Practical Bioinformatics

Wagner Section

Exercise Block 3. Biomass growth in a genome-scale metabolic network

Exercise 3.1. Identify the molecular components of a cell (protein, DNA, RNA, lipids) whose precursors need the greatest amounts of resources for their synthesis.

The metabolites in the biomass (objective) function for *E. coli* are the metabolites that are required for the growth of the cell. They comprise amino acids (in proteins), deoxynucleotide-triphosphates (DNA), nucleotide-triphosphates (RNA), lipids, and various cofactors. Their relative proportions are known from experiments. The four metabolic models that are part of this exercise contain different biomass objective functions that have been edited such that each is missing one major constituent. Use FBA to compute the growth flux for each biomass function in a glucose-minimal environment.

*A **minimal environment** or **minimal medium** is a medium which contains only one source of each essential element (commonly carbon, nitrogen, oxygen, phosphorus and sulfur). In a glucose-minimal medium, glucose is the only source of carbon.*

Compare your values with the wild type biomass function. Why do you think there is a change in the growth flux? In which of the four classes do you observe the greatest change in the growth flux values between the alternative and the wild type biomass functions? Why? (Hint: Think of the resources required for a major constituent's synthesis.)

The files necessary for this exercise are listed below:

File name	**Description of the biomass function**
iAF1260.xml	All biomass constituents of the wild type *E. coli*
iAF1260_no_AA.xml	All biomass constituents of the wild type *E. coli* except amino acids
iAF1260_no_Lipids.xml	All biomass constituents of the wild type *E. coli* except lipids
iAF1260_no_RNA.xml	All biomass constituents of the wild type *E. coli* except RNA
iAF1260_no_DNA.xml	All biomass constituents of the wild type *E. coli* except DNA

Exercise 3.2. Identify essential reactions and the pathways they belong to for the synthesis of the four major biomass constituents (amino acids, lipids, DNA, RNA) of a cell.

Amino acids (in proteins), deoxynucleotide-triphosphates (in DNA), nucleotide-triphosphates (in RNA), lipids, and various cofactors are the metabolite precursors that are required for the growth of a cell. The four metabolic models that are part of this question all contain different biomass objective functions, each of which contains only one major constituent at a time (as opposed to the previous exercise, where the functions contained all *but* one constituent).

The files necessary for this part of the exercise are listed below:

File name	**Description**
iAF1260_AA.xml	Only amino acids in the biomass function
iAF1260_Lipids.xml	Only lipids in the biomass function
iAF1260_RNA.xml	Only RNA in the biomass function
iAF1260_DNA.xml	Only DNA in the biomass function
iAF1260_subsystems.csv	Reactions and their respective subsystems

Find the essential reactions for each biomass function in a glucose-minimal environment. Then, find the metabolic pathways (or subsystem) that each set of these reactions belongs to. Do the metabolic pathways designations match with those of the major biomass constituents whose biosynthesis they are essential for? Why or why not? (Optional: Could you think of a way to plot these results so you can visually compare the usage of each pathway between the different models?).

A semicolon-separated .csv file (see the table above) containing all *E. coli* model reactions and their respective subsystems is provided. A few reactions will have subsystems with NA values as they are artificial external reactions, however, this will not affect your results. You can read the content of a .csv file like a .txt file and iterate over its lines using:

```
file = open("filename.csv")
for line in file.readlines():
    print(line)
```

This will give you each `line` as a string, which you can separate into parts using the string method `.split()`.

Exercise 3.3. How does the number of essential reactions increase, on average, when increasing the number of metabolites included in the biomass reaction?

The biomass reaction of the metabolic network of wild type *E. coli* contains 67 metabolites. 293 essential reactions are required for the synthesis of all biomass metabolites.

To tackle the main question of the exercise, we have created various biomass reactions with randomly chosen biomass metabolites. First, we randomly chose 20 metabolites and calculated the number of essential reactions that were needed for biomass production with glucose as the carbon source. We repeated this 5 times. Afterwards, we repeated the analysis but selecting 30 metabolites for the biomass reaction. We did this one more time but selecting 40 metabolites for the biomass reaction.

The file biomass_essential_reactions.tsv is a tab-separated file which contains the results of the analysis (Note: The tab character is `"\t"`). The number of essential reactions is given. The different columns correspond to different numbers of metabolites included in the biomass reaction (that is, 20, 30, 40 and 67). The first row is a header indicating the information of each column. Rows 2 to 6 contain the results of the independent sampling of biomass metabolites. You can use the same method described in exercise 3.2. for .csv files to access the file content.

Calculate the average number of essential reactions and the standard deviation for each using the `numpy` functions `numpy.mean()` and `numpy.std()`, both of which take a list of values as an argument.
Plot the number of essential reactions against the number of molecules to be synthesized (20, 30, 40, 67). You can use the `matplotlib.pyplot` function `errorbar(x_values, means, yerr=standard_deviations)`. What is the pattern you observe?