Systems biology's role in leveraging microalgal biomass potential: Current status and future perspectives

Emanuel Cunha, Vítor Sousa, Pedro Geada, José A. Teixeira, António A. Vicente, Oscar Dias

PII: S2211-9264(22)00334-4

DOI: https://doi.org/10.1016/j.algal.2022.102963

Reference: ALGAL 102963

To appear in: *Algal Research*

Received date: 31 January 2022

Revised date: 14 November 2022

Accepted date: 24 December 2022

Please cite this article as: E. Cunha, V. Sousa, P. Geada, et al., Systems biology's role in leveraging microalgal biomass potential: Current status and future perspectives, *Algal Research* (2022), https://doi.org/10.1016/j.algal.2022.102963

Systems Biology's role in leveraging microalgal biomass potential: current status and future perspectives

Emanuel Cunha[1a,b], Vítor Sousa[1a,b], Pedro Geada[2a,b], José A. Teixeira[2a,b], António A. Vicente[2a,b], Oscar Dias[2a,b]

[1]These authors contributed equally to the article.

[2]Senior authors

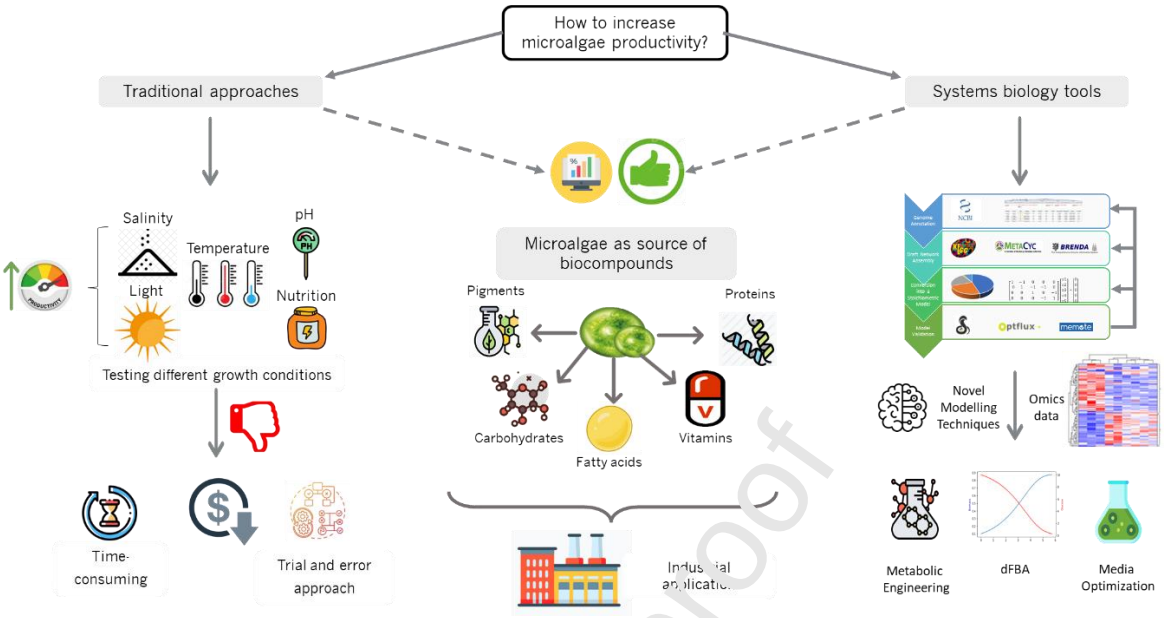
[a]Centre of Biological Engineering, Universidade do Minho, 4710-057 Braga, Portugal

[b]LABBELS – Associate Laboratory, Braga, Guimarães, Portugal

Abstract

Microalgae are an eco-friendly and alternative source of several compounds that can be applied in numerous biotechnological branches such as cosmetics, food, or pharmaceutical industries. However, the commercialization of products from microalgae has still some drawbacks, among which it is possible to highlight the low biomass and compounds' productivities at industrial level. The common strategies to improve the process cost-effectiveness of microalgae cultivation are essentially based on optimising nutrient needs and several environmental factors (e.g., temperature, light intensity, salinity) that impact both their growth and biochemical composition. In this regard, genome-scale metabolic (GSM) models allow understanding the metabolic processes that lead to a final phenotype since they contain all known pathways, reactions, and metabolites of the organism, based on genomic information, available literature, and experimental data. This review provides an overview of the most important factors that need to be considered in microalgae cultivation, providing strategies to improve process cost-effectiveness. Particular emphasis will be given to the *in silico*-guided optimization as a real alternative, using GSM models to enhance the production of defined compounds and biomass.

Graphical Abstract:



Keywords: Microalgae, Upstream processing, Environmental factors, Systems Biology, Genome-scale Metabolic Model.

1 Introduction

1.1 Impact of different parameters on microalgae growth and biochemical composition

Recently, the interest in microalgae has grown due to their high biotechnological potential as a result of being a novel source of a wide range of compounds, such as pigments, proteins, fatty acids, or carbohydrates, with several commercial applications in different industrial fields [1]. Besides not requiring arable land for their production, microalgae can also grow using non-potable water sources (e.g., saline, hypersaline, and wastewater), thus reducing the pressure placed on fresh water consumption [2]. Although the ancient existence of microalgae, these microorganisms are still rather unexplored. The total number of species is estimated to range between hundreds of thousands and several million; however, only approximately 73000 are currently identified and a very small part of them is cultivated at industrial scale for commercial purposes [1, 3, 4]. *Dunaliella salina* and *Haematococcus pluvialis*, as sources of β-carotene and astaxanthin, respectively, are some of the (few) microalgae already established commercially [5].

According to their nutritional requirements, microalgae can adopt different strategies **(Figure 1)**. These microorganisms can grow either in autotrophic, mixotrophic or heterotrophic conditions. Regardless of the adopted nutritional mode, most microalgae require a variety of elements, such as carbon, oxygen, hydrogen, nitrogen, potassium, calcium, magnesium, iron, sulphur, phosphorus, and trace elements, such as copper, manganese, selenium, or zinc. Within this group of nutrients, carbon, nitrogen, and phosphorus are usually regarded as the most important for the cultivation of this group of photosynthetic microorganisms, being the key players for shifting cells' metabolic pathways [3, 6, 7]. The carbon fixed by the microalgae can be used in three different physiological functions: i) for cell respiration (e.g., CO_2); ii) as an energy source (e.g., glucose); and iii) as raw material for the formation of new cells [6]. Nitrogen is an essential constituent of all functional and structural proteins of microalgae, being considered the second most important nutrient in their cultivation [3]. Nitrogen concentration has an important role in cell composition since it influences the metabolites profile in terms of polyunsaturated fatty acids (PUFAs), polysaccharides, carotenoids, and chlorophylls. The phosphorus content on the culture medium, on the

other hand, has an important role in microalgal growth and different cellular processes due to its participation in DNA biosynthesis. In line with the case of nitrogen, phosphorus content has also impact on the accumulation of some biocompounds, especially lipids and carbohydrates [3].

Some abiotic factors play also an important role over microalgal growth and biochemical composition. Illumination is the most important factor influencing biomass composition [8]. Indeed, the effects of different light intensities and cycles (light and dark) have a predominant impact on the metabolite profile obtained during microalgae cultivation [9]. The light used in this process should be within the photosynthetically active radiation (PAR) range so that microalgae can obtain energy through the photosynthesis process. The range of this parameter varies from 400 nm to 700 nm, which corresponds to the visible light spectrum [8]. Under different light availabilities, microalgae can adapt using diverse mechanisms to maintain normal biochemical functions. Changes in the quantity and type of pigments produced, growth rate, dark respiration rate, and/or the accumulation of some compounds, such as fatty acids are frequently observed. Morphological changes, such as the variation of cell volume or the number and density of thylakoid membranes [6] are also common. Temperature impacts directly the microalgae's metabolism, having a significant role in growth rate, cell size, metabolites profile, and nutritional needs [6, 10]. Based on these facts, temperature influences microalgae productivity at different levels: ***i*)** high temperature contributes to protein degradation, which leads to cell mortality; ***ii*)** temperature affects culture medium solubility, contributing to changes in the CO_2 and O_2 mass transfer rates; and ***iii*)** temperature has a central role on the photoinhibition, as some mechanisms are temperature-dependent, for example, under low temperatures, there are fewer electron transporters, resulting in less CO_2 fixation and decreased CO_2 availability for photosynthesis [6, 11]. However, low temperatures also protect Photosystem II (PSII), inhibiting the active oxygen species responsible for photoinhibition [6, 11]. The pH value of the culture medium is an important factor to consider in microalgae cultivation as it affects the solubility and availability of nutrients, cell's influx and efflux of cations and anions, liquid-gas transfer phenomena (affecting, for example, CO_2 availability), and metabolic processes – due to its impact on the performance of enzymes involved in metabolic pathways [6]. Particularly, in the carbonaceous species – that have a primary need for carbon, namely inorganic carbon such as the CO_2 –, pH has an important role

in microalgae growth [6]. Salinity is another environmental factor that has an important impact on microalgae growth and composition. Each microalga has its own optimal range of salinity, under which cells have the highest biomass productivity [2]. Salt stress induces three different types of stress on microalgae: ionic, osmotic, and oxidative. The ionic stress is caused by the influence of salinity over the influx and efflux of ions, where Na^+ competes with K^+ and the deficiency of K^+ inside the cell is frequently the outcome [2, 12]. Osmotic stress, on the other hand, is caused by the presence of high amounts of salt in the extracellular environment that reduces the water uptake by microalgae due to the lower osmotic potential [12]. Finally, oxidative stress is generated by the increase of reactive oxygen species (ROS), which contributes to a lower photosynthetic efficiency [12, 13].

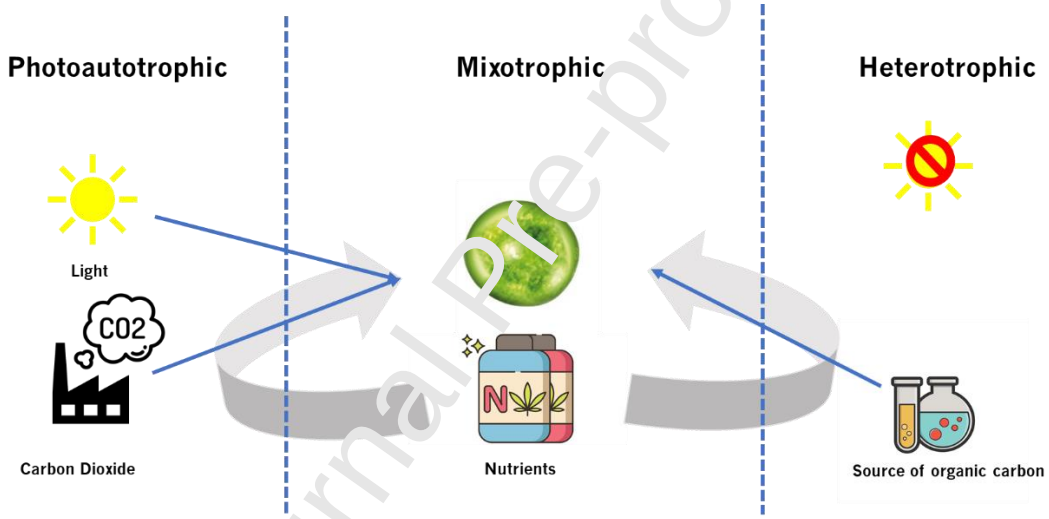


Figure 1- Microalgae classification according to their nutritional requirements.

1.2 Current cultivation strategies to enhance biomass and metabolites productivity

Different approaches are currently adopted to improve microalgae cultivation process cost-effectiveness. One of the traditional strategies employed leverages stress conditions. Under nutrient-stress conditions, microalgae cells might use different strategies. Based on this fact, it is possible to control cell metabolism to obtain a higher content of the desired compounds. The application of nutritional stress strategies shows some advantages, such as reducing the production costs of microalgae cultivation and increasing the cellular content of lipids (mainly triacylglycerol (TAG)), carbohydrates (mostly starch), and carotenoids. However, this strategy has drawbacks associated with a lower growth rate, which might lead to a lower final biomass concentration.

Microalgae exposed to an extended period of nutrient depletion are frequently reported to consume storage products (e.g., lipids), reducing their accumulation and yield, and contributing to microalgal biomass devaluation [85]. However, adopting a correct nutrients stress strategy, that allows an increase of (certain) high-value compounds content, can be an economically interesting alternative for microalgae valorisation.

Nitrogen stress conditions induce an increase of storage compounds on microalgae biomass. Guihéneuf and Stengel (2017) demonstrated that nitrogen starvation leads to increased synthesis and accumulation of lipids in *Pavlova lutheri*, namely TAGs and PUFAs, such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). The authors suggest that starvation conditions are the most important prerequisite for TAG accumulation, which occurs due to a potential lipid remodelling where the membrane's lipids are broken down and rearranged into storage lipids. This is the case of *P. lutheri*, in which the lipid remodelling observed during nitrogen starvation results in the accumulation of storage lipids (TAG) and PUFAs, containing 12-13 % and 5-6 % of EPA and DHA, respectively. Although the regulation of pigments was also studied under nitrogen starvation, no significant differences were found between the assays [14]. With the lack of nitrogen in the medium, it was expected that secondary pigments (e.g., astaxanthin, β-carotene) accumulation would increase, which was not the case. On the contrary, in a study carried out by Wu *et al.* (2016), the authors assessed the impact of different nitrogen concentrations on the accumulation of β-carotene in *Dunaliella salina* and concluded that greater contents of the pigment are obtained using lower nitrogen concentrations. This can be explained by the necessity of the microalgae to maintain the normal metabolic performance in the absence of nitrogen; under these conditions, *D. salina* uses the available carbon and hydrogen for the synthesis of non-nitrogen-induced pigments, as the case of β-carotene [15].

As observed in nitrogen stress conditions, the accumulation of storage compounds by microalgae cells was also enhanced under phosphorus stress conditions. Xin *et al.* (2010), for instance, demonstrated that phosphorus limitation results in increasing lipid content of *Scenedesmus sp.* The authors demonstrated that using initial total phosphorus between 0.2 and 2.0 $mg \cdot L^{-1}$ leads to a lipid content of 23-28 %, whereas under phosphorus limitation (0.1 $mg \cdot L^{-1}$), the amount of lipids reaches 53 %. However, under the same growth conditions, a negative impact was reported over biomass

concentration, with an algal biomass productivity of 0.15 $g \cdot L^{-1}$ under phosphorus limitation and 0.35 $g \cdot L^{-1}$ under non-stress conditions [16].

The application of light stress conditions – as high radiation – leads to greater production of certain compounds, such as lipids, carbohydrates or carotenoids [17, 18]. Lipid content under this stress increases due to the excess of energy supplied to microalgae, which promotes the photoinhibition phenomenon (blocking the microalgae growth), being that excessive energy used to synthesize lipids [17, 18]. Kim *et al.* (2019), for instance, achieved the highest lipid productivity – 0.25 $g \cdot L^{-1} \cdot d^{-1}$ – under 900 $\mu mol_{photons} \cdot m^{-2} \cdot s^{-1}$ and the lowest, less than 0.05 $g \cdot L^{-1} \cdot ^{-1}$, using 50 $\mu mol_{photons} \cdot m^{-2} \cdot ^{-1}$ [17]. Besides the changes in lipid content, pigment production is also affected when microalgae are exposed to greater light irradiation. In this case, a decrease of the content of chlorophylls is observed, whereas the carotenoid content increases [9, 19]. He *et al.* (2015) suggested that quantum yield (F_v/F_m) – ratio between variable (emitted photons) and maximum (absorbed photons) fluorescence – declines when microalgae are subject to high light intensities due to the inactivation of photosystem II (PS II), which contributes to increased oxidative stress of cells and, consequently, greater carotenoids and lipid contents. This increase is explained by the important antioxidant role of carotenoids, while lipid accumulation results from the fact of being a receptor to dissipate the excess of electrons, mitigating the oxidative stress inside the cells [19]. Kim *et al.* (2019) refer to the combination of two different stress conditions – high light intensity and nutrient limitation – as a promising strategy to improve lipid content. Considering these findings, the design of multiple combinations of stress conditions can contribute to an increase of lipid productivity, arising as a more economical approach [17]. The adoption of a strategy combining light intensity with the right photoperiod can also contribute to high growth rates of microalgae and, at the same time, enhance the productivity of some compounds that are produced under light-stress conditions. Xi *et al.* (2020) demonstrated that, depending on the light regime adopted during microalgae cultivation, different stress levels are observed. The authors evaluated the F_v/F_m fluctuating patterns under different light regimes and determined that, in continuous mode, this parameter decreases and remains low. On the other hand, when light/dark cycles are applied, quantum yield decreases during the light phase, and increases under dark conditions. Periodical exposure to light promotes a better accumulation of β-carotene than continuous exposure, as demonstrated by Xi *et al*. (2020). In their study,

the authors reported 0.53 % (in dry weight) of carotenoids in *Dunaliella salina* grown under light:dark cycles of 14:10, while, under a continuous light regime, the content of this group of pigments was just 0.35 % (in dry weight), evidencing the benefits of cyclical irradiation on the accumulation of products that are synthesized in response to light stress conditions [9].

Under cold stress, microalgae cells usually have a nearly round shape and present their maximum size [20]. Ferro *et al.* (2018) suggested the accumulation of storage products, such as lipids or starch, in the form of granules, is responsible for increasing cell's size [20]. The stress induced by temperature also impacts the accumulation of other compounds. For instance, microalgae growing at low temperatures increase the content of unsaturated fatty acids within the membrane, consequently increasing its fluidity. In a study conducted by Camacho-Rodríguez *et al.* (2014), it was found that higher PUFAs contents were obtained under low temperatures, whereas saturated fatty acids production was favoured using high temperatures. As happens in the case of lipids, pigments' contents are also affected by temperature. Under high temperatures, chlorophylls' content decreases [11] and the amount of carotenoids increases [10]. These changes in pigments' contents were essentially related to the increase of oxidative stress caused by higher temperatures, which increases photosynthetic rate. Oxidative stress damages PSII, which contributes to the decrease of chlorophylls and an increase in carotenoid content, as a consequence of their important antioxidant role [6, 20].

Recent findings suggest that microalgae present low concentration of lipids under extreme pH, having a total amount per cell 3-fold higher when grown under the optimal pH [21]. Considering this, the adoption of a two-stage strategy where microalgae grow under the optimal pH in the first stage and then, for a short period of time, the pH of the culture medium decreases to enhance lipids production, seems to be an interesting approach to obtain higher lipid productivities. On the other hand, certain studies point out the impact of extreme pH on the enhancement of carotenoid content [21–23]. Indeed, Minyuk *et al.* (2016) suggested that acidic conditions promote an increase in carotenoids/chlorophyll (Car/Chl) mass ratio in *Coelastrella rubescens* cultures [21]. Conversely, Guedes *et al.* (2011) demonstrated that using *Scenedesmus obliquus* cultures, higher pH improves the accumulation of lutein and carotene, presenting 203.57 and 18.20 $mg \cdot mL^{-1}$, respectively [22]. These findings suggest that, under extreme pH

conditions, a secondary metabolism contributing to limited chlorophyll production and an increase in carotenoid content is triggered.

Moreover, several studies have determined that pigment content decreases drastically under salinity stress conditions. Srivastava *et al* (2017) reported that chlorophyll content declined from 2.54 $mg.ml^{-1}$, using control conditions, to 2.35 $mg.ml^{-1}$, under the presence of 5 $mmol \cdot L^{-1}$ of NaCl. On the other hand, the Car/Chl ratio increases with the increase of salinity, being an indication that salty conditions contribute to oxidative stress, which is, in turn, responsible to reduce the activity of PSII [12, 13]. The same authors have also concluded that the reduction of chlorophylls content occurs due to a reduction in the utilization of CO_2 and nutrients, which results in a deceleration of NADPH formation, leading to metabolic modifications that affect the production of this pigment [24]. In contrast to pigments, lipid contents are positively influenced by salinity stress [13, 24, 25]. As reported by Pal *et al* (2011), for a growth medium containing 13 $g \cdot L^{-1}$ of NaCl, the total fatty acids fraction represented 13.5 % of the biomass (in dry weight), while this percentage increased to 16.6 % using a high-salinity environment (40 $g \cdot L^{-1}$ of NaCl) [25]. Under salt stress conditions, microalgae increase their lipidic content as storage energy material to overcome unfavourable conditions [24]. This type of condition also contributes to changes in the lipid profile. As an example, Pal *et al.* (2011) demonstrated that the increase in salinity contributes to a higher accumulation of saturated and monounsaturated fatty acids but also to a decrease of polyunsaturated fatty acids, such as EPA [25].

Finally, although traditional cultivation strategies have been demonstrated to enable greater microalgal biomass concentration and overproduction of high-value compounds, these approaches present some problems related to the consumption of resources during the optimization process, consequent high energy consumption, and equipment wear out **(Figure 2)**. This optimization process requires a certain period of time and, contrary to traditional approaches, Systems Biology allows reconstructing a model from genome information that enables the optimization of the overall metabolic mechanisms, instead of focusing on a single compound**.** Incidentally, Systems Biology presents itself as a promising strategy that can contribute to establishing a landmark in decreasing production process costs of microalgae biomass.

1.3 Systems Biology

Unlike traditional components biology, which is directed to individual components involved in biological processes, systems biology aims to depict biological systems, such as a cell, tissue, or organism, quantifying and analysing each component and the interactions between them [26]. This approach involves mathematical models that simulate cell metabolism and predict such behaviour in different environmental and genetic conditions. The development of metabolic models usually follows two different methodologies: stoichiometric modelling and dynamic (kinetic) modelling. Stoichiometric models require less information, representing only the system's structure: compounds and reactions, as well as their respective stoichiometry and reversibility. A pseudo-steady state is assumed: for all metabolites, the fluxes leading to the production of a given metabolite are balanced with the fluxes leading to its consumption, meaning that there is no net accumulation of metabolites. Stoichiometric models are virtually applicable to any species with a sequenced genome and are usually developed at the genome scale. On the other hand, dynamic models account not only for structural details and the kinetic parameters of the process under study but also the concentration of metabolites over time. Such information is usually not available at the genome scale; thus, these models are often targeted to a lower number of metabolic pathways [27].

The development of next-generation sequencing methods and biological databases allows an increasing availability of genome sequences, which are used to study biological networks at the systems level. The reconstruction of genome-scale metabolic (GSM) models is one of the most relevant biological networks' modelling approaches. Such models comprise genes, reactions, and metabolites identified in the target organism, combining data retrieved from the genome sequence, biological databases, and literature. Due to the complexity and laborious work of reconstructing a metabolic model, several tools, like *merlin* [26, 28], were developed in the last decade to assist this process. GSM models have a wide range of applications depending on the study objectives, including drug targeting, metabolic engineering, media optimization, and analysis of cellular phenotypes under different environmental and genetic conditions.

The reconstruction of GSM models has been applied to organisms belonging to all kingdoms of life, including microalgae. The first genome-scale metabolic models for

these organisms were reconstructed for *Chlamydomonas reinhardtii* [29] and *Arthrospira platensis* [30], in 2008 and 2003, respectively. From then, metabolic models for other species, such as *Chlorella vulgaris* or *Synechocystis* sp., were developed and repeatedly updated with new information. However, the reconstruction of GSM models for microalgae is still limited to model organisms, being their main focus on the production of desirable compounds, such as hydrogen and lipids. In addition, biomass growth under different light conditions, network analysis, and metabolic engineering methods have also been studied using these models.

This review focuses on the recent developments in the reconstruction and application of GSM models for microalgae and how the retrieved information can be applied to leverage the cost-effectiveness of the experimental procedure. In addition to the Systems Biology approach, a review of the common strategies utilized to improve biomass cultivation and metabolites' production is also performed.

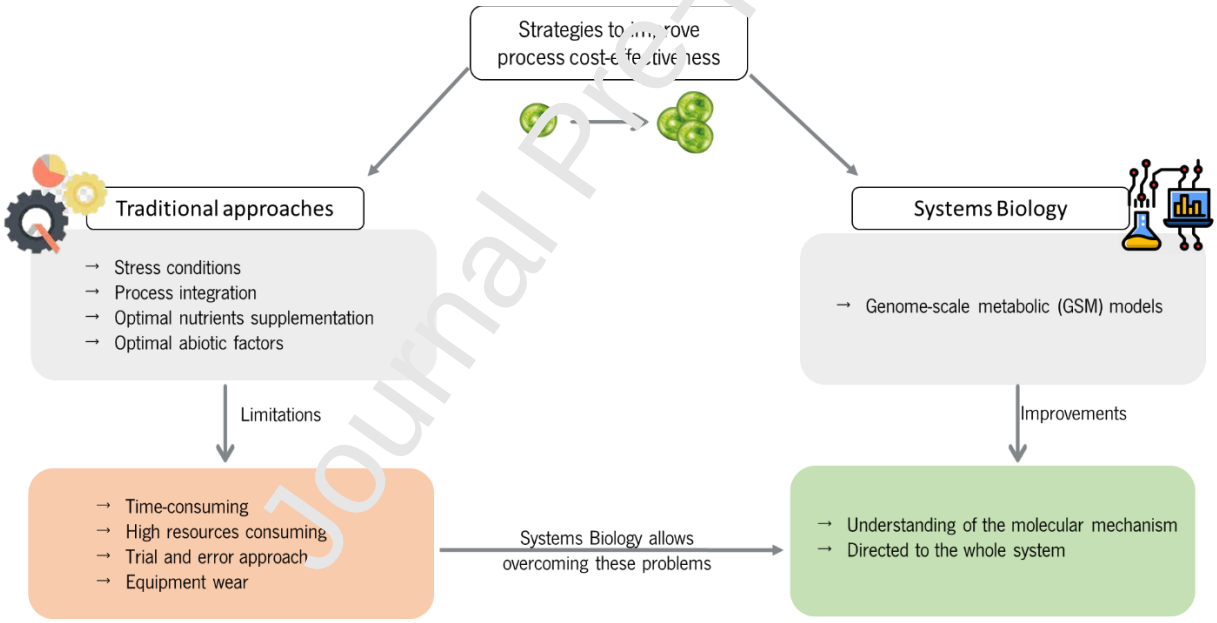


Figure 2 Strategies to improve process cost-effectiveness.

2 *In silico*-guided optimization as an accurate alternative

2.1 Systems Biology tools

In the last decade, systems biology has been providing frameworks to improve the cost-effectiveness of microalgae cultivation. Together with omics data, these methods result in the generation and management of large amounts of biological data. Genome-scale metabolic models, for example, usually contain hundreds or thousands of reactions and

metabolites based on genomes that can have more than 20,000 genes. Hence, managing this data with bioinformatics tools is highly advisable. These tools can often automatically fetch and simplify information available in biological databases or other online resources, accelerating the reconstruction process and reducing the associated costs (**Table 1**).

Table 1: Tools available for assisting the reconstruction of GSM models and generic biological databases and microalgae-specific databases with relevant information for this task

Tool Name	Description	Reference/URL
	GSM model reconstruction tools	
CarveMe	Command-line python-based tool designed to create GSM models from a BiGG-based manually curated universal template with a gap-filling algorithm.	[31]
merlin	Java application with a graphical interface that supports the reconstruction of both prokaryotic and eukaryotic organisms. It provides frameworks for genome re-annotation, manual curation, generation of transport reactions, gap identification, and network visualization.	[26, 28]
ModelSEED	Web-resource that allows the reconstruction and analysis of GSM models for microorganisms and plants. It uses RAST to annotate the genome and performs gap-filling based on the growth medium selected by the user. ModelSeed also provides biochemical data, including reaction's reversibility and the pKa of metabolites.	[32, 33]
Pathway Tools	Pathway Tools is a comprehensive bioinformatics software package that spans enterprise genome data management, systems biology, and omics data analysis. The software has been licensed by more than 12,000 groups and powers BioCyc.org and several other Pathway/Genome Database websites.	[34]
RAVEN	RAVEN (Reconstruction, Analysis and Visualization of Metabolic Networks) is a command-line tool compliant with COBRA Toolbox v3, thus running in proprietary software MATLAB. It provides functionalities to reconstruct models from scratch or use template models, merge networks from different databases and curate from the command line.	[35]
	Simulation & Optimization Tools	
COBRA	The COBRA Toolbox is a MATLAB software suitable	[36]

Toolbox	for predicting cellular and multicellular biochemical networks with constraint-based modelling. It implements a comprehensive collection of basic and advanced modelling methods, including reconstruction and model generation and biased and unbiased model-driven analysis methods.	
OptFlux	OptFlux is an open-source and modular software with a graphical interface to support *in silico* metabolic engineering tasks.	[37]
MEWpy	Python workbench for metabolic engineering, which covers a wide range of metabolic and regulatory modelling approaches, as well as phenotype simulation and computational strain optimization algorithms	[38]
Generic Online Databases		
NCBI	Repository of biological databases, presenting resources for analysis and visualization of biomedical, genomic, taxonomic, proteomic, and other biological data.	[39]
JGI	The Joint Genome Institute is a resource providing genomic sequences, as well as data management systems and specialized analytical capabilities to manage and interpret complex genomic data sets.	[40]
KEGG	Collection of databases comprising information on the biological system including the cell, the organism, and the ecosystem. It provides genomic and metabolic information, including genome sequences, reactions, pathways, and chemical compounds relevant to cellular processes.	[41]
BioCyc	BioCyc is a collection of 20,025 Pathway/Genome Databases (PGDBs) for model eukaryotes and for thousands of microbes, plus software tools for exploring them. The EcoCyc and MetaCyc databases are freely available, but access to the remaining BioCyc databases requires a paid subscription.	[42]
BRENDA	Repository of manually curated enzyme functional data providing functional and molecular information of enzymes, such as kinetics, substrates/products, inhibitors/activators, and cofactors, using the Enzyme Commission (EC) system.	[43]
BiGG Models	Biochemical, genetic, and genomic knowledge database containing more than 85 high-quality manually curated	[44]

	GSM models. This resource allows users to browse, search, and visualize models connected to genome annotations and external databases.	
MetaNetX	MetaNetX/MNXref is a reconciliation of metabolites and biochemical reactions providing cross-links between major public biochemistry and metabolic models databases. It also allows uploading and analysing GSM models.	[45]
TCDB	The Transporter Classification Database (TCDB) incorporates both functional and phylogenetic information on membrane transport proteins to provide the Transporter Classification (TC), an extensive approved classification method for transporter proteins.	[46]
Microalgae-specific Databases		
AlgaeBase	Database of information on algae that includes terrestrial, marine, and freshwater organisms.	[47] www.algaebase.org
AlgaTerra	An information system for microalgal biodiversity, containing taxonomic, molecular, and ecological information.	[48] www.algaterra.org
Chlamydomonas Resource Center	Central repository to receive, catalogue, preserve, and distribute high-quality and reliable wild type and mutant cultures of the green alga *Chlamydomonas reinhardtii*, as well as useful molecular reagents and kits for education and research.	[49] www.chlamycollection.org
CyanoDB	The primary goal of CyanoDB 2.0 is to be a reference tool for taxonomists and other persons interested in the taxonomy and diversity of Cyanobacteria. It provides information on primary descriptions of taxa and references molecular data available on descriptions and revisions.	[50, 51] www.cyanodb.cz
NanDeSyn	*Nannochloropsis* Design & Synthesis Initiative focuses on creating a "Scalable Photosynthetic Yeast" chassis by combining community effort. It currently centralizes databases of genetic and genomic data for *Nannochloropsis* spp.	[52] nandesyn.single-cell.cn/
Other tools		
BRAKER2	BRAKER2 is an extension of BRAKER1, which allows for fully automated training of the gene prediction tools GeneMark-EX and AUGUSTUS from RNA-Seq and/or protein homology information, and that integrates the	[53]

	extrinsic evidence from RNA-Seq and protein homology information into the prediction.	
MAKER	MAKER is a configurable genome annotation pipeline that identifies repeats, aligns ESTs and proteins to a genome, produces ab-initio gene predictions and automatically synthesizes these data into gene annotations with evidence-based quality values. MAKER can be used cyclically: outputs of preliminary runs can be used to automatically retrain its gene prediction algorithm, producing higher quality gene-models on subsequent runs.	[54]
BUSCA	BUSCA (Bologna Unified Subcellular Component Annotator) is a web-server for predicting protein subcellular localization. BUSCA integrates different tools to predict localization-related protein features (DeepSig, TPpred3, PredGPI, BetAware and ENSEMBLE3.0) as well as tools for discriminating subcellular localization of both globular and membrane proteins (BaCelLo, MemLoci and SChloro). It presents specific predictors for animal, plant, fungi, and bacterial proteins, allowing also using proteins from other taxonomic groups.	[55]
DeepLoc	Protein location prediction tool that uses deep learning algorithms to predict the subcellular location of proteins.	[56]
LocTree3	Protein location prediction tool for prokaryotes, eukaryotes, and archaea based on support vector machine and annotation by sequence homology searches.	[57]
PredAlgo	Sequence analysis tool dedicated to the prediction of protein subcellular localization in green algae. It uses a neural network trained with carefully curated sets of *C. reinhardtii* proteins. PredAlgo predicts the localization in one of three compartments: the mitochondrion, the chloroplast, and the secretory pathway within the cell.	[58]
PSORTb	Prokaryotic protein subcellular localization prediction tool. PSORTb v3.0 handles archaeal sequences as well as Gram-positive and Gram-negative bacterial sequences, allowing using whole genomes as input (suite for cyanobacteria).	[59]
WoLF PSORT	Protein location prediction tool for fungi, animals, and plants. It is an extension of the PSORTb family, making	[60]

	predictions based on both known sorting signal motifs and some correlative sequence features, such as amino acid content.	
eQuilibrator	A resource offering a database with thermodynamic information and a tool for standard Gibbs energy prediction. It offers a web service and a python API.	[61]
ChemAxon pKa plug-in	ChemAxon offers several non-open-source calculators and predictors for chemoinformatics. The pKa plug-in has a Java API and allows determining the pKa of a metabolite in a given pH.	www.chemaxon.com

The reconstruction of GSM models can follow very different approaches. Tools like CarveMe can develop a functional model in a few minutes, using manually curated models as a template. However, this approach has some limitations as it does not account for the metabolic specificities of the target organism and can only be used for prokaryotic organisms. Other tools, such as *merlin*, offer more freedom to the user to adapt the metabolic model to the target organism and the purpose of the work. The COBRA toolbox has been widely used for the reconstruction, analysis, simulation, and optimization of microalgae's metabolic models [62, 63, 72–78, 64–71]. This MATLAB package offers basic and advanced modelling methods that can be used for the reconstruction, analysis, and simulation of metabolic models. The RAVEN toolbox [79–81], Pathway Tools [82–85], and ModelSEED [86, 87], were also used for the reconstruction of such models. The selection of an adequate reconstruction tool must consider different aspects, namely, the target organism (e.g., some tools are only suite for prokaryotes), the user programming skills (some tools, such as the COBRA toolbox, require programming in MATLAB, while others, like *merlin*, offer a graphical interface), the time required for the reconstruction, and the level of curation desired (automatic tools allow accelerating the process, often introducing errors in the model, while other tools allow the user to introduce their expertise in the target organism's metabolism).

As mentioned above, the COBRA toolbox, and its corresponding python package COBRApy [88], allows applying advanced simulation and optimization methods. However, most other reconstruction tools offer none or only simple simulation methods. Thus, a specific software can be used to use advanced approaches. For example, strain

optimization methods and integration of gene expression data can be supported by OptFlux or MEWpy.

Regardless of the reconstruction approach, biological databases are required to retrieve genomic, proteomic, metabolic, and enzymatic information. Organism-specific databases present information often curated and focused on the target organism, presenting its known metabolic and physiological specificities. Nevertheless, such information is not always available, thus more generic databases need to be used.

Most reconstruction tools allow retrieving metabolic data (namely, pathways, reactions, metabolites, and proteins) from a few biological databases (usually one or two). This decreases the complexity of the software development and allows standardizing reactions and metabolites' identifiers, which is helpful to compare and update metabolic models. For example, Pathway Tools uses BioCyc information, ModelSEED has its own internal database, CarveMe uses BiGG Models models as template, and *merlin* retrieves data from KEGG, allowing also reconstructing models from BiGG Models. Hence, the reaction and metabolites' identifiers are dependent on the reconstruction tool used. Mapping such identifiers is not always straightforward due to different types of data organisation and redundancy. MetaNetX is suite to get cross-links between metabolic modelling and biochemistry databases, including KEGG, BioCyc, BiGG Models, and ModelSEED.

KEGG, BioCyc, and BiGG Models are often used as the major source of metabolic information for the reconstruction of microalgal GSM models [62, 63, 74–79, 81–84, 64, 85, 89–97, 65, 98–105, 67–71, 73]. The information available at KEGG and BioCyc is organised by pathway, which is helpful to analyse small portions of metabolic networks. These databases provide information at the reaction, enzyme, gene, and metabolite level, making them suite to be used as template for reconstruction tools. KEGG and BioCyc also contain organism-specific information, usually obtained from automatic genome annotations. However, BioCyc databases require a paid subscription, except for EcoCyc (*Escherichia coli* database) and MetaCyc (non-organism specific database). BiGG Models is a repository of highly curated GSM models, being the information available oriented to the development, analysis, and application of metabolic models. However, the information available here is often limited to model organisms. For example, *Chlamydomonas reinhardtii*, *Synechococcus elongatus* PCC

7942, and *Synechocystis sp.* PCC 6803 are the only microalgae represented in the database at the moment.

Although one database can be used as template, other resources are used to get additional information, allowing filling gaps in the network and generating a more complete metabolic model. Besides the above-mentioned databases, BRENDA and TCDB are usually used to retrieve information regarding enzymes and transporters, respectively. The PlantCyc repository shows an organisation similar to BioCyc, containing several organism-specific genome-scale metabolic pathway databases of 126 algae and plants. The green algae available here include *C. reinhardtii* (ChlamyCyc), *Volvox carteri* (VcarteriCyc), *Ostreococcus lucimarinus* (OlucimarinusCyc), *Chromochloris zofingiensis* (CzofingiensisCyc), *Cyanidioschyzon merolae* (MpusillaCyc), *Micromonas commode* (McommodaCyc), *Micromonas pusilla* (MpusillaCyc), and *Chlorella variabilis* (CvariabilisCyc). The *C. reinhardtii* database (ChlamyCyc) has been used as a source of information to develop GSM models for this organism [64, 65, 93, 94]. These databases can be useful as they include a few organism-specific pathways, such as "carbon concentration mechanism (*Chlamydomonas*)" and "ergosterol biosynthesis II (*Chlamydomonas*)".

As shown in **Table 1**, other resources can be used throughout the reconstruction process. Several tools allow the prediction of protein subcellular locations, often based on machine learning algorithms. Some tools, like LocTree3, combine these algorithms with homology and annotation based inference [106]. These tools show different performances depending on the organism taxa of the target sequence and on the organelle. For example, LocTree3 and BUSCA show better performance predicting chloroplastic proteins, while DeepLoc has better results on animals/fungi extracellular proteins [55]. On the other hand, PredAlgo allows predicting protein subcellular location specifically for green algae. However, it only predicts three compartments: the mitochondrion, the chloroplast, and the secretory pathway, making this tool incompatible with a genome-scale reconstruction. Selecting a protein location prediction tool for GSM reconstruction purposes must also account for its compatibility with reconstruction tools. For example, RAVEN allows predicting compartments with WoLF PSORT, while *merlin* integrates results from WoLF PSORT, LocTree3, and PSORTb.

Bioinformatics tools like eQuilibrator and ChemAxon, which retrieve thermodynamic and chemical properties of reactions and metabolites, respectively, are also useful to develop high-quality metabolic models.

2.2 GSM models for microalgae

The reconstruction of GSM models has been extensively described [27, 107] and often comprises four main steps: genome annotation, assembly of a draft network, conversion into a stoichiometric model, and model validation. Nevertheless, each one of these stages comprises several minor steps in an iterative process (Figure 3).

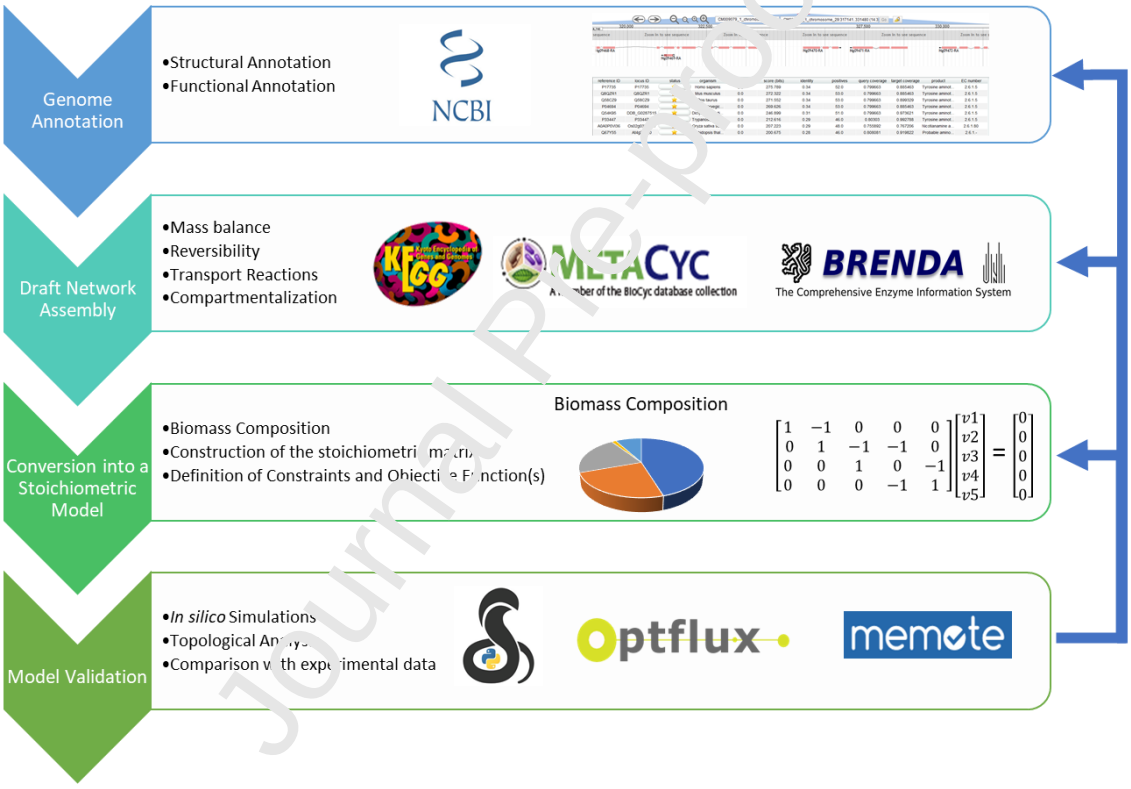


Figure 3: Representation of the four major stages of reconstructing a GSM metabolic model. This iterative process requires utilising several data sources and tools for genome annotation, draft network assembly and curation, conversion into a stoichiometric model, and model validation.

2.2.1 Genome annotation

The first step of reconstructing a GSM model is obtaining the genome structural and functional annotation, which can be found in organism-specific databases or databases comprising collections of genome annotations, such as NCBI or KEGG [107]. Nevertheless, only well-studied microalgae, such as *Chlamydomonas reinhardtii*, have

available genome annotations in the most-known public databases. Organism-specific databases, such as the ones listed in **Table 1**, can be used to search annotated genomes. However, it might be necessary for less-studied organisms to perform gene prediction tasks before the functional annotation step.

Structural annotation involves searching and identifying all features present in the genome, including coding sequences, promoters, and different types of RNA. Gene prediction can follow different methodologies: *ab initio* prediction, transcript-based approaches, and homology-based approaches [108].

Gene finders, like Augustus [109], SNAP [110], and GeneMark-ES [111], use statistical models to predict protein-coding genes in genomic sequences. This approach, often called *ab initio* prediction, is required for genes not represented – or weakly represented – in RNA-seq libraries and, consequently, do not have enough similarity to known genes or proteins.

Transcript-based approaches are based on the spliced alignment of transcript sequences (usually from RNAseq) of the organism or a phylogenetically close organism. Tools like TopHat [112] or RNA STAR [113] align these sequences against the organism's genome, allowing to retrieve information about the location and structure of transcripts: introns, exons, and alternative splicing possibilities. On the other hand, transcriptome assemblers like Cufflinks [114] and StringTie [115] build a set of transcripts from the alignments used by gene finders as hints to identify gene structures.

Homology-based approaches have the assumption that protein sequence and gene structures are conserved across different organisms. These methods use a set of protein sequences or a representation of a protein family obtained from multiple sequence alignments to identify gene structures from alignments against the genome.

Multiple approaches are often applied to increase the number and accuracy of the identified genomic features. MAKER v3 and BRAKER2 tools, for instance, can accept and process transcriptomics data, Expression Sequence Tags, protein sequences, and trained *ab initio* predictors. Moreover, MAKER v3 can also identify alternative splicing, use Evidence Modeler to refine gene predictions, and add tRNAs using tRNAscan [116].

Genome functional annotation comprises the assignment of functions to all genes identified in the genome. Homology search methods, such as the Basic Local Alignment Search Tool (BLAST) [117], HMMER [118], and DIAMOND [119], are usually used to perform functional annotation. For the reconstruction of GSM models, enzyme and transporter-encoding genes have particular relevance since the metabolic network will be based on reactions catalysed by proteins encoded by such genes. Enzymes encoding genes are usually annotated with Enzyme Commission (EC) numbers, while transporter-encoding genes can also be associated with different identifiers, namely Transport Classification (TC) numbers or TranSyt identifiers (https://identifiers.org/transyt) proposed by the authors of this review [120].

Nevertheless, similarity approaches have limitations. For example, BLAST-based methods often assume the one-to-one recognition principle, meaning that the annotation of a query gene is based on the annotation of a single known gene.

C. reinhardtii is often used as a reference organism for microalgae genome functional annotation. The most recent genome assembly (GenBank assembly accession GCA_000002595.3) presents a genome size of approximately 110 Mbp and 19,500 proteins. Though only 365 reviewed protein records are available at UniProt, and only 261 of them have experimental evidence. The *Plantae* model organism *A. thaliana* presents almost 11,500 proteins with experimental evidence, being this number higher than the case of *Saccharomyces cerevisiae* – 5,500 proteins. Consequently, a considerable number of *C. reinhardtii* genes were annotated using homology approaches. Hence, annotating the genome of non-model microalgae based on the annotation of *C. reinhardtii* can cause the propagation of annotation errors. Moreover, microalgae are a wide diverse phylogenetic group with representation in the *Plantae* and *Protista* kingdoms.

A more recent approach with a one-to-many strategy is Argot2 [121], which combines results of BLAST, HMMER, and an assessment of semantic similarities of Gene Ontology (GO) terms. FFPred3 [122] is another approach not merely based on sequence similarity, accounting for factors such as peptide signals and compartmentalization and applying machine learning methods.

The assignment of subcellular protein locations is one of the main differences between prokaryotic and eukaryotic GSM models. While in prokaryotes, subcellular locations

are usually limited to the cytoplasm and extracellular space, eukaryotic organisms can account for several compartments. In microalgae, compartmentalized metabolic models usually account for the cytosol, mitochondria, chloroplast, peroxisome, extracellular environment, nucleus, Golgi apparatus, endoplasmic reticulum, vacuole, and cell wall. This step is central for obtaining an accurate metabolic model, allowing to create pools of cofactors like NADH and NADPH in different compartments, mimicking metabolic shunts, and differentiating the protonation state of the same metabolite at compartments with different pH values. Tools like BUSCA, DeepLoc, LocTree3, WoLF PSORT, and PSORTb (for prokaryotic organisms) can predict protein location. Nevertheless, primary literature and compartment-specific tools, such as TargetP [123] and DeepMito [124], can provide additional information.

2.2.2 Manual Curation

The identifiers used for genome annotation, namely EC numbers, can help assemble a draft network with a set of reactions catalysed by enzymes encoded in the genome. Non-enzymatic and spontaneous reactions can also be added to the network when necessary. The automatically generated network usually contains wrong information, such as missing reactions and reactions that should not be included. Hence, manual refinement of the network is necessary to obtain a high-quality GSM model.

One of the critical requirements of a metabolic network is that reactions must be mass-balanced. However, reactions retrieved from databases like KEGG can be unbalanced, often due to metabolites without or incorrect chemical formulae and missing protons or water molecules.

The assignment or correction of the metabolites formulae can be performed by consulting biological databases. Nevertheless, metabolites with variable formulae can represent a challenge performing this task. These compounds include tRNAs, acyl-carrier protein, ferredoxin, quinones, lipids, and polysaccharides. An approach to balance reactions containing tRNA, acyl-carrier protein (ACP), and ferredoxin is to assign generic formulas like “R” or “T” to such metabolites. This approach is only viable for metabolites that are recycled inside the network. For example, ACP is “consumed” at the beginning of fatty acid biosynthesis and “produced” at the end of this pathway. On the other hand, polysaccharides are usually consumed from a culture medium or secreted to it. Hence, the quantity of each monomer comprised in such

polymers must be determined using available experimental data and biological databases.

Thermodynamic constraints can be included in the model by restraining the reversibility and direction of each reaction. This kind of information is usually available in databases such as MetaCyc, BRENDA, or BiGG Models. Otherwise, the standard Gibbs free energy of the reaction can be determined and used to infer the direction and reversibility of the reaction. Tools like eQuilibrator can also be useful to determine such thermodynamic parameters. These constraints must be carefully analysed as these can completely change posterior simulations. Having too many reversible reactions can create energy-generating cycles. On the other hand, excessively restricting reactions' reversibility can overfit the model and render it unable to predict metabolic behaviour under different environmental or genetic conditions.

Transport reactions are essential to keep network connectivity by allowing the exchange of metabolites across different compartments. Such reactions enforce different mechanisms, such as simple diffusion, proton symport, ABC cassettes, or the phosphotransferase system. The generation and inclusion of transport reactions can be based on the genome annotation performed in the previous stage. Nevertheless, this process includes searching for transporter-encoding genes, determining the transport mechanism and reversibility, and the substrates that the transporter can accept. Tools like TranSyT can be used to identify transporter-coding genes based on similarity searches against TCDB, automatically creating the respective transport reactions. Simple diffusion reactions need to be included manually, depending on the specificity of the metabolic model.

Photosynthesis plays a major role in GSM models of photosynthetic organisms, providing energy and carbon sources in photo- and mixotrophic conditions. Light is usually included in the metabolic models as an ordinary metabolite, named as “photon”. Nevertheless, the light wavelength and intensity can induce changes in the metabolic behaviour of microalgae. Chang *et al.* [91] developed a GSM model for *C. reinhardtii* where a modelling approach was devised enabling quantitative growth prediction for a given light source accounting for the wavelength and photon flux. In this approach, the total light spectrum was divided into effective bandwidths with associated coefficients, allowing to calculate flux distribution for these reactions. Hence, this approach enables models to account for reactions simulating different light sources, such as solar,

incandescent, fluorescent, and light-emitting diodes. However, it has been shown that the bandwidth coefficients can vary for different organisms, as well as with culture size and growth vessels used [125]. Thus, this approach must be refined for the target organism to obtain GSM models with improved capabilities.

2.2.3 Conversion into a stoichiometric model

Converting the network into a stoichiometric model requires formulating a reaction representing the biomass composition of the organism. This reaction accounts for the most abundant macromolecules (DNA, RNA, protein, etc.) and their respective precursors (e.g., ATP, dATP, L-Alanine). The biomass composition can be retrieved from literature, experimental data, or using bioinformatics tools. If no data is available for the organism under study, information from closely related species can be used. Finkel *et al*. [126] analysed the median macromolecular composition of microalgae from eight different taxonomic groups, showing that proteins, lipids, and carbohydrates are the most representative macromolecules in these organisms. Nevertheless, biomass composition can show significant differences depending on the species, strain, and environmental conditions.

DNA, RNA, and protein composition can be inferred from genomic and transcriptomics data using bioinformatics tools [127], which is very useful for less-studied organisms. Carbohydrates comprise both monomers, such as glucose, and polymers, like cellulose. The inclusion of monomers is straightforward as their chemical formulae are well-defined. Polymers, however, require additional consideration because of their variable qualitative and quantitative compositions, which can be overcome by defining a fixed chemical composition to the polymer, according to experimental data.

As mentioned before, microalgae are generally able to accumulate lipids, especially under nitrogen deficiency conditions. These macromolecules create an additional challenge to metabolic modelling because of their high diversity. For example, phospholipids can have different head groups combined with fatty acyl chains with multiple lengths. The most straightforward approach is to determine the average fatty acid composition of the cell and create a metabolite representing that average. This method is simple and adequate when the model is not used to study lipid metabolism. However, modelling lipid composition and production yields is often a target for microalgae modelling. Recent tools, such as BOIMMG

(https://boimmg.bio.di.uminho.pt/), create structurally defined lipids and the respective biosynthetic reactions, while SLIMEr [128] generates lipids biosynthetic reactions but requires the lipids to be already structurally defined.

Although cofactors and vitamins do not represent a significant content in the overall biomass composition, their inclusion in the biomass reaction guarantees that the model can produce them. Common cofactors and vitamins in GSM models include NADH, NADPH, S-adenosyl-methionine, pyridoxal 5-phosphate, coenzyme A, thiamine, folates, and quinones. Nevertheless, organism-specific data must be used to include additional compounds.

2.2.4 Model validation

The final stage of the reconstruction of a GSM model consists of verifying, evaluating, and validating the model. A variable number of quantitative and qualitative tests can be applied to validate the model: spontaneous growth, autotrophies, growth rate assessment, secreted products, alternative element sources, or gene essentiality. These tests are based on *in silico* simulations that employ linear, mixed-integer, and quadratic programming optimization problems.

If fluxes of external exchange reactions are available, the total flux distribution can be determined using Metabolic Flux Analysis (MFA). This approach is straightforward and useful to study the central metabolic pathways by using 13C labelled substrates (^{13}C-MFA). Such experimental measurements can be used by tools, such as OptFlux or WUFlux [129], to quantify fluxes in a complex metabolic network. Nevertheless, this type of data is not always available for less studied organisms and is often restricted to the central carbon metabolic pathways, limiting the potential for the application of MFA to metabolic models.

Flux Balance Analysis (FBA) [130] is the most frequently used approach for analysing biochemical networks. This method, based on linear programming, can be formulated as:

$$\begin{aligned} &\text{Maximize} && Z \\ &\text{subject to} && S \cdot v = 0 \\ & && \alpha_j \leq v_j \leq \beta_j, \qquad j = 1, \ldots, N \end{aligned}$$

where *S* is the stoichiometric matrix and *v* is the flux vector. The objective function Z can be chosen according to the purpose of the optimization: maximize cell growth (biomass formulation), maximize/minimize ATP production, maximize production of target metabolites, or minimize substrate uptake. Despite this, FBA provides a single solution that is not necessarily unique since other optimal flux distributions can exist. Therefore, parsimonious flux balance analysis (pFBA) can be used instead. This method optimizes the objective function Z, and then, minimizes the total sum of flux through the model, providing a unique optimal solution.

2.2.5 Additional approaches

A generic GSM model includes all the reactions catalyzed by enzymes encoded by metabolic genes identified in the organism's genome. Nevertheless, these models do not account for the regulatory network and gene expression present in the organism under different environmental conditions.

The development of omics technologies (genomic, transcriptomic, proteomic, lipidomic, and metabolomics) allows monitoring molecular components at the cellular and the genome-scale level. This type of biological data has been increasingly available in the last decade, especially for well-studied organisms. Hence, efforts have been made to develop methods to integrate omics data into GSM models. Such integration can be based on a "switch" approach, where a reaction is kept *On* or turned *Off* depending on the expression of the associated gene. On the other hand, a "valve" approach determines flux constraints of reactions based on the quantitative gene expression data.

Omics data from microalgae have been used to improve GSM models. Imam *et al*. (2015) [93] integrated transcriptomics data in a *C. reinhardtii* GSM model to identify changes in the central pathways in response to nitrogen starvation and light availability. This type of data was also combined with GSM model data to study the circadian rhythm in *Synechocystis* [131].

Some compounds with commercial interest, such as lipids and pigments, can be accumulated by microalgae in intracellular structures. Since stoichiometric models assume that there is no net accumulation of intracellular metabolites, the utilization of traditional simulation methods, like FBA, is not straightforward. Dynamic flux balance analysis (dFBA) [132] methods allow the prediction of metabolite consumption and production over time. Two approaches can be followed: the dynamic optimization

approach (DOA) and the static optimization approach (SOA). DOA involves optimising the entire time-period of interest to obtain time profiles of fluxes and metabolite levels. The dynamic optimization problem is transformed into a nonlinear programming (NLP) problem, which is solved once. Conversely, SOA involves dividing the batch time into several time intervals and solving the instantaneous optimization problem with LP, similar to FBA [132]. In recent years, dFBA approaches have been applied to microalgae GSM models to analyse the production of lipids and pigments between dark and light phases [133, 134]. Dynamic modelling can provide important insights for microalgae cultivation, including multi-stage cultivation (e.g., starting with a nitrogen source and then removing it to induce stress conditions) and photoperiod modulation.

The industrial utilization of an organism imposes requirements regarding efficient biomass yields and compound production. Extensive genetic engineering tasks, which can be a time-consuming and high-cost process, are required to improve the feasibility of using microalgae for industrial purposes.

GSM models can be used to predict the phenotypic behaviour of an organism under different environmental and genetic conditions. There are three major types of approaches for strain design: *de novo* pathway assembly, enzyme modulation, and enzyme deletion. OptStrain [135] can be used for pathway design, identifying heterologous reactions to improve the production of a target compound. On the other hand, OptGene [136] can identify gene and reaction targets for a knockout. Finally, algorithms like OptReg [137] and OptForce [138] can identify over/under-expression targets. Such algorithms are usually available in metabolic modelling optimization tools such as OptFlux and MEWpy.

Metabolic engineering using GSM models is already well-established for simple prokaryotic organisms. Nevertheless, metabolic engineering approaches with phototrophs is still limited to models of the cyanobacteria *Synechocystis* PCC6803 [139, 140] and *Synechococcus* PCC7002 [141].

Elementary flux mode (EFM) analysis is a useful constraint-based approach for metabolic engineering to identify all the genetically independent pathways inherent to a metabolic network. In simple words, an EFM analysis represents the minimal set of reactions between input and output metabolites at steady-state [90]. Thus, this approach is often used in metabolic engineering tasks to improve the production of desired

compounds. However, this strategy is more suitable for small networks due to the computational complexity of the process. For example, Rügen *et al*. (2012) [92] applied EFM analysis to a core model of *C. reinhardtii* to analyse this species' photoautotrophic metabolism. On the other hand, Baroukh *et al*. (2014) [142] used EFM to study the accumulation of lipids and carbohydrates in the microalgae *Tisochrysis lutea* under light/dark cycles.

In recent years, machine learning (ML) has started to be applied to reconstruct and analyse GSM models, improving their predictive performance and data coverage. ML algorithms can be applied in different stages of the reconstruction: genome annotation, identification of errors and gap-filling, gene essentiality, integration of regulatory data, and kinetic parameterization of enzymes [143, 144]. Only a few examples of this type of approach could be identified in cyanobacteria published GSM models. Vijayakumar *et al* (2020) [145] combined a GSM model of *Synechococcus* sp. PCC 7002 with ML to retrieve information from omics data regarding the metabolic responses to light intensity and salinity. *Saini et al*. [146], in turn, used a GSM model together with ML to formulate a multi-objective optimization problem to enhance biomass and phycobiliproteins production by the cyanobacteria *Nostoc* sp. CCC-403. The combination of metabolic models and ML is expected to increase in the next years, providing models with improved prediction capabilities and allowing overcome some restrictions of the traditional constraint-based modelling [143].

Alike ML applications, new formulations of metabolic models have been developed recently, such as Genome-scale metabolic models with Enzymatic Constraints using Kinetics and Omics (GECKO) models [147] or Metabolic and gene Expression (ME) models [148]. GECKO models are an extension of the traditional GSM models by incorporating detailed descriptions of enzymatic demands for metabolic reactions. Here, enzymes are considered as metabolites in reactions, whose stoichiometry is the inverse of the enzyme turnover number (k_{cat}). Integrating proteomics data, the abundancy of each enzyme can be used as an additional constraint. This approach has only been applied to model organisms, such as *S. cerevisiae* [147] and *E coli* [149], and no GECKO models have been found for microalgae. Obtaining k_{cat} values for all enzymes in an organism is clearly a limiting factor to develop these models. Nevertheless, ML methods have been developed recently to predict such parameters based on the protein sequence and substrate chemical structures [150]. Additionally, ME-models can also be

viewed as an extension of GSM models by including the biosynthesis of the gene expression mechanisms, enabling the computation of the entire proteome in a growing cell.

2.3 Successful examples

In the last decade, the number of GSM models for microalgae has increased exponentially. Before 2010, only eight published metabolic models were identified, and most of them were core models or draft networks. From 2010 until 2021, this number increased to 45 models, with a quite diverse set of methodologies being applied to these models, although limited to the better-studied species (**Table 2**).

Table 2: Published metabolic reconstructions of microalgae, as well as their general properties and examples of applications. Rxns: reactions; Mets: metabolites; Cmps: compartments

Organism	Reference	Genes	Rxns	Mets	Cmps	Application/Focus
Anabaena sp.	[78]	862	897	777	6	Evaluation of exchanges between heterocyst and vegetative cells
	[151]	1004	1035	1635	7	Analysis of amino acids and secondary metabolism
Anabaena variabilis	[105]	953	983	926	4	Integration of transcriptomics data to heterocyst and vegetative cells; Hydrogen production
Arthrospira platensis	[30]	-	121	134	2	Study of a shunt of phosphoenolpyruvate to pyruvate
	[82]	692	875	837	2	Gene essentiality
	[108]	620	746	673	3	Network design for glycogen and ethanol
	[72]	888	1096	994	-	Glycogen production
Chlamydomonas reinhardtii	[62]	1069	-	-	-	Combination of metabolomics and proteomics methods to obtain a draft metabolic network
	[63]	-	259	467	6	Integrated transcripts for verification of enzyme encoding open reading frames
	[89]	-	484	458	3	Application of FBA for estimation of intracellular

					fluxes and biomass yield in mixo-, hetero-, and photoautotrophic conditions
	[64]	-	280	278	-	Utilization of mixed-integer linear programming methods to determine flux distributions, with focus on light-driven respiration and photon flux density
	[91]	1080	2190	1068	10	Developed a new approach to account for light conditions in GSM models
	[65]	2249	1725	1862	4	Hydrogen Production
	[92]	-	280	278	-	Application of EFM analysis to analyse phototrophic metabolism
	[152]	-	160	154	2	Determination of energy requirements
	[66]	1106	2445	1959	10	Refinement of the iRC10810 model by integrating phenotype microarray data
	[93]	1155	2394	1133	10	Integration of transcriptomics data to improve growth rate predictions and to study the accumulation of triacylglycerols
	[67]	-	139		3	^{13}C-MFA to study central carbon metabolism
	[94]	-	3,726	2,436	10	Application of dynamic approaches to evaluate phenotypes at different CO_2 levels
***Chlorella* sp. FC2IITG**	[95]	-	114	161	-	Utilization of FBA and dFBA to study lipid accumulation
Chlorella protothecoides	[68]	627	1963	2115	14*	Hydrogen production
Chlorella pyrenoidosa	[97]	-	67	-	0	Energy demands and light/dark cycle metabolism
Chlorella variabilis	[86]	526	1455	1236	5	Analysis of different light conditions
Chlorella vulgaris	[79]	843	2294	1770	6	Pathway and flux

						distribution analysis during nitrogen starvation
	[80]	946	2294	1770	6	Dynamic metabolic demands during photosynthesis
***Cyanothece* sp.**	[153]	773	946	811	5	Comparison between *Synechocystis* sp. and *Cyanothece* sp.; gene essentiality
Cyanothece sp.	[102]	806	719	587	1	System analysis of light-driven metabolism
***Nannochloropsis* sp.**	[96]	383	987	1024	6	TAG production
Nannochloropsis gaditana	[69]	1321	1918	1862	4	Flux distribution under different light conditions and CO_2 uptakes
Nannochloropsis salina	[70]	934	2345	-	10	Optimization of TAG production
Ostreococcus lucimarinus	[71]	-	964	1100	2	Analysis of gap-filling strategies
Ostreococcus tauri	[71]		871	1014	2	Analysis of gap-filling strategies
Phaeodactylum tricornutum	[154]	151	88		5	Identification of genes involved in C4-like photosynthesis
	[155]	680	318	355	5	EFM analysis
	[156]	-	-	-	2	Analysed metabolic responses to light intensity variation using the model developed by Hunt *et al.* [155]
	[81]	1027	4456	2172	6	Indication of a glutamine-ornithine shunt
Synechocystis sp.	[157]	-	29	-	-	^{13}C-MFA to study the C4 pathway in cyanobacteria
	[125]	-	70	46	1	Analysis of the glyoxylate shunt and other important metabolic reactions; gene deletion analysis
	[99]	-	43	-	-	MFA for hydrogen production
	[83]	669	882	790	2	Analysis of gene regulation during light-shifting growth regimes

	[73]	337	380	291	1	Impact of photon uptake in photorespiration
	[100]	376	493	465	2	MFA to study the central carbon metabolism and ethanol production
	[84]	811	956	911	2	Flux coupling analysis and integration of transcriptomics data; ethanol production
	[74]	678	863	795	3	Analysis of different light and inorganic carbon availabilities and genetic perturbations
	[153]	731	1156	996	5	Comparison between *Synechocystis* sp. and *Cyanothece* sp.; gene essentiality
	[75]	677	759	601	6	Analysis of light/dark cycles in cyanobacterial metabolism
	[101]	816	1045	925	7	Flux distribution analysis and implementation of dFBA for diurnal cycles
	[158]	778	1163	1005	5	Periodic dynamic model spaning the metabolism over 12 time point models
	[159]	678	864	795	4	FBA analysis for overproduction of organic acids under dark anoxic condition
***Synechococcus* sp.**	[103]	611	552	542	2	Analysis of metabolic differences in cyanobacteria
	[141]	708	602	581	2	Biofuel production using gene knock-out strategies
	[104]	728	742	696	7	Gene essentiality and synthetic lethality analysis
	[87]	821	744	777	2	Increased glycogen and lipid synthesis under nitrogen depletion
	[76]	706	908	900	4	Production of glycogen and polyglucans
Synechococcus elongatus	[85]	715	851	838	-	Growth in autotrophic conditions

	[77]	785	850	768	7	Phototrophic growth and gene essentiality
Tisochrysis lutea	[142]	-	157	162	2	New dynamic metabolic modelling framework that handles non-balanced growth conditions
Trichodesmium erythraeum	[160]	647	973	988	2	Prediction of optimal cellular composition with different nitrogen sources

The first metabolic reconstructions were focused on the central carbon metabolism and photosynthetic pathways. Such networks were developed through ^{13}C-MFA and EFM analysis to study the metabolism of microalgae.

The increasing availability of genome sequences allowed the reconstruction of metabolic models at the genome-scale. The most common objectives are related to *i*) improving the production of desirable compounds, such as lipids, hydrogen, and bioethanol; *ii*) analysing the growth and metabolism under mixo-, hetero-, and photoautotrophic conditions; and *iii*) studying the response to different light conditions and light/dark cycles.

Microalgae can produce and accumulate high amounts of lipids, especially TAGs. Imam *et al*. [93] integrated high-resolution time series transcriptomics data to identify dynamic changes in central and TAGs' production pathways in response to nitrogen starvation and light availability. This study allowed verifying how carbon flux was redirected to TAG biosynthesis instead of biomass. Since lipids are usually accumulated intracellularly, dFBA approaches have been applied in microalgae GSM models to study lipid metabolism and production. Loira *et al*. [70] applied dFBA to study TAG production over time, designing knock-out strategies for strain optimization as a source of TAG. Muthuraj *et al*. [95] captured the light-dark metabolism of *Chlorella* sp. FC2 IITG using dFBA. Some of the most significant findings of this study include the following: the oxidative pentose phosphate pathway and Krebs cycle are relatively inactive under photoautotrophic conditions; redirecting the carbon flux from polysaccharide and neutral lipid resulted in an up-regulation of Krebs cycle in the dark phase; significantly active phosphorylation in the light phase was able to satisfy cellular energy requirement without the need of oxidative pentose phosphate pathway.

Hydrogen and ethanol are other compounds with particular interest. Dal'Molin *et al*. [65] developed a model for *C. reinhardtii* comprising the pathway for hydrogen production, identifying new targets to improve the production yield. Mekanik *et al*. [68], on the other hand, proposed changes in the NADH metabolism aiming at enhancing the hydrogen production in *A. protothecoides.* In another study [105], transcriptomics data was also integrated into a GSM model of *A. variabilis* to improve the hydrogen production by this species. Yoshikawa *et al*. [100] have evaluated ethanol production in *Synechocystis* sp. PCC6803 at different photon and nitrate uptake rates. The ethanol production by this species has also been studied by integrating transcriptomics data in a metabolic model [84]. Using *A. platensis* NIES-39, Yoshikawa *et al*. [98] performed *in silico* knockout simulations indicating that the deletion of genes related to the respiratory chain, such as NADH dehydrogenase and cytochrome *c* oxidase, could enhance ethanol production. Vu *et al*. [141] applied gene knockout algorithms to get insights into the potential for production of a series of compounds, such as acetate, alanine, succinate, fatty acids, and hydrogen, by *Synechococcus* 7002.

3 Future challenges

The cultivation of microalgae for industrial purposes has received more attention in the last years. Nevertheless, challenges must be overcome to increase the economic and environmental feasibility of processes using these organisms at the industrial scale. Due to the estimated population growth over the forthcoming years, microalgae arise as an excellent alternative to animal-based proteins, with all the advantages associated with microalgae cultivation. The increased demand for microalgae metabolites requires decreasing production costs and optimising the production of desired compounds, such as proteins, some lipids (e.g., PUFAs), or carotenoids (e.g., β-carotene). A common strategy to promote the overproduction of high-value compounds is to cultivate microalgae under stress conditions, which impairs growth and limits biomass formation. It is, therefore, necessary to find a balance between the production of target compounds and biomass formation. In addition, the implementation of other strategies within the circular bioeconomy concept, such as the bioremediation of wastewater or the biorefinery approach, to improve process cost-effectiveness has been providing promising indications in microalgae's cultivation process cost-effectiveness.

Emerging fields, such as Systems Biology, offer tools to optimise these processes at the systems level. GSM models have been used to study metabolic traits of microalgae and predict the respective phenotypes, providing insights to increase biomass yields and production of desired compounds. However, these approaches have some challenges and limitations. First, reconstructing a high-quality GSM model demands a considerable amount of time, and requires intensive literature research and extensive expertise regarding the metabolism of the target organism. The continuous development and improvement of bioinformatics tools and databases allow directing such efforts from the reconstruction process to the application and analysis of metabolic models, leading to more expedited knowledge extraction. Secondly, these metabolic models do not include physical properties like temperature and pressure, or gene regulation and expression information, which has been overcome by integrating omics and regulatory data, originating condition-specific GSM models. However, genomic, proteomic, metabolomic, and physiological data are not always available, especially for less-studied organisms. Although using information from close-related organisms is a very common approach in metabolic modelling, this can decrease the accuracy and specificity of the model. Thus, generating more experimental data, as well as finding new approaches to adapt and infer data from other organisms, is mandatory to obtain high-quality models for less-studied organisms. Finally, traditional approaches in systems biology are based on a pseudo-steady state, not allowing the net accumulation of intracellular metabolites that can occur *in vivo*, for instance, with compounds like lipids, carbohydrates, and pigments. The implementation of dynamic approaches has been useful to evaluate microalgae's potential for producing such compounds over time, including at the industrial level. In recent years, new modelling approaches have been developed, including GECKO and ME models. Applying these new-generation modelling methods individually or in combination can provide new information and predicting capabilities unattainable through traditional systems biology approaches. The availability of experimental data is still a challenge, although it can be overcome, in part, through ML approaches.

4 Conclusions

Microalgae biomass is a noble source of high-value products. The commercialization of these compounds is dependent on improving microalgae cultivation's cost-effectiveness. Traditional approaches based on understanding the interactions between environmental

conditions and nutritional factors rely on trial-and-error methodologies and are lengthy. Systems biology has shifted this paradigm through tools like GSM, an excellent surrogate that allows devising strategies to optimize the production of biomass and high-value products. Implementing these tools reduces the time and costs of microalgae cultivation process optimization, potentially facilitating the implementation of a wider range of products derived from microalgae at the industrial level.

Acknowledgements

Emanuel Cunha and Vitor Sousa acknowledge the Foundation for Science and Technology (FCT) for their fellowship (DFA/BD/8076/2020 and UI/BD/151238/2021, respectively). This study was supported by the Portuguese Foundation for Science and Technology (FCT) under the scope of the strategic funding of UIDB/04469/2020 unit, by the I&D&I AgriFood XXI project (NORTE-01-0145-FEDER-000004) co-funded by the European Regional Development Fund under the scope of Norte2020 (Programa Operacional Regional do Norte 2014/2020) and BioTecNorte operation (NORTE-01-0145-FEDER-000004) funded by the European Regional Development Fund under the scope of Norte2020 - Programa Operacional Regional do Norte. This research work was supported by ALGAVALOR - Lisboa-01-0247- FEDER-035234, supported by Operational Programme for Competitiveness and Internationalization (COMPETE2020), Lisbon Portugal Regional Operational Programme (Lisboa 2020) and Algarve Regional Operational Programme (Algarve 2020) under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). Oscar Dias acknowledge FCT for the Assistant Research contract obtained under CEEC Individual 2018.

Competing Interests

The authors declare that they have no competing interests.

Author contributions

Emanuel Cunha: Conceptualization, Writing-Original draft; Vítor Sousa: Conceptualization, Writing-Original draft; Pedro Geada: Conceptualization, Writing-Review & Editing; José Teixeira; Writing-Review & Editing; António Vicente: Writing-Review & Editing; Oscar Dias: Conceptualization, Writing-Review & Editing.

References

1. Geada P, Rodrigues R, Loureiro L, et al (2018) Electrotechnologies applied to microalgal biotechnology – Applications, techniques and future trends. Renew Sustain Energy Rev 94:656–668. https://doi.org/10.1016/j.rser.2018.06.059

2. Ishika T, Moheimani NR, Bahri PA (2017) Sustainable saline microalgae co-cultivation for biofuel production: A critical review. Renew Sustain Energy Rev 78:356–368. https://doi.org/10.1016/j.rser.2017.04.110

3. Geada P, Vasconcelos V, Vicente A, Fernandes B (2017) Microalgal Biomass Cultivation. Algal Green Chem Recent Prog Biotechnol 257–284. https://doi.org/10.1016/B978-0-444-63784-0.00013-8

4. Abu-Ghosh S, Dubinsky Z, Verdelho V, Iluz D (2021) Unconventional high-value products from microalgae: A review. Bioresour Technol 329:124895. https://doi.org/10.1016/j.biortech.2021.124895

5. Brennan L, Owende P (2010) Biofuels from microalgae—A review of technologies for production, processing, and extractions of biofuels and co-products. Renew Sustain Energy Rev 14:557–577. https://doi.org/10.1016/j.rser.2009.10.009

6. Juneja A, Ceballos RM, Murthy GS (2013) Effects of environmental factors and nutrient availability on the biochemical composition of algae for biofuels production: A review. Energies 6:4607–4638. https://doi.org/10.3390/en6094607

7. Costa SS, Miranda AL, Andrade RB, et al (2018) Influence of nitrogen on growth, biomass composition, production, and properties of polyhydroxyalkanoates (PHAs) by microalgae. Int J Biol Macromol 116:552–562. https://doi.org/10.1016/j.ijbiomac.2018.05.064

8. Fernandes BD, Dragone GM, Teixeira JA, Vicente AA (2010) Light regime characterization in an airlift photobioreactor for production of microalgae with high starch content. Appl Biochem Biotechnol 161:218–226. https://doi.org/10.1007/s12010-009-8783-9

9. Xi Y, Wang J, Chu Y, et al (2020) Effects of different light regimes on Dunaliella salina growth and β-carotene accumulation. Algal Res 52:102111. https://doi.org/10.1016/j.algal.2020.102111

10. Minhas AK, Hodgson P, Barrow CJ, Adholeya A (2016) A review on the assessment of stress conditions for simultaneous production of microalgal lipids and carotenoids. Front Microbiol 7:546. https://doi.org/10.3389/FMICB.2016.00546/BIBTEX

11. Serra-Maia R, Bernard O, Gonçalves A, et al (2016) Influence of temperature on Chlorella vulgaris growth and mortality rates in a photobioreactor. Algal Res 18:352–359. https://doi.org/10.1016/j.algal.2016.06.016

12. Chokshi K, Pancha I, Ghosh A, Mishra S (2017) Salinity induced oxidative stress alters the physiological responses and improves the biofuel potential of green

microalgae Acutodesmus dimorphus. Bioresour Technol 244:1376–1383. https://doi.org/10.1016/j.biortech.2017.05.003

13. Pancha I, Chokshi K, Maurya R, et al (2015) Salinity induced oxidative stress enhanced biofuel production potential of microalgae Scenedesmus sp. CCNM 1077. Bioresour Technol 189:341–348. https://doi.org/10.1016/j.biortech.2015.04.017

14. Guihéneuf F, Stengel DB (2017) Interactive effects of light and temperature on pigments and n-3 LC-PUFA-enriched oil accumulation in batch-cultivated Pavlova lutheri using high-bicarbonate supply. Algal Res 23:113–125. https://doi.org/10.1016/j.algal.2017.02.002

15. Wu Z, Duangmanee P, Zhao P, Ma C (2016) The effects of light, temperature, and nutrition on growth and pigment accumulation of three *Dunaliella salina* strains isolated from saline soil. Jundishapur J Microbiol 9:1–9. https://doi.org/10.5812/jjm.26732

16. Xin L, Hong-ying H, Ke G, Ying-xue S (2010) Effects of different nitrogen and phosphorus concentrations on the growth, nutrient uptake, and lipid accumulation of a freshwater microalga Scenedesmus sp. Bioresour Technol 101:5494–5500. https://doi.org/10.1016/j.biortech.2010.02.016

17. Kim DW, Shin WS, Sung MG, et al (2019) Light intensity control as a strategy to improve lipid productivity in Chlorella sp. HS2 for biodiesel production. Biomass and Bioenergy 126:211–219. https://doi.org/10.1016/j.biombioe.2019.05.014

18. Kumar G, Nguyen DD, Huy N, et al (2019) Effects of light intensity on biomass, carbohydrate and fatty acid compositions of three different mixed consortia from natural ecological water bodies. J Environ Manage 230:293–300. https://doi.org/10.1016/j.jenvman.2018.09.026

19. He Q, Yang H, Wu L, Hu C (2015) Effect of light intensity on physiological changes, carbon allocation and neutral lipid accumulation in oleaginous microalgae. Bioresour Technol 191:219–228. https://doi.org/10.1016/j.biortech.2015.05.021

20. Ferro L, Gorzsás A, Gentili FG, Funk C (2018) Subarctic microalgal strains treat wastewater and produce biomass at low temperature and short photoperiod. Algal Res 35:160–167. https://doi.org/10.1016/j.algal.2018.08.031

21. Minyuk GS, Chelebieva ES, Chubchikova IN, et al (2016) pH and CO2 effects on Coelastrella (Scotiellopsis) rubescens growth and metabolism. Russ J Plant Physiol 63:566–574. https://doi.org/10.1134/S1021443716040105

22. Guedes AC, Amaro HM, Pereira RD, Malcata FX (2011) Effects of temperature and pH on growth and antioxidant content of the microalga Scenedesmus obliquus. Biotechnol Prog 27:1218–1224. https://doi.org/10.1002/btpr.649

23. Jin E, Polle J, Lee HK, et al (2003) Xanthophylls in Microalgae: From Biosynthesis to Biotechnological Mass Production and Application. J Microbiol Biotechnol 37:165–174

24. Srivastava G, Nishchal, Goud V V. (2017) Salinity induced lipid production in microalgae and cluster analysis (ICCB 16-BR_047). Bioresour Technol 242:244–252. https://doi.org/10.1016/j.biortech.2017.03.175

25. Pal D, Khozin-Goldberg I, Cohen Z, Boussiba S (2011) The effect of light, salinity, and nitrogen availability on lipid production by Nannochloropsis sp. Appl Microbiol Biotechnol 90:1429–1441. https://doi.org/10.1007/s00253-011-3170-1

26. Dias O, Rocha M, Ferreira EC, Rocha I (2015) Reconstructing genome-scale metabolic models with merlin. Nucleic Acids Res 43:3899–3910. https://doi.org/10.1093/nar/gkv294

27. Rocha II, Förster JJ, Nielsen J (2008) Design and Application of Genome-Scale Reconstructed Metabolic Models. In: Humana Press Inc. (ed) Methods in Molecular Biology, vol. 416: Gene Essentiality. pp 409–431

28. Capela J, Lagoa D, Rodrigues R, et al (2022) merlin , an improved framework for the reconstruction of high-quality genome-scale metabolic models. Nucleic Acids Res 50:6052–6066. https://doi.org/10.1093/nar/gkac459

29. Christian N, May P, Kempa S, et al (2009) An integrative approach towards completing genome -scale metabolic networks. Mol Biosyst 5:1889–1903. https://doi.org/10.1039/B915913B

30. Cogne G, Gros J-B, Dussap C-G (2003) Identification of a metabolic network structure representative of arthrospira (spirulina) platensis metabolism. Biotechnol Bioeng 84:667–676. https://doi.org/10.1002/BIT.10808

31. Machado D, Andrejev S, Tramontano M, Patil KR (2018) Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res 46:7542–7553. https://doi.org/10.1093/nar/gky537

32. Henry CS, DeJongh M, Best AA, et al (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28:977–982. https://doi.org/10.1038/nbt.1672

33. Seaver SMD, Liu F, Zhang Q, et al (2021) The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. Nucleic Acids Res 49:D575–D588. https://doi.org/10.1093/NAR/GKAA746

34. Karp PD, Midford PE, Billington R, et al (2021) Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. Brief Bioinform 22:109–126. https://doi.org/10.1093/BIB/BBZ104

35. Agren R, Liu L, Shoaie S, et al (2013) The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. PLoS Comput Biol 9:e1002980. https://doi.org/10.1371/journal.pcbi.1002980

36. Vlassis N, Pacheco MP, Sauter T (2014) Fast Reconstruction of Compact Context-Specific Metabolic Network Models. PLoS Comput Biol.

https://doi.org/10.1371/journal.pcbi.1003424

37. Rocha I, Maia P, Evangelista P, et al (2010) OptFlux: An open-source software platform for in silico metabolic engineering. BMC Syst Biol 4:45. https://doi.org/10.1186/1752-0509-4-45

38. Pereira V, Cruz F, Rocha M (2021) MEWpy: a computational strain optimization workbench in Python. Bioinformatics. https://doi.org/10.1093/bioinformatics/btab013

39. Sayers EW, Barrett T, Benson DA, et al (2009) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 38:5–16. https://doi.org/10.1093/nar/gkp967

40. Grigoriev I V., Nordberg H, Shabalov I, et al (2012) The Genome Portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res 40:D26–D32. https://doi.org/10.1093/NAR/GKR947

41. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res.

42. Karp PD, Billington R, Caspi R, et al (2019) The BioCyc collection of microbial genomes and metabolic pathways. 20:1085–1093. https://doi.org/10.1093/BIB/BBX085

43. Chang A, Jeske L, Ulbrich S, et al (2021) BRENDA, the ELIXIR core data resource in 2021: new developments and updates. Nucleic Acids Res 49:D498–D508. https://doi.org/10.1093/nar/gkaa1025

44. King ZA, Lu J, Dräger A, et al (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res 44:D515–D522. https://doi.org/10.1093/nar/gkv1049

45. Moretti S, Tran VDT, Mehl F, et al (2021) MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. Nucleic Acids Res 49:D570–D574. https://doi.org/10.1093/NAR/GKAA992

46. Saier MH, Reddy VS, Moreno-Hagelsieb G, et al (2021) The Transporter Classification Database (TCDB): 2021 update. Nucleic Acids Res 49:D461–D467. https://doi.org/10.1093/nar/gkaa1004

47. Guiry MD, Guiry GM, Morrison L, et al (2014) AlgaeBase: An On-line Resource for Algae. https://doi.org/107872/crya.v35.iss22014105 35:105–115. https://doi.org/10.7872/CRYA.V35.ISS2.2014.105

48. Kusber W-H (2007) AlgaTerra Information System: Types data and data types. In: Proceedings of the 1st Central European Diatom Meeting 2007. Botanic Garden and Botanical Museum Berlin-Dahlem, pp 97–100

49. Lefebvre PA, Laudon M, Silflow C (2019) The chlamydomonas resource center. Biol Resour Model Org 95–111. https://doi.org/10.1201/9781315100999-6/CHLAMYDOMONAS-RESOURCE-CENTER-PAUL-LEFEBVRE-MATTHEW-LAUDON-CAROLYN-SILFLOW

50. Hauer T, Komárek J (2021) CyanoDB 2.0 - On-line database of cyanobacterial genera. - World-wide electronic publication. Univ South Bohemia Inst Bot AS CR

51. Hauer T, Komárek J (2020) CyanoDB 2.0 - On-line database of cyanobacterial genera. In: World-wide Electron. Publ.

52. Gong Y, Kang NK, Kim YU, et al (2020) The NanDeSyn database for *Nannochloropsis* systems and synthetic biology. Plant J 104:1736–1745. https://doi.org/10.1111/TPJ.15025

53. Brůna T, Hoff KJ, Lomsadze A, et al (2021) BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics Bioinforma 3:. https://doi.org/10.1093/nargab/lqaa108

54. Campbell MS, Holt C, Moore B, Yandell M (2014) Genome Annotation and Curation Using MAKER and MAKER-P. Hoboken, NJ, USA

55. Savojardo C, Martelli PL, Fariselli P, et al (2018) BUSCA: an integrative web server to predict subcellular localization of proteins. Nucleic Acids Res 46:W459–W466. https://doi.org/10.1093/NAR/GKY320

56. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, et al (2017) DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics 33:3387–3395. https://doi.org/10.1093/bioinformatics/btx431

57. Goldberg T, Hecht M, Hamp T, et al (2014) LocTree3 prediction of localization. Nucleic Acids Res 42:W350. https://doi.org/10.1093/nar/gku396

58. Tardif M, Atteia A, Specht M, et al (2012) Predalgo: A new subcellular localization prediction tool dedicated to green algae. In: Molecular Biology and Evolution

59. Yu NY, Wagner JR, Laird MR, et al (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615. https://doi.org/10.1093/bioinformatics/btq249

60. Horton P, Park KJ, Obayashi T, et al (2007) WoLF PSORT: Protein localization predictor. Nucleic Acids Res. https://doi.org/10.1093/nar/gkm259

61. Beber ME, Gollub MG, Mozaffari D, et al (2022) eQuilibrator 3.0: a database solution for thermodynamic constant estimation. Nucleic Acids Res 50:D603–D609. https://doi.org/10.1093/NAR/GKAB1106

62. May P, Wienkoop S, Kempa S, et al (2008) Metabolomics- and Proteomics-Assisted Genome Annotation and Analysis of the Draft Metabolic Network of *Chlamydomonas reinhardtii*. Genetics 179:157–166. https://doi.org/10.1534/genetics.108.088336

63. Manichaikul A, Ghamsari L, Hom EFY, et al (2009) Metabolic network analysis integrated with transcript verification for sequenced genomes. Nat Methods 6:589–592. https://doi.org/10.1038/nmeth.1348

64. Cogne G, Rü M, Bockmayr A, et al (2011) A Model-Based Method for Investigating Bioenergetic Processes in Autotrophically Growing Eukaryotic Microalgae: Application to the Green Algae Chlamydomonas reinhardtii. Am Inst Chem Eng Biotechnol Prog 27:631–640. https://doi.org/10.1002/btpr.596

65. Gomes de Oliveira Dal'Molin C, Quek L-EE, Palfreyman RW, et al (2011) AlgaGEM - a genome-scale metabolic reconstruction of algae based on the Chlamydomonas reinhardtii genome. BMC Genomics 12:S5. https://doi.org/10.1186/1471-2164-12-S4-S5

66. Chaiboonchoe A, Dohai BS, Cai H, et al (2014) Microalgal metabolic network model refinement through high-throughput functional metabolic profiling. Front Bioeng Biotechnol 2:. https://doi.org/10.3389/fbioe.2014.00068

67. Boyle NR, Sengupta N, Morgan JA (2017) Metabolic flux analysis of heterotrophic growth in Chlamydomonas reinhardtii. PLoS One 12:e0177292. https://doi.org/10.1371/journal.pone.0177292

68. Mekanik M, Motamedian E, Fotovat R, Jafarian V (2019) Reconstruction of a genome-scale metabolic model for Auxenochlorella protothecoides to study hydrogen production under anaerobiosis using multiple optimal solutions. Int J Hydrogen Energy 44:2580–2591. https://doi.org/10.1016/j.ijhydene.2018.12.049

69. Shah AR, Ahmad A, Srivastava S, Jaffar Ali BMM (2017) Reconstruction and analysis of a genome-scale metabolic model of *Nannochloropsis gaditana*. Algal Res 26:354–364. https://doi.org/10.1016/j.algal.2017.08.014

70. Loira N, Mendoza S, Paz Cortés M, et al (2017) Reconstruction of the microalga *Nannochloropsis salina* genome-scale metabolic model with applications to lipid production. BMC Syst Biol 11:. https://doi.org/10.1186/s12918-017-0441-1

71. Krumholz EW, Yang H, Weisenhorn P, et al (2012) Genome-wide metabolic network reconstruction of the picoalga Ostreococcus. J Exp Bot 63:2353–2362. https://doi.org/10.1093/jxb/err407

72. Klanchui A, Dulsawat S, Chaloemngam K, et al (2018) An Improved Genome-Scale Metabolic Model of Arthrospira platensis C1 (iAK888) and Its Application in Glycogen Overproduction. Metabolites 8:84. https://doi.org/10.3390/metabo8040084

73. Knoop H, Zilliges Y, Lockau W, Steuer R (2010) The Metabolic Network of Synechocystis sp. PCC 6803: Systemic Properties of Autotrophic Growth 1[C][W]. Plant Physiol Ò 154:410–422. https://doi.org/10.1104/pp.110.157198

74. Nogales J, Gudmundsson S, Knight EMM, et al (2012) Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Natl Acad Sci 109:. https://doi.org/10.1073/pnas.1117907109

75. Knoop H, Gründel M, Zilliges Y, et al (2013) Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of Synechocystis sp. PCC 6803. PLoS Comput Biol 9:. https://doi.org/10.1371/journal.pcbi.1003081

76. Ahmad A, Pathania R, Srivastava S (2020) Biochemical Characteristics and a Genome-Scale Metabolic Model of an Indian Euryhaline Cyanobacterium with High Polyglucan Content. Metabolites 10:177. https://doi.org/10.3390/metabo10050177

77. Broddrick JT, Rubin BE, Welkie DG, et al (2016) Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. Proc Natl Acad Sci 113:E8344–E8353. https://doi.org/10.1073/pnas.1613446113

78. Malatinszky D, Steuer R, Jones PR (2017) A Comprehensively Curated Genome-Scale Two-Cell Model for the Heterocystous Cyanobacterium Anabaena sp. PCC 7120. Plant Physiol 173:509–523. https://doi.org/10.1104/pp.16.01487

79. Zuñiga C, Li C-T, Huelsman T, et al (2016) Genome-Scale Metabolic Model for the Green Alga Chlorella vulgaris UTEX 395 Accurately Predicts Phenotypes under Autotrophic, Heterotrophic, and Mixotrophic Growth Conditions. Plant Physiol Ò 172:589–602. https://doi.org/10.1104/pp.16.00593

80. Zuñiga C, Levering J, Antoniewicz MR, et al (2018) Predicting Dynamic Metabolic Demands in the Photosynthetic Eukaryote Chlorella vulgaris 1. Plant Physiol Ò 176:450–462. https://doi.org/10.1104/pp.17.00605

81. Levering J, Broddrick J, Dupont CL, et al (2016) Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. PLoS One 11:. https://doi.org/10.1371/journal.pone.0155038

82. Klanchui A, Khannapho C, Phodee A, et al (2012) IAK692: A genome-scale metabolic model of Spirulina platensis C1. BMC Syst Biol 6:. https://doi.org/10.1186/1752-0509-6-71

83. Montagud A, Navarro E, Fernández de Córdoba P, et al (2010) Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. BMC Syst Biol 4:156. https://doi.org/10.1186/1752-0509-4-156

84. Montagud A, Zelezniak A, Navarro E, et al (2011) Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium Synechocystis sp. PCC6803. Biotechnol J 6:330–342. https://doi.org/10.1002/biot.201000109

85. Triana J, Montagud A, Siurana M, et al (2014) Generation and Evaluation of a Genome-Scale Metabolic Network Model of Synechococcus elongatus PCC7942. Metabolites 4:680–698. https://doi.org/10.3390/metabo4030680

86. Juneja A, Chaplen FWR, Murthy GS (2016) Genome scale metabolic reconstruction of Chlorella variabilis for exploring its metabolic potential for biofuels. Bioresour Technol 213:103–110. https://doi.org/10.1016/j.biortech.2016.02.118

87. Qian X, Kim MK, Kumaraswamy GK, et al (2017) Flux balance analysis of photoautotrophic metabolism: Uncovering new biological details of subsystems involved in cyanobacterial photosynthesis. Biochim Biophys Acta - Bioenerg

1858:276–287. https://doi.org/10.1016/j.bbabio.2016.12.007

88. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol 7:74. https://doi.org/10.1186/1752-0509-7-74

89. Boyle NR, Morgan JA (2009) Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii. BMC Syst Biol 3:. https://doi.org/10.1186/1752-0509-3-4

90. Schuster S, Dandekar T, Fell DA (1999) Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering. Trends Biotechnol. 17:53–60

91. Chang RL, Ghamsari L, Manichaikul A, et al (2011) Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. Mol Syst Biol 7:518. https://doi.org/10.1038/msb.2011.52

92. Rügen M, Bockmayr A, Legrand J, Cogne G (2012) Network reduction in metabolic pathway analysis: Elucidation of the key pathways involved in the photoautotrophic growth of the green alga Chlamydomonas reinhardtii. Metab Eng 14:458–467. https://doi.org/10.1016/j.ymben.2012.01.009

93. Imam S, Schäuble S, Valenzuela J, et al (2015) A refined genome-scale reconstruction of *Chlamydomonas* metabolism provides a platform for systems-level analyses. Plant J 84:1239–1256. https://doi.org/10.1111/tpj.13059

94. Fernando González Barrios A, Alejandra Mora Salguero D, Fernández-Niño M, et al (2018) Development of a Chlamydomonas reinhardtii metabolic network dynamic model to describe distinct phenotypes occurring at different CO2 levels. PeerJ 2018:e5528. https://doi.org/10.7717/peerj.5528

95. Muthuraj M, Palabhanvi B, Misra S, et al (2013) Flux balance analysis of Chlorella sp. FC2 IITG under photoautotrophic and heterotrophic growth conditions. Photosynth Res 118:167–179. https://doi.org/10.1007/s11120-013-9943-x

96. Pham N, Hohmann-Marriott MF (2016) Genome-scale constraint-based metabolic modeling and analysis of *Nannochloropsis Sp*

97. Yang C, Hua Q, Shimizu K (2000) Energetics and carbon metabolism during growth of microalgal cells under photoautotrophic, mixotrophic and cyclic light-autotrophic/dark-heterotrophic conditions. Biochem Eng J 6:87–102. https://doi.org/10.1016/S1369-703X(00)00080-2

98. Yoshikawa K, Aikawa S, Kojima Y, et al (2015) Construction of a Genome-Scale Metabolic Model of Arthrospira platensis NIES-39 and Metabolic Design for Cyanobacterial Bioproduction. PLoS One 10:e0144430. https://doi.org/10.1371/journal.pone.0144430

99. Navarro E, Montagud A, Fernández de Córdoba P, Urchueguía JF (2009) Metabolic flux analysis of the hydrogen production potential in Synechocystis sp. PCC6803. Int J Hydrogen Energy 34:8828–8838. https://doi.org/10.1016/j.ijhydene.2009.08.036

100. Yoshikawa K, Kojima Y, Nakajima T, et al (2011) Reconstruction and verification of a genome-scale metabolic model for Synechocystis sp. PCC6803. Appl Microbiol Biotechnol 92:347–358. https://doi.org/10.1007/s00253-011-3559-x

101. Joshi CJ, Peebles CAM, Prasad A (2017) Modeling and analysis of flux distribution and bioproduct formation in Synechocystis sp. PCC 6803 using a new genome-scale metabolic reconstruction. Algal Res 27:295–310. https://doi.org/10.1016/J.ALGAL.2017.09.013

102. Vu TT, Stolyar SM, Pinchuk GE, et al (2012) Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium Cyanothece sp. ATCC 51142. PLoS Comput Biol 8:. https://doi.org/10.1371/journal.pcbi.1002460

103. Hamilton JJ, Reed JL (2012) Identification of functional differences in metabolic networks using comparative genomics and constraint-based models. PLoS One 7:. https://doi.org/10.1371/journal.pone.0034670

104. Hendry JI, Prasannan CB, Joshi A, et al (2016) Metabolic model of Synechococcus sp. PCC 7002: Prediction of flux distribution and network modification for enhanced biofuel production. Bioresour Technol 213:190–197. https://doi.org/10.1016/j.biortech.2016.03.128

105. Malek Shahkouhi A, Motamedian E (2020) Reconstruction of a regulated two-cell metabolic model to study biohydrogen production in a diazotrophic cyanobacterium Anabaena variabilis ATCC 29413. PLoS One 15:e0227977. https://doi.org/10.1371/journal.pone.0227977

106. Barberis E, Marengo E, Manfredi M (2021) Protein Subcellular Localization Prediction. Methods Mol Biol 2361:197–212. https://doi.org/10.1007/978-1-0716-1641-3_12/COVER

107. Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc 5:93–121. https://doi.org/10.1038/nprot.2009.203

108. Hoff KJ, Stanke M (2015) Current methods for automated annotation of protein-coding genes. Curr. Opin. Insect Sci. 7:8–14

109. Keller O, Kollmar M, Stanke M, Waack S (2011) A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics 27:757–763. https://doi.org/10.1093/bioinformatics/btr010

110. Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5:. https://doi.org/10.1186/1471-2105-5-59

111. Brůna T, Lomsadze A, Borodovsky M (2020) GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genomics Bioinforma 2:. https://doi.org/10.1093/nargab/lqaa026

112. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111. https://doi.org/10.1093/BIOINFORMATICS/BTP120

113. Dobin A, Davis CA, Schlesinger F, et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

114. Trapnell C, Williams BA, Pertea G, et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010 285 28:511–515. https://doi.org/10.1038/nbt.1621

115. Pertea M, Pertea GM, Antonescu CM, et al (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33:290–295. https://doi.org/10.1038/nbt.3122

116. Chan PP, Lowe TM (2019) tRNAscan-SE: Searching for tRNA genes in genomic sequences. In: Methods in Molecular Biology

117. Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

118. Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763. https://doi.org/10.1093/bioinformatics/14.9.755

119. Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12:59–60

120. Lagoa D, Liu F, Faria JP, et al (2021) TranSyT, an innovative framework for identifying transport systems. bioRxiv doi 101101/20210429441738. https://doi.org/10.1101/2021.04.29.441738

121. Falda M, Toppo S, Pescarolo A, et al (2012) Argot2: A large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. BMC Bioinformatics 13:1–9. https://doi.org/10.1186/1471-2105-13-S4-S14

122. Cozzetto D, Minneci F, Currant H, Jones DT (2016) FFPred 3: feature-based function prediction for all Gene Ontology domains. Sci Rep 6:31865. https://doi.org/10.1038/srep31865

123. Armenteros JJA, Salvatore M, Emanuelsson O, et al (2019) Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance 2:. https://doi.org/10.26508/LSA.201900429

124. Savojardo C, Bruciaferri N, Tartari G, et al (2020) DeepMito: accurate prediction of protein sub-mitochondrial localization using convolutional neural networks. Bioinformatics 36:56–64. https://doi.org/10.1093/bioinformatics/btz512

125. Shastri AA, Morgan JAJA (2005) Flux Balance Analysis of Photoautotrophic Metabolism. Wiley Online Libr 21:1617–1626. https://doi.org/10.1021/bp050246d

126. Finkel Z V., Follows MJ, Liefer JD, et al (2016) Phylogenetic Diversity in the Macromolecular Composition of Microalgae. PLoS One 11:e0155977. https://doi.org/10.1371/journal.pone.0155977

127. Santos S, Rocha I (2016) Estimation of biomass composition from genomic and

transcriptomic information. J Integr Bioinform 13:1–14. https://doi.org/10.1515/jib-2016-285

128. Sánchez BJ, Li F, Kerkhoven EJ, Nielsen J (2019) SLIMEr: Probing flexibility of lipid metabolism in yeast with an improved constraint-based modeling framework. BMC Syst Biol 13:4. https://doi.org/10.1186/s12918-018-0673-8

129. He L, Wu SG, Zhang M, et al (2016) WUFlux: an open-source platform for 13C metabolic flux analysis of bacterial metabolism. BMC Bioinformatics 17:444. https://doi.org/10.1186/s12859-016-1314-0

130. Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl Environ Microbiol 60:3724–3731. https://doi.org/10.1128/AEM.60.10.3724-3731.1994

131. Saha R, Liu D, Hoynes-O'Connor A, et al (2016) Diurnal regulation of cellular processes in the cyanobacterium Synechocystis sp. strain PCC 6803: Insights from transcriptomic, fluxomic, and physiological analyses. MBio. https://doi.org/10.1128/mBio.00464-16

132. Mahadevan R, Edwards JS, Doyle FJ (2002) Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. Biophys J 83:1331–1340. https://doi.org/10.1016/S0006-3495(02)73903-9

133. Flassig RJ, Fachet M, Höffner K, et al (2016) Dynamic flux balance modeling to increase the production of high-value compounds in green microalgae. Biotechnol Biofuels 9:165. https://doi.org/10.1186/s13068-016-0556-4

134. Baroukh C, Steyer JP, Bernard O, Chachuat B (2016) dynamic Flux Balance Analysis of the Metabolism of Microalgae under a Diurnal Light Cycle. IFAC-PapersOnLine 49:791–796. https://doi.org/10.1016/j.ifacol.2016.07.285

135. Pharkya P (2004) OptStrain: A computational framework for redesign of microbial production systems. Genome Res 14:2367–2376. https://doi.org/10.1101/gr.2872004

136. Rocha I, Maia P, Rocha M, Ferreira EC (2008) OptGene-a framework for in silico metabolic engineering. In: 10th International Conference on Chemical and Biological Engineering. Braga, pp 218–219

137. Pharkya P, Maranas CD (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng 8:1–13. https://doi.org/10.1016/j.ymben.2005.08.003

138. Ranganathan S, Suthers PF, Maranas CD (2010) OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol 6:. https://doi.org/10.1371/journal.pcbi.1000744

139. Erdrich P, Knoop H, Steuer R, Klamt S (2014) Cyanobacterial biofuels: New insights and strain design strategies revealed by computational modeling.

Microb Cell Fact 13:. https://doi.org/10.1186/s12934-014-0128-x

140. Nogales J, Gudmundsson S, Thiele I (2013) Toward systems metabolic engineering in cyanobacteria. Proc Natl Acad Sci U S A 4:22308420. https://doi.org/10.4161/bioe.22792

141. Vu TT, Hill EA, Kucek LA, et al (2013) Computational evaluation of Synechococcus sp. PCC 7002 metabolism for chemical production. Biotechnol J 8:619–630. https://doi.org/10.1002/biot.201200315

142. Baroukh C, Muñoz-Tamayo R, Steyer J-PP, Bernard O (2014) DRUM: A New Framework for Metabolic Modeling under Non-Balanced Growth. Application to the Carbon Metabolism of Unicellular Microalgae. PLoS One 9:e104499. https://doi.org/10.1371/journal.pone.0104499

143. Kim Y, Kim GB, Lee SY (2021) Machine learning applications in genome-scale metabolic modeling. Curr. Opin. Syst. Biol. 25:42–49

144. Zampieri G, Vijayakumar S, Yaneske E, Angione C (2019) Machine and deep learning meet genome-scale metabolic modeling. PLoS Comput. Biol. 15:e1007084

145. Vijayakumar S, Rahman PKSM, Angione C (2020) A Hybrid Flux Balance Analysis and Machine Learning Pipeline Elucidates Metabolic Adaptation in Cyanobacteria. iScience 23:101818. https://doi.org/10.1016/j.isci.2020.101818

146. Saini DK, Rai A, Devi A, et al (2021) A multi-objective hybrid machine learning approach-based optimization for enhanced biomass and bioactive phycobiliproteins production in Nostoc sp. CCC-403. Bioresour Technol 329:124908. https://doi.org/10.1016/j.biortech.2021.124908

147. Sánchez BJ, Zhang C, Nilsson A, et al (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol Syst Biol 13:935. https://doi.org/10.15252/MSB.20167411

148. O'Brien EJ, Lerman JA, Chang RL, et al (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol Syst Biol 9:693. https://doi.org/10.1038/MSB.2013.52

149. Ye C, Luo Q, Guo L, et al (2020) Improving lysine production through construction of an Escherichia coli enzyme-constrained model. Biotechnol Bioeng 117:3533–3544. https://doi.org/10.1002/BIT.27485

150. Li F, Yuan L, Lu H, et al (2022) Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. Nat Catal 2022 58 5:662–672. https://doi.org/10.1038/s41929-022-00798-z

151. Norena-Caro DA, Zuniga C, Pete AJ, et al (2021) Analysis of the cyanobacterial amino acid metabolism with a precise genome-scale metabolic reconstruction of Anabaena sp. UTEX 2576. Biochem Eng J 171:108008. https://doi.org/10.1016/j.bej.2021.108008

152. Kliphuis AMJJ, Klok AJ, Martens DE, et al (2012) Metabolic modeling of Chlamydomonas reinhardtii: Energy requirements for photoautotrophic growth and maintenance. J Appl Phycol 24:253–266. https://doi.org/10.1007/s10811-011-9674-3

153. Saha R, Verseput AT, Berla BM, et al (2012) Reconstruction and Comparison of the Metabolic Potential of Cyanobacteria Cyanothece sp. ATCC 51142 and Synechocystis sp. PCC 6803. PLoS One 7:. https://doi.org/10.1371/journal.pone.0048285

154. Kroth PG, Chiovitti A, Gruber A, et al (2008) A model for carbohydrate metabolism in the diatom Phaeodactylum tricornutum deduced from comparative whole genome analysis. PLoS One 3:. https://doi.org/10.1371/journal.pone.0001426

155. Hunt KA, Folsom JP, Taffs RL, et al (2014) Complete enumeration of elementary flux modes through scalable demand-based subnetwork definition. 30:1569–1578. https://doi.org/10.1093/bioinformatics/btu021

156. Singh D, Carlson R, Fell D, Poolman M (2015) Modelling metabolism of the diatom Phaeodactylum tricornutum. Biochem Soc Trans 43:1182–1186. https://doi.org/10.1042/BST20150152

157. Yang C (2002) Metabolic Flux Analysis in Synechocystis Using Isotope Distribution from 13C-Labeled Glucose. Metab Eng 4:202–216. https://doi.org/10.1006/mben.2002.0226

158. Sarkar D, Mueller TJ, Liu D, et al (2019) A diurnal flux balance model of Synechocystis sp. PCC 6803 metabolism. PLOS Comput Biol 15:e1006692. https://doi.org/10.1371/JOURNAL.PCBI.1006692

159. Japhalekar K, Srinivasan S, Viswanathan G, Venkatesh KV (2022) Flux balance analysis for overproduction of organic acids by Synechocystis sp. PCC 6803 under dark anoxic condition. Biochem Eng J 178:108297. https://doi.org/10.1016/J.BEJ.2021.108297

160. Gardner JJ, Boyle NR (2017) The use of genome-scale metabolic network reconstruction to predict fluxes and equilibrium composition of N-fixing versus C-fixing cells in a diazotrophic cyanobacterium, Trichodesmium erythraeum. BMC Syst Biol 11:. https://doi.org/10.1186/s12918-016-0383-z

Highlights

-Microalgae are eco-friendly and alternative sources of several high-value compounds

-Microalgae biomass and compounds' productivity are low at the industrial level

-Metabolic models provide valuable insights into the metabolism of an organism

-Metabolic models can be used to optimise the growth conditions of microalgae

-Optimising the culture conditions will allow improving the cost-effectiveness