Development of methane mitigation strategies using the microalga *Chlorella vulgaris*

Gonçalo Apolinário, Emanuel Cunha, Pedro Geada, and Oscar Dias

CEB-Centre of Biological Engeneering, University of Minho, Braga, Portugal
pg49834@alunos.uminho.pt

Abstract. *Chlorella vulgaris* is a green microalga with great biotechnological potential in industries such as food, cosmetics and pharmaceuticals. More recently, it gathered the attention of this investigation group regarding its capability of possibly mitigating methane (CH_4). Developing a genome-scale metabolic (GSM) model for this organism can serve as a valuable tool to assess the extent to which *Chlorella vulgaris* can contribute to methane mitigation and addressing the broader challenge of global warming. Given the challenges associated with reconstructing a GSM model, the utilization of *merlin* will provide valuable assistance throughout the entire process. The genomic information of the assembly of strain 221/11P was incorporated into *merlin*. Enzymes annotation was carried out using BLAST and transporters were incorporated into the model. Additionally, the biomass composition was determined. Subsequently, the manual curation process started, focusing on gap-filling reactions within the amino acid and nucleotide metabolisms. The model is so far consists of 1,636 genes, 792 enzymes, 2,682 reactions and 2,340 metabolites. This information is subject to change as the model get refined.

Keywords: *Chlorella vulgaris*· Methane · Systems Biology · Genome-scale methabolic model· *Merlin* · Bio-based mitigation strategy.

1 Introduction

1.1 Microalgae

The first evidence of the existence of microalgae dates back to 3.4 billion years ago and, even though their structure has persisted over time, only now are we beginning to fully understand their mechanisms. Microalgae are single or multicellular microorganisms that can be classified as prokaryotic or eukaryotic, encompassing over 73,000 identified species. In light of this diversity, microalgae can grow in almost every environmental condition, from aquatic to terrestrial surroundings, and it is estimated the existence of hundreds of thousands to several million extant species [1].

Over the last few years, the interest in these microorganisms has been growing due to their biotechnological potential in a wide range of applications. This sudden attention arose because these cells are capable of synthesizing several

metabolites of interest such as lipids, carbohydrates, proteins, pigments, vitamins, and minerals through their complex pathways. To name but a few, cosmetics, biofuel production, pharmaceuticals, and food industries are some examples of these organisms usefulness in today's society. In addition, it is known that they can be grown sustainably since they don't require potable water nor arable land, and are photosynthetic beings capable of absorbing massive amounts of carbon dioxide (CO_2)[2].

In general, microalgae can be classified as autotrophic, mixotrophic and heterotrophic depending on how they manage to fulfill their material and energy requirements. Autotrophic cultures are the most commonly used for mass cultivation and use solar radiation and CO_2 to produce energy for growth. On the contrary, heterotrophic microalgae use organic compounds produced by other organisms for growth like glucose. With respect to mixotrophic microalgae, they use both organic compounds and CO_2. Regarding nutritional needs, microalgae require various elements for their growth such as carbon, oxygen, hydrogen, nitrogen, potassium, calcium, magnesium, iron, and trace elements like copper, manganese, and zinc [3][4].

1.2 Methane

The problematic of greenhouse gas emissions is a global environmental challenge that has been ever-present in our lives in the past few decades. The increase in atmospheric concentrations of CO_2, methane, and other damaging gases can lead to catastrophic repercussions to life as we know it. The primary sources of these emissions include burning fossil fuels, industrial processes, and agriculture. Reducing greenhouse gas emissions and fomenting strategies for their mitigation are crucial steps that need to be made to meet the Paris Agreement and ultimately ensure a sustainable future for our planet and those that inhabit it [5].

There is extensive research on the possibility of using microalgae - including *Chlorella vulgaris*, the focus of this study - as an alternative to other photosynthetic organisms, such as plants, in the mitigation of carbon dioxide. It is thought that plants only contribute between 3-6% to fossil fuel emissions mitigation. Combining their faster growth rates with their higher photosynthetic activity, microalgae represent an alternative and sustainable approach to mitigating CO_2 [6].

Recently, the more overlooked greenhouse gas, methane, seized the interest of this investigation group regarding its possible mitigation by *Chlorella vulgaris* similarly to what happens with CO_2. Although methane only contributes to roughly 20% of the total anthropogenic greenhouse emissions, compared to CO_2's alarming 74%, due to its structure, methane traps more heat in the atmosphere per molecule than CO2, meaning that it has 80 times more warming power. This highlights the crucial need to devise strategies to reduce its impact [7].

Currently, the strategies involved in the mitigation of methane are too broad and difficult to implement like adjusting dietary composition for livestock, reducing waste that ends up in landfills, improving the detection of methane leaks

in oil and gas facilities, or resorting to anaerobic digestion. It is in this context that microalgae could turn the tables on the common endeavour of preventing global warming [8].

Early experiments indicated that *Chlorella vulgaris* might be able to directly mitigate a great amount of methane, even though further investigation is necessary.

2 Objective

The long-term objective of the work is to evaluate the viability of the microalga *Chlorella vulgaris* as an ecological and cost-effective solution for mitigating methane emissions by employing a range of bioinformatics tools. However, the main goal of this stage is to develop a draft genome-scale metabolic network for *Chlorella vulgaris* 211/11B using *merlin*. The model will be used later to conduct *in silico* simulations and optimizations aiming at modelling the CH_4 metabolization by the microalga.

3 State of the art

3.1 Traditional approach

The cultivation of microalgae has evolved over the years with the development of new and optimized growth media. The selection of the appropriate medium depends on the type of microalgae and the intended purpose of cultivation, whether it is for obtaining the maximum biomass, enhancing the production of some metabolite, or some other objective. Finally, the medium composition chosen will naturally also affect the metabolism adopted by these microorganisms [3] .

There have been many attempts to improve the production and optimization of this high-value asset but still, the high production costs and the poor productivity indexes are limiting factors from an industrial perspective. The traditional way of boosting the yield of certain metabolites is based on a series of stress conditions, *id est*, according to the nutrient-stress conditions microalgae adapt their *modus operandi* to produce different concentration levels of compounds. This particular strategy has the downside of usually undermining the total biomass, as a result of a low growth rate [9].

Notwithstanding the fact that these strategies can enhance both the total biomass production and the concentration of desired metabolites, it is nonetheless a time-consuming and overall resource-demanding strategy that is unable to keep up with current requirements.

3.2 Systems Biology

Over the last 20 years, metabolic models have been widely used as a source of information for metabolic engineering, drug targeting, metabolic pathway analysis, and process optimization. Currently, there are two different approaches in

this area: stoichiometric and dynamic modeling. While stoichiometric models only use structural information (*e.g.*, metabolite's stoichiometries, reaction's reversibility), dynamic models account for the variation of the concentrations over time, requiring information regarding kinetic parameters [10]. Considering the scarcity of kinetic and regulatory data needed for assembling a dynamic model as well as its inherent complexity and time constraints, stoichiometric models are often used at the genome-scale.

GSM models allow the *in silico* simulation and prediction of metabolic fluxes on a large scale, providing a powerful tool for optimizing and designing metabolic engineering methods. By integrating high-throughput data with genome-scale models, a comprehensive understanding of cellular metabolism and identification of strategies to improve a determined objective function can be obtained. The significance of this emerging technology for the industry lies in the faster and cost-effectiveness of this approach when compared with the traditional one [11] .

The first GSM model, regarding the bacteria *Haemophilus influenza*, was published in 1999 [12]. Since then, genome sequencing techniques have undergone significant advancements, resulting in a remarkable reduction in sequencing costs. This unprecedented amount of sequenced genomes is allowing the reconstruction of GSM models for numerous organisms, from prokaryotes to eukaryotes. A GSM model computationally describes a whole set of stoichiometry-based, mass-balanced metabolic reactions in an organism using gene-protein-reaction (GPR) associations whose formulations are based on genome annotation data and experimentally obtained information. Since the very first model was released, the number of GSM models has grown exponentially, and in 2021 there were 45 published metabolic reconstructions of microalgae [13].

3.3 Previously published models

In 2016, Cristal Zuñiga *et al.* [14] published the first and only GSM model for the microalga *C. vulgaris* to date. The paper describes the reconstruction of a model for the strain UTEX-395 with the aid of the software RAVEN Toolbox [15]. The model was validated against experimental data and is able to accurately predict phenotypes under autotrophic, heterotrophic, and mixotrophic conditions. Nevertheless, this model was developed seven years ago with a different genome sequence, thus it probably misses new information that was published since. Moreover, this model does not include metabolic pathways for the metabolization of methane, making it not suitable for the objective of this work.

In addition to this, the reconstruction process of a GSM model for *C. vulgaris* can benefit from valuable insights derived from other existing models. Models developed for the reference organism *C. reinhardtii* [16] and the closely related *C. variabilis* [17] can be used to establish comparisons and extract relevant information for the reconstruction.

3.4 Genome-Scale Metabolic Models reconstruction

The process of reconstructing GSM models is a well-documented procedure that typically involves four main steps: genome annotation, assembly of a draft network, conversion into a stoichiometric model, and model validation (figure 1). The reconstruction process is iterative, as demonstrated by the metabolic network of *Escherichia coli*, whose reconstruction has been updated over the last 19 years. Building a metabolic network is a time-consuming process that involves a great deal of manual curation. This happens in order to attain a model increasingly accurate and trustworthy, eventually resulting in a high-quality model.

The main steps of the reconstruction are well described in literature, and involve the genome annotation, draft network assembly, conversion into a stoichiometric model, and model validation [18, 10].

Genome annotation The first step involves obtaining the genome sequence of the organism of interest and its structural and functional annotation. Considering that the model draft will be based on this particular step, it is of vital importance that the annotation is up-to-date and reliable. Structural annotation consists in defining all the features present in the genome such as genes, coding sequences, open reading frames and so on. Functional annotation, on the other hand, consists of associating every gene with a function. For these models, it is of particular relevance to obtain the enzyme commission (EC) numbers of genes encoding enzymes and the identifiers for membrane transporters on account that the network will be based on reactions catalyzed by such genes.

Draft network assembly In this step, the information acquired from genome annotation is used to assemble a draft metabolic network. The information pertaining to genes encoding enzymes and transporters is not always present or correct which demands the completion of the set of reactions (gap-filling) by revising biochemistry literature and through homology search methods. Another way of reviewing functional annotation is by making use of reference organisms like *C. reinhardtii* (reference organism for microalgae genome functional annotation). While about 90% of all reactions are derived by the annotated genome relatively quickly, typically accounting for 10% of the reconstruction time, whereas the remaining 10% of reactions require 90% of the time. Moreover, some cases that do not go by the rule of one gene-one protein-one reaction need to be taken into account, which is the case of enzymatic complexes, isoenzymes, and promiscuous enzymes. The main difference between GSM models from prokaryotic and eukaryotic organisms lies in compartmentalization, essential for discerning the localization of metabolites and their exchanges across different compartments. During this step, constraints are imposed on the network by limiting the reversibility and direction of reactions.

Conversion into a stoichiometric model Once the draft network is assembled, it needs to be converted into a stoichiometric model. When building a

metabolic model, it is essential to incorporate reactions that describe the formation of the organism's biomass. In microalgae, the most representative macromolecules are proteins, lipids and carbohydrates. Cofactors and vitamins are also incorporated in the biomass reaction even though their presence is not significant. This information is mainly obtained in literature, through experimental data or by utilizing bioinformatics tools. Furthermore, adenosine triphosphate (ATP) consumption for maintenance purposes has to be considered (ATP molecules needed per gram of biomass synthesized).

Model validation The final step in building a GSM model is its validation. To do so we can resort to various tests. The most used one is flux balance analysis(FBA) [19]. FBA is based on the concept that, for all intracellular metabolites, the fluxes of reactions producing a given metabolite are balanced with the fluxes of reactions consuming metabolite, therefore, there is no net accumulation of metabolites. In FBA, the metabolic network is characterized as a set of linear equations:

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

where Z stands for the objective function, S represents the stoichiometric matrix, and V is the flux vector. So, for a given objective function, FBA provides a single best solution that is not exclusive since there can be a number of optimal flux distributions.

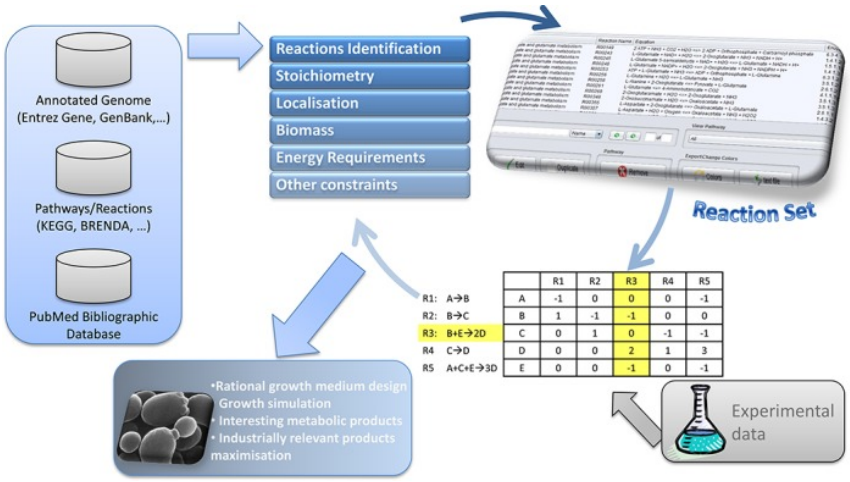

Fig. 1: Ilustration of the process of reconstruction of a GSM model. Adapted from [13].

Usually, the reconstruction of these high-quality models can take over a year due to their dependability on countless bioinformatics tools to perform the different stages of the process. Besides, many of these steps require manual curation and data extrapolated from literature. Hence, the need to automate some of the processes. There are a number of software to aid in the various stages of the process. Tools like ModelSEED [20], AuReMe [21], Pathway Tools [22], and CarveMe [23] have been around for some time and can be used resourcefully for the reconstruction of these models .

Nonetheless, this project will focus on the tool Metabolic Models Reconstruction Using Genome-Scale Information (*merlin*) to elaborate a GSM model for *C. vulgaris*.

3.5 *merlin*

Merlin is a user-friendly Java application for reconstructing genome-scale metabolic models of both eukaryotic and prokaryotic organisms with sequenced genomes. It provides a graphical interface throughout the whole process and includes tools for performing all steps of the reconstruction. *Merlin* features four independent modules: the Load internal database, the Enzymes annotation, the Transporters annotation, and the Compartments prediction. Lastly, *merlin* accelerates the conversion process of genome-scale data to SBML metabolic models, providing an efficient way to obtain an initial view of the biochemical network [24, 25].

4 Results and Discussion

To initiate the reconstruction process, it was required to upload the full genome of *C. vulgaris* into *merlin*. This was accomplished by directly uploading the assembly file from NCBI (National Center for Biotechnology Information) [26], utilizing the respective Taxonomy ID (3077). The chosen assembly, associated with strain 211/11P, was selected once it was the only entry that contained the essential files for the reconstruction.

4.1 Enzymes Annotation

Enzyme annotation was conducted using a homology search employing Basic local alignment search tool (BLAST) [27]. The homology search was performed twice, first against UniProtKB/SwissProt (reviewed) with 6,086 matches, and then against UniProtKB/TrEMBL (unreviewed) with 5,055 matches.

Table 1 illustrates the structure of the implemented automated workflow. The choice of species and genus for inclusion in the workflow was determined by two criteria: similarity to *C. vulgaris* (codes A, B and C), and in the cases of *A. thaliana* and *O. japonica* due to the abundance of information available about these organisms in online databases.

Table 1: Design of the Automatic Workflow

Automatic Workflow		
Category	Species	Code
Species	Chlorella vulgaris	A
Genus	Chlorella	B
Species	Chlamydomonas reinhardtii	C
Species	Arabidopsis thaliana	D
Species	Oryza sativa subsp. japonica	E

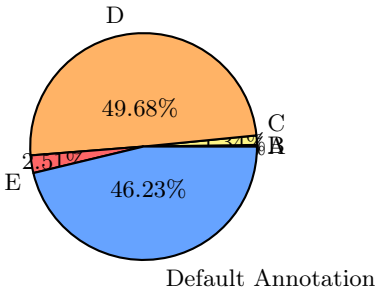

Fig. 2: Confidence levels distribution

Figure 2 illustrates that a significant portion of the annotations obtained through the automatic workflow pertains to *Arabidopsis thaliana* (49.68%). This is likely due to the extensive research and abundance of information available for this well-studied organism in the databases.

Another big part of the annotation (46.23%) corresponds to default annotation made by *merlin* that happens when no matches are found with any of the species/genus in the workflow. Only one entry was associated with the genus *Chlorella* and seven for the species *C. vulgaris* showing that there is few curated information for this organism and its genus.

4.2 Transport Reactions

Merlin comes incorporated with the tool TranSyt which automatically retrieves and processes information from TCDB (Transporter Classification Database) and incorporates reactions in the model. In total, TranSyt added 468 transport reactions belonging to four distinct transport mechanisms as detailed in table 2.

Table 2: Transport mechanisms automatically provided by the tool TranSyt

Transporters classification	
Class	Percentage (%)
Electrochemical Potential-driven	41.3
Primary Active	46.2
Channels	8.7
Group Translocators	3.8

As table 2 shows, the majority of transport mechanisms present in the model belong to electrochemical potential-driven and primary active with 41.3% and 46.2%, respectively.

In addition to the reactions automatically incorporated by TranSyt, four reactions needed to be manually added. These manually added reactions were essential to account for the missing compounds required for the composition of the medium. By including the reactions for H_2O, CO_2, photon, and oxygen, the total number of transporters reactions in the model comes up to 472.

4.3 Biomass Composition

The biomass composition was determined by recurring to a wide range of published articles containing the experimental data pertaining to each e-metabolite. The values for each e-metabolite are detailed in the table below.

Table 3: Biomass Composition

Biomass Composition			
e-Metabolite	*C. vulgaris* 211/11P	*C. reinhardtii*	*C. vulgaris UTEX-395*
e-Protein	0.457	0.460	0.480
e-DNA	0.001	0.001	0.006
e-RNA	0.032	0.087	0.080
e-Lipids	0.180	0.126	0.163
e-Carbohydrates	0.274	0.295	0.248
e-Pigments	0.045	0.027	0.026
e-Cofactors	0.011	0.002	-
Reference	[28–32]	[33]	[14]

Table 3 presents the biomass composition in gMM/gDW, which was obtained by calculating the average and then normalizing to one the values gathered from experimental data (Supplementary Table 2). The composition is compared between this model, a model of *C. reinhardtii*, and the previously reconstructed model of *C. vulgaris*. Upon analysis, we can see that the values for each metabolite are more or less consistent across the three models. The same procedure was done to determine the composition of each metabolite constituting every e-metabolite (Supplementary Tables 3-5).

4.4 Manual Curation

KEGG metabolic data was uploaded into the model, along with the integration of enzymes annotation, with 1,636 encoding genes. This resulted in 121 reactions provenient from KEGG bringing the total number of reactions in the model to 2,682. Conventionally, all reactions from KEGG are reversible so *merlin* fix this by using templates from KBase [34]. Additionally, unbalanced reactions were found using Biocyc [35] and BRENDA [36] which retrieve information on metabolites formula and reaction stoichiometry.

Despite completing all these steps, there were still missing some reactions in the network (gaps). As a result, it was crucial to manually curate every pathway. *C. variabilis* and *C. reinhardtii* were used as reference organisms for this purpose. So far, gap-filling was performed for the nucleotide and amino acid metabolisms, and additional reactions were added to the following pathways.

Table 4: Nº of added reactions in the process of gap-filling

Pathways	**Nº of reactions added**
Purine metabolism	16
Pyrimidine Metabolism	5
Cysteine and methionine metabolism	8
Valine, leucine and isoleucine degradation	5
Histidine metabolism	2
Tyrosine metabolism	5
Phenylalanine metabolism	2
Phenylalanine, tyrosine, and tryptophan biosynthesis	1

Out of the 15 pathways that were manually curated, Table 4 highlights the pathways where additional reactions were incorporated, amounting to a total of 45 added reactions.

Since neither of the reference organisms, *C. variabilis* and *C. reinhardtii*, includes the methane pathway in KEGG, it is crucial to investigate and determine how this pathway specifically functions in *C. vulgaris*. The methane pathway in the model is substantially incomplete as stated in Supplementary Figure 2. The goal is to evaluate the active enzymes and the specific reactions that characterize the methane metabolism in this microalga to determine how methane is consumed. This will be conducted by assessing relevant literature pertaining to this topic.

The process of reconstructing a GSM model for *C. vulgaris* is still at a preliminary stage with a lot of curation and additional analysis to be performed. Table 5, provides an overview of the current state of the model, along with the finished models of *C. vulgaris* UTEX-395 and *C. reinhardtii*.

So far, the model includes 1,636 enzyme-encoding genes, which accounts for approximately 15% of the genome. Out of the 2682 reactions, 970 are reversible and the remaining 1,712 are irreversible. This substantial number of components is expected to significantly enhance our understanding of the metabolism of *C. vulgaris*, especially when compared to the previous model. The integration of compartments into the model will only be performed when the gap-filling process is over.

Table 5: Components of the models.

Overview of model components			
Model component	*C. vulgaris 211/11P*	*C. reinhardtii*	*C. vulgaris UTEX-395*
Genes	1,636	1,086	843
Reactions	2,682	2,191	2,286
Metabolites	2,340	1,706	1,770
Compartments	*	10	6
Reference	This work	[16]	[14]

*Not yet defined

5 Conclusions and Future Perspectives

Developing a genome-scale metabolic model for *C. vulgaris* can serve as a valuable tool for exploring the microalga's potential in methane mitigation. Merlin streamlines the entire reconstruction process by providing a user-friendly interface and incorporating essential tools, that otherwise would lead to more time spent. Even though there already exists a model for *C. vulgaris*, this does not meet the requirements for the work in question. Understanding how the methane metabolism functions is of major importance, but it can prove to be a challenging step considering the scarcity of information about this topic available in scientific literature.

Thus far, the model has successfully incorporated metabolites, reactions, enzymes, and transporters. Additionally, the reversibility of reactions has been corrected, and any unbalanced reactions have been addressed. Furthermore, manual curation has been conducted for amino acid and nucleotide metabolisms.

The subsquent steps of reconstruction will involve finalizing manual curation, incorporate compartments using the tool LocTree3 [37] or DeepLoc [38] and assembling gene-protein-reactions (GPR) relationships. Once these steps are complete, it will be necessary to validate the model in order to start making simulations to address the focus of this work of finding methane mitigation strategies.

References

1. C. Safi, B. Zebib, O. Merah, P.-Y. Pontalier, and C. Vaca-Garcia, "Morphology, composition, production, processing and applications of chlorella vulgaris: A review," *Renewable and Sustainable Energy Reviews*, vol. 35, pp. 265–278, 2014.
2. S. Abu-Ghosh, Z. Dubinsky, V. Verdelho, and D. Iluz, "Unconventional high-value products from microalgae: A review," *Bioresource technology*, vol. 329, p. 124895, 2021.
3. P. Geada, V. Vasconcelos, A. Vicente, and B. Fernandes, "Microalgal biomass cultivation," in *Algal green chemistry*, pp. 257–284, Elsevier, 2017.

4. P. Geada, R. Rodrigues, L. Loureiro, R. Pereira, B. Fernandes, J. A. Teixeira, V. Vasconcelos, and A. A. Vicente, "Electrotechnologies applied to microalgal biotechnology–applications, techniques and future trends," *Renewable and Sustainable Energy Reviews*, vol. 94, pp. 656–668, 2018.
5. E. Nisbet, R. Fisher, D. Lowry, J. France, G. Allen, S. Bakkaloglu, T. Broderick, M. Cain, M. Coleman, J. Fernandez, *et al.*, "Methane mitigation: methods to reduce emissions, on the path to the paris agreement," *Reviews of Geophysics*, vol. 58, no. 1, p. e2019RG000675, 2020.
6. M. Anjos, B. D. Fernandes, A. A. Vicente, J. A. Teixeira, and G. Dragone, "Optimization of co2 bio-mitigation by chlorella vulgaris," *Bioresource technology*, vol. 139, pp. 149–154, 2013.
7. S. A. Montzka, E. J. Dlugokencky, and J. H. Butler, "Non-co2 greenhouse gases and climate change," *Nature*, vol. 476, no. 7358, pp. 43–50, 2011.
8. S. E. Hook, A.-D. G. Wright, and B. W. McBride, "Methanogens: methane producers of the rumen and mitigation strategies," *Archaea*, vol. 2010, 2010.
9. E. Cunha, V. Sousa, P. Geada, J. A. Teixeira, A. A. Vicente, and O. Dias, "Systems biology's role in leveraging microalgal biomass potential: Current status and future perspectives," *Algal Research*, p. 102963, 2022.
10. I. Rocha, J. Förster, and J. Nielsen, "Design and application of genome-scale reconstructed metabolic models," *Microbial Gene Essentiality: Protocols and Bioinformatics*, pp. 409–431, 2008.
11. P. Maia, M. Rocha, and I. Rocha, "In silico constraint-based strain optimization methods: the quest for optimal cell factories," *Microbiology and Molecular Biology Reviews*, vol. 80, no. 1, pp. 45–67, 2016.
12. J. S. Edwards and B. O. Palsson, "Systems properties of the haemophilus influenzaerd metabolic genotype," *Journal of Biological Chemistry*, vol. 274, no. 25, pp. 17410–17416, 1999.
13. C. Gu, G. B. Kim, W. J. Kim, H. U. Kim, and S. Y. Lee, "Current status and applications of genome-scale metabolic models," *Genome biology*, vol. 20, pp. 1–18, 2019.
14. C. Zuñiga, C.-T. Li, T. Huelsman, J. Levering, D. C. Zielinski, B. O. McConnell, C. P. Long, E. P. Knoshaug, M. T. Guarnieri, M. R. Antoniewicz, *et al.*, "Genome-scale metabolic model for the green alga chlorella vulgaris utex 395 accurately predicts phenotypes under autotrophic, heterotrophic, and mixotrophic growth conditions," *Plant physiology*, vol. 172, no. 1, pp. 589–602, 2016.
15. H. Wang, S. Marcišauskas, B. J. Sánchez, I. Domenzain, D. Hermansson, R. Agren, J. Nielsen, and E. J. Kerkhoven, "Raven 2.0: A versatile toolbox for metabolic network reconstruction and a case study on streptomyces coelicolor," *PLoS computational biology*, vol. 14, no. 10, p. e1006541, 2018.
16. R. L. Chang, L. Ghamsari, A. Manichaikul, E. F. Hom, S. Balaji, W. Fu, Y. Shen, T. Hao, B. Ø. Palsson, K. Salehi-Ashtiani, *et al.*, "Metabolic network reconstruction of chlamydomonas offers insight into light-driven algal metabolism," *Molecular systems biology*, vol. 7, no. 1, p. 518, 2011.
17. A. Juneja, F. W. Chaplen, and G. S. Murthy, "Genome scale metabolic reconstruction of chlorella variabilis for exploring its metabolic potential for biofuels," *Bioresource Technology*, vol. 213, pp. 103–110, 2016.
18. I. Thiele and B. Ø. Palsson, "A protocol for generating a high-quality genome-scale metabolic reconstruction," *Nature protocols*, vol. 5, no. 1, pp. 93–121, 2010.
19. A. Varma and B. O. Palsson, "Metabolic flux balancing: basic concepts, scientific and practical use," *Bio/technology*, vol. 12, no. 10, pp. 994–998, 1994.

20. S. M. Seaver, F. Liu, Q. Zhang, J. Jeffryes, J. P. Faria, J. N. Edirisinghe, M. Mundy, N. Chia, E. Noor, M. E. Beber, *et al.*, "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 research*, vol. 49, no. D1, pp. D575–D588, 2021.
21. M. Aite, M. Chevallier, C. Frioux, C. Trottier, J. Got, M. P. Cortés, S. N. Mendoza, G. Carrier, O. Dameron, N. Guillaudeux, *et al.*, "Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models," *PLoS computational biology*, vol. 14, no. 5, p. e1006146, 2018.
22. P. D. Karp, P. E. Midford, R. Billington, A. Kothari, M. Krummenacker, M. Latendresse, W. K. Ong, P. Subhraveti, R. Caspi, C. Fulcher, *et al.*, "Pathway tools version 23.0 update: software for pathway/genome informatics and systems biology," *Briefings in bioinformatics*, vol. 22, no. 1, pp. 109–126, 2021.
23. D. Machado, S. Andrejev, M. Tramontano, and K. R. Patil, "Fast automated reconstruction of genome-scale metabolic models for microbial species and communities," *Nucleic acids research*, vol. 46, no. 15, pp. 7542–7553, 2018.
24. O. Dias, M. Rocha, E. C. Ferreira, and I. Rocha, "Reconstructing genome-scale metabolic models with merlin," *Nucleic acids research*, vol. 43, no. 8, pp. 3899–3910, 2015.
25. J. Capela, D. Lagoa, R. Rodrigues, E. Cunha, F. Cruz, A. Barbosa, J. Bastos, D. Lima, E. C. Ferreira, M. Rocha, *et al.*, "merlin, an improved framework for the reconstruction of high-quality genome-scale metabolic models," *Nucleic Acids Research*, vol. 50, no. 11, pp. 6052–6066, 2022.
26. E. W. Sayers, J. Beck, E. E. Bolton, D. Bourexis, J. R. Brister, K. Canese, D. C. Comeau, K. Funk, S. Kim, W. Klimke, *et al.*, "Database resources of the national center for biotechnology information," *Nucleic acids research*, vol. 49, no. D1, p. D10, 2021.
27. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of molecular biology*, vol. 215, no. 3, pp. 403–410, 1990.
28. M. Muthuraj, B. Palabhanvi, S. Misra, V. Kumar, K. Sivalingavasu, and D. Das, "Flux balance analysis of chlorella sp. fc2 iitg under photoautotrophic and heterotrophic growth conditions," *Photosynthesis research*, vol. 118, pp. 167–179, 2013.
29. S. Weber, P. M. Grande, L. M. Blank, and H. Klose, "Insights into cell wall disintegration of chlorella vulgaris," *PLoS One*, vol. 17, no. 1, p. e0262500, 2022.
30. Y. Panahi, A. Y. Khosroushahi, A. Sahebkar, and H. R. Heidari, "Impact of cultivation condition and media content onchlorella vulgaris composition," *Advanced pharmaceutical bulletin*, vol. 9, no. 2, p. 182, 2019.
31. J. A. Coronado-Reyes, J. A. Salazar-Torres, B. Juarez-Campos, and J. C. Gonzalez-Hernandez, "Chlorella vulgaris, a microalgae important to be used in biotechnology: a review," *Food Science and Technology*, vol. 42, 2020.
32. Y. Liang, N. Sarkany, and Y. Cui, "Biomass and lipid productivities of chlorella vulgaris under autotrophic, heterotrophic and mixotrophic growth conditions," *Biotechnology letters*, vol. 31, pp. 1043–1049, 2009.
33. S. Imam, S. Schäuble, J. Valenzuela, A. López García de Lomana, W. Carter, N. D. Price, and N. S. Baliga, "A refined genome-scale reconstruction of chlamydomonas metabolism provides a platform for systems-level analyses," *The Plant Journal*, vol. 84, no. 6, pp. 1239–1256, 2015.
34. A. P. Arkin, R. W. Cottingham, C. S. Henry, N. L. Harris, R. L. Stevens, S. Maslov, P. Dehal, D. Ware, F. Perez, S. Canon, *et al.*, "Kbase: the united states department

of energy systems biology knowledgebase," *Nature biotechnology*, vol. 36, no. 7, pp. 566–569, 2018.
35. P. D. Karp, R. Billington, R. Caspi, C. A. Fulcher, M. Latendresse, A. Kothari, I. M. Keseler, M. Krummenacker, P. E. Midford, Q. Ong, *et al.*, "The biocyc collection of microbial genomes and metabolic pathways," *Briefings in bioinformatics*, vol. 20, no. 4, pp. 1085–1093, 2019.
36. A. Chang, L. Jeske, S. Ulbrich, J. Hofmann, J. Koblitz, I. Schomburg, M. Neumann-Schaal, D. Jahn, and D. Schomburg, "Brenda, the elixir core data resource in 2021: new developments and updates," *Nucleic acids research*, vol. 49, no. D1, pp. D498–D508, 2021.
37. T. Goldberg, M. Hecht, T. Hamp, T. Karl, G. Yachdav, N. Ahmed, U. Altermann, P. Angerer, S. Ansorge, K. Balasz, *et al.*, "Loctree3 prediction of localization," *Nucleic acids research*, vol. 42, no. W1, pp. W350–W355, 2014.
38. J. J. Almagro Armenteros, C. K. Sønderby, S. K. Sønderby, H. Nielsen, and O. Winther, "Deeploc: prediction of protein subcellular localization using deep learning," *Bioinformatics*, vol. 33, no. 21, pp. 3387–3395, 2017.

Supplementary Material

https://github.com/Apolinario8/UC_projeto