Assignment 3

CompSysBio

BT5240

Completed on
Mar 22, 2025

Prepared by
Shivan Ajay Iyer b...

The following submission is a report accompanying the MATLAB code and details the key findings and results of the metabolic analysis.

Models Involved

All models were downloaded from the **Bigg Models Database** as **SMBL files** and loaded into Matlab using the **readCbModel** functionality provided by the NanoCobraToolbox. Following are the models that were used

- Geobacillus icigianus model
- E. coli core model
- iNJ661 model of Mycobacterium tuberculosis
- iCN718 model of Acinetobacter baumanii

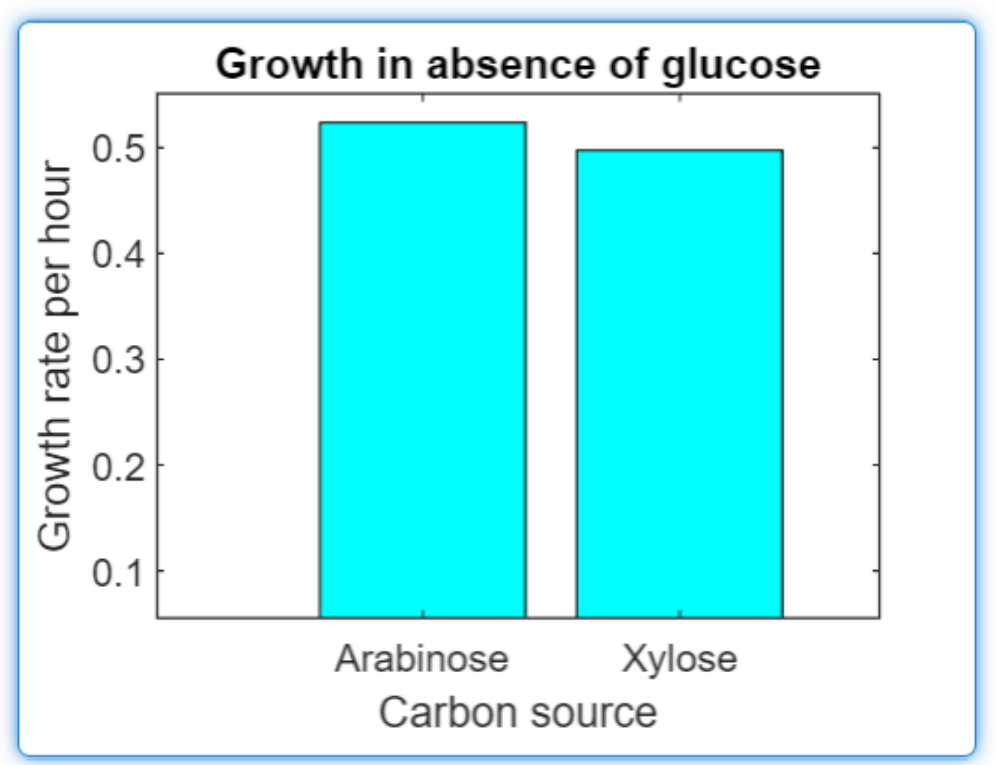
Problem 1 Part i and ii

Following is an overview of the method employed in the code. Additionally, the key assumptions are also listed below.

1. The oxygen uptake rate was changed to 0.03125 mmol gDC/Wh by changing the lower bound of the EX_o2_e reaction to -0.03125 as directed in the question.
2. The growth rate was simulated in the absence of glucose but in the presence of arabinose or xylose by setting the glucose uptake rate (EX_glc__D_e) to zero. The carbon source uptake rate to be measured was set to the given value (-20 and -19), and the other was set to zero.

- The growth rate for EX_o2_e reaction to -0.03125 condition was found to be 0.413268 per hour
- The growth rate was 0.523773 per hour using just Arabinose as a Carbon source

- The growth rate was 0.497584 per hour using just Xylenose as a Carbon source
- Thus, based on the assumptions of the simulation, Arabinose is a marginally better carbon source in the absence of glucose



Problem 1 Part iii

This part aims to improve the production of 2,3-butanediol (EX_btd_RR_e) while simultaneously maximising the growth rate. Therefore, the problem can be cast as Flux scanning based on enforced objective flux. The code and function used were taken from: https://github.com/RamanLab/co-FSEOF/blob/main/FSEOF.m

3. The FSEOF function returns reactions that can be overexpressed or deleted to improve the production of the desired product (2,3-butanediol) with the additional constraint of maximising biomass.

4. Since the function returns the reactions, **findGenesFromRxns in NanoCobraToolbox** was used to get the mapping to corresponding genes.

The figure below shows the key targets that can be used to improve 2,3-butanediol production. Please refer to the code output and solution variables for the finer details.

- 10 reactions were overexpression targets
- 389 reactions were deletion targets to improve the production
- 21 genes were identified as overexpression targets
- 606 genes were identified as knockout targets.

Problem 2 Part I

The **Mycobacterium tuberculosis model** has 661 genes, 1025 reactions, and 825 metabolites, while the **E. coli core model** has only 137 genes, 95 reactions, and 72 metabolites.

- By comparing the model metadata, we can conclude that the number of reactions and metabolites increases much faster as the number of genes increases.
- This is expected as more genes imply more proteins being encoded. Subsequently, these enzymes can

catalyse more reactions and interact with each other to form new pathways and byproducts.

Problem 2 Part II

The uptake reactions were found using findExcRxns, and the reactions were mapped to metabolites using findMetsFromRxns. Both the functions are available in the NanoCobraToolbox library. This analysis revealed that M. tuberculosis takes up 17 essential nutrients below.

Carbon Sources	Glucose, Glycerol, Citrate
Nitrogen Sources	Ammonium, Glutamate
Mineral Sources	H+, phosphate, sulphate, K+, bicarbonate, Na+, Fe3+, Cu+, Cl−, Ca2+
Miscellaneous	O2, H2O

The essential nutrients were found by iteratively performing singleRxnDeletion on all 17 uptake reactions. The following 5 reactions were found to be essential.

- Essential carbon source- Glycerol
- Nitrogen sources were not essential, suggesting the cell has alternate pathways to produce metabolites involving N2
- Essential minerals- phosphate, sulphate, Fe3+
- Oxygen was also found to be essential

When abundantly available, the best nutrient sources were taken to be the ones that resulted in the highest growth rate using flux balance analysis.
This was practically coded by setting all uptake rates of the 17 nutrients to -1 and iteratively setting the uptake rate of the nutrient under study to -1000.

- Glucose and Glycerol are the best C sources. They give a growth rate of 0.30597 per hour when abundantly present separately!
- Ammonium was the best N2 source, giving a growth rate of 0.053113 per hour when abundantly present separately. This again suggests that N2 is not essential for the cell's growth.
- All the mineral sources are equally good (or bad) as each being separately maximally present achieves a growth rate of 0.05255 per hour.
- Thus having a high amount of Carbon is essential for high cell growth.

Growth rate with nutrient made maximally availible:

Nutrient	GrowthRate
{'glyc[e]' }	0.30597
{'glu__L[e]'}	0.30597
{'cit[e]' }	0.14403
{'ca2[e]' }	0.053113
{'so4[e]' }	0.05255
{'k[e]' }	0.05255
{'h2o[e]' }	0.05255
{'h2co3[e]' }	0.05255
{'glc__D[e]'}	0.05255
{'nh4[e]' }	0.05255
{'na1[e]' }	0.05255
{'fe3[e]' }	0.05255
{'cu2[e]' }	0.05255
{'cl[e]' }	0.05255
{'o2[e]' }	0.05255
{'h[e]' }	0.05255
{'pi[e]' }	0.05255

Combinations of different nutrients being abundantly available were by creating pairs of abundantly available nutrients and simulating with Flux Balance analysis similar to as described above, i.e., the uptake rate of the nutrient pair under study is set to -1000, and all others are set to -1.

- Glucose and Ammonium are the only maximally available nutrients, so the growth rate is 0.9023 per hour.
- With Citrate and Glutamate being the only maximally available nutrients, the growth rate is only 0.05255 per hour.

- Glucose and ammonium being available in excess is much more beneficial for the cell.

Problem 2 Part III

Essential genes in M. tuberculosis were identified using **singleGeneDeletion**. Please see the code and solution variables for the gene IDs. Additionally, I have attached a CSV file containing the gene IDs and the gene names.

- 188 genes out of 688 genes were identified as essential
- Potentially, targeting any of these genes would kill the M tuberculosis cell.
- Only 174 of the genes were annotated, i.e. had a RefSeq ID. Therefore, there are 14 essential genes whose functions are unknown.

The **Polyprenol-monophosphomannose (PPM) synthase Ppm1 gene** (Rv2051c in the model structure) in *Mycobacterium tuberculosis* is crucial for the synthesis of PPM, a key intermediate in the biosynthesis of lipoglycans, such as lipoarabinomannan (LAM) and arabinogalactan.

These lipo glycans are components of the mycobacterial cell envelope and are responsible for structural integrity and protecting the bacterium from environmental stresses
Source: https://pmc.ncbi.nlm.nih.gov/articles/PMC5062980/

Problem 2 Part IV

The Acinetobacter baumanii model has 709 genes, 1015 reactions, and 888 metabolites.
In contrast, the Mycobacterium tuberculosis model has 661 genes, 1025 reactions, and 825 metabolites, while the E. coli core model has only 137 genes, 95 reactions, and 72 metabolites.
Essential genes in A. baumanii were again identified using **singleGeneDeletion.**

- Enzyme promiscuity, alternate reaction pathways, and dead-end metabolic pathways make it impossible to predict the number of reactions and metabolites simply based on the number of genes.
- We see that Acinetobacter baumanii has more genes but fewer reactions.
- Only 74 out of 709 genes were found to be essential, which is unexpectedly low.
- Only 2 of the 74 essential genes were annotated, i.e. had a RefSeq ID, implying that this organism is not well studied.

The following genes are essential in both M. tuberculosis and A. baumanii:

{'adk' }Fo {'aroB'}{'asd' }{'cdsA'} {'dapA'}
{'dapD'}{'dapE'}{'dapF'} {'dxr' } {'glmS'}
{'glmU'} {'gmk' }{'ispD'} {'ispE'} {'ispF'}
{'metK'} {'murA'} {'murB'} {'murC'} {'murD'}
{'murE'} {'murF'} {'murG'} {'ppa' } {'psd' }
{'pssA'} {'purB'} {'pyrG'} {'serC'} {'tkt' }

Out of the above list, PyrG is well studied and is known to code for the enzyme orotidine 5'-phosphate decarboxylase, essential to form uridine monophosphate (UMP). UMP is a key metabolite in RNA biosynthesis and is vital for both these species.

Surprisingly conserved genes

Following are some surprisingly conserved genes:

1. Isoprenoid Biosynthesis Genes (dxr, ispD, ispE, ispF)
   The conservation of the complete MEP pathway genes is remarkable, considering the following:
   a. These bacteria belong to different phyla (Actinomycetota vs. Pseudomonadota)
   b. They have fundamentally different cell envelope structures

2. Phospholipid Biosynthesis Genes (cdsA, psd, pssA)
Despite dramatically different membrane compositions:
   M. tuberculosis has a complex membrane containing mycolic acids. A. baumannii has a typical Gram-negative envelope

3. Diaminopimelate Pathway Genes (dapA, dapD, dapE, dapF)
   The presence of all four genes is unexpected because:
   Diaminopimelate plays different structural roles in each bacterium's cell wall and other bacteria have alternate pathways for lysine biosynthesis. Further, the pathway involves multiple and it is expected that they would evolve differently.

Common Essentiality Explanations

Common essentiality is a

1. The Mur ligases (MurA-G) represent a universal requirement because:
   Despite peptidoglycan forming ~90% of Gram-positive cell walls but only ~10% of Gram-negative walls, it remains structurally critical for both

2. Nucleotide Metabolism Genes

Adenylate kinase (adk) is essential because it "regulates adenine nucleotide homeostasis which is crucial in cellular viability and cell energy

3. Amino Acid Biosynthesis Genes
   Aspartate-semialdehyde dehydrogenase (asd) forms an early branch point in the metabolic pathway forming lysine, methionine, leucine and isoleucine from aspartate. This pathway also produces diaminopimelate which plays an essential role in bacterial cell wall formation

4. Methyl Cycling Function
   S-Adenosylmethionine synthetase (metK) produces SAM, a methyl donor for transmethylation" that allows DNA methylation and switches the genes off," thus controlling gene expression

Closing thoughts

The conservation of these 30 essential genes across such divergent bacteria reinforces the concept of a bacterial ”minimal genome”.
The most surprising conserved genes relate to isoprenoid biosynthesis, phospholipid metabolism, and cell wall precursor formation, given that these processes might have been expected to diverge more dramatically based on the structural differences between these organisms. However, their conservation highlights how fundamental biochemical pathways can remain essential despite billions of years of divergent evolution

Sources

Crystal structure of IspF from Bacillus subtilis and absence of protein complex assembly amongst IspD/IspE/IspF enzymes in the MEP pathway | Bioscience Reports | Portland Press
New Insight into Isoprenoids Biosynthesis Process and Future Prospects for Drug Designing in Plasmodium
Perturbations of Phosphatidate Cytidylyltransferase (CdsA) Mediate Daptomycin Resistance in Streptococcus mitis/oralis by a Novel Mechanism

Architecture and genomic arrangement of the MurE–MurF bacterial cell wall biosynthesis complex | PNAS
Architecture and genomic arrangement of the MurE–MurF bacterial cell wall biosynthesis complex | PNAS
Adenylate kinase from Streptococcus pneumoniae is essential for growth through its catalytic activity - PMC
Aspartate-semialdehyde dehydrogenase - Wikipedia
S-Adenosylmethionine synthetase enzyme - Wikipedia