Computational Systems Biology (BT5240)

Assignment 3

Anirudh Rao (BE21B004)

NOTE: All simulations for this assignment were performed using the `nanoCobratoolbox`, instead of using the entire COBRA Toolbox. Please ensure that `nanoCobratoolbox` is in the same directory as the MATLAB scripts provided as part of the submission for this assignment while running the scripts.

Problem 1

Sub-Problem (i)

The *Geobacillus icigianus* model was loaded in MATLAB. The uptake rate of oxygen was changed to 0.03125 mmol gDCW^{-1} h^{-1} by changing the lower bound of `EX_o2_e` to -0.03125. The growth rate under this condition was found to be 0.413268 h^{-1}.

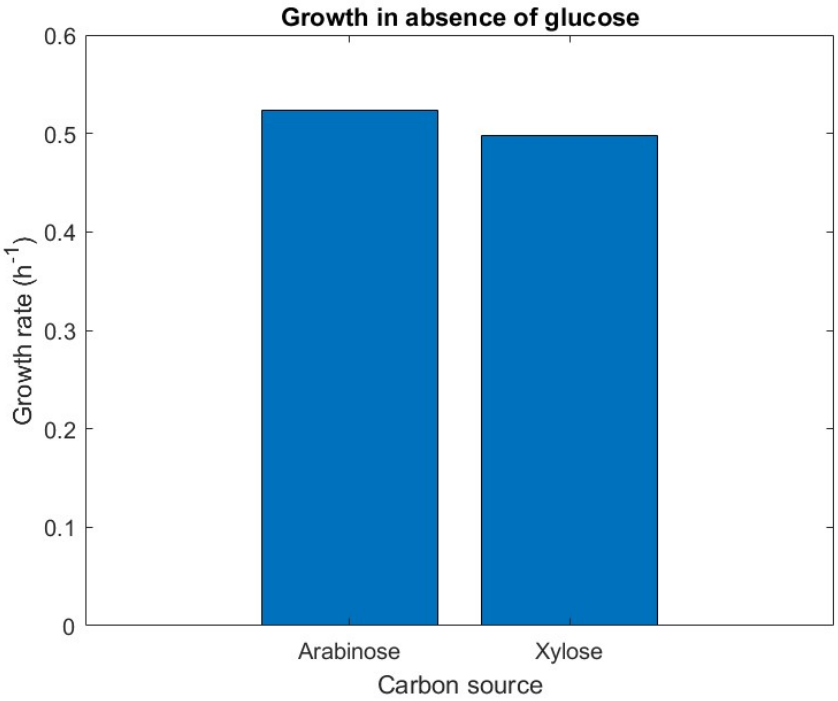
Sub-Problem (ii)

The growth of the bacterium in the absence of glucose, but in the presence of arabinose or xylose was simulated. The uptake rate of glucose (`EX_glc__D_e`) was set to 0.

First, the uptake rate of arabinose (`EX_arab__L_e`) was set to -20 and the uptake rate of xylose (`EX_xyl__D_e`) was set to 0. Under this condition, the growth rate was 0.523773 h^{-1}.

Next, the uptake rate of xylose (`EX_xyl__D_e`) was set to -19 and the uptake rate of arabinose (`EX_arab__L_e`) was set to 0. Under this condition, the growth rate was 0.497584 h^{-1}.

Based on these simulations, we can see that arabinose is a better carbon source than xylose for the growth of *Geobacillus icigianus* in the absence of glucose.



Sub-Problem (iii)

We wish to improve the production of 2,3-butanediol (`EX_btd_RR_e`) while maximising the biomass. To do this, FSEOF was performed using an implementation of the algorithm that was available on the Raman Lab's GitHub page.

FSEOF identifies reactions that can be overexpressed or deleted in order to improve the given bioengineering objective while maximising the biomass. It was found that 10 reactions were overexpression targets and 389 reactions were deletion targets to improve the production of 2,3-butanediol by *Geobacillus icigianus*.

To identify the genes that can be overexpressed or deleted to improve the production of 2,3-butanediol, the `findGenesFromRxns` function in the `nanoCobratoolbox` was used to map the reactions identified by FSEOF to the genes they are related to. From this, 21 genes were identified as overexpression targets and 606 genes were identified as knockout targets.

Problem 2

Sub-Problem (i)

The XML version of the `iNJ661` model of *Mycobacterium tuberculosis* was downloaded from BiGG Models and loaded into MATLAB. This has 661 genes, 1025 reactions, and 825 metabolites. In comparision, the *E. coli* core model has 137 genes, 95 reactions, and 72 metabolites.

Based on this comparison alone, we can arrive at the conclusion that as the number of genes incorporated in a model increases, so does the number of reactions and metabolites. As more genes are included, we can expect to have more protein products (i.e., enzymes). These enzymes can catalyse a larger number of reactions, allowing the model to incorporate information about more metabolites that undergo these reactions. However, enzyme multifunctionality and pathway redundancy may lead to non-linear changes in the number of reactions and metabolites as the number of genes increases.

Sub-Problem (ii)(a)

To identify the nutrients being taken up by the bacterium, the uptake reactions were first found using the `findExcRxns` function in the `nanoCobratoolbox`. These were mapped to nutrients using `findMetsFromRxns`. This revealed that *M. tuberculosis* takes up 17 nutrient sources for its growth:

- Carbon sources - Glucose, glycerol, citrate
- Nitrogen sources - Glutamate, ammonium
- Mineral sources - H^+, phosphate, sulphate, K^+, bicarbonate, Na^+, Fe^{3+}, Cu^+, Cl^-, Ca^{2+}
- Other sources - Water, oxygen

To identify which of these nutrients are absolutely essential for the survival of the bacterium, `singleRxnDeletion` was performed on the 17 uptake reactions. Of the 17 nutrients taken up by *M. tuberculosis*, it absolutely requires 5 key nutrient sources for survival, without which it will die. These are:

- Essential carbon sources - Glycerol
- Essential nitrogen sources -
- Essential mineral sources - phosphate, sulphate, Fe^{3+}
- Other essential sources - Oxygen

Interestingly, none of the nitrogen sources are essential.

Sub-Problem (ii)(b)

To identify the "best" nutrient sources, the uptake rates of all nutrients were initially set to -1. Next, for each of the 17 nutrients, the uptake rate was iteratively set to -1000 to simulate the abundance of that nutrient and the low availability of the other nutrients. FBA was performed to compute the growth rate.

- The best carbon sources were found to be glucose and glycerol, both of which achieve a growth rate of 0.30597 h^{-1} when present separately.
- The best nitrogen source was found to be ammonium, which achieves a growth rate of 0.053113 h^{-1} when present separately.
- All the mineral sources are equally good as each of them achieves a growth rate of 0.05255 h^{-1} when present separately.

Sub-Problem (ii)(c)

Combination effects were studied by creating two combinations of a carbon source and a nitrogen source and finding the growth rate when these sources are present abundantly in the medium. To simulate this, the carbon source and nitrogen source that are part of the combination being studied are set to have an uptake rate of -1000, but the uptake rates of the remaining nutrients are set to -1.

- When the carbon source is glucose and the nitrogen source is ammonium, the growth rate is 0.9023 h^{-1}.
- When the carbon source is citrate and the nitrogen source is glutamate, the growth rate is 0.05255 h^{-1}.

From this, we can say that the combination of glucose and ammonium is better than the combination of citrate and glutamate.

Sub-Problem (iii)

Essential genes in *M. tuberculosis* were identified using `singleGeneDeletion`. This identified that 188 genes (out of a total of 661 genes) are essential. These genes can be targetted to stop the growth of the bacterium. Of the 188 essential genes, only 174 genes have a RefSeq annotation provided in `model.geneisrefseq_nameID`.

One of the genes identified as essential is Rv1338. Using Mycobrowser, it was found that this gene codes for a glutamate racemase. This provides the (R)-glutamic acid required to synthesise the peptidoglycan that forms the cell wall of the bacterium. Targeting this gene can possibly inhibit cell wall formation, inducing the death of the bacterium.

Sub-Problem (iv)

The XML version of the `iCN718` model of *Acinetobacter baumanii* was downloaded from BiGG Models and loaded into MATLAB. This has 709 genes, 1015 reactions, and 888 metabolites. In comparison, the *M. tuberculosis* model has 661 genes, 1025 reactions, and 825 metabolites. Although *A. baumanii* has more genes and metabolites than *M. tuberculosis*, it has fewer reactions. Thus, we can never really predict how the number of reactions and metabolites while change as the number of genes is changed. As mentioned earlier, this can be attributed to phenomena like enzyme multifunctionality/promiscuity and the existence of alternative or redundant metabolic pathways.

Essential genes in *A. baumanii* were identified using `singleGeneDeletion`. This identified that 74 genes (out of a total of 709 genes) are essential. Of the 74 essential genes, only 2 genes have a RefSeq annotation provided in `model.geneisrefseq_nameID`.

To identify the common essential genes in *M. tuberculosis* and *A. baumanii*, an attempt was made to compare the gene names directly. However, this proved to be infeasible as the naming convention in the two models is drastically different (e.g., 'Rv0046c' in *M. tuberculosis*, 'ABAYE_RS00740' in *A. baumanii*).

Although the RefSeq annotations in both models are incomplete, they were used to identify at least some of the common essential genes, if not all. This revealed that pyrG is the only annotated essential gene common to both bacterial species. According to UniProt, this codes for a orotidine 5'-phosphate decarboxylase that is involved in the synthesis of uridine monophosphate (UMP). This is a key metabolite in RNA biosynthesis, proving that pyrG can indeed be a common essential gene to both species.

However, we can expect that some of the unannotated essential genes may be common to both species. It is currently not feasible to identify these in the scope of this assignment.