Advanced computational methods (SS'21)
Problem 3: Metabolic Flux Balance Analysis

After investigating the genetic regulation of *E.coli* in case-study 1, we are now interested in its metabolism. One method to investigate metabolism is Flux Balance Analisys (FBA), which allows for simulation of metabolite fluxes through different reactions in a network. Usually, the goal of a FBA is to maximize (or minimize) the flux towards a specific metabolite. The stoichiometric matrix N, which contains the stoichiometric information of all reactions inside the cellular network, is required to perform a FBA. The columns of the matrix N refer to the reactions, whereas the rows of N refer to the metabolites. Consider the following reactions as example: $\xrightarrow{\mathrm{r_1}} A \xrightarrow{\mathrm{r_2}} B \xrightarrow{\mathrm{r_3}} C \xrightarrow{\mathrm{r_4}}$

$$N = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{cccc} r_1 & r_2 & r_3 & r_4 \end{array} \\ \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \end{array} \tag{1}$$

Under the hypothesis that the dilution of metabolites and the growth (in size) of the cells can be neglected, the Mass Balance equation can be written as $\dot{\underline{c}} = N \cdot \underline{r}$. During the exponential growth phase, the intracellular concentration of the metabolites is assumed to be constant (Steady state). In this case, the previous equation can be rewritten as $N \cdot \underline{r} = 0$. Thanks to this simplified relation, it is possible to represent and analyse large networks using only the stoichiometric matrix.

Background case-study

In this case-study, polyhydroxyalkanoate (PHA) synthesis by *E.coli* is investigated again. However, this time the focus lies on studying and optimizing the intracellular reactions leading to PHB synthesis and the reactions involved in cellular growth using FBA. The metabolic network is presented in figure 1. The network has been subdivided into smaller parts. The stoichiometric matrix including all reactions will be built step by step in the different tasks.

Task 1

First, a very simplified version of glycolysis is investigated. For this first task, reducing equivalents($FAD/FADH_2$, NAD/NADH), ATP and CO_2 will not be considered. Also, for the moment, all framed processes (Glycerin metabolism, Redox / ATP- synthesis, PHB-synthesis, Gas transport, Maintenance/ Growth and Calvin cycle) can be ignored. All reactions and metabolites are numbered; please stick to this nomenclature for the construction of the stoichiometric matrix in all tasks.

1. In line with the previous set of information, write the stoichiometric matix corresponding to the simplified glycolysis pathway (for convenience, call the matrix N_1) . Assume that in this simplified version of the network, Acetyl-CoA flows away in r_{13}.

2. To test the correctness of the matrix, a reaction vector $\underline{r}$ that fulfills the steady state condition ($N \cdot \underline{r} = 0$) will be defined. To do this, fix the values of the vector $\underline{r}$ to 1 for the first six reactions (reactions from glucose to GAP). From GAP on, all fluxes represented in vector $\underline{r}$ should be fixed to 2. By looking at the representation in figure 1, can you explain the reason behind this doubled flux? With the defined vector $\underline{r}$, check whether the steady state conditions is fulfilled.

3. What is the rang of the stoichiometric matrix?
   Determine the null space. What information can we infer from the null space (from a biological point of view)?

Task 2

The equation $N \cdot \underline{r} = 0$ can be rewritten as a system of linear equations, but in a typical real-life network several fluxes are undetermined, thus making the system unsolvable. In addition, the values of the fluxes and the metabolites can vary in a range whose maxima and minima are subject to themodynamic - and biophysical constraints. Within the defined boundaries, the rates can adopt positive or negative values: positive fluxes are commonly associated to forward reactions, negative fluxes to backwards reactions. Fluxes that can be either positive or negative potentially represent reversible reactions. This set of boundaries is then summarized in the following condition: $\underline{lb} < \underline{r} < \underline{ub}$.
This means that even if a reaction rate is not known, the variation of the rate can still be constrained by setting the upper and the lower limit.

To define the distribution of fluxes in a steady state, it is possible to apply a very simple algorithm. If one or more flux(es) are to be maximized, the following optimization problem

can be formulated:

$$\begin{aligned} max \quad & c^T \cdot \underline{r} \\ s.t. \quad & N\underline{r} = 0, \\ & \underline{lb} < \underline{r} < \underline{ub} \end{aligned} \tag{2}$$

c^{T} indicates the weights/priority of each flux in the overall optimization. In case only one flux is maximized, the vector c^{T} will be all zeros except for the column that represent the flux that is sought to be maximized. *s.t.* stands for *subject to* and indicates the conditions that must be fulfilled. In case the system of equations is mostly undetermined and not solvable, the algorithm here mentioned describes the feasibility of the rates rather than the actual rates.

By using the MATLAB function ***r = linprog(f,...)***, the optimization problem can be solved in MATLAB. The input to the *linprog* function is a vector $\underline{f}$ that indicates the weights of the fluxes. This vector corresponds to the formed described vector c. The function returns a distribution of fluxes in the vector $\underline{r}$ so that $\underline{f}^T \cdot \underline{r}$ is minimal. This formulation can therefore only solve minimization problems. To solve the above mentioned optimization problem in which the fluxes are maximized, the flux weight should be made negative (eg -1 instead of 1).

Now, a more elaborated version of the metabolic network will be investigated. The following information should be considered to set the boundaries correct: **(i)** the glucose uptake rate cannot be greater than 1, **(ii)** all reactions where hydrogen is involved remain constant at 0 **(iii)** and all remaining irreversible reactions cannot be greater than 100. Since CO_2 can both be taken up and excreted, the boundaries for this gas are fixed between ± 100.

1. Define a new stoichiometric matrix (N_2) including CO_2 (r_{13}) and the following modules: Gas transport, Maintenance/Growth and Redox / ATP-synthesis. Please note that at this point, you should also include the reducing equivalents and ADP/ATP synthesis for the glycolysis module (reactions r_1 until r_{13}) in your stoichiometric matrix.

2. How do you correctly construct all the vectors for the *linprog* MATLAB function? Look at the FBA Lecture and MATLAB help for more information. Define the boundary vectors accordingly to the previously given information .

3. Maximize the biomass production rate r_μ (reaction 14) by using the stoichiometric matrix N_2. Report the amount of ATP produced per mol glucose (glucose uptake rate is fixed by the boundaries). Compare this value with literature and draw you conclusion.

4. Now, the previous optimization problem is "reversed". How many moles of glucose are necessary in order to produce 50 ATPs? In this case, glucose uptake should be minimized in the optimization problem.

Task 3

In this task, the production of polyhydroxybutyrate (PHB) will be studied by including the PHB-synthesis module in the stoichiometric matrix. The synthesis rate of PHB is indicated as r_{24}.

1. Implement an optimization problem with a new stochiometeric matrix that includes the module for PHB synthesis (N_3). The matrix N_3 is provided on Moodle. The newly added rates can vary in the range between 0 and 100, while the boundaries of the other rates do not change compared to the previous exercise. In this case, the PHB synthesis rate r_{PHB} (r_{24}) has to be maximized rather than growth. How many moles of PHB can be produced from one mole of glucose?

2. Implement a second optimization problem where r_μ and r_{PHB} are simultaneously optimized. In this case, it is necessary to prioritize the rates (*prioritize* in this context equals *set an optimal weight value for vector f*). r_{PHB} constitutes the main priority, while r_μ is the secondary one; this means that the weight value for $r_{PHB} > r_\mu$ .Report the values of the rates of r_{PHB} and r_μ when:
(i) r_{PHB} is maximized and r_μ is minimized
(ii) both r_{PHB} and r_μ are maximized but with different priorities (meaning r_{24} has a higher weight value in vector f than r_{14}).
What do you observe?

3. The simultaneous optimization of two criteria can be assessed with an optimization function by varying the priorities as in the previous exercise, or one can find the best combination of the priorities by using the so-called 'Pareto frontier'. Look for the description in literature, then draw the Pareto front of the problem 2. This can be done as followed:
Fix the priority of r_{PHB} as your primary one, then let the priority of r_μ vary between ± 1. Plot r_{PHB} against r_μ. This is the Pareto front.

   What is the form of the Pareto front? How do you interpret it?

4. Show in a separate plot the values of r_{PHB} and r_μ against the priority used in the previous exercise.

Task 4

Polyhydroxypropionate (PHP) is a polyhydroxyalkanoate like PHB, and it can be synthesized using glycerine as a C-source by *E. coli* in only a few steps. In this task, you will compare the production of the two different biopolymers using glycerine. The matrix N_4, which includes the glycerin module, is provided on Moodle.

For this taks, assume that glycerine is the only substrate ($r_1 = 0$). The boundaries remain the same as in the previous tasks. Additionally, the new reactions can range from 0 to 100.

1. When maximizing PHP synthesis (r_{30}), how many moles of PHP can be synthesized per mole of glycerine?
2. How many moles of PHB are produced per mole of glycerine, when PHB production (r_{24}) is maximized?
3. Comparing the previous results, which of the polymers has the highest **mass-yield** on glycerine ?

Task 5

Glycerine is a cheap substrate, since it is a waste product accumulated during biodiesel production. However, given the increasing (environmental) concern about the production of biodiesel, alternatives to PHP synthesis from glycerine should be investigated.

1. According to the Kegg-database, what other synthesis pathways exist for the production of 3-Hydroxy-Propionyl-CoA (the precusor of PHP)?
2. Choose **one** of the pathways and, using Kegg, give the reactions that lead to the synthesis of 3-Hydroxy-Propionyl-CoA.
3. Construct a new stochiometric matrix (N_5), by including the reactions of task 5.2 in the previous stoichiometric matrix (N_4) and run an optimization problem aiming to produce a 1:1 ratio co-polymer made up of PHB and PHP.
   Use the same values of the boundaries of the reactions as in the previous exercises. PHP synthesis from glycerine is not considered ($r_{25} = 0$).

Task 6

During heterotrophic catabolism, like the one we simulated in thet tasks before, carbon is lost in the form of CO_2. To simulate the recovery of CO_2, it is necessary to include the Calvin-Cycle (as is found in autotrophic organisms). When we include this cycle, we are no longer simulating the metabolism of *E.coli*, but that of, for example, *Cupriavidus necator*. *Cupriavidus necator* is a gram-negative, hydrogen utilizing "knallgas" bacterium and a natural producer of PHB. *C. necator* is a mixotrophe, meaning it can grown on glucose, as well as on CO_2 and hydrogen.

In the upcoming simulations, hydrogen will need to be considered. The hydrogen uptake rate is limited between 0 and 100, and the same limitation is applied to the reduction of NAD^+ (r_{18}). The stoichiometric matrix that includes the Calvin-Cycle (N_6) can be found on Moodle.

1. How much hydrogen is necessary per glucose, in order to convert all the CO_2 that is normally produced under heterotrophic circumstances to PHB?

2. Investigate the yield $Y_{\frac{PHB}{glucose}}$ in relation to the variation of $\frac{Hydrogenuptake}{Glucoseuptake}$. Suppose that no additional CO_2 is supplied in the cultivation. (*Suggestion*: let hydrogen upper boundary vary in a given range)

3. Now, the PHB production in a synthetic gas fermentation will be simulated, meaning that only CO_2 ,H_2 and O_2 are substrates, and glucose is no longer considered. In this case, where the first part of glycolysis becomes irrelevant and the Calvin-Cycle is the only means of carbon metabolism, **reaction 7 and 8 are not active**. Please plot the minimal required gas mixture CO_2 / H_2 / O_2 against the energy requirement r_μ, for the case where PHB production should be maximized. How much oxygen is required per mol of formed Acetyl-CoA, when $r_\mu = 0$?

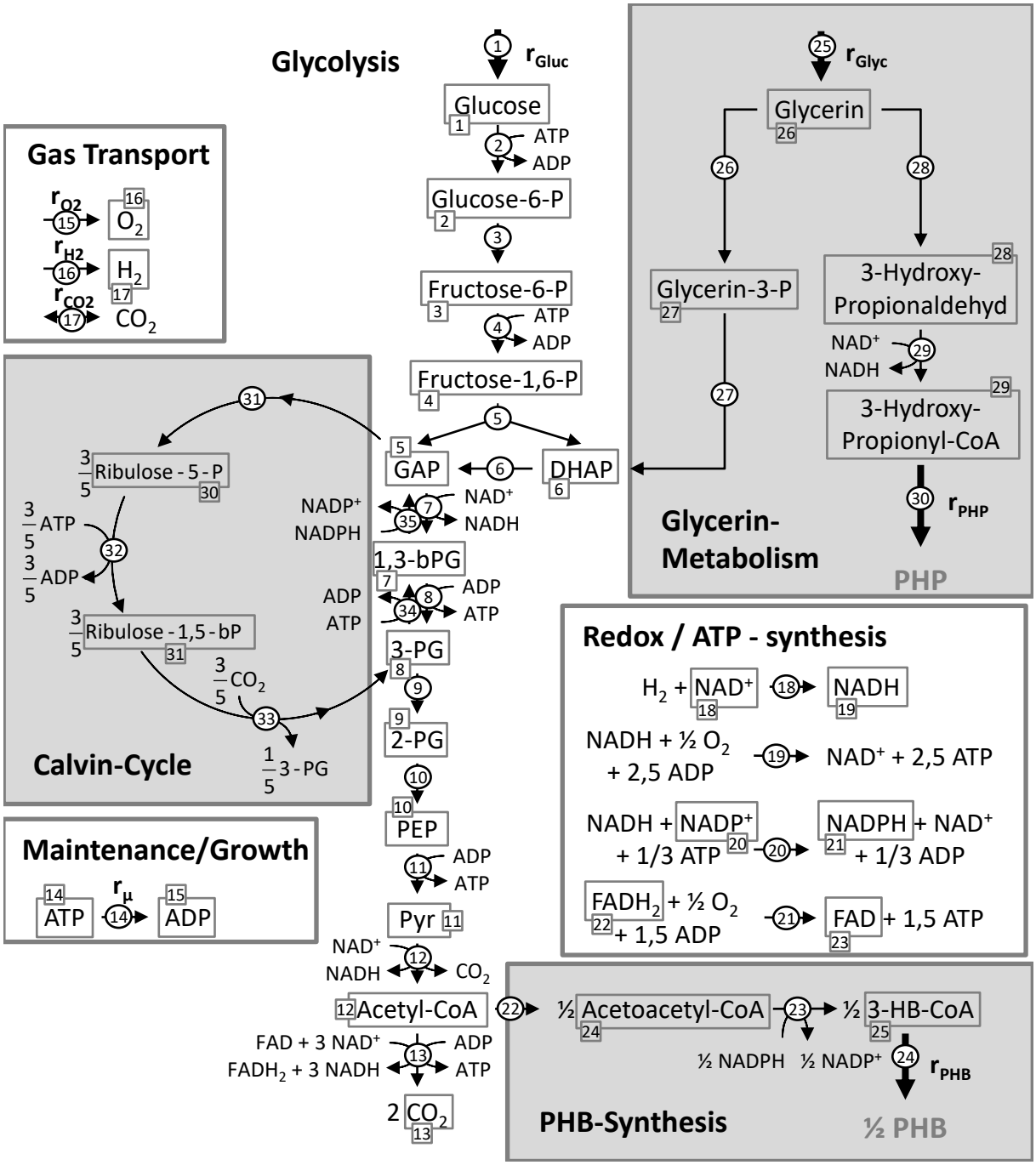

Figure 1: Simplified metabolic network in *C. necator* and *E. coli*