Systems Biology
MSc Bioinformatics

Exercise – implementation of FBA problem

Requirements:
The exercise can be done in a jupyter notebook or with python script(s). For jupyter notebooks please keep the outputs when submitting it. If you use python script(s), attach the results to the scripts using a pdf file. You can paste screenshots of the results there. For repetitive tasks, such as calculating BPCY, or setting growth media conditions, it is recommended to use functions, instead of repeating code.

What will you need:

- Python with COBRApy (pip install cobrapy) and MEWpy (pip install mewpy) installed.
- CPLEX installed on the computer (check CPLEX file to see how to install it). This is only necessary for exercise 2.
- Install ray (pip install ray) package (only necessary for exercise 3).

If exercise 3 is taking too long to run (>3h), ask us for help!

Algae have garnered significant interest in recent years for their potential commercial applications in biofuels and nutritional supplements. Among eukaryotic microalgae, *Chlamydomonas reinhardtii* has arisen as the hallmark, model organism. *C. reinhardtii* has been widely used to study photosynthesis, cell motility and phototaxis, cell wall biogenesis, and other fundamental cellular processes. Commercial use and basic scientific research of photosynthetic organisms could benefit from better understanding of how light is absorbed and affects cellular systems. The quality of light sources implemented in photobioreactors largely determines the efficiency of energy usage in industrial algal farming. Light spectral quality also affects how photon absorption induces various metabolic processes: photosynthesis, pigment and vitamin synthesis, and the retinol pathway required for phototaxis.
Consider the stoichiometric model of *Chlamydomonas reinhardtii* (https://doi.org/10.1038/msb.2011.52). This alga has a robust metabolism that allows it to adapt to different environmental conditions. The simplest growth conditions are autotrophic conditions, where only light, water, and minerals are available. In these conditions, the alga uses photosynthesis to produce biomass, fixing carbon dioxide and releasing oxygen. However, it can also grow in heterotrophic conditions, where it uses organic carbon sources to produce biomass. In some conditions, the algae can also produce additional metabolites, such as hydrogen, and organic acids (ethanol, succinate, formate, lactate, etc). These metabolites can be used industrially to produce biofuels and/or other products. Thus, it is important to understand how the algae produces these metabolites, and how can we control their production.

In this exercise, you will study the metabolism of *C. reinhardtii* under different conditions, aiming at predicting and improving environmental and genetic conditions to maximize the production of one desirable compound.
The compound of interest for your group is available in
Table 1.

Table 1- Group-metabolite association and conditions for omics integration (exercise 2).

Group	Metabolite	Condition for omics integration
1	Succinate	Microaerobic heterotrophic
2	Lactate	Microaerobic heterotrophic
3	Ethanol	Aerobic mixotrophic
4	Formate	Aerobic heterotrophic
5	H2	Microaerobic mixotrophic

The alga grows in autotrophic (using light as energy source and CO2 as carbon source), heterotrophic (using acetate/starch as carbon/energy source), and mixotrophic (using both light and organic carbon sources) conditions. Since the biomass composition of the organism changes according to these conditions, the authors of the model defined different biomass equations.
The alga achieves higher growth rates when oxygen is present, but it can also grow in microaerobic conditions, adapting its metabolism to redirect the carbon flux to fermentative pathways, instead of the citrate cycle.
Light is often represented in GSM models by an artificial metabolite named "photon". To limit the light uptake, the exchange rate of photons must be restrained to zero as usual, and all reactions in the model whose id starts with "PRISM" must be restrained to zero (both lower and upper bound). When light is present, the uptake rate of photons must be limited to a maximum of 2000 mmol/gDW/h, and the bounds of the "PRISM" reactions must be kept as default (those already defined when you read the model). These conditions are summarized in Table 2.

Table 2- Environmental conditions, objective functions, and uptake restrictions. Note that the uptake column refers to the **maximum** amount that can be **consumed** by the organism.

Condition	Objective Function	Metabolite	Uptake (mmol/gDW/h)
Aerobic autotrophic	"BIOMASS_Chlamy_auto"	photon	-2000
		CO_2	-11.16

		O2	-10
		Acetate	0
		Starch	0
Microaerobic autotrophic		photon	-2000
		CO_2	-11.16
		O2	-0.01
		Acetate	0
		Starch	0
Aerobic heterotrophic	"BIOMASS_Chlamy_hetero"	photon	0
		CO_2	-11.16
		O2	-10
		Acetate	-10
		Starch	$-1.72e^{-04}$
Microaerobic heterotrophic		photon	0
		CO_2	-11.16
		O2	-0.01
		Acetate	-10
		Starch	$-1.72e^{-04}$
Aerobic mixotrophic	"BIOMASS_Chlamy_mixo"	photon	-2000
		CO_2	-11.16
		O2	-10
		Acetate	-10
		Starch	$-1.72e^{-04}$
Microaerobic mixotrophic		photon	-2000
		CO_2	-11.16
		O2	-0.01
		Acetate	-10
		Starch	$-1.72e^{-04}$

The evaluation of the production of a given target depends on the production rate of that compound, but also on the substrate consumption rate, and on the growth of the organism. Hence, to evaluate the production of the compound, use the BPCY formulation:

$$\mathrm{BPCY} = \frac{Product \cdot Growth}{Substrate}$$

, where the product is the compound of your group, the growth rate is the flux of the biomass reaction, and the substrate is the sum of the consumption of all carbon sources (acetate, CO2, etc). Note that all consumption/production fluxes must be multiplied by the number of carbons of the respective metabolite. For instance, if the substrate is glucose and the production flux is 1 mmol/gDW/h, the final value of Substrate for glucose is 6x1 = 6 mmolC/gDW/h. If the substrate is CO_2 and the production flux is 1 mmol/gDW/h, the final value of Substrate for CO_2 is 1x1 = 1 mmolC/gDW/h. Note that if you have multiple carbon sources you need to sum the. In these case, Substrate = 6 +1 =7 mmolC/gDW/h.

If the compound is H_2, the substrate is the flux of the photon exchange reaction, and no multiplication is required.

1) The following questions must be repeated for **all** conditions described before.

    a) What is the maximum growth rate achieved by the organism in each condition? Which metabolites are consumed and produced?
    b) What is the wild-type production of the compound of interest?
    c) Access the robustness of the presented solutions using the Flux Variability Analysis approach.
    d) What are the maximum compound production capabilities, guaranteeing a minimum growth rate of 20% of the wild type?
    e) Plot a production envelope showing the production of the compound and the growth rate.
    f) Try to improve the production of the compound by changing the update rates and/or add/remove compounds in the media.

2) In stress conditions, microalgae often change their metabolism to fight the alterations in the environment. In these scenarios, some carbon can be secreted in the form of organic acids. Although some stress conditions can be replicated directly in GSM models, others, such as temperature, pH, salinity, require the addition of additional information, such as gene expression data. The file "expression_data.tsv" contains the normalized gene expression profile of *C. reinhardii* in two conditions: control and stress.

    a) Integrate the expression data for both conditions using the eFLUX algorithm. If your metabolite is H2, use the column "Stress_h2" instead of "Stress". The environmental condition for each group/metabolite is available in
    b) Table 1. Compare the production of your compound in both conditions and with the results obtained in exercise 1 b).

3)

a) Determine the essential genes and reactions of the organism.

b) Using gene knockout optimization strategies, identify sets of solutions to improve the production of the compound. Through the optimization process account for the BCPY and WYIELD to evaluate the solutions.