TBT4165 - Project 3

Introduction

In this project, you will begin working with COBRA (COnstraint-Based Reconstruction and Analysis) methods using the Python library cobrapy. Originally implemented in MATLAB as the COBRA Toolbox, cobrapy is an object-oriented programming framework for the reconstruction and analysis of constraint-based metabolic models. Although still not as comprehensive as the COBRA Toolbox when it comes to modeling capabilities and implemented algorithms, it provides ample functionality to cover our needs for this course. Before starting with this project, I strongly encourage you to take a look at the documentation and API reference where you will find useful descriptions and minor tutorials on how to work with the library. We will also be using the web-based tool Escher for the visualization of metabolic pathway maps and calculated flux phenotypes.

1.1 Toy model

Consider the following toy metabolic reaction network with 8 reactions (R_1, ..., R_8) and 8 metabolites (A, B, C, D, E, ATP, ADP, and P_i)

$$\xrightarrow{R_1} A$$
$$A + ADP + P_i \xrightarrow{R_2} B + C + ATP$$
$$C + ATP \xrightarrow{R_3} B + ADP + P_i$$
$$B \xrightarrow{R_4} D$$
$$A + D + ADP + P_i \xrightarrow{R_5} E + ATP$$
$$E \xrightarrow{R_6} C$$
$$ATP \xrightarrow{R_7} ADP + P_i$$
$$D \xrightarrow{R_8}$$

(i) Specify which metabolites are transported across the cell boundaries (e.g., cell membrane), and the direction of transport.

(ii) Write down the stoichiometric matrix **S** using the ordering of reactions and metabolites as defined above. How many degrees of freedom does this reaction system have and what is the dimensionality of the solution space (i.e. null space of **S**)?

(iii) Given an upper flux bound for R_1 of 10 mmol gDW^{-1} h^{-1}, what is the maximal attainable flux through reaction R_8 and the corresponding flux distribution? What is the net production of ATP (i.e., the flux through R_7)? Implement the model using cobrapy and verify your answer by selecting R_8 as the objective and maximizing its flux.

The `objective` is an attribute of the Model object, while the lower and upper flux bounds of a reaction is given by the attributes `lower_bound` and `upper_bound` of the corresponding Reaction object, respectively.

(iv) It has been shown that the maximization of ATP yield in certain instances is a realistic cellular objective. Given the same flux bound for R_1 as in (iii), explain and discuss the maximal feasible net production of ATP. Verify your answer using cobrapy.

(v) Assume that the flux of reaction R_6 is to be constrained to zero. You may implement this in the stoichiometric matrix by adding a new row to **S** where all column entries are zero except in column 6, which is 1. Explain why this will constrain the flux of reaction R_6 to zero. What is the dimensionality of this new stoichiometric matrix?

1.2 *Escherichia coli* core model

(i) Download `ecoli_core_model` from Blackboard and read the model using the `read_sbml_model` function in cobrapy. Give a description of its content (i.e., number of reactions, metabolites, genes, etc.). Which metabolic subsystems are implemented in the model?

Hint: the subsystems are found in the Model attribute `groups`.

(ii) Simulate the optimal growth phenotype on aerobic, minimal glucose media, by setting the lower bound of glucose uptake ('EX_glc__D_e') to a biologically reasonable uptake rate of -18.5 mmol gDW^{-1} h^{-1} and the oxygen uptake ('EX_o2_e') to -1000 mmol gDW^{-1} h^{-1} [i]. What is the maximal specific growth rate and what are the uptake fluxes of glucose, ammonia, oxygen, and inorganic phosphate in the optimal solution?

(iii) What is the secretion profile of anaerobic growth on glucose? Compare with that of aerobic growth on glucose (ii) and give a biochemical explanation for their differences.

(iv) Visualize the reaction fluxes of both the aerobic and anaerobic flux phenotypes using the *E. coli* core model pathway map (found here) by creating a dictionary of reaction ids and corresponding flux values, then writing this to a json file. Import the data into Escher by clicking Data → Load reaction data, then select your json file. Describe and discuss the difference in flux distributions.

(v) Setting the maximal substrate uptake rate to 20 mmol gDW^{-1} h^{-1}, maximize growth using each of the carbon sources listed in Table 1 individually under both aerobic and anaerobic conditions.

Table 1: Various carbon substrates and corresponding exchange reaction identifiers.

Substrate	Exchange reaction
acetate	EX_ac_e
acetaldehyde	EX_acald_e
2-oxoglutarate	EX_akg_e
ethanol	EX_etoh_e
D-fructose	EX_fru_e
fumarate	EX_fum_e
D-glucose	EX_glc__D_e
L-glutamine	EX_gln__L_e
L-glutamate	EX_glu__L_e
D-lactate	EX_lac__D_e
L-malate	EX_mal__L_e
pyruvate	EX_pyr_e
succinate	EX_succ_e

[i] Note that uptake reaction flux bounds by default are negative, which is due to how these traditionally are defined in constraint-based models of metabolism. A boundary metabolite X is taken up by the system using the following format for the exchange reaction X <=>, where a positive and negative flux denotes secretion and uptake, respectively.