Introduction

The problematic of greenhouse gas (GHG) emissions is a global environmental challenge that has raised concerns in the past few decades. Particularly, the increase in atmospheric concentrations of carbon dioxide, methane, and other damaging gases can lead to catastrophic repercussions to life as we know it. Therefore, reducing GHG emissions and fomenting strategies for their mitigation are crucial steps that need to be taken in order to meet the Paris Agreement and ultimately ensure a sustainable future for our planet and those that inhabit it. Taking this into account, our work aims at identifying the metabolic capabilities of photosynthetic microorganisms to reduce methane emissions.​

Possivelmente referir o porquê de se ter pensado nisto – **Wet-lab experiments**:

* Chlorella/*Synechocystis,*dies upon the depletion of CO2
* Chlorella/*Synechocystis, exhibits* neither growth nor mortality in the presence of CH4
* *Synechocystis* produces methane

**O porquê de se utilizar modelos metabólicos para estudar isto:**

Genome-scale metabolic (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 GSM models, a comprehensive understanding of cellular metabolism and identification of strategies to improve a certain objective function can be obtained. The importance of this emerging technology in industry stems from its ability to offer a faster and more cost-effective approach, surpassing the efficiency of traditional methods.​

​

Herein, we describe the reconstruction of GSM models for the microalga *Chlorella vulgaris sp*. – *i*GA1305 –, and for the cyanobacterium *Synechocystis sp*. – *i*JG708.  Both GSM models provide a powerful tool for metabolic improvement, allowing predictions and simulations of CH4 metabolism in response to different culture conditions and genetic modifications.

**Why should we reconstruct a model for a organism that already has a model?**

The first step towards this work was to review the literature. This was done in order to, firstly, determine whether other models for these organisms already existed and secondly, to better understand their metabolism. In this search, we found an existing *Chlorella vulgaris* model (iCZ843) and 3 *Synechocystis*models (iJN678, iSynCJ816 e iSynCJ669). Despite this and considering that genomic and proteomic data is constantly updated, along with the fact that they don't incorporate methane in their models , we reconstructed from scratch models for both organisms.

**Methods**

**Draft reconstruction:**

The reconstruction process begins with the genome sequence of the organism, which was retrieved from NCBI. For this work, we decided to use *merlin* (developed here at Uminho), one of many existing tools to aid in the reconstruction of GSM models. We chose merlin because it contains many other tools that expedite the reconstruction and curation of the model (plugins) and because it can handle both eukaryotic and prokaryotic organisms.

Upon loading the genome to merlin, metabolic data is extracted from KEGG. Then, an homology search using BLAST (chlorella) and DIAMOND(syne) was allowed to perform the enzymes annotation (isto foi feito contra Swiss-prot e Trembl). Isto significa associar genes a enzima(s) -EC number(s). With these steps, we obtain an **initial draft.**

**Manual curation:**

This initial model is still very incomplete as, for example, all reactions extracted from KEGG are reversible thus their correction was required. *Merlin* allows doing this automatically, using templates from ModelSEED database for that purpose. In this step, reversibility (Kbase, MetaCyc), directionality, and gaps are corrected (gap-filling). Transport reactions are generated using Transyt (also a tool developed at Uminho and within *merlin*), all reactions and metabolites are regarding their mass balance (BioCyc, BRENDA). Compartmentalization is performed using third party software (loctree/psort) and then integrated into *merlin*, dividing reactions into their predicted locations (organneles).

Despite completing all these steps, there were still missing some reactions in the network (gaps).

The possible gaps are, for example, metabolites that are not consumed or produced by any reaction, resulting in a ”dead-end” metabolite, and consequently in “blocked” reactions. In certain scenarios, the flux through the reaction during simulations is blocked, which can influence model performance. To address this, *merlin* has a tool that identifies these reactions and helps study disconnected reactions. However, there is currently no automatic tool available to correct these gaps, and manual curation is required.

As a result, it was crucial to manually curate every pathway to ensure that biomass was produced.

**Conversion into a stoichiometric model:**

A major step is to assemble the biomass equation. This is important because it represents every building block necessary for cell growth and consequently enabling predictions of cell growth and phenotypic behavior. Esta parte é feita maioritariamente através da literatura -dados experimentais, na qual definimos as quantidades associadas a cada building block genérico (dna, rna…) . Depois para cada metabolito especifico há abordagens diferentes. O *merlin* calcula para o DNA (dATP, dCTP…), RNA() e proteínas(). No caso dos lípidos, por exemplo, extraiu-se de outros models já existentes.

The growth medium was defined by adding exchange reactions (Exchange reactions represent the growth medium in a metabolic model, including both consumed and produced metabolites). The carbon source was selected (glucose for heterotrophic) and CO2 for photoautotrophic.

The energy requirements, both GAM (growth associated maintenance) and NGAM (non growth associated maintenance) were retrieved from other models.

**Model validation:**

The last step serves to identify inconsistencies and assess if the model is working properly and according to existing information. To do so, cobrapy and optflux were used to perform simulations. Predicted values for growth rates, uptake/secretion were compared to experimental data as well as other models. Spontaneous growth was performed (retira-se o meio-drains e vemos se temos biomassa). Critical genes and reactions were determined using optflux.

This is an iterative process and is often mingled with the model curation, as found discrepancies lead to changes within the model.

**TABELAS – chlorella**

Tabela 1:

Comparison between the predicted growth rates of the model with other models and with experimental data for all conditions.

The values for the heterotrophic growth are within the expected experimental values. The same does not happen for photoautotrophic (and consequently) mixotrophic growths. The difference was found to be connected with the values associated with the quantity of starch in the biomass equation. The previous model considers a much larger quantity of starch and thus the model allocates more energy towards the storage of starch instead of maximizing growth. When we use the previous starch quantity, we get values for growths rates much closer to the experimental values.

Tabela 2:

Comparison of the properties of the model with other models.

Nao ha grande coisa a dizer aqui

**TABELAS – *Synechocystis***

Tabela 3:

Comparison between the predicted growth rates of the model with other models and with experimental data for all conditions.

Based on the results obtained, it can be observed that the reconstructed GSM model (iJG709) generally shows consistency with experimental results and with other previously made models (iJN678), particularly in terms of Synechocystis growth rate under different metabolic conditions, including heterotrophy, mixotrophy, and autotrophy. Although there are some minor discrepancies between the values predicted by the model and the experimental values, the differences are generally small and within an acceptable range of variation. Therefore, we can conclude that the iJG709 model is valid and adequately represents the metabolic behavior of Synechocystis under different growth conditions. This suggests that the model can be a useful tool for investigating and predicting the metabolism of this species in biological and biotechnological contexts.

Tabela 4:

As shown in Table , 709 genes (19.85% of the genome) were designated as protein-encoding genes. This model has a higher number of genes than other models, which can be explained by the number of transport systems identified and by the information available at the time of the reconstruction of each model. The number of metabolites and reactions follows the same pattern. Additionally, the number of reactions in the model is 2172 after compartmentalization: 603 reversible and the remaining 1569 irreversible.

**Conclusion:**

-The reconstruction process did not reveal any genomic evidence of enzymes associated with methane metabolism​.

Similarity search of enzymes associated with methane metabolism were performed and no evidence was found that these are present in the model.

-Both models were validated regarding spontaneous growth, growth rates in different conditions, and uptake/secretion values​;

This means that the models are validated, id est, they can be used as accurate representations of the metabolism of the organisms.

-The models are subject to change as new approaches are employed towards fine-tuning​

-Next steps will involve the incorporation of possible methane-utilizing pathways in the models.

Nao sei se se fala aqui sobre aquelas 4 vias do artigo.