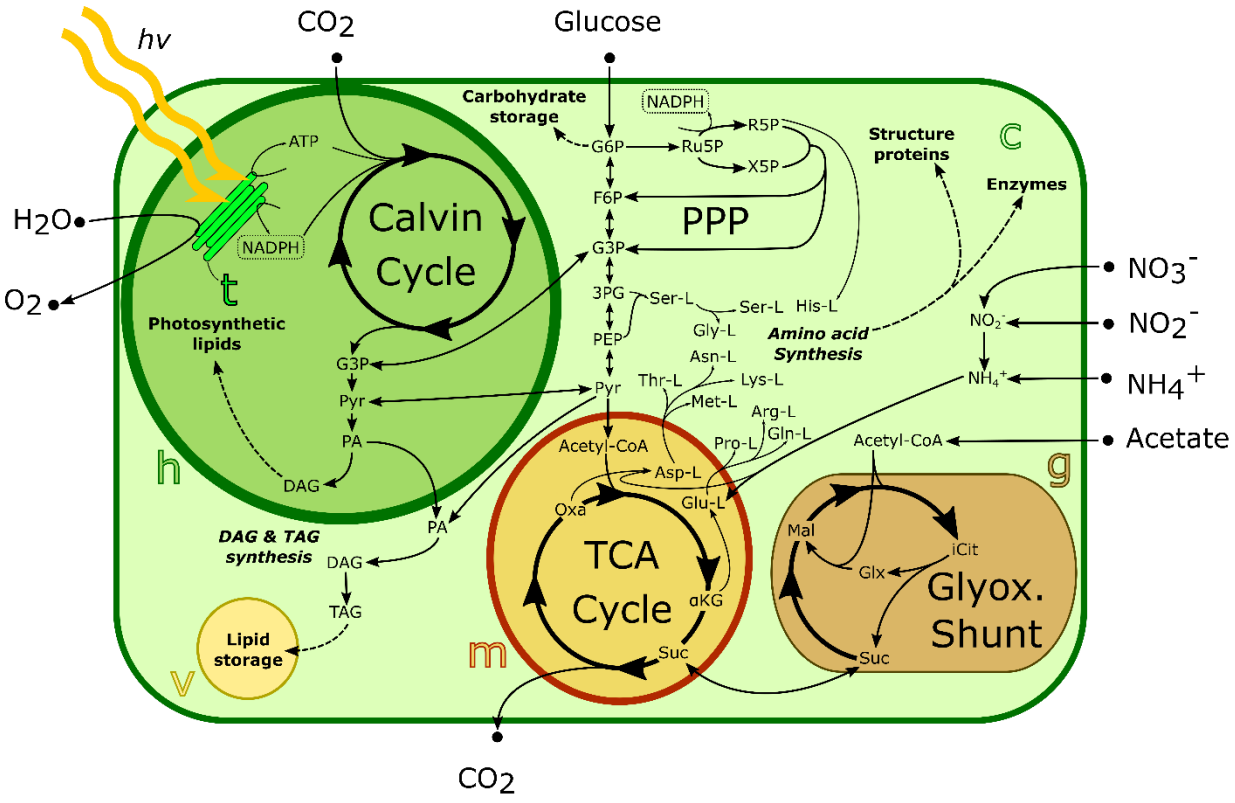

Figure 2.3. Central metabolism in eukaryotic microalgae. The main compartments of active metabolism, i.e. the chloroplast (h), thylakoid lumen (t), vacuole (v), mitochondria (m), glyoxysome (g) and cytosol (c) and shown.

2.4.1. Tricarboxilic acid cycle

The TCA accounts for the highest carbon fluxes and number of active reactions under heterotrophic growth conditions [32,35]. Under this mode, the percentage of the total carbon input flux into the TCA cycle was reported to be 35.6% in *C. reinhardtii* grown with acetate [22] and 18.7% in *C. protothecoides* grown with glucose. However, under autotrophic and mixotrophic conditions only half of the activity has been reported [79], with only 8-10 out of 22 reactions carrying flux for both microorganisms [32]. The role of the TCA under these conditions shifts to the production of biosynthetic precursors rather

than energy production. Figure 2.4 shows complete and possible bypass variations of the TCA cycle observed in different photosynthetic microorganisms. Cogne et al. [27] reported that under autotrophic growth the TCA cycle in *C. reinhardtii* was operating as two branches with production of 2-oxoglutarate on one end, and malate on the other, with an input through oxaloacetate via the anaplerotic activity of the phosphoenol pyruvate carboxylase (Figure 2.4).

Zero flux was found through the enzymes 2-oxoglutarate dehydrogenase, succinyl-CoA synthetase, and succinate dehydrogenase, since energy demands can be supplied by the PPP and the glyoxylate shunt. Other studies have also reported such similarities between prokaryotic and eukaryotic organisms [88,89], in which prokaryotic microalgae, like *Synechocystis sp.* and *Synechococcus elongatus,* do not possess a complete TCA cycle. These bacteria lack the alpha-ketoglutarate (2-oxoglutarate) dehydrogenase and succinyl CoA synthetase [17,78]. Knoop et al. [41] have claimed a bypass via the succinate-semialdehyde dehydrogenase to compensate for the lack of 2-oxoglutarate dehydrogenase as shown in Figure 2.4. The bypass replenishes intermediaries in the TCA cycle linked to lipids biosynthesis such as acetyl-CoA.

2.4.2. Reductive/oxidative pentose phosphate pathway

The oxidative and reductive phases of the PPP serve different purposes. While the oxidative phase is serving as catabolic pathway for NADPH production from the oxidation of G6P, the reductive phase represents an anabolic pathway of biosynthesis of 5-carbon

carbohydrates for synthesis of nucleic acid, coenzyme A, ATP, and other biomass and lipids biosynthetic

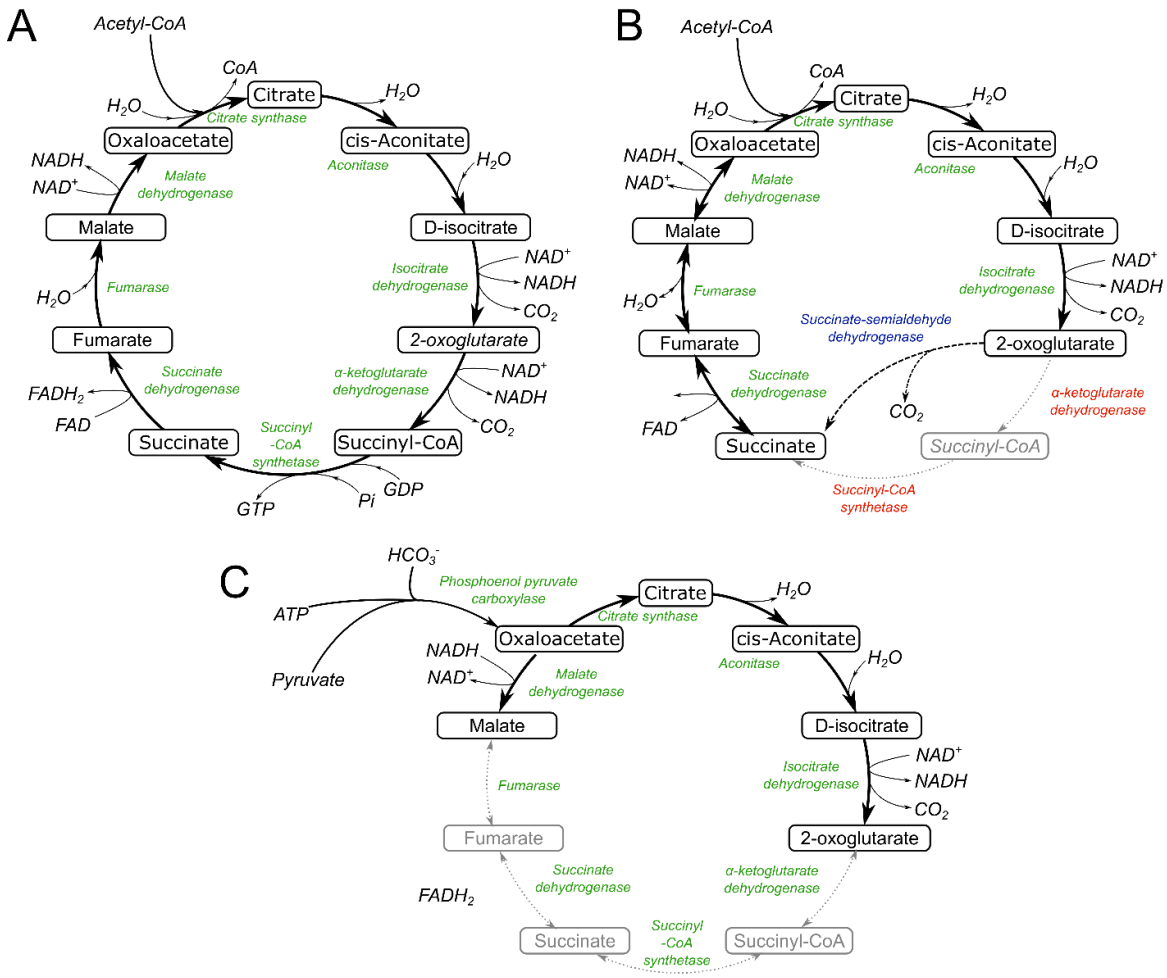

Figure 2.4. Variations of the TCA cycle in photosynthetic microorganisms. A) Complete and fully functional TCA cycle. B) TCA cycle observed in microalgae, such as *Synechococcus* sp., which lacks the enzymes succinyl-CoA synthetase and *α*-ketoglutarate dehydrogenase (enzymes highlighted in red). A bypass observed in *Synechocystis sp.* via succinate-semialdehyde dehydrogenase. C) Split TCA cycle as reported for *C. reinhardtii* [30]. The two branches producing 2-oxoglutarate and malate for downstream biosynthesis. Oxaloacetate in this split TCA cycle is provided via anaplerotic activity of phosphoenol pyruvate carboxylase [46].

precursors [79]. High flux through the oxidative PPP means that energy production is being used for maintenance rather than growth [35]. It has been reported that depending on the growth conditions, either one phase can appear active. However, for the diatom *P. tricornutum* a low flux through the whole PPP pathway was determined. The reduced flux is explained with this organisms' ability to obtain 5-carbon carbohydrates via phosphopentose epimerase [53].

As a general result for non-compartmentalized models, energy-yielding oxidative PPP appears inactive during autotrophic growth, since the model predicts energy comes from the ETC in the form of NADPH rather than the dissimilatory pathways [35,39]. As stated above, NADPH demand outside the chloroplast should be supplied by the PPP rather than the ETC. However, the compartmentalized models of Cogne et al. [27] and Boyle & Morgan [23] predicted inactivation of the oxidative PPP for *C. reinhardtii* under autotrophic conditions. In the latter study, cells were found to prefer indirect energy transport by taking G3P from the chloroplast to the mitochondria and degrading it to 3-phosphoglycerate (3PG), releasing both ATP and NADH [23]. Furthermore, the fact that *C. reinhardtii* uses acetate as a carbon source instead of glucose greatly affects its phenotypic behavior and flux distribution under heterotrophy. Since the input to the PPP is G6P, incoming acetate would have to be transformed through several reactions in the glyoxylate shunt to oxaloacetate and then to G6P (Figure 2.3). For this reason, NADPH production in *C. reinhardtii* is preferably taking place via the ETC under autotrophic growth, while it is produced mainly through the glyoxylate shunt under heterotrophic growth [22,23,31,34,35,39–41,79]. Limitation in the transport or consumption of G6P or 3PG can

result in metabolite accumulation, leading to the synthesis of certain types of lipids. For example, *C. reinhardtii* produces mainly triglyceride lipids.

Apart from growth conditions, other external factors have been reported to alter the flux distribution through the PPP. Wu et al. [31] found that increased oxygen availability in *C. protothecoides* decreases the flux through the PPP and instead enhances flux through the TCA cycle, thus producing more energy and yielding more CO_2. Moreover, increased synthesis of storage compounds under nitrogen-depletion conditions were shown to increase PPP fluxes due to increased demand of NADPH for biosynthesis [34].

2.4.3. Glyoxylate shunt

The ability of the glyoxylate shunt of transforming acetyl-CoA into succinate for biosynthetic purposes renders it vital for the metabolism of acetate independent of its source, i.e. extracellular environment. However, the glyoxylate shunt has been found to be inactive under heterotrophic [31,34,79], autotrophic [39,40,79], or mixotrophic growth conditions [42] for various organisms, e.g. *Synechococcus* sp. In *C. reinhardtii* and *P. tricornutum* however, the glyoxylate shunt has been reported to be active for all tested heterotrophic conditions [22,23,25]. The inactive glyoxylate shunt under autotrophic growth can be explained by the cell not taking up acetate from the environment, but rather synthetizing storage compounds, such as lipids and carbohydrates, that represent desirable bioproducts [40,80].

2.4.4. Calvin cycle

Reducing equivalents and ATP formed in the ETC under autotrophic conditions are used later in the Calvin cycle to produce triose phosphates (G3P) for further synthesis of carbohydrates, which can be assimilated or turned into back-bone structures of lipids. It has been reported in green algae that the Calvin cycle fixes CO_2 in the form of 3PG, which gets converted to dihydroxyacetone phosphate (DHAP) subsequently [79]. Naturally, the Calvin cycle is inactive in the dark. When microalgae are subjected to mixotrophic conditions, carbohydrate demand poses a competition between uptake of external organic carbon sources and the Calvin cycle (i.e. inorganic carbon uptake). In *C. reinhardtii* the majority of carbon flux was found to be directed towards the Calvin cycle, rather than glycolysis and TCA under mixotrophic growth [23]. The cyanobacterium *Synechocystis sp.* however was found to be dominated completely by the organic carbon uptake before a specific threshold of light intensity was surpassed. After this verge of irradiance, rubisco-dependent carboxylation and oxygenation were increased immediately and all Calvin cycle reactions were activated [41].

2.5. Conclusions

Great advances have been made in constraint-based modeling of photosynthetic microorganisms over the last two decades. Metabolic modeling has been proven critical for our understanding of complex metabolism in microalgae. Model-driven approaches have helped to identify boundaries for light and nutrient conditions as well as suitable genetic targets to increase lipid productivity. Metabolic models have progressed from core models

to genome-scale metabolic models, which now include detailed compartmentalization and light uptake. Furthermore, the dynamic behavior and rapidly changing phenotypes due to changing environmental parameters is an important trait of these organisms and has now been included in model simulations. Those recent extensions and improvements allow to elucidate phenotypic behavior under different growth and culture conditions over time. In addition, these new models provide a high-quality standard for the improvements of existing metabolic models as well as for future reconstructions. Despite extensive efforts on refinement and manual curation of metabolic models, there are still open questions regarding the central metabolism in microalgae, such as the activity of the glyoxylate shunt and ornithine shunt as well as pathways coupling across different compartments.

2.6. References

1. Falkowski PG, Raven JA. An introduction to photosynthesis in aquatic systems. In: Elworthy S, editor. Aquatic photosynthesis. Princeton: Princeton University Press; 2013. p. 1–43.

2. Gavrilescu M, Chisti Y. Biotechnology - a sustainable alternative for chemical industry. Biotechnol Adv. 2005;23:471–499.

3. Ghasemi Y, Rasoul-Amini S, Fotooh-Abadi E. The biotransformation, biodegradation, and bioremediation of organic compounds by microalgae. J Phycol. 2011;47:969–980.

4. EPA. Atmospheric concentrations of greenhouse gases. Environ. panel agency. https://www.epa.gov/climate-indicators/climate-change-indicators-atmospheric-concentrations-greenhouse-gases. Accessed April 2 2018.

5. Etheridge DM, Steele LP, Langenfelds RJRL, Francey RJ, Barnola JM, Morgan VI. Historical CO_2 records from the Law Dome DE08, DE08-2, and DSS ice cores. A compendium of data on global change. http://cdiac.ess-dive.lbl.gov/trends/co2/lawdome.html. Accessed December 6 2017.

6. NASA. Climate change: Vital signs of the planet: Carbon dioxide. 2017. https://climate.nasa.gov/vital-signs/carbon-dioxide/. Accessed December 6 2017.

7. Marland G, Boden TA, Andres RJ, Brenkert AL, C.A. J. Global, regional, and national fossil fuel CO_2 emissions. A compendium of data on global change. http://cdiac.ess-dive.lbl.gov/trends/emis/overview.html. Accessed December 6 2017.

8. Brennan L, Owende P. Biofuels from microalgae - a review of technologies for production, processing, and extractions of biofuels and co-products. Renew Sustain Energy Rev. 2010;14:557–577.

9. EPA. The sources and solutions: Fossil fuels. 2015. https://www.epa.gov/nutrientpollution/sources-and-solutions-fossil-fuels. Accessed November 15 2017.

10. European Academies Science Advisory Council. The current status of biofuels in the European Union, their environmental impacts and future prospects.

http://www.easac.eu/fileadmin/PDF_s/reports_statements/Easac_12_Biofuels_Complete.pdf. Accessed December 6 2017.

11. Rulli MC, Bellomi D, Cazzoli A, De Carolis G, D'Odorico P. The water-land-food nexus of first-generation biofuels. Sci Rep. 2016;6:22521.

12. Chisti Y. Biodiesel from microalgae. Trends Biotechnol. 2008;26:126–131.

13. Kumar S. Sub- and supercritical water technology for biofuels. In: Lee JW, editor. Advanced Biofuels and Bioproducts. New York: Springer; 2013. p. 147–83.

14. Greenwell HC, Laurens LML, Shields RJ, Lovitt RW, Flynn KJ. Placing microalgae on the biofuels priority list: a review of the technological challenges. J R Soc Interface. 2009;7:703–726.

15. Sheehan J, Dunahay T, Benemann J, Roessler P. Look back at the U.S. Department of Energy's Aquatic Species Program: Biodiesel from algae. https://www.nrel.gov/docs/legosti/fy98/24190.pdf. Accessed November 30 2017.

16. Muylaert K, Bastiaens L, Vandamme D, Gouveia L. Harvesting of microalgae: overview of process options and their strengths and drawbacks. Microalgae-based biofuels and bioproducts. Lisbon: Woodhead Publishing; 2017. p. 113–132.

17. Lee RE. Phycology. New York: Cambridge University Press; 2008. p. 3–30.

18. Serrano-Bermúdez LM, Serrano Bermúdez LM. Estudio de cuatro cepas nativas de microalgas para evaluar su potencial uso en la producción de biodiesel. 2012. http://www.bdigital.unal.edu.co/7825/1/299883.2012.pdf. Accessed June 10 2017.

19. May P, Wienkoop S, Kempa S, Usadel B, Christian N, Rupprecht J, et al. Metabolomics- and proteomics-assisted genome annotation and analysis of the draft metabolic network of *Chlamydomonas reinhardtii*. Genetics. 2008;179:157–166.

20. Christian N, May P, Kempa S, Handorf T, Ebenhöh O. An integrative approach towards completing genome-scale metabolic networks. Mol Biosyst. 2009;5:1889.

21. Imam S, Schäuble S, Valenzuela J, Lõpez García De Lomana A, Carter W, Price ND, et al. A refined genome-scale reconstruction of *Chlamydomonas* metabolism provides a platform for systems-level analyses. Plant J. 2015;84:1239–1256.

22. Boyle NR, Sengupta N, Morgan JA. Metabolic flux analysis of heterotrophic growth in *Chlamydomonas reinhardtii*. PLoS One. 2017;12.

23. Boyle NR, Morgan JA. Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. BMC Syst Biol. 2009;3:4.

24. Manichaikul A, Ghamsari L, Hom EFY, Lin C, Murray RR, Chang RL, et al. Metabolic network analysis integrated with transcript verification for sequenced genomes. Natl Inst Heal. 2009;6:589–592.

25. Dal'Molin CG de O, Quek L-E, Palfreyman RW, Nielsen LK. AlgaGEM - A genome-scale metabolic reconstruction of algae based on the *Chlamydomonas reinhardtii* genome. BMC Genomics. 2011;12 Suppl 4:S5.

26. Chang RL, Ghamsari L, Manichaikul A, Hom EFY, Balaji S, Fu W, et al. Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. Mol Syst Biol. 2011;7.

27. Cogne G, Rügen M, Bockmayr A, Titica M, Dussap CG, Cornet JF, et al. A model-based method for investigating bioenergetic processes in autotrophically growing eukaryotic microalgae: Application to the green algae *Chlamydomonas reinhardtii*. Biotechnol Prog. 2011;27:631–640.

28. Kliphuis AMJ, Klok AJ, Martens DE, Lamers PP, Janssen M, Wijffels RH. Metabolic modeling of *Chlamydomonas reinhardtii*: Energy requirements for photoautotrophic growth and maintenance. J Appl Phycol. 2012;24:253–266.

29. Rügen M, Bockmayr A, Legrand J, Cogne G. Network reduction in metabolic pathway analysis: Elucidation of the key pathways involved in the photoautotrophic growth of the green alga *Chlamydomonas reinhardtii*. Metab Eng. 2012;14:458–467.

30. Chaiboonchoe A, Dohai BS, Cai H, Nelson DR, Jijakli K, Salehi-Ashtiani K. Microalgal metabolic network model refinement through high-throughput functional metabolic profiling. Front Bioeng Biotechnol. 2014;2:1–12.

31. Wu C, Xiong W, Dai J, Wu Q. Genome-based metabolic mapping and ^{13}C Flux Analysis reveal systematic properties of anoleaginous microalga *Chlorella protothecoides*. Plant Physiol. 2015;167:586–599.

32. Zuñiga C, Li C-T, Huelsman T, Levering J, Zielinski DC, McConnell BO, et al. Genome-scale metabolic model for the green alga *Chlorella vulgaris* UTEX 395 accurately predicts phenotypes under autotrophic, heterotrophic, and mixotrophic growth Conditions. Plant Physiol. 2016;172:589–602.

33. Zuñiga C, Levering J, Antoniewicz MR, Guarnieri MT, Betenbaugh MJ, Zengler K, et al. Predicting dynamic metabolic demands in the photosynthetic eukaryote *Chlorella vulgaris*. Plant Physiol. 2017;176:450–462.

34. Xiong W, Liu L, Wu C, Yang C, Wu Q. ^{13}C-tracer and gas chromatography-mass spectrometry analyses reveal metabolic flux distribution in the oleaginous microalga *Chlorella protothecoides*. Plant Physiol. 2010;154:1001–1011.

35. Yang C, Hua Q, Shimizu K. Energetics and carbon metabolism during growth of microalgal cells under photoautotrophic, mixotrophic and cyclic light-autotrophic/dark-heterotrophic conditions. Biochem Eng J. 2000;6:87–102.

36. Pham N. Genome-scale constraint-based metabolic modeling and analysis of *Nannochloropsis sp.* 2016. https://brage.bibsys.no/xmlui/bitstream/handle/11250/2399905/12511_FULLTEXT.pdf?sequence=1&isAllowed=y. Accessed December 4 2017.

37. Loira N, Mendoza S, Paz Cortés M, Rojas N, Travisany D, Genova A Di, et al. Reconstruction of the microalga *Nannochloropsis salina* genome-scale metabolic model with applications to lipid production. BMC Syst Biol. 2017;11:66.

38. Shah AR, Ahmad A, Srivastava S, Jaffar Ali BM. Reconstruction and analysis of a genome-scale metabolic model of *Nannochloropsis gaditana*. Algal Res. 2017;26:354–364.

39. Shastri A, Morgan J. Flux balance analysis of photoautotrophic metabolism. Biotechnol Prog. 2005;21:1617–1626.

40. Navarro E, Montagud A, Fernández de Córdoba P, Urchueguía JF. Metabolic flux analysis of the hydrogen production potential in *Synechocystis sp.* PCC6803. Int J Hydrogen Energy. 2009;34:8828–8838.

41. Knoop H, Zilliges Y, Lockau W, Steuer R. The metabolic network of *Synechocystis sp.* PCC 6803: Systemic properties of autotrophic growth. Plant Physiol. 2010;154:410–422.

42. Yoshikawa K, Kojima Y, Nakajima T, Furusawa C, Hirasawa T, Shimizu H. Reconstruction and verification of a genome-scale metabolic model for *Synechocystis sp.* PCC6803. Appl Microbiol Biotechnol. 2011;92:347–358.

43. Montagud A, Zelezniak A, Navarro E, de Córdoba PF, Urchueguía JF, Patil KR. Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis sp.* PCC6803. Biotechnol J. 2011;6:330–342.

44. Montagud A, Navarro E, Fernández de Córdoba P, Urchueguía JF, Patil K. Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. BMC Syst Biol. 2010;4:156.

45. Knoop H, Gründel M, Zilliges Y, Lehmann R, Hoffmann S, Lockau W, et al. Flux balance analysis of cyanobacterial metabolism: The metabolic network of *Synechocystis sp.* PCC 6803. PLoS Comput Biol. 2013;9:e1003081.

46. Nogales J, Gudmundsson S, Knight EM, Palsson BO, Thiele I. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Proc Natl Acad Sci. 2012;109:2678–2683.

47. Lim DKY, Schuhmann H, Thomas-Hall SR, Chan KCK, Wass TJ, Aguilera F, et al. RNA-Seq and metabolic flux analysis of *Tetraselmis sp.* M8 during nitrogen starvation reveals a two-stage lipid accumulation mechanism. Bioresour Technol. 2017;244:1281–1293.

48. Bogen C, Al-Dilaimi A, Albersmeier A, Wichmann J, Grundmann M, Rupp O, et al. Reconstruction of the lipid metabolism for the microalga *Monoraphidium neglectum* from its genome sequence reveals characteristics suitable for biofuel production. BMC Genomics. 2013;14:926.

49. Krumholz EW, Yang H, Weisenhorn P, Henry CS, Libourel IGL. Genome-wide metabolic network reconstruction of the picoalga *Ostreococcus*. J Exp Bot. 2012;63:2353–2362.

50. Baroukh C, Muñoz-Tamayo R, Steyer JP, Bernard O. A new framework for metabolic modeling under non-balanced growth. Application to carbon metabolism of unicellular microalgae. PLoS One. 2013;12:107–112.

51. Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, et al. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. PLoS One. 2008;3:e1426.

52. Singh D, Carlson R, Fell D, Poolman M. Modelling metabolism of the diatom *Phaeodactylum tricornutum*. Biochem Soc Trans. 2015;43:1182–1186.

53. Kim J, Fabris M, Baart G, Kim MK, Goossens A, Vyverman W, et al. Flux balance analysis of primary metabolism in the diatom *Phaeodactylum tricornutum*. Plant J. 2016;85:161–176.

54. Levering J, Broddrick J, Dupont CL, Peers G, Beeri K, Mayers J, et al. Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. PLoS One. 2016;11:1–22.

55. Banerjee A, Banerjee C, Negi S, Chang J-S, Shukla P. Improvements in algal lipid production: a systems biology and gene editing approach. Crit Rev Biotechnol. 2017;38:369–385.

56. Čuperlović-Culf M, Čuperlović-Culf M. In: Nakamura K, editor. NMR Metabolomics in Cancer Research. Oxford: Woodhead Publishing; 2013. p. 365–383.

57. Tan J, Zuñiga C, Zengler K. Unraveling interactions in microbial communities - from co-cultures to microbiomes. J Microbiol. 2015;53:295–305.

58. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010;28:245–248.

59. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. 2010;5:93–121.

60. Feist AM, Palsson BØ. The growing scope of application of genome-scale metabolic reconstructions: the case of *E. coli*. Nat Biotechnol. 2008;26:659–667.

61. Overbeek R, Begley T, Butler RM, Choudhuri J V, Chuang HY, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 2005;33:5691–5702.

62. Snyder EE, Kampanya N, Lu J, Nordberg EK, Karur HR, Shukla M, et al. PATRIC: The VBI PathoSystems Resource Integration Center. Nucleic Acids Res. 2007;35:D401–406.

63. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. 2008;9:75.

64. Fritzemeier CJ, Hartleb D, Szappanos B, Papp B, Lercher MJ. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. PLoS Comput Biol. 2017;13:1–14.

65. Terzer M, Maynard ND, Covert MW, Stelling J. Genome-scale metabolic networks. Wiley Interdiscip Rev Biol Med. 2009;1:285–297.

66. Villadsen J, Nielsen J, Lidén G. Biochemical reaction networks. In: Nielsen J, Villadsen J, editors. Bioreaction engineering principles. 3rd ed. New York: Springer; 2011. p. 151–214.

67. Price ND, Papin JA, Schilling CH, Palsson BO. Genome-scale microbial *in silico* models: The constraints-based approach. Trends Biotechnol. 2003;21:162–169.

68. Chen K, Gao Y, Mih N, O'Brien EJ, Yang L, Palsson BO. Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation. Proc Natl Acad Sci. 2017;114:11548–11553.

69. Levering J, Dupont CL, Allen AE, Palsson BO, Zengler K. Integrated regulatory and metabolic networks of the marine diatom *Phaeodactylum tricornutum* predict the response to rising CO_2 levels. mSystems. 2017;2:e00142-16.

70. Takahashi O, Park Y-I, Nakamura Y. Biotechnology of microalgae, based on molecular biology and biochemistry of eukaryotic algae and cyanobacteria. FEBS Lett. 2009;583:3882–3890.

71. Wan MX, Wang RM, Xia JL, Rosenberg JN, Nie ZY, Kobayashi N, et al. Physiological evaluation of a new *Chlorella sorokiniana* isolate for its biomass production and lipid accumulation in photoautotrophic and heterotrophic cultures. Biotechnol Bioeng. 2012;109:1958–1964.

72. Toledo-Cervantes A, Garduño Solórzano G, Campos JE, Martínez-García M, Morales M. Characterization of *Scenedesmus obtusiusculus* AT-UAM for high-energy molecules accumulation: deeper insight into biotechnological potential of strains of the same species. Biotechnol Reports. 2018;17:16–23.

73. Cabello J, Morales M, Revah S. Dynamic photosynthetic response of the microalga *Scenedesmus obtusiusculus* to light intensity perturbations. Chem Eng J. 2014;252:104–111.

74. Dikicioglu D, Kırdar B, Oliver SG. Biomass composition: the "elephant in the room" of metabolic modelling. Metabolomics. 2015;11:1690–1701.

75. García Sánchez CE, Torres Sáez RG. Comparison and analysis of objective functions in flux balance analysis. Biotechnol Prog. 2014;30:985–991.

76. Serrano-Bermúdez LM, González Barrios AF, Maranas CD, Montoya D. *Clostridium butyricum* maximizes growth while minimizing enzyme usage and ATP production: Metabolic flux distribution of a strain cultured in glycerol. BMC Syst Biol. 2017;11:1–13.

77. Feist AM, Palsson BO. The biomass objective function. Curr Opin Microbiol. 2010;13:344–349.

78. Broddrick JT, Rubin BE, Welkie DG, Du N, Mih N, Diamond S, et al. Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. Proc Natl Acad Sci. 2016;113:E8344–53.

79. Muthuraj M, Palabhanvi B, Misra S, Kumar V, Sivalingavasu K, Das D. Flux balance analysis of *Chlorella sp.* FC2 IITG under photoautotrophic and heterotrophic growth conditions. Photosynth Res. 2013;118:167–179.

80. Mahadevan R, Edwards JS, Doyle FJ. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys J. 2002;83:1331–1340.

81. Cuthrell JE, Biegler LT. On the optimization of differential-algebraic process systems. AIChE J. 1987;33:1257–1270.

82. Bordbar A, Yurkovich JT, Paglia G, Rolfsson O, Sigurjónsson ÓE, Palsson BO. Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics. Sci Rep. 2017;7:1–12.

83. Wiechert W. ^{13}C metabolic flux analysis. Metab Eng. 2001;3:195–206.

84. Trinh CT, Wlaschin A, Srienc F. Elementary mode analysis: A useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl Microbiol Biotechnol. 2009;81:813–826.

85. Baroukh C, Muñoz-Tamayo R, Bernard O, Steyer JP. Mathematical modeling of unicellular microalgae and cyanobacteria metabolism for biofuel production. Curr Opin Biotechnol. 2015;33:198–205.

86. Nikolaev E V., Burgard AP, Maranas CD. Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions. Biophys J. 2005;88:37–49.

87. Nelson DL, Cox MM. Glycolysis, gluconeogenesis, and the pentose phosphate pathway. In: Palgrave Macmillan, editor. Lehninger Principles of Biochemistry. 4th ed. New York: W. H. Freeman; 2008. p. 521–60.

88. Coronil T, Lara C, Guerrero MG. Shift in carbon flow and stimulation of amino-acid turnover induced by nitrate and ammonium assimilation in *Anacystis nidulans*. Planta. 1993;189:461–467.

89. Rai AK. Symbiotic systems with cyanobacteria-cyanobioses. In: Rai AK, editor. Cyanobacterial nitrogen metabolism and environmental biotechnology. New Delhi: Narosa Pub. House; 1997. p. 299.

90. Juneja A, Chaplen FWR, Murthy GS. Genome-scale metabolic reconstruction of *Chlorella variabilis* for exploring its metabolic potential for biofuels. Bioresour Technol. 2015;213:103–110.

91. Hamilton JJ, Reed JL. Identification of functional differences in metabolic networks using comparative genomics and constraint-based models. PLoS One. 2012;7.

92. Vu TT, Hill EA, Kucek LA, Konopka AE, Beliaev AS, Reed JL. Computational evaluation of *Synechococcus sp.* PCC 7002 metabolism for chemical production. Biotechnol J. 2013;8:619–630.

93. Qian X, Kim MK, Kumaraswamy GK, Agarwal A, Lun DS, Dismukes GC. Flux balance analysis of photoautotrophic metabolism: Uncovering new biological details of subsystems involved in cyanobacterial photosynthesis. Biochim Biophys Acta - Bioenerg. 2016;1858:276–287.

94. Hendry JI, Prasannan CB, Joshi A, Dasgupta S, Wangikar PP. Metabolic model of *Synechococcus sp. PCC 7002*: Prediction of flux distribution and network modification for enhanced biofuel production. Bioresour Technol. 2016;213:190–197.

95. Yang C. Metabolic flux analysis in *Synechocystis* using isotope distribution from ^{13}C-labeled glucose. Metab Eng. 2002;4:202–216.

2.7. Tables

Table 2.1. Characteristics of current metabolic models of oleaginous microalgae. Metabolic models are classified in two different groups: Genome-scale metabolic models (GSM) and core models (CM); whereas the analyses were classified in: Flux Balance Analysis (FBA), Dynamic FBA (dFBA), Elementary Modes (EM), Metabolic Flux Analysis (MFA), MFA using ^{13}C Tracer (^{13}C MFA), and their combinations.

Organism	Metabolic model (ID)	Analysis	Genes	Reactions	Metabolites	Compartments	Citations [Reference]
Chlamydomonas reinhardtii	GSM	-	1069	-	-	-	143 [19]
Chlamydomonas reinhardtii	GSM	-	-	1500	1200	-	53 [20]
Chlamydomonas reinhardtii	GSM	FBA	-	484	458	3	292 [23]
Chlamydomonas reinhardtii	GSM	FBA	-	259	-	10	82 [24]
Chlamydomonas reinhardtii	GSM (AlgaGEM)	FBA	2249	1725	1862	4	96 [25]
Chlamydomonas reinhardtii	GSM (iRC1080)	FBA	1080	2190	1068	10	231 [26]
Chlamydomonas reinhardtii	CM	FBA	-	280	278	-	47 [27]
Chlamydomonas reinhardtii	GSM	FBA	-	160	164	2	100 [28]
Chlamydomonas reinhardtii	GSM	FBA	-	280	278	0	12 [29][1]
Chlamydomonas reinhardtii	GSM (iBD1106)	FBA	1106	2445	1959	10	10 [30][2]
Chlamydomonas reinhardtii	GSM (iCre1355)	FBA	1355	2394	1133	10	12 [21]
Chlamydomonas reinhardtii	GSM	FBA/^{13}C MFA	-	139	-	3	2 [22]
Chlorella protothecoides	CM	^{13}C MFA	-	24	19	0	83 [34]
Chlorella protothecoides	GSM	FBA/^{13}C MFA	461	272	-	4	0 [31]
Chlorella pyrenoidosa	CM	MFA	-	67	-	0	258 [35]
Chlorella sp.	CM	dFBA	-	114	161	-	31 [79]

Chlorella variabilis	GSM (iAJ526)	FBA	526	1455	1236	5	10 [90]
Chlorella vulgaris UTEX 395	GSM (iCZ843)	FBA	843	2294	1770	6	14 [32]
Chlorella vulgaris UTEX 396	GSM (iCZ947)	dFBA	946	2294	1770	6	2 [33][3]
Nannochloropsis gaditana	GSM (iRJ1321)	FBA	1321	1918	1862	4	1 [38]
Nannochloropsis salina	GSM (iNS934)	dFBA	934	2345	-	10	4 [37]
Nannochloropsis sp.	GSM	FBA	383	987	1024	6	0 [36]
Ostreococcus lucimarinus	GSM	FBA	-	964	1100	2	38 [49]
Ostreococcus tauri	GSM	FBA	-	871	1014	2	38 [49]
Phaeodactylum tricornutum	GSM	-	151	88	-	5	289 [51]
Phaeodactylum tricornutum	GSM	FBA	-	-	-	2	12 [52]
Phaeodactylum tricornutum	GSM	FBA	607	849	587	6	27 [53]
Phaeodactylum tricornutum	GSM (iLB1027)	FBA	1027	4456	2172	6	24 [54]
Synechococcus elongatus PCC7942	GSM (iJB785)	FBA	785	850	768	7	13 [78]
Synechococcus sp. PCC 7002	GSM (iSyp611)	FBA	611	552	542	2	39 [91]
Synechococcus sp. PCC 7002	GSM (iSyp708)	FBA	708	646	581	2	39 [92]
Synechococcus sp. PCC 7002	GSM (iSyp821)	FBA	821	792	777	3	3 [93]
Synechococcus sp. PCC 7002	GSM (iSyp728)	FBA	728	742	696	7	22 [94]
Synechocystis sp. PCC 6803	CM	^{13}C MFA	-	29	-	-	181 [95]
Synechocystis sp. PCC 6803	CM	FBA	-	70	46	2	165 [39]

Synechocystis sp. PCC 6803	CM	FBA	-	43	-	-	43 [40]
Synechocystis sp. PCC 6803	GSM	FBA	-	380	291	6	159 [41]
Synechocystis sp. PCC 6803	GSM	FBA	669	882	790	2	113 [44]
Synechocystis sp. PCC 6803	GSM (iSyn811)	FBA	811	956	911	2	59 [43]
Synechocystis sp. PCC 6803	GSM	FBA/^{13}C MFA	-	493	465	2	51 [42]
Synechocystis sp. PCC 6803	GSM (iJN678)	FBA	678	863	795	3	206 [46]
Synechocystis sp. PCC 6803	GSM	FBA	677	759	601	6	143 [45]
Tetraselmis sp.	GSM	FBA	2249	1725	1862	4	2 [47][4]
Tisochrysis lutea	CM	EM	-	157	162	2	2 [50]

[1] Modified the metabolic model of C. reinhardtii from Cogne et al. [27].

[2] Modified the metabolic model of C. reinhardtii from Chang et al. [26].

[3] Used the genome-scale model of C. vulgaris from Zuñiga et al. [32].

[4] Used the genome-scale model of C. reinhardtii from Dal'Molin et al. [25] with constraints for *Tetraselmis* sp.

Chapter 3. Objectives

General: Predict the dynamic behavior of cellular concentration and composition of an oleaginous microalga in a photobioreactor.

Specific:

1. Generate a spatiotemporal genome-scale metabolic model for *Chlorella vulgaris* based on the model *iCZ947*, which includes the influence of spatial distribution of photon flux, phenomena of photoinhibition and recovery of photosystem II, as well as nutrient uptake kinetics for the transient prediction of growth and cellular composition.

2. Validate growth and cellular composition predictions of the model by contrasting them with previously reported experimental kinetic data of *Chlorella vulgaris*.

3. Calculate the conditions of light intensity, photoperiod and culture timespan which maximize the global lipid productivity of an internally-illuminated stirred-tank photobioreactor.

Chapter 4. Multiscale metabolic modeling of the oleaginous microalga *Chlorella vulgaris* in a photobioreactor

The content of this chapter will be submitted as a research paper with authors: Juan D. Tibocha-Bonilla, Cristal Zuñiga, Jared T. Broddrick, Karsten Zengler, Rubén D. Godoy-Silva under the same title.

4.1. Abstract

Metabolic modeling of microalgae has been a resourceful tool to predict and analyze metabolic behavior of organisms for almost one decade. In the case of oleaginous microalgae, some genome-scale metabolic models have been generated and improved to study their metabolism. However, little effort has been made on applying metabolic models to control large-scale cultures of industrial interest. For such purpose, nutrients starvation as well as light uptake and attenuation have been identified as drivers of the process performance. In this work, we combined the latest genome-scale metabolic model of *Chlorella vulgaris* with kinetic models that consider light uptake, photoinhibition, nitrogen and carbon uptake, metabolite-specific carbon allocation (carbohydrates, lipids, and nucleotides), and reactor geometry. We successfully predicted growth under different growth conditions and light intensities. The developed model's robustness enabled to design strategies upon different light sources as well as culture timing for increased lipid productivity.

Key words – Photobioreactor, reactor design and optimization, microalgae bioengineering, genome-scale metabolic modeling, multiscale modeling.

4.2. Introduction

Microalgae hold an important position in biological and industrial research, as they serve as model organisms for the study of photosynthetic pathways and interspecific interactions, as well as the production of biofuels. Oleaginous microalgae are capable of accumulating lipids beyond 20% of dry cell weight in reduced culture areas and without the requirement of organic carbon input, which renders them remarkable for third and fourth generation biofuel production, especially *Chlorella vulgaris* [1–3].

However, several drawbacks are yet to be overcome for these biofuels to be economically feasible. First and most importantly, the process of lipid extraction often accounts for about 50% of the final cost, as water adds up to around 99% of the mass content of the separated algal biomass. Moreover, the high protein content of the algal cell turns into high nitrogen and sulfur content of the final product, which ends up lowering the quality of the fuel and deteriorating fungibility. Therefore, the maximization of lipid productivity in microalgae bioreactors is necessary, as it has been reported that a lipid content of 20 – 40% is required for profitability [4], depending on culture conditions. main

The maximization of lipid productivity in microalgae has been a main field of study for several decades, mainly by varying light, nitrogen and growth mode conditions. Yet, experimental determination of optimal conditions is extremely time- and resource-

intensive, which makes modeling a resourceful alternative for photobioreactor design and optimization.

4.2.1. Metabolic modeling of oleaginous microalgae

The industrial application and evolutionary importance of microalgae have rendered them a focus of systems biology for the past decade. Early GSM models of oleaginous microalgae consisted of non-compartmentalized networks [5], such as *Chlamydomonas* [6] and *Synechocystis* [7]. Further improvement of the models allowed the study of organelle functionality and pathway coupling, as well as the interactions between photosynthetic pathways with the rest of the network under different light wavelengths [8].

However, the importance that has been given to the metabolic modeling of large-scale bioreactors is little, and non-existent for photobioreactors. So far, the metabolic modeling of oleaginous microalgae has focused on carbon allocation, organelle functions, and community modeling. These, although having been vital for the formulation and curation of the models, are not enough for predicting reactor scale-up and optimization.

4.2.2. Modeling of large-scale photobioreactors

Reactor scale-up comes with a series of improvements on productivity, unit cost, and profitability, but at the expense of amplified operational inefficiencies. In microalgae photobioreactors, such issues include shading (light attenuation), photoinhibition and mixing. Other drawbacks that are inherent to microalgae cannot be ignored for the correct prediction of a complete growth curve, namely nutrient depletion, and the tradeoff between growth and lipid production. In this work, we generated the first-ever multiscale

metabolic model of a photobioreactor, by including the modeling of sub-optimal growth phenomena in the latest genome-scale metabolic model of *C. vulgaris* (*iCZ947*) as follows: light attenuation and uptake, photoinhibition, nitrogen and carbon uptake kinetics, and a carbon allocation algorithm (carbohydrate and lipid accumulation and consumption).

4.3. Results and Discussion

4.3.1. Prediction of biomass composition and growth

In their study, Adesanya et al. [9] cultivated *Chlorella vulgaris* at two different initial nitrogen concentrations and tracked macromolecular content of the cells through the culture time. Initial nitrate concentrations of 0.35 and 1.89 mM were employed to evaluate the impact of nitrogen availability on carbon allocation and growth. This allowed us to test our model for its capability of predicting intracellular concentrations of molecules of interest, specifically triacylglycerols (TAGs). As explained in *Methods*, our model includes a mechanistic approach to the modeling of carbon distribution across the cell, which is bound to provide it with the ability of predicting the microscopic and macroscopic effect of nitrogen concentration.

We used one set of reported data (Figure 4.1A) for the regression of strain-specific parameters (shown in Table 4.2) and simulated a second scenario (Figure 4.1B) to test for predictive capability. In the first scenario, a relatively low initial nitrate concentration in the media (exactly half that of standard BBM medium [10]) caused the size of the internal nitrogen pool to decrease steadily throughout the time of culture (Figure 4.1D). Since the

microalga was not able to replenish its nitrogen reserves, lipid accumulation was triggered 100 h after nitrogen was depleted in the medium. More accurately, nitrogen depletion from the medium signifies the beginning of the end of exponential growth, rather than the end itself. A similar behavior was obtained by Mansouri et al. [11] under a comparable setup. They reported that exponential growth was maintained for the first 96 h of growth, time after which growth gradually stopped until their last recorded instance at 168 h.

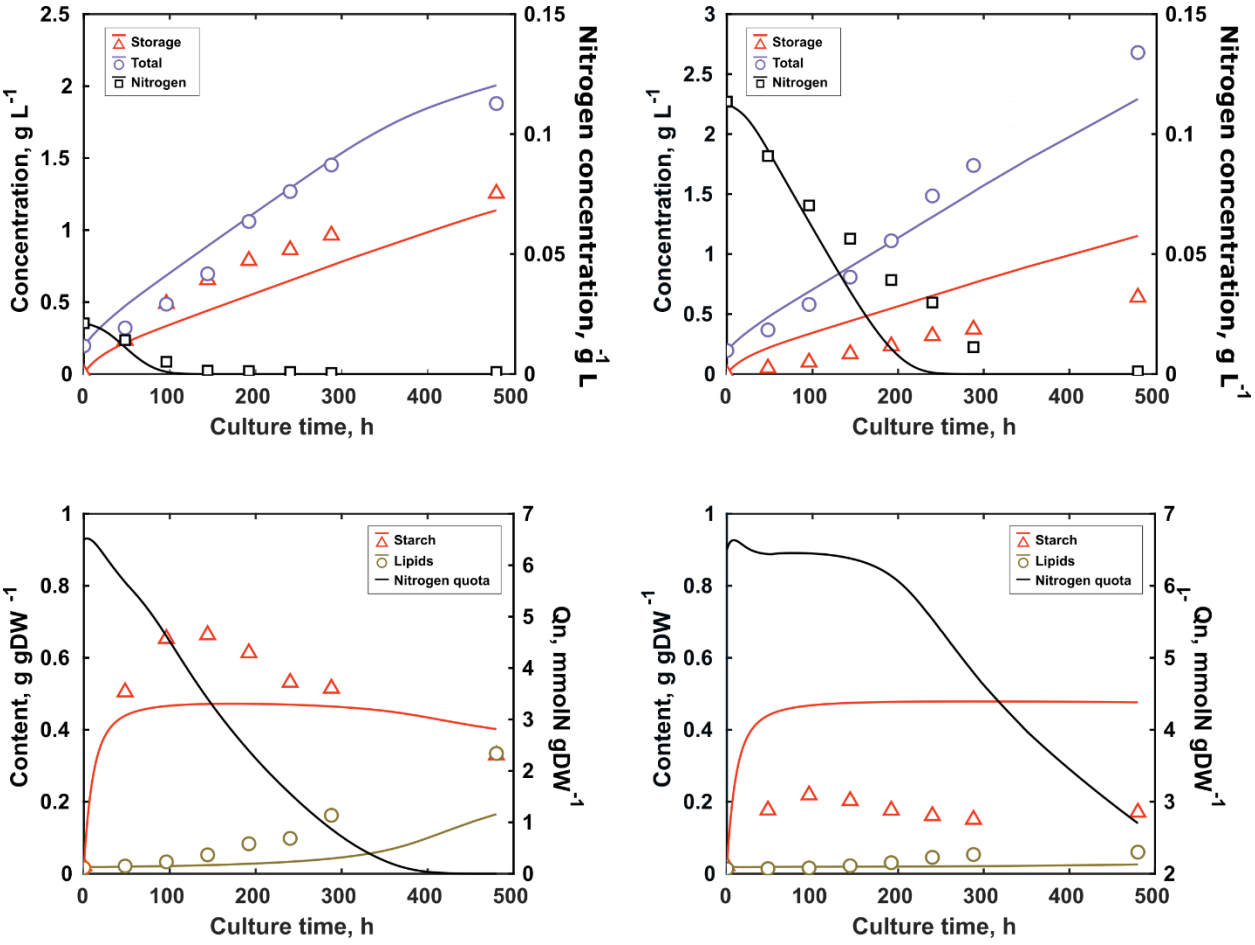

Figure 4.1. Simulated and reported data of the culture of *Chlorella vulgaris* at two different initial nitrogen concentrations. (A) and (C) correspond to the experiment with an initial nitrogen concentration of 0.021 g L^{-1}, while (B) and (D) were recorded under one of 0.124 g L^{-1}. All experiments were reported under a continuous irradiance of 80 $\frac{\mu E}{m^2 s}$. Continuous lines and markers represent predicted data by our model and reported data by Adesanya et al. [9], respectively.

Under the growth conditions of the second scenario, nitrogen availability was increased six-fold. However, growth rate was only amplified from an average of 0.0032 to 0.0038 h^{-1} (19% increase), which was still one order of magnitude lower than previously reported maximum growth rates of 0.039 [3] and 0.033 h^{-1} [11,12]. This was mainly caused by the comparatively reduced irradiance used by Adesanya et al. [9], as opposed to working irradiances of 648 [2,3], 20 – 1400 [13], and 30 – 848 [12] $\frac{\mu E}{m^2 s}$. As reported and shown by our predictions, low light irradiance hindered the overall growth rate of the microalga, which, in addition to its higher nitrogen availability, allowed it to replenish its nitrogen reserves for the first 200 h, and caused it to deplete nitrogen from the medium at 230 h, as well as its internal nitrogen reserves 250 h later. This relative nutrient sufficiency caused storage molecule (lipid and starch) production to drop and rendered lipid accumulation almost non-present. A visible over-estimation of starch content in the second scenario was mainly caused by the prioritization of starch consumption in the dark period of the topology of our carbon allocation algorithm (see Figure S4.2), which induces error when trying to predict a permanently-illuminated culture, as used by Adesanya et al. [9]. Further work on the generation of a GSM model for the mixotrophic growth of *C. vulgaris* will be necessary for this model to properly include starch consumption in the light period, as carbon allocation and the differential destination of carbon sources under mixotrophy is not yet well understood and accounted for in the models.

4.3.2. Simulation at different light intensities

Data reported by Kim et al. [12] at an irradiance of 848 $\mu E\, m^{-2}\, s^{-1}$ was used for the regression of parameters (see Methods), while data at 30, 55, 80, 197 and 476 $\mu E\, m^{-2}\, s^{-1}$ were employed for model validation. Simulation results of the regression at an irradiance of 848 $\mu E\, m^{-2}\, s^{-1}$ are shown in Figure 4.2.

Experimental total biomass concentrations were followed closely by our model, and both internal and external concentrations of important nutrients and macromolecules were predicted. Even though intracellular concentrations were not measured by Kim et al., the model could be used to theorize about microscopic and macroscopic phenomena that lay underneath. As opposed to the case of Adesanya et al. [9], an elevated irradiance of 848 $\frac{\mu E}{m^2 s}$ made impossible for the microalga to maintain the intracellular nitrogen levels after the nitrogen was depleted from the medium at 100 h. This caused the lipid production to be triggered around the same time point, as intracellular nitrogen levels were already decreasing sharply (Figure 4.2B).

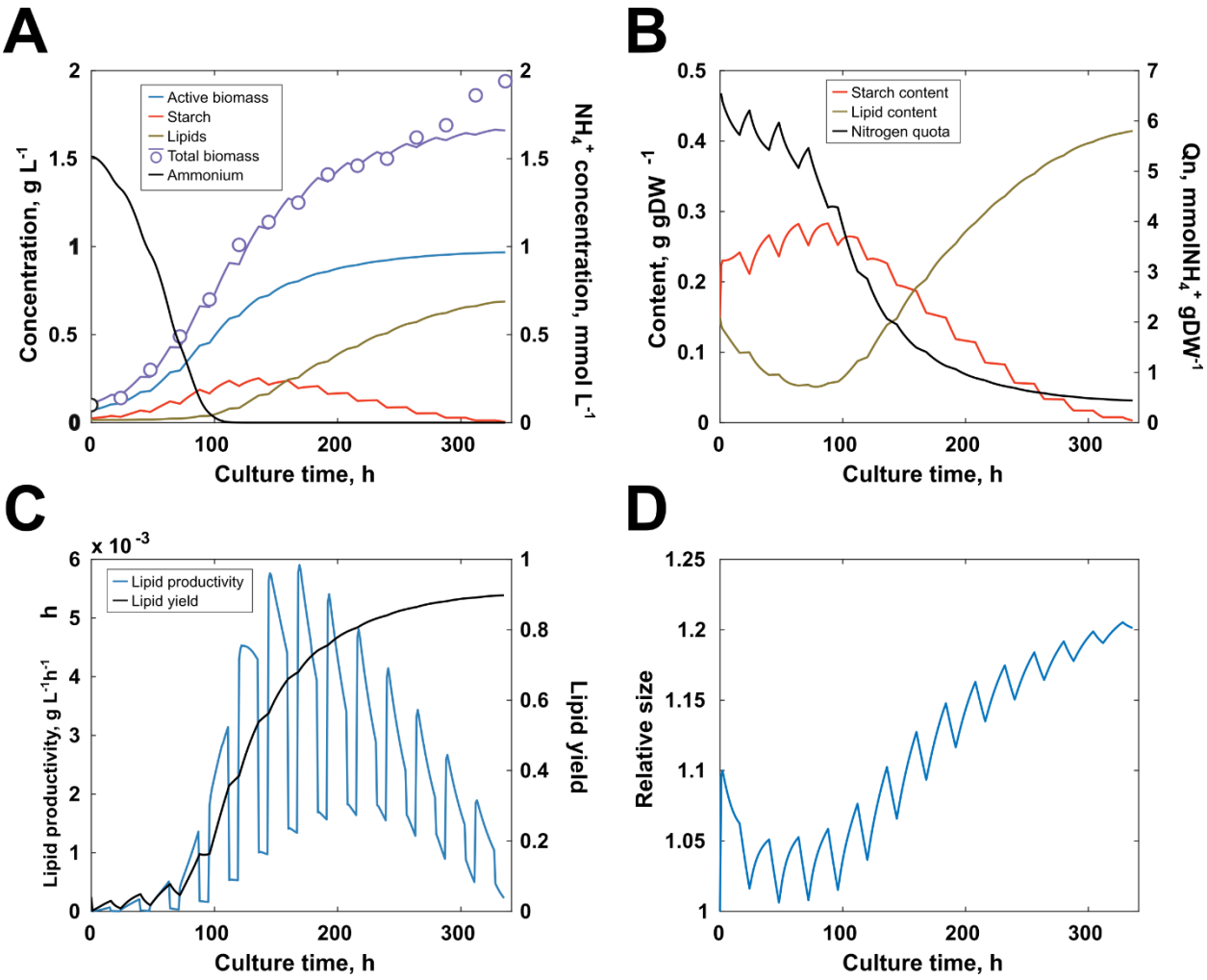

Figure 4.2. Simulation of results of the culture of *Chlorella vulgaris* contrasted with data reported by Kim et al [12] at 848 $\frac{\mu E}{m^2 s}$ and using a 16:8 light/dark strategy. (A) Global reactor concentrations of active (non-storage) biomass, starch, lipids, total biomass and nitrate contrasted with reported data of total biomass. (B) Intracellular content of starch, lipids and nitrogen. (C) Contrast of lipid productivity with lipid yield (% of carbon input directed to lipid production). (D) Variation of cell size. Continuous lines and markers represent predicted data by our model and reported data, respectively.

Circadian clock oscillations are also evident in all monitored variables. Figure 4.2B and Figure 4.2D show the starch accumulation-consumption cycles, along with a macroscopic interchange of carbon flow between starch and lipids after nitrogen depletion. Other oleaginous microalgae, such as *Synechococcus elongatus* [14], *Chlorella sorokoniana* [15] and *Dunalliela salina* [16], have been shown to exhibit such circadian clock oscillations. In

one study, *S. elongatus* exhibited a peak in ADP-glucose pyrophosphorylase activity, as well as in glycerol-3-phosphate (G3P) production from ribulose biphosphate (RuBP) coming from the reductive pentose-phosphate pathway (PPP) close to dawn, implying high starch production in the light period [14]. The same study found the exact opposite behavior in the dark, with peak activities of glygocen debranching enzyme (glgX).

Moreover, Figure 4.2C and Figure 4.2D exhibit the well-known trade-off behavior between specific lipid bioproduction and growth rate. According to our simulations, the optimum lipid productivity was achieved shortly after nitrogen was consumed from the medium but rapidly decreased afterward. It is worthy to note that even though nitrogen depletion from the medium was achieved at 100 h, peak global lipid productivity of the photobioreactor took place at 168 h, when internal reserves were running low but were not completely depleted yet. Overall lipid productivity decreased steadily for the following 167 h, as growth was increasingly hindered by internal nitrogen depletion (where $q = q_n$) and the cell size reaching its limit.

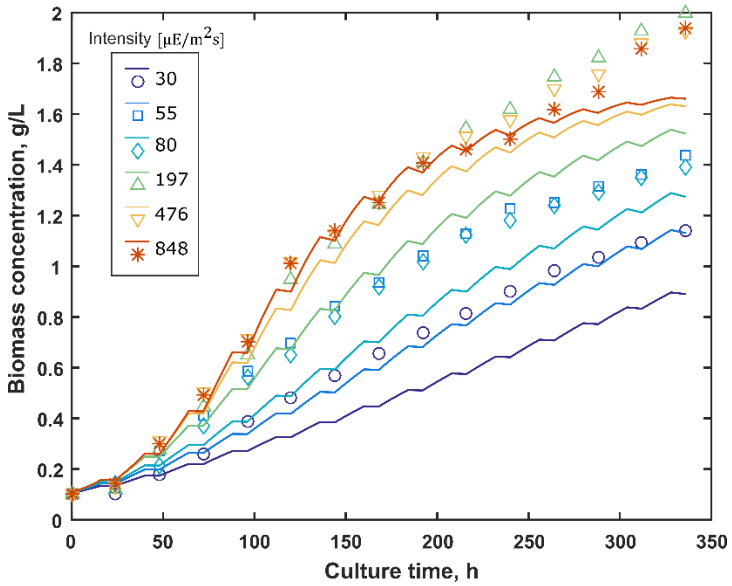

Figure 4.3. Simulation results of the experimental conditions reported by Kim et al. [12] at different irradiance conditions. Lines represent model simulations while markers show reported experimental data.

Figure 4.3 shows simulation results of the multiscale metabolic model of this jacketed cylindrical photobioreactor. The regressed model successfully predicted growth behavior in this geometry at different irradiance conditions. Even though it showed a sub-estimation of biomass production at low intensities, the overall growth trends were predicted accurately. Low intensities slowed down carbon dioxide uptake (data not shown) and increased overall shading in the photobioreactor, thus lowering the steepness of the growth curves. Photoinhibition and light attenuation influences were visible at high intensities since an increase from 476 to 848 μE m^{-2} s^{-1} did not signify an improvement in the overall culture growth.

4.3.3. Optimization of the lipid productivity in an internally-illuminated stirred-tank photobioreactor

In order to illustrate the model's aptitude for process design and optimization, we predicted the optimal light strategy to maximize lipid productivity in an internally-illuminated stirred-tank photobioreactor reactor with six radially-distributed fluorescent lamps.

In general, five variables were manipulated to search for the optimal global lipid productivity condition: lamp irradiance at time zero I_0, lamp irradiance at the end of the culture I_f, photoperiod p, culture duration t_f, and shape of irradiance temporal profile (see *Methods*) represented by the coefficient b_I. A hypothetical base case was given to the model as the initial condition of the optimization. Figure S4.3 shows a summary of the simulation results of the base case, and Table 4.1 summarizes the base case and optimized values of the manipulated variables.

Table 4.1. Initial and final values of the manipulated variables of the optimization.

	I_0 $\mu E\ m^{-2}\ s^{-1}$	I_f $\mu E\ m^{-2}\ s^{-1}$	p h	t_f h	b_I
Base case	400	600	16	300	0.5
Optimized	966	966	17	374	0.0

As shown in Figure 4.4, a final biomass concentration of 1.83 g/L is achieved after 374 h, which is roughly a 7% increase in biomass for 25% more culture time. Even though this does not seem as much of an improvement, final global lipid concentration increased 46%.

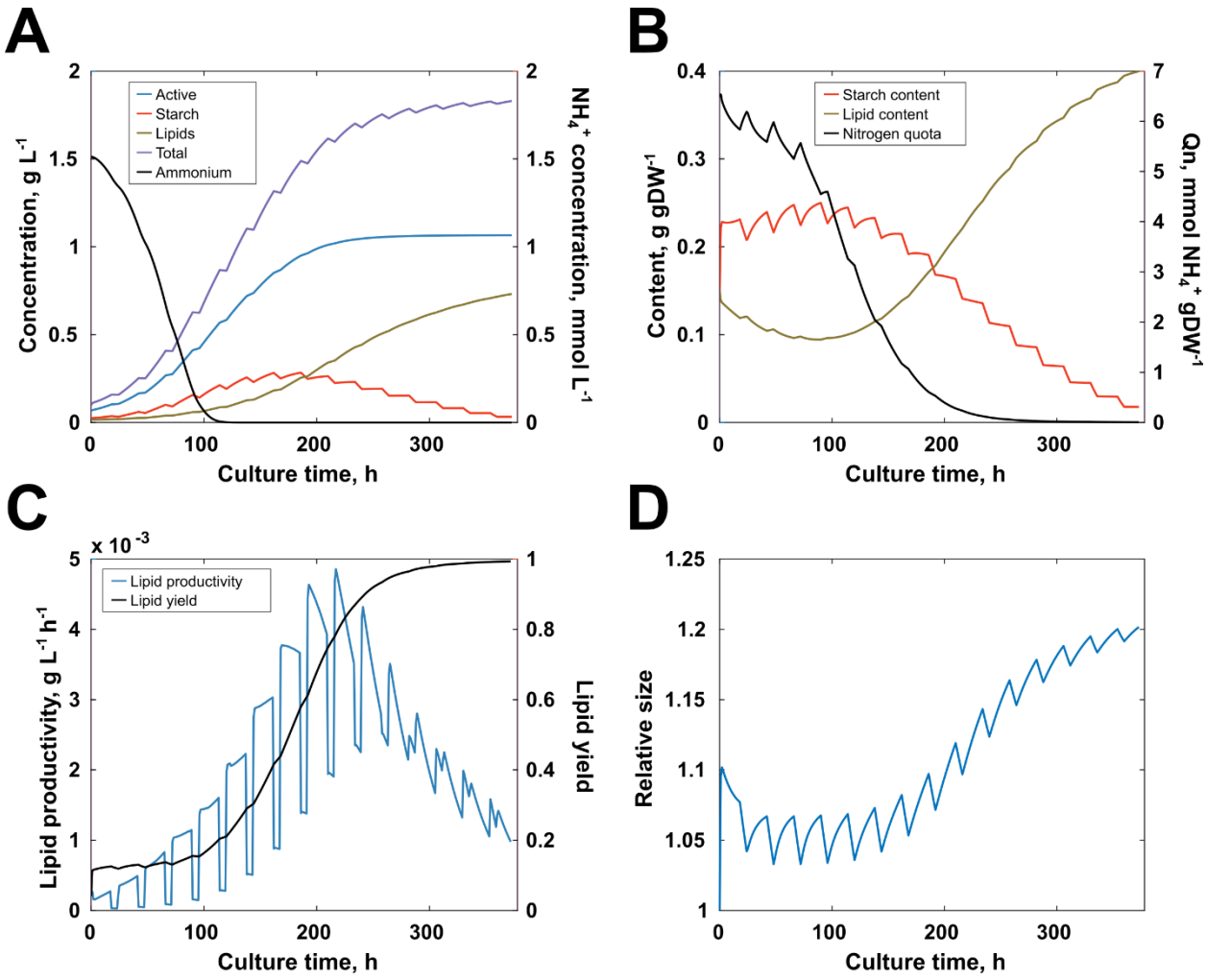

Figure 4.4. Simulation results of the maximization of lipid productivity by varying light strategy, photoperiod and culture time.

As opposed to previous trials on light strategies which irradiance increase in a stairstep fashion [17], our model predicted an optimal global light productivity at a constant lamp irradiance of 966 μE m^{-2} s^{-1}. Even though that means that at the early stage of the culture (0 – 100 h) a fraction of the culture is subjected to an irradiance of around 3000 μE m^{-2} s^{-1}, a large portion of it is under lower irradiances but still relatively high irradiances of 200 μE m^{-2} s^{-1}, which along with high nitrogen availability, favors higher growth rates in such a way that photoinhibition is compensated. Moreover, the optimization showed that

a photoperiod of 17:7 is enough for full photoreparation to take place without further negative impact on growth and finally lipid productivity.

As early as 25 h into the culture, the highest irradiance inside the culture lowers to 2400 $\mu E\ m^{-2}\ s^{-1}$ which, according to Pfendler et al., is the limit above which photoinhibition seriously hinders light uptake in *C. vulgaris* [18]. During the medium-growth stage (100 - 200 h) shading rapidly diminishes photoinhibition from approximately 80% to 25%, as highest irradiances are of only 800 $\mu E\ m^{-2}\ s^{-1}$, and the average drops to 200 $\mu E\ m^{-2}\ s^{-1}$.

At the low-growth stage (200+ h), shading is so substantial that the average irradiance drops to 142 $\mu E\ m^{-2}\ s^{-1}$ and stabilizes there for the rest of the culture. Moreover, as exhibited in Figure 4.5, at this point light uptake has almost halted throughout the majority of the reactor, with an average of only 70 $mmol\ gDW^{-1}\ h^{-1}$, as opposed to an average of 1000 $mmol\ gDW^{-1}\ h^{-1}$ at the high-growth stage. At this point, increasing the irradiance of the lamps hardly alters the global light availability and uptake, but critically increases irradiance in the vicinity of the lamps, where nitrogen-deplete microalgae are no longer capable of compensating photoinhibition with amplified metabolic activity. This means that, for this case, an individual lamp irradiance of 966 $\mu E\ m^{-2}\ s^{-1}$ is not high enough to halt growth at the early stage and is not low enough to do it at the late stage, which renders it optimal for overall growth and lipid production in the photobioreactor.

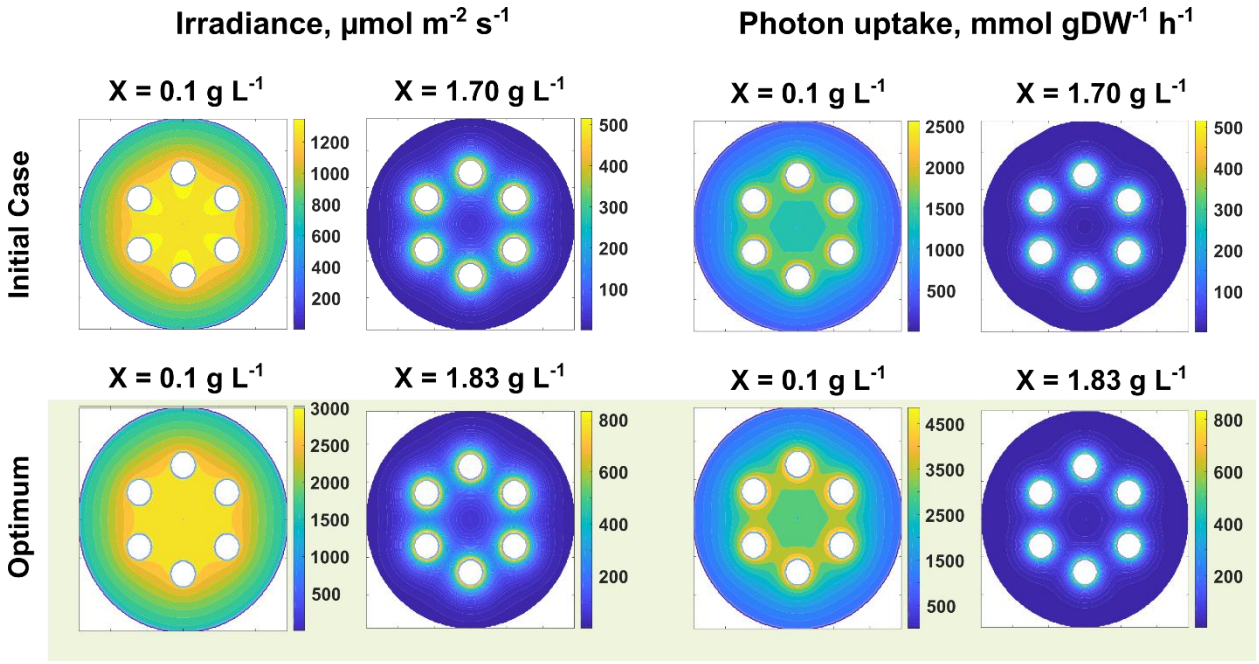

Figure 4.5. Light and photon uptake distributions at the beginning and end of the culture for both the initial and the final case (optimum).

In addition, it is worth noting that our model does not yet include neither the modeling of heat transfer mechanisms between the lamps, medium and surroundings, nor mixing phenomena which causes it to assume every cell is at a fixed position and subjected to the same temperature. With this, the found optimum is only attainable under a cooling system that is efficient enough to maintain overall temperature between 22 – 26 °C [19].

4.4. Methods

All simulations were carried out within the MATLAB 2016b (MathWorks Inc.) environment and using the COBRA Toolbox v3.0 [20]. Dynamic Flux Balance Analysis (dFBA) was used for time-course flux distribution calculations and concentration updates, and GUROBI

7.5.2 was employed as the solver for the linear optimization problems. A more detailed explanation of the model's algorithms is shown in this section.

4.4.1. The multiscale metabolic model

At the core of our calculations lies the genome-scale metabolic model of the oleaginous microalga *Chlorella vulgaris*: *iCZ947* [2], with previously proposed modifications for both heterotrophic and autotrophic growth [2,3]. Overall, the GSM model was solved using COBRA Toolbox (dFBA) for metabolic flux distributions. In addition, a set of additional models were included to account for secondary phenomena which constrained the solution space of the Linear Programming (LP) system (GSM model in Figure 4.6). Phenomena were included according to previous reports of specific physical and physiological mechanisms significantly affecting growth, as well as our expertise in microalgae culture. Included mechanisms were light attenuation, light uptake, photoinhibition, nitrogen and carbon uptake kinetics, and carbon allocation (carbohydrate and lipid accumulation and consumption). Mixing, heat and mas transfer phenomena were not included in the present model. A simplified representation of the general numerical algorithm is presented in Figure 4.6.

4.4.2. Light attenuation

Several studies have focused on light attenuation of microalgae [21–26], with a few solely on *Chlorella vulgaris* [21,22]. In this work, we decided to use the model for light absorption and scattering proposed by Naderi et al. [22], which allowed to accurately predict light

distribution at low and high cell densities. The intensity profile function is shown in Eq. (1).

$$I(r, X) = I_0 \exp\left(-r \cdot K_a \cdot X \frac{r^w}{p_k + r^w}\right) \quad (1)$$

For internally-lit reactors, the distance *r* was computed as the distance between the edge of the light source and any given point inside the culture. Several internal sources were accounted for by calculating individual light distributions and adding them up. For externally-lit (jacketed) reactors *r* was calculated as the distance between the illuminated edge of the reactor and any given point inside the culture.

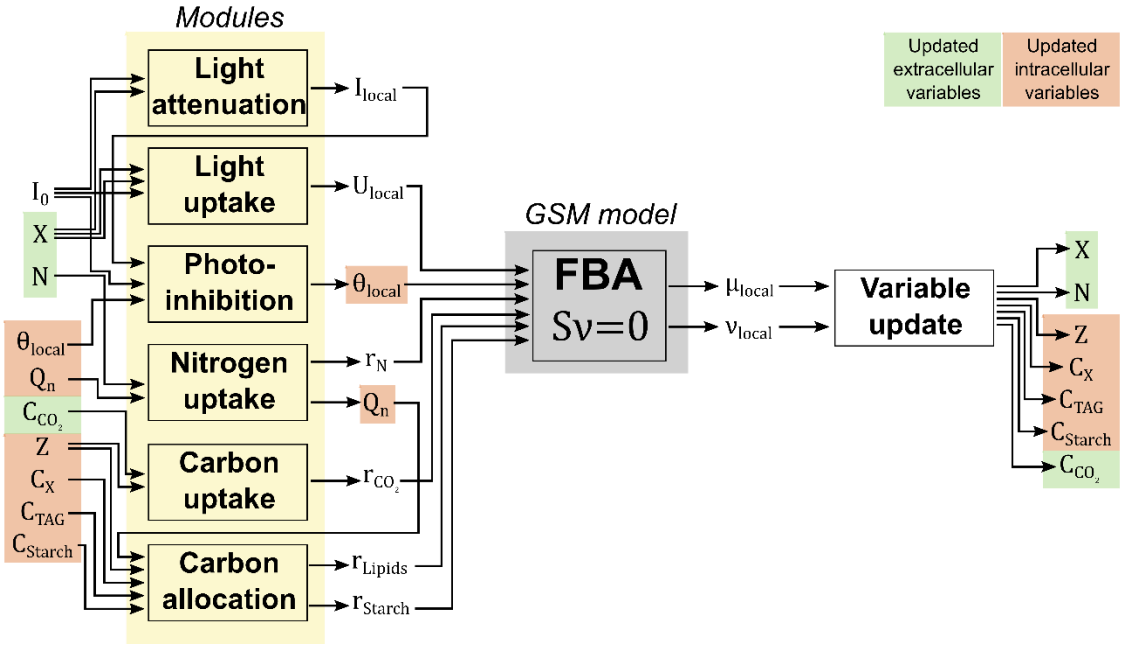

Figure 4.6. Schematic representation of the numerical algorithm employed in a single timestep and light interval.

The initial intensity I_0 corresponds to either the nominal or measured intensity of the light source, whichever was reported in the studies. Absorption and scattering coefficients were

left unchanged throughout the culture time, although further studies can compute time-specific coefficients from absorption spectrum data.

For increased computation speed, we divided the photobioreactor in a finite number of zones with the same light intensity and calculated the overall reaction rates as a volume-weighted average of the individual intervals. The number of light intervals (N_I) were determined in a logarithmic fashion, as shown in Eq. (2). Mesh dependence analyses showed that 10 active (with non-zero irradiance) intervals were enough for the simulations to be independent of the number of intervals.

$$I_i = 10^{\left(I_{min} + \frac{(I_{max} - I_{min})}{N_I} * i\right)} \qquad \forall i \ni \{1, \dots, N\} \tag{2}$$

4.4.3. Light uptake

To date, metabolic modeling studies on the photoautotrophic growth of microalgae have assumed that the available intensity at any given point in the culture matches the amount of light that enters microalgae [2,3,5,8,27]. Even though this causes unit inconsistency, it has been useful as an approximation while the inclusion of light uptake mechanisms in GSM models improved. However, for accurate predictions of large-scale vessels, it is necessary to accurately define the amount of uptaken photons per cell throughout the culture, in standard GSM model units of mmol gDW^{-1} h^{-1}, as opposed to formerly-used flux units of µE m^{-2} s^{-1}. Therefore, we defined a photon conservation balance over a differential element (Figure S4.1), as shown in Eq. (3).

$$(I\,A)_r - (I\,A)_{r+\Delta r} = U\,X\,\Delta V \tag{3}$$

The conservation balance is readily converted to the differential equation shown in Eq. (4), and a cellular uptake profile (U) is obtained in Eq. (5). The magnitude U is at this point a unit-consistent input to the GSM model of the microalga, which represents the upper bound of specific photon uptake rate.

$$\frac{\partial I}{\partial r} + \frac{1}{r} I + U X = 0 \tag{4}$$

$$U(r, X) = -\frac{1}{X} \cdot \left[\frac{\partial I}{\partial r} + \frac{1}{r} I\right] \tag{5}$$

A similar procedure for a planar reactor yields the homologous expression shown in Eq. (6) .

$$U(r, X) = -\frac{1}{X} \frac{dI}{dr} \tag{6}$$

4.4.4. Photoinhibition

Photoinhibition is a major drawback in photobioreactors, as it restricts the maximum amount of light a culture can be subjected to, as well as the duration of exposure. Although several species have been shown to adapt to high light conditions in the long term [28,29], photoinhibition still significantly diminishes the growth capability of phototrophs [17,30,31], and specifically of *Chlorella vulgaris* above 2400 $\mu E\ m^{-2}\ s^{-1}$ [18].

A group of studies have focused their attention on the modeling of photoinhibition in microalgae [32], which allowed us to include a photoinhibition model that described the fraction of active chlorophyll. Photosystem II (PSII) has been shown to be the bottleneck in the photosynthetic pathways as a consequence of the photodamage of the protein D1. Therefore, the effect of photoinhibition in the model was represented by the fraction of

active PSII, as proposed by Han [32], with the coefficients reported by Baroli et al. [30] (see Eqs. (7) and (8)). The magnitude θ represented the fraction of available light that was used by the metabolic network.

$$\frac{d\theta}{dt} = k_r(1-\theta) - k_d\, I\, \theta \tag{7}$$

$$k_d = m_k\, I + b_k \tag{8}$$

4.4.5. Nitrogen and carbon uptake kinetics

Nitrogen availability in the media directly influences carbon allocation in oleaginous microalgae, since its depletion hinders protein, nucleic acid and pigment biosynthesis. Nitrogen-induced stress conditions trigger in these phototrophs the accumulation of storage molecules, namely fatty acids. The uptake rate of nitrogen is a function of nitrogen quota (Q_n) and extracellular nitrogen concentration (N), as proposed by Adesanya et al. [9] and shown in Eq. (9) .

$$r_N = \frac{q_{nm} - Q_n}{q_{nm} - q_n}\left[\frac{\nu_{nm}\, N}{N + \nu_{nh}}\right] \tag{9}$$

A simple mass balance on including growth-induced depletion and replenishment yields Eq. (10).

$$\frac{dQ_n}{dt} = r_N - Q_n\mu \tag{10}$$

Similarly, we used the inorganic carbon uptake kinetics model proposed by Filali [33] to calculate the maximum carbon uptake rate at any given moment, as shown in Eq. (11).

$$r_{CO_2}^{max} = r_{CO_2}^{max,GSM}\left(\frac{C_{CO_2}}{C_{CO_2} + K_C * \frac{X}{Z_{min}T}}\right) \quad (11)$$

4.4.6. Carbon allocation

Nutrient availability in the media directly alters the way carbon is distributed across the cell. During nutrient-sufficient conditions, microalgae tend to allocate carbon on amino acid and nucleic acid biosynthesis (herein *active biomass* or X); on the other hand, nutrient-depletion, and in general stress conditions, causes metabolism to shift carbon flow towards lipid elongation. In photobioreactors, the nitrogen poses as bottleneck for overall growth, but also as trigger for lipid accumulation [2,3,9,11,34].

Light, as the energy source, induces differential phenotypes as well. Under light conditions, microalgae prioritize synthesis of active biomass, and rapid-use storage molecules, namely starch. During the dark period, the latter is used as organic carbon source for further growth and maintenance of the cell [14].

For this rather intricate behavior, we proposed a simple flow distribution algorithm, with cell size (Z) and nitrogen quota (Q_n) as coefficients for the estimation of carbon allocation. Increased nitrogen quota favored biosynthesis of active biomass and starch, whereas low nitrogen levels shifted carbon flow towards lipid production. We defined a magnitude n, which played the role of a penalty function on active biomass production and followed the Michaelis-Menten-type function shown in Eq. (12).

$$n = 1 - \frac{q}{q + q_h} \quad (12)$$

Where q is the relative nitrogen level $q = Q_n/q_{nm}$, and q_h is the half-saturation coefficient. In a similar fashion, decreased cell sizes favored the uptake of inorganic carbon and the accumulation of storage molecules, while bigger cells were assumed to lower carbon uptake levels, as previously reported by the studies of Taguchi et al. [35] and Thompson et al. [36]. Therefore, we defined a penalty function z on inorganic carbon uptake, presented in Eq. (13).

$$z = \frac{T - 1}{T_{max} - 1} \tag{13}$$

Where T is the size increase, calculated as a function of the intracellular content of starch (x_{starch}) and lipids (x_{lipid}), as shown in Eq. (14).

$$T = \frac{1}{1 - x_{starch} - x_{lipid}} \tag{14}$$

Finally, storage starch consumption is limited by a third penalty function based on the intracellular starch concentration.

$$s = \frac{C_{starch}}{C_{starch} + K} \tag{15}$$

In the end, the penalty functions were used for GSM model constraints as shown in Eq. (16) to (19). Every variable with superscript *max* is internally calculated in the algorithm as the maximum possible value at any given time point and light interval. During light and dark periods, the objective functions were, respectively, starch accumulation and biomass production, following previous reports of peak activities of starch production and consumption in light and dark periods, respectively [14–16]. An overview of the carbon allocation algorithm is illustrated in Figure S4.2.

$$r_{TAG} = r_{TAG}^{max} \cdot n \quad (16)$$

$$r_{CO_2}^{light} = r_{CO_2}^{max} \cdot (1 - z) \quad (17)$$

$$r_{Starch}^{dark,max} = r_{Starch}^{max} \cdot (z) \cdot (s) \quad (18)$$

$$\mu_{light} = \mu^{max} \cdot (1 - n) \cdot (z) \quad (19)$$

4.4.7. Parameter estimation

Metabolic capabilities across species and even strains rarely remain constant. This has been one of the most significant drawbacks when trying to generate a wide-spectrum biological model. However, in this work we were able to identify five strain-specific parameters which are assumed to be inherent to the microorganism: maximum size increase z_{max}, maximum oxygen evolution $r_{O_2}^{max}$ maximum carbon uptake $r_{CO_2}^{max}$, nitrogen quota half-saturation coefficient q_h, and starch accumulation half-saturation coefficient K. Parameter estimation was done with MATLAB Optimization Toolbox, using the *active-set* algorithm. As a result, this model is capable of predicting the macroscopic outcome of a photobioreactor under different conditions for a single strain if these parameters are known. For each study we used one of the available sets of kinetic data to calculate these parameters, specifically data at an initial nitrate concentration of 0.35 mM for Adesanya et al. [9] and data at an irradiance of 848 μmol m^{-2} s^{-1} for Kim et al. [12]. Regression parameter values are shown in Table 4.2, and a summary of all other parameters is shown Table 4.3.

Table 4.2. Parameter regression results for both studies included in this work.

Data	z_{max}	$r_{O_2}^{max}$ **mmol gDW^{-1} h^{-1}**	$r_{CO_2}^{max}$ **mmol gDW^{-1} h^{-1}**	q_h **g gDW^{-1}**	**K g L^{-1}**	**Ref.**

Adesanya et al.	2.94	5.29	-4.64	0.040	0.034	[9]
Kim et al.	1.74	8.84	-5.18	0.049	0.124	[12]

4.4.8. Maximization of lipid productivity

Maximization of lipid productivity in microalgae has been known to be possible through different growth modes [1,9], nutrient-related stress [37,38] and time-dependent irradiance strategies [17,39]. Although some genome-scale modeling has been used for giving insight into the first two options in oleaginous microalgae [5], no such strategy has been employed to assess the latter. Therefore, we employed the model to maximize the lipid productivity of a photobioreactor as an illustration of the model's capability of predicting optimal operating conditions for process design and optimization. The photobioreactor in question is owned by the Chemical and Biochemical Processes research group of the Universidad Nacional de Colombia, which renders further experimentation possible for the validation of model predictions. Five variables were manipulated to search for the optimal global lipid productivity (R_L in Eq. (20)) condition: initial lamp irradiance I_0, final lamp irradiance I_f, photoperiod p, culture duration t_f, and a coefficient b_I which represents the shape of the light profile, as shown in Eq. (21).

$$R_L = \frac{C_{Lipids}}{t_f} \tag{20}$$

$$I(t) = a_I * t^{b_I} + I_0 \tag{21}$$

In Eq. (21), only b_I is used as an optimization variables, as a_I is dependent on the variables I_0, I_f and b_I itself, as presented in Eq. (22).

$$a_I = \frac{I_f - I_0}{\left(t_f\right)^{b_I}} \tag{22}$$

Table 4.3. Summary of model parameters.

Parameter	*Symbol*	*Definition*	*Units*	Ref.
Light intensity profile	I	$I(r, X)$	$\mu mol\ m^{-2}\ s^{-1}$	[22]
Light uptake rate	U	$U(r, X)$	$mmol\ gDW^{-1}\ h^{-1}$	–
Fraction of active photo-system II (PSII)	θ	$\theta(\theta, I, t)$	–	[30]
Maximum attenuation coefficient	$K_{a,max}$	1041	m^{-1}	[22]
Light modeling parameters	b	1.03	$kg\ m^{-3}$	[22]
	w	-0.3128	–	[22]
	p_k	12.66	–	[22]
First-order PSII photoreparation coefficient	k_r	0.7	h^{-1}	[30]
First-order PSII photodamage coefficient	k_d	$k_d(I)$	$m^2\ s\ \mu mol^{-1}\ h^{-1}$	[30]
Photoinhibition parameters	m_k	0.00042	$m^4\ s^2\ \mu mol^{-2}\ h^{-2}$	[30]
	b_k	0.05	$m^2\ s\ \mu mol^{-1}\ h^{-1}$	[30]
Nitrogen uptake rate	r_N	$r_N(Q_n. N)$	$mmol\ gDW^{-1}\ h^{-1}$	[9]
Nitrogen quota	Q_n	$Q_n(r_N, Q_n, \mu)$	$mmol\ gDW^{-1}$	[9]
Maximum nitrogen quota	q_{nm}	6.78	$mmol\ gDW^{-1}$	[9]
Minimum nitrogen quota	q_n	2.29	$mmol\ gDW^{-1}$	[9]
Maximum nitrogen uptake rate	ν_{nm}	2.02	$mmol\ gDW^{-1}\ h^{-1}$	[9]
Nitrogen uptake half-saturation coefficient	ν_{nh}	4.29	mM	[9]
Carbon uptake half-saturation coefficient	K_C	0.0128	$mmolN\ cell^{-1}$	[33]
Nitrogen-dependent penalty function	n	$n(q)$	–	–
Relative nitrogen quota	q	$q(Q_n)$	–	–
Size-dependent penalty function	z	$z(Z)$	–	–
Cell size	Z	$Z(x_{starch}, x_{lipid})$	$pg\ cell^{-1}$	
Minimum cell size	Z_{min}	75	$pg\ cell^{-1}$	[40]
Cell size increase	T	$T(Z)$	–	–
Intracellular starch mass fraction	x_{starch}	–	–	–
Intracellular lipid mass fraction	x_{lipid}	–	–	–
Starch content-dependent penalty function	s	$s(C_{starch})$	–	–
Lipid production rate	r_{TAG}	$r_{TAG}(n)$	$mmol\ gDW^{-1}\ h^{-1}$	–
Maximum CO_2 uptake in GSM	$r_{CO_2}^{max,GSM}$	–	$mmol\ gDW^{-1}\ h^{-1}$	[3]
Maximum lipid production rate	r_{TAG}^{max}	–	$mmol\ gDW^{-1}\ h^{-1}$	–
Starch consumption rate	$r_{Starch}^{dark,max}$	$r_{Starch}^{dark}(s, z)$	$mmol\ gDW^{-1}\ h^{-1}$	–
Maximum starch production rate	r_{Starch}^{max}	–	$mmol\ gDW^{-1}\ h^{-1}$	–
CO_2 consumption rate	$r_{CO_2}^{light}$	$r_{CO_2}^{light}(s, z)$	$mmol\ gDW^{-1}\ h^{-1}$	–
Maximum CO_2 consumption rate	$r_{CO_2}^{max}$	–	$mmol\ gDW^{-1}\ h^{-1}$	–
Specific active biomass production rate	μ^{light}	$\mu^{light}(n, z)$	h^{-1}	–
Global lipid productivity	R_L	$R_L(C_{lipids}, t_f)$	$g\ L^{-1}$	–
Global lipid concentration	C_{lipids}	–	$g\ L^{-1}$	–

Global CO_2 concentration	C_{CO_2}	–	$g\ L^{-1}$	–

4.5. References

1. Serrano Bermúdez LM. Estudio de cuatro cepas nativas de microalgas para evaluar su potencial uso en la producción de biodiesel. 2012;173.

2. Zuñiga C, Li C-T, Huelsman T, Levering J, Zielinski DC, McConnell BO, et al. Genome-Scale Metabolic Model for the Green Alga *Chlorella vulgaris* UTEX 395 Accurately Predicts Phenotypes under Autotrophic, Heterotrophic, and Mixotrophic Growth Conditions. Plant Physiol. 2016;172:589–602.

3. Zuñiga C, Levering J, Antoniewicz MR, Guarnieri MT, Betenbaugh MJ, Zengler K, et al. Predicting dynamic metabolic demands in the photosynthetic eukaryote Chlorella vulgaris. Plant Physiol. 2017;176:450–62.

4. Muylaert K, Bastiaens L, Vandamme D, Gouveia L. Harvesting of microalgae: overview of process options and their strengths and drawbacks. Microalgae-based biofuels and bioproducts. Lisbon: Woodhead Publishing; 2017. p. 113–32.

5. Tibocha-Bonilla JD, Zuñiga C, Godoy-Silva RD, Zengler K. Advances in metabolic modeling of oleaginous microalgae. Biotechnol Biofuels. 2018;11:241.

6. May P, Wienkoop S, Kempa S, Usadel B, Christian N, Rupprecht J, et al. Metabolomics- and proteomics-assisted genome annotation and analysis of the draft metabolic network of *Chlamydomonas reinhardtii*. Genetics. 2008;179:157–66.

7. Knoop H, Zilliges Y, Lockau W, Steuer R. The metabolic network of *Synechocystis sp.* PCC 6803: Systemic properties of autotrophic growth. Plant Physiol. 2010;154:410–22.

8. Chang RL, Ghamsari L, Manichaikul A, Hom EFY, Balaji S, Fu W, et al. Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. Mol Syst Biol. 2011;7.

9. Adesanya VO, Davey MP, Scott SA, Smith AG. Kinetic modelling of growth and storage

molecule production in microalgae under mixotrophic and autotrophic conditions. Bioresour Technol. 2014;157:293–304.

10. Bold HC. The Morphology of Chlamydomonas chlamydogama, Sp. Nov. Bull Torrey Bot Club. 1949;76:101.

11. Mansouri M. Predictive modeling of biomass production by Chlorella vulgaris in a draft-tube airlift photobioreactor. Adv Environ Technol. 2017;3:119–26.

12. Kim J, Lee JY, Lu T. A model for autotrophic growth of Chlorella vulgaris under photolimitation and photoinhibition in cylindrical photobioreactor. Biochem Eng J. 2015;99:55–60.

13. Bialon J, Rath T. Growth rates and photon efficiency of Chlorella vulgaris in relation to photon absorption rates under different LED-types. Algal Res. 2018;31:204–15.

14. Diamond S, Jun D, Rubin BE, Golden SS. The circadian oscillator in Synechococcus elongatus controls metabolite partitioning during diurnal growth. Proc Natl Acad Sci. 2015;112:E1916–25.

15. Bose A, Chakraborty S. Mathematical modelling of the effects of circadian rhythm on microalgal growth in phototrophic and mixotrophic cultures. Chem Eng Trans. 2016;52:955–60.

16. Xu Y, Ibrahim IM, Harvey PJ. The influence of photoperiod and light intensity on the growth and photosynthesis of Dunaliella salina (chlorophyta) CCAP 19/30. Plant Physiol Biochem. 2016;106:305–15.

17. Sun Y, Liao Q, Huang Y, Xia A, Fu Q, Zhu X, et al. Application of growth-phase based light-feeding strategies to simultaneously enhance Chlorella vulgaris growth and lipid accumulation. Bioresour Technol. 2018;256:421–30.

18. Pfendler S, Alaoui-sossé B, Alaoui-sossé L, Bousta F, Aleya L. Effects of UV-C radiation on Chlorella vulgaris, a biofilm-forming alga. J Appl Phycol. 2018.

19. Serra-Maia R, Bernard O, Gonçalves A, Bensalem S, Lopes F. Influence of temperature on *Chlorella vulgaris* growth and mortality rates in a photobioreactor. Algal Res.

2016;18:352–9.

20. Becker SA, Feist AM, Mo ML, Hannum G, Palsson B, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox. Nat Protoc. 2007;2:727–38.

21. Yun YS, Park JM. Attenuation of monochromatic and polychromatic lights in *Chlorella vulgaris* suspensions. Appl Microbiol Biotechnol. 2001;55:765–70.

22. Naderi G, Znad H, Tade MO. Investigating and modelling of light intensity distribution inside algal photobioreactor. Chem Eng Process Process Intensif. 2017;122:530–7.

23. Salleh SF, Kamaruddin A, Uzir MH, Mohamed AR, Shamsuddin AH. Modeling the light attenuation phenomenon during photoautotrophic growth of A. variabilis ATCC 29413 in a batch photobioreactor. J Chem Technol Biotechnol. 2017;92:358–66.

24. Suh IS, Lee SB. A light distribution model for an internally radiating photobioreactor. Biotechnol Bioeng. 2003;82:180–9.

25. Csögör Z, Herrenbauer M, Schmidt K, Posten C. Light distribution in a novel photobioreactor - Modelling for optimization. J Appl Phycol. 2001;13:325–33.

26. Csögör Z, Herrenbauer M, Perner I, Schmidt K, Posten C. Design of a photo-bioreactor for modelling purposes. Chem Eng Process Process Intensif. 1999;38:517–23.

27. Levering J, Dupont CL, Allen AE, Palsson BO, Zengler K. Integrated regulatory and metabolic networks of the marine diatom *Phaeodactylum tricornutum* predict the response to rising CO_2 levels. mSystems. 2017;2:e00142-16.

28. Masojídek J, Torzillo G, Koblížek M, Kopecký J, Bernardini P, Sacchi A, et al. Photoadaptation of two members of the Chlorophyta (Scenedesmus and Chlorella) in laboratory and outdoor cultures: Changes in chlorophyll fluorescence quenching and the xanthophyll cycle. Planta. 1999;209:126–35.

29. Garmyn M, Young AR, Miller SA. Mechanisms of and variables affecting UVR photoadaptation in human skin. Photochem Photobiol Sci. 2018;17:1932–40.

30. Baroli I, Melis A. Photoinhibition and repair in Dunaliella salina acclimated to different growth irradiances. Planta. 1996;198:640–6.

31. Han BP. A mechanistic model of algal photoinhibition induced by photodamage to photosystem-II. J Theor Biol. 2002;214:519–27.

32. Han BP, Virtanen M, Koponen J, Straskraba M. Effect of photoinhibition on algal photosynthesis: a dynamic model. J Plankton Res. 2000;22:865–85.

33. Filali R, Tebbani S, Dumur D, Isambert A, Pareau D, Lopes F. Growth modeling of the green microalga *Chlorella vulgaris* in an air-lift photobioreactor. IFAC Proc Vol. 2011;18:10603–8.

34. Lv JM, Cheng LH, Xu XH, Zhang L, Chen HL. Enhanced lipid production of Chlorella vulgaris by adjustment of cultivation conditions. Bioresour Technol. 2010;101:6797–804.

35. Taguchi S. Relationship Between Photosynthesis and Cell Size of Marine Diatoms. Journal of Phycology. 1976. p. 185–9.

36. Thompson PA, Harrison PJ, Parslow JS. Influence of Irradiance on Cell Volume and Carbon Quota for Ten Species of Marine Phytoplankton. Journal of Phycology. 1991. p. 351–60.

37. Zuñiga C, Zaramela L, Zengler K. Elucidation of complexity and prediction of interactions in microbial communities. Microb Biotechnol. 2017;10:1500–22.

38. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, et al. Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances. Plant J. 2008;54:621–39.

39. Atta M, Idris A, Bukhari A, Wahidin S. Intensity of blue LED light: A potential stimulus for biomass and lipid content in fresh water microalgae *Chlorella vulgaris*. Bioresour Technol. 2013;148:373–8.

40. Chioccioli M, Hankamer B, Ross IL. Flow cytometry pulse width data enables rapid and sensitive estimation of biomass dry weight in the microalgae *Chlamydomonas reinhardtii* and *Chlorella vulgaris*. PLoS One. 2014;9:1–12.

4.6. Supplementary Figures

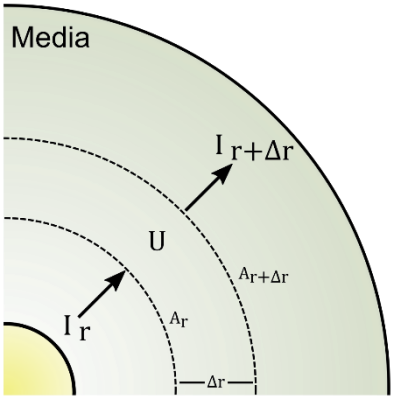

Figure S4.1. Differential photon conservation balance in a photobioreactor.

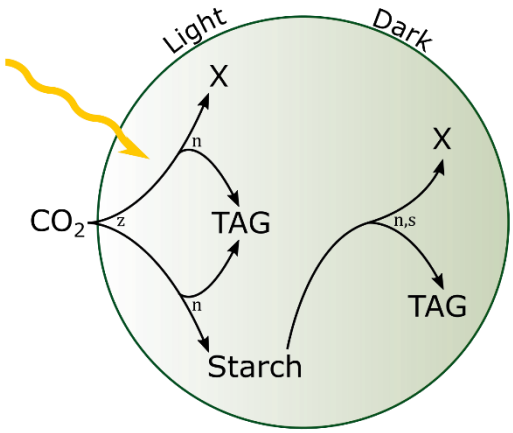

Figure S4.2. Illustration of carbon allocation algorithm.

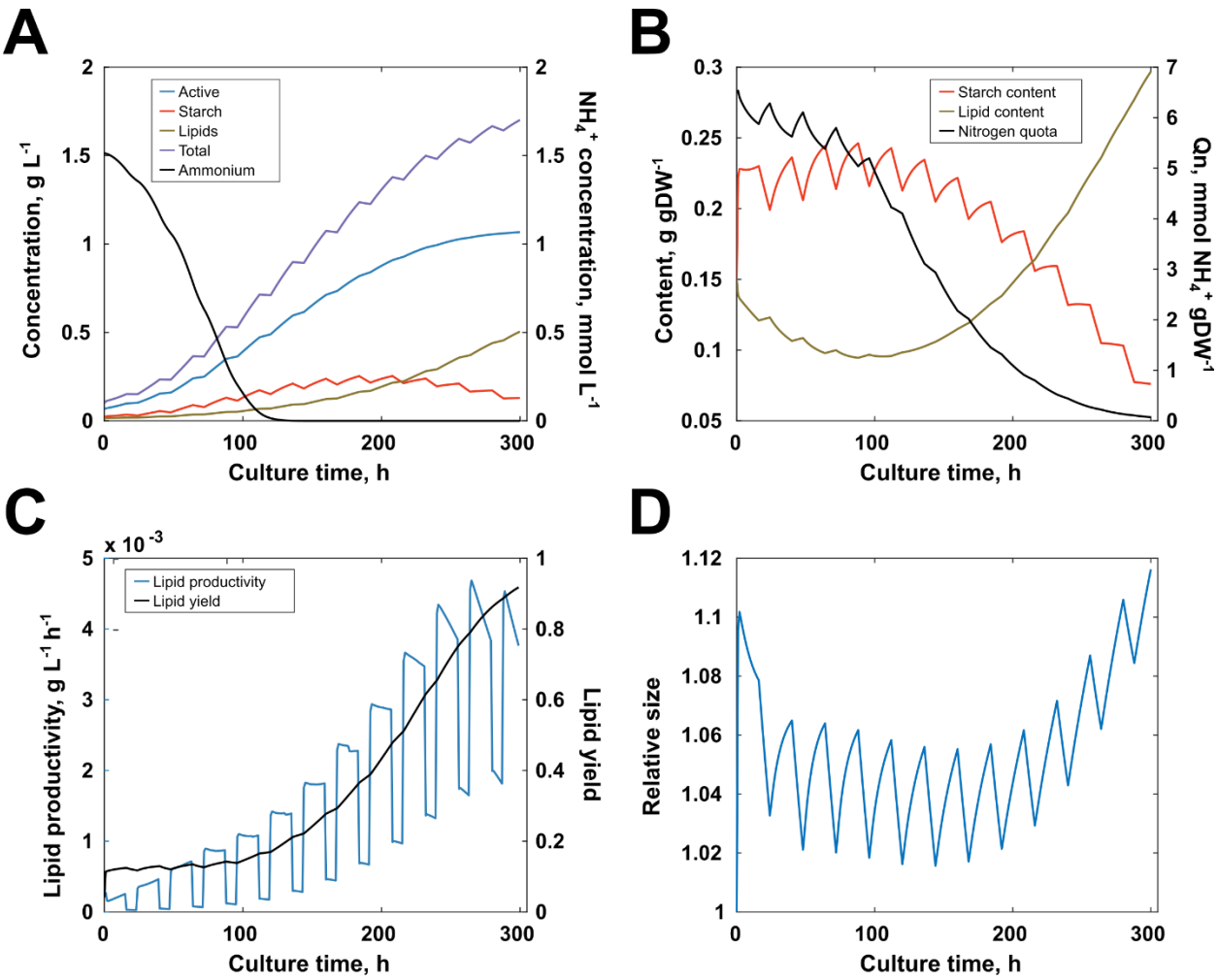

Figure S4.3. Simulation results of the growth of *Chlorella vulgaris* at a hypothetical initial condition for the optimization of lipid productivity. (A) Global reactor concentrations of active biomass, starch, lipids, total biomass and ammonium. (B) Intracellular content of starch, lipids and nitrogen. (C) Contrast of lipid productivity with lipid yield. (D) Variation of cell size.

Chapter 5. Conclusions and recommendations

5.1. Conclusions

In this work, we obtained the first-ever multiscale metabolic model for the prediction of the dynamic growth of *Chlorella vulgaris* in a photobioreactor. This model included substrate uptake and light attenuation phenomena acting as modules that allowed for the accurate estimation of global biomass production, as well as intracellular concentrations of macromolecules of industrial interest, namely triacylglycerols. The fact that at the core lies a genome-scale metabolic bases our model on species-specific *omics* data which rendered our model capable of predicting metabolic flux distributions at every time of the culture in a photobioreactor. With it, we were able to predict light strategy conditions which would increase the global lipid productivity of an internally-illuminated stirred-tank photobioreactor in roughly 47%, exhibiting the potential this model has for process design and optimization. In the future, this model could help decide upon gene-knockout strategies that not only favor specific lipid production rates, but also global lipid productivities of the whole photobioreactor and an industrial plant.

5.2. Recommendations

The model allows for several applications we did not include in this study, such as:

- Prediction of continuous cultures, as well as estimation of best culture conditions for optimal steady-state lipid productivities.
- Adaptation to other oleaginous microalgae species, by changing species-specific models and coefficients to appropriate ones.

- Prediction of promissory gene-knockout mutants that maximize lipid productivity in the photobioreactor, or profit in a whole-plant scale.

On the other hand, future work will be necessary for this model's flexibility to be even greater. We identified several key points that need to be addressed for this purpose:

- The model needs to account for the influence of temperature on cell mortality, as well as a heat transfer model that includes convection and radiation phenomena.
- Starch consumption in the light period should be included in the model, as no-dark strategies are often suggested for increased lipid productivity, and the model overestimates starch accumulation as a result of the current carbon allocation algorithm.
- Light attenuation coefficients can vary greatly with biomass composition. Since our model is already capable of predicting biomass composition, time-dependent absorption coefficients could be calculated by employing absorption spectrum data.
- Our model does not yet include shading due to the presence of an internal lamp. This phenomenon could prove significant at even larger scales than tested in this study.
- Photoinhibition is a complex phenomenon, and the kinetic model we used is, although mechanistic and the best available, very simple. As the understanding of photoinhibiton and photoreparation/photoadaptation improves, better models might arise giving opportunity for this model to be even more accurate in high-irradiance cases.

Appendix 1. The genome-scale metabolic model

A1.1. A brief reminder of genome-scale metabolic modeling

As previously explained in *Chapter 2*, a genome-scale metabolic (GSM) model is an omics-based network of metabolic reactions, which describe the metabolism of an organism. GSM models can be used in various ways to predict metabolic flux distributions, pathway coupling, interspecific interactions, and to analyze the impact of nutrient-stress and gene deletions these aspects. The most widely used strategy to compute these distributions is called Flux Balance Analysis (FBA), which is based on the following steady-state mass balance:

$$S \cdot v = 0$$

Where S is the matrix containing all the stoichiometric information of the network and v is the vector of metabolic fluxes corresponding to each metabolic reaction included. Said reactions not only include chemical transformations, but also transport between compartments and exchange reactions with the medium. Commonly, a genome-scale metabolic model can account for over 2000 reactions and 1500 metabolites. In any case, the number of reactions is always greater than that of metabolites, rendering the mass balance under-determined. Hence, FBA uses linear programming to compute the values inside vector v following an objective function, which usually corresponds to but is not restricted to the maximization of biomass production.

Overall, the generic solution algorithm of FBA is as follows:

(Definition of objective function) ..... $\max \mu \qquad s.t.$

(Mass balance at steady state) ..... $S \cdot v = 0$

(Upper and lower boundaries of reactions) ..... $v_i^{min} \leq v_i \leq v_i^{min}$

A1.2. The GSM model *i*CZ843

The model *i*CZ843, available online at http://www.plantphysiol.org/cgi/pmidlookup?view=long&pmid=27372244, contains a total of 2294 reactions, 1770 metabolites and 843 genes. This model was generated using RAVEN Toolbox and curated both manually and semi-automatically using COBRA Toolbox within MATLAB. The high quality of this model allowed it to predict growth under several different carbon sources (including heterotrophy, autotrophy and mixotrophy), along with outstanding definition of amino acid and lipid metabolism pathways. These characteristics rendered the model suitable for further work on prediction of dynamic behavior, as proposed and developed in this work.

Appendix 2. Further details on methods

A2.1. MATLAB code, scripts and functions

The complete set of functions (herein "the toolbox") and scripts is available in GitHub at https://github.com/jdtibochab/pbr. Overall, the toolbox consists of 4 subgroups of functions which correspond to: prediction, parameter regression, photobioreactor optimization and visualization. Naturally, regression and optimization functions refer to predictive functions for their calculations, and visualization functions reproduce the graphs used to generate the figures showed in Chapter 4.

A2.2. Predictive functions

Predictive functions compose the multiscale metabolic model, and are spread across the following folders: *calculation, consumptionKinetics, experimentalData, light,* and *lipids*. Within these folders functions for solution algorithms, nutrient consumption kinetics, retrieval of reported data, light distribution calculation and carbon allocation are available, respectively.

A2.2.1. FBAcalc

The core calculation of the multiscale metabolic model is contained in **FBAcalc**, following the algorithm shown in Figure 4.6. **FBAcalc** receives the struct variables **solution** and **PBR**, both containing a pre-allocated solution and setup variables, respectively, and returns an updated version of **solution** to be used by **sFBAcalc** to update time-dependent variables.

PBR is generated by **PBRgeneration**, after the latter has called the data-containing script **data**, which contains all the case-specific information. This script also contains calls to other data-containing scripts, such as: **experimentalData** (for experimental time-course data) and **reactorData** (for reactor geometry data).

The algorithm used by **FBAcalc** can be broken down as follows:

1. Retrieve values of time-dependent variables at previous timestep.
2. Calculate irradiance and light uptake distribution, as well as generate different light intervals.
3. Calculate global photoinhibition at present time as the weight-averaged photoinhibition and assume all light intervals at present time are at the same photoinhibition (so as to account for macromixing during one timestep).
4. Calculate maximum possible CO_2 uptake rate from carbon uptake kinetics.
5. If during light period, for each light interval do as follows. If not, skip to 6.
    a. Calculate "ideal" metabolic flux distribution at current light interval.
    b. Calculate maximum possible light uptake.

$$U^{max} = U^{ideal} * \theta$$

    c. Calculate maximum biomass and lipid production rates by solving for "photoinhibited" metabolic flux distributions.
    d. Calculate active biomass and lipids production rates using the carbon allocation algorithm.
    e. Solve for final metabolic flux distributions at current light interval by fixing active biomass and lipid production rates.

6. If during dark period do as follows:
    a. Calculate maximum starch consumption rate from starch consumption kinetics.
    b. Solve for metabolic flux distributions to determine maximum possible lipid productivity.
    c. Calculate lipid production rate from carbon allocation algorithm.
    d. Solve for final metabolic flux distributions by setting upper boundary of starch consumption rate and lipid fixing production rate.
7. Return metabolic flux rates and variables at present time through **solution** struct.

A2.2.2. sFBAcalc

This function carries out the whole calculation by calling **FBAcalc** and updating variables after obtaining production and consumption rates at each timestep. This function receives the setup struct variable **PBR** and returns the solution struct variable **solution**. Moreover, it contains code for calculation monitoring and data saving.

A2.2.3. Nutrient consumption kinetics

In this model, nutrient uptake kinetics are restricted to CO_2, starch and nitrogen source. Kinetic models are included in separate scripts in the folder *consumptionKinetics*.

A2.2.4. Light distribution

Irradiance and light uptake distributions are calculated in the folder *light* with biomass concentration-dependent functions, with the light distribution model at its core, along with functions for both internal and external lamps.

A2.2.5. Carbon allocation

The carbon allocation algorithm in this model is contained in both *lipids* and *consumptionKinetics* folders. The latter being mainly due to the **nitrogenUptake** functions, which serves the purpose of calculating intracellular nitrogen availability and transforming it into the nitrogen-dependent penalty function. In addition, inside the folder *lipids* lie the modification of the GSM model for the inclusion of lipid and starch demand reactions.

A2.2.6. How to run

All data regarding the prediction of culture growth is contained in the script **data**, which is called by **PBRgeneration** to create the struct variable with all the relevant information called **PBR**. To run a calculation, define appropriately the variables and parameters in **data**, and run the main script **run**.

A2.3. Regression functions

Parameter regression functions are contained in the folder *regression*. There lies the calculation of the error function which is evaluated by the optimization function **fmincon**.

The cases of study included in this work are located inside the folder *cases*. One dedicated folder of data-defining functions belongs to each case of study, in this work Adesanya et al. and Kim et al. For each case, basic settings for prediction are defined in the script called **data** and are called by the function **run**, which generates the setup struct variable **PBR** and a pre-allocated solution variable **solution**, both of which are inputs to **sFBAcalc**.

For parameter regression calculations, the script **data** was slightly changed to allow for optimization variables to be changed by the main regression script **regressionModel.** The regression data script is called **data_reg**. Parallelization is recommended for this computation (6+ cores).

A2.4. Optimization functions

These functions are contained in the folder *optimization*. The optimization followed the same strategy of calculation as the regression, which required the script **data** to be altered into **data_opt**. The data script is called by the main script **optimizationPBR**, which uses a fitness function in the same folder.

A2.5. Visualization functions

A set of suggested visualization functions are provided in the folder *visualization*. All the results figures in Chapter 4 were generated using these functions. For a quick summary of results, call **summary(solution)**, or for a more detailed report, call **report(solution).**