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 lacks one major constituent. Use FBA to compute the growth flux for each biomass function in a glucose minimal environment. 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:

aa biggest increase, since if aa is excluded from biomass reaction (bc it's imported from the outside), the lots of resources are freed and it can produce more of other stuff
DNA had smallest increase, therefore if dna is imported from the outside (and not produced from within), it opens up fewer resources

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). Find the essential reactions for each biomass function in a glucose minimal environment.

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

File name	**Description of the biomass function**
iAF1260_AA.xml	Only amino acids
iAF1260_Lipids.xml	Only lipids
iAF1260_RNA.xml	Only RNA
iAF1260_DNA.xml	Only DNA

Then find the metabolic pathways 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?

You can access the pathway to which a reaction belongs by using the 'subsystem' command, e.g.:

```
In[1]: model.reactions[9].subsystem
Out[1]: 'Transport, Outer Membrane Porin'
```

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 63 metabolites. 256 essential reactions are required for the synthesis of all biomass metabolites.
In order to tackle the main question of the exercise we have created various biomass reactions with randomly chosen biomass metabolites. We first chose 20 metabolites, randomly and calculated the number of essential reactions that were needed for biomass production with glucose as 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. 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 63). 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.
Calculate the average number of essential reactions and the standard deviation for each. Plot the number of essential reactions against the number of molecules to be synthesized (20, 30, 40, 63). What is the pattern you observe?