REI601M **Homework assignment 6**

Out: 13.2.2018 **Due: 19.2.2018**

1. (Short essay, 200 – 250 words, 25%) Read the review paper by McCloskey et al. *Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli* (http://msb.embopress.org/content/9/1/661.long). Select one of the categories in figure 3 and write a description of a small project that you would like to work on. Your description should include, **what** the project is about, **why** it is worthy of study (personal interest, importance, novelty, does it generate new knowledge? etc) and **how** you plan to accomplish it (a rough sketch at this stage). You are not restricted to using *E. coli*. Assume that you possess Matlab/COBRA/Python/CobraPy superpowers but aim for a realistic project proposal.

2. (Flux under aerobic and anaerobic conditions, 20%) In this problem you use flux balance analysis to study the aerobic and anaerobic metabolism in the *E. coli* core network when glucose is used as a carbon source. To investigate pathway usage, you should try to visualize the flux values for each condition (Note 1).

a) Aerobic conditions: Set the lower bounds on EX_o2(e) to a large negative value, e.g. -1000. i) What is the uptake of oxygen (in mol / mol glucose) under these conditions? ii) What metabolic pathways are used in aerobic conditions?
b) Anaerobic conditions: Now set the oxygen uptake rate to zero. i) How is growth affected by no oxygen? ii) What metabolic pathways are used in anaerobic conditions? iii) What by-products are secreted?
c) Summarize your findings in a single figure containing a flux map for both conditions. Incude a descriptive figure caption.

3. (Sensitivity analysis, 30%) Flux variability analysis (FVA) is a form of sensitivity analysis that is frequently used to study the flexibility of metabolic networks. Perform FVA on the *E. coli* core network using glucose as a carbon source under **anaerobic conditions** and a lower bound on growth corresponding to 75% of the maximum (Note 2). Use the results to

a) Determine the number of reactions that can only proceed in the forward direction (i.e. are irreversible).
b) Identify the carbon compounds, besides carbon dioxide, that the network is forced to secrete (if any).
c) Identify blocked reactions, i.e. reactions that are unable to carry flux (if any).
d) Identify reactions with very large flux ranges. What can explain such high values?
e) Prepare a figure showing the flux ranges for each reaction (Note 3).

4. (Gene essentiality, 25%) Here you study gene essentiality in metabolic reconstructions of five different organisms, *Thermotoga maritima*, *Synechocystsis, E. coli* and *S. cerevisiae*. (Note 1). An *essential gene* is one that when knocked out reduces the maximum growth rate below a given threshold.

a) Create a figure which illustrates the fraction of essential genes in each organism. You can assume that a growth rate below 10% of the maximum is lethal (Note 4). Use the default model constraints.
b) Is the fraction of lethal genes similar in all the organisms or is there a clear difference? Can you find any biological reasoning for consistencies or lack thereof?

Notes

Note 1 (Flux visualization) Flux values in the *E. coli* core model can be conveniently visualized using the Escher package (http://escher.github.io/). After determining the flux values with

```
>> sol=optimizeCbModel(model)
```

you write them to a text file on the form

```
ID, FLUX
GADP,1.000
PYK, 0.000
```
etc.

Matlab users should use the `write_escher_file` function included in the zip file which takes care of converting IDs in the Matlab model to the Escher format (COBRApy users only need to create a text file)

```
>> write_escher_file(model, sol.x, 'flux.txt');
```

Select "Centeral metabolism" in Escher, then Data/Load Reaction Data and select the file flux.txt. You fine tune the scaling and color scheme under View/Settings. If the the flux values span a very large range, a log-transform is useful. It can also be helpful to split reactions into flux / no flux, e.g. by (abs(sol.x)>1e-4).

You can look up individual reactions (based on their IDs displayed in the map) using the BiGG Models database (http://bigg.ucsd.edu/) or simply by using the printRxnFormula COBRA function.

Note 2 (Flux variability analysis) Flux variability analysis is performed using the fluxVariability function, a function used in homework 5. The second argument (optPercentage) corresponds to the lower bound placed on the cellular objective (growth rate in the *E. coli* core model by default) and is a value between 0 and 100.

Note 3 (Flux variability analysis – cont.): Consider having reactions IDs on the vertical axis and flux ranges on the horizontal axis. There are a few reactions with very wide ranges and for that reason you should adjust the limits of the vertical axis (e.g. between -25 and 35). In Matlab: Use the “set“

command to tweak the labels on the horizontal axis by modifying the 'XTick' and 'XTickLabels' properties.

Note 4 (Gene deletions in COBRA) To carry out gene deletions you can use the singleGeneDeletion function in Matlab (COBRApy: single_gene_deletion).

Tips on creating figures for the report:

Create a PDF document containing the figure with

```
>> print –dpdf figname
```

or EPS file using

```
>> print –depsc figname
```

instead of taking a screenshot directly from Matlab.

Increase the font size of axis labels

```
>> hx=xlabel ('blah');
>> set(hx, 'fontsize',14)
>> set(gca, 'fontsize',14) % increase size of axis numbering
```