BT5240 - Assignment 4

RA Keerthan
CH17B078

1)

I selected the paper named "*RobOKoD: microbial strain design for (over)production of target compounds*" [1] for the purpose of a critical study. The following is the summary that is sectioned as per the requirement.

Brief abstract on the study:

Designing strategies to over-produce desirable product is of vital importance, given the constraints related to the viability and sustainability of the cell that harbors the productivity of the product/target. Popular techniques such as OptKnock and RobustKnock, that existed before the proposal of RobOKoD primarily aimed to design a strategy for over production solely based on gene-knockouts. However, these techniques did not account for over-expression or dampening of certain genes, which could also favor over-production of desirable product/compound. To address this, the authors of RobOKoD proposed a method which is based on Flux Variability Analysis to identify potential knockouts, over-/under-expression of genes. Results using RobOKoD were compared against OptKnock and RobustKnock to design an E. coli strain with a reverse β-oxidation cycle for butanol production. This comparison showed the strategies predicted by RobOKoD are more suited for over-production of target compounds than OptKnock and RobustKnock.

Methods used to arrive at a particular strategy:

The proposed RobOKoD method is an iterative algorithm to rank/select modifications that can be made in the metabolic system inorder to enhance the target productivity. The iteration involves three key steps:

i. Computation of *Metabolite Consumption Test* (MCT) score which identifies if a given metabolite in the target production pathway could lead to flux loss to biomass production. In the event of flux loss, all reactions that consume the given metabolite is preferred for potential knock-out.
ii. *Flux variability Analysis profiling* (FVAp), which is similar to a profile analysis as done in FSEOF, is done to each reaction. From the profile, a score is computed which would signify the importance of a reaction for target production and cell growth.
iii. Preferential knock-out of reactions which either divert the carbon flux away from the desired product production or consumes a metabolite that enhances flux loss from the production of target.

After step iii), we go back to step i) thereby starting a new iteration. Using the results of steps i.) and ii.), the potential modifications/strategies are ranked. Candidate modifications would include

a) altering the conditions of the environment
b) gene-deletion
c) gene over-expression
d) gene-dampening

The ranking is based on attributing each reaction to a knockout score. Higher knockout score would imply that reaction can be knocked-out (provided the reaction is non-lethal). If there is a tie in knockout score, the reaction with high MCT score is prioritized for

knockout. Apart from knockout score-based ranking, the targets which are to be over-expressed or damped are also ranked based on computation of over-expression score. Ranking these modifications would give user the flexibility on which strategy to choose for increasing target productivity.

Main findings:

For validation of the proposed approach, experiments were done on E.Coli metabolic model. The MCT score identified that pyruvate was a key metabolite which could lead to flux loss to biomass production. Based on this, 11 reactions which consume pyruvate was identified and were preferred for a potential knock-out.

The results also showed that RobOKoD predicts that certain genes that encodes relevant transport proteins are to be dampened inorder to enhance productivity of butanol (product). This could be treated as an additional strain improvement strategy.

While comparing the efficiency of RobOKod with that of OptKnock and RobustKnock, it was observed that with the case of OptKnock and RobustKnock, the knockout predictions were not ranked and were deterministic i.e, different/equivalent knockouts which could give rise to the same phenotype could not be predicted. However, RobOKod would quantitively yield the essentiality/significance of different modifications using which the user can verify if modifications having similar score give rise to same phenotypes. Thereby, RobOKod makes it possible to rank as well as to capture the set of modifications which corresponds to the same phenotype.

Main challenges they might have faced during their research work:

I could think of three main challenges that they might have faced during their research

- The optimal target production has a direct tradeoff with the sustainability of the cell. Hence the task of finding the pathways that could improve target productivity while minimizing the cell decay is pretty challenging. While re-routing fluxes to pathways that enhance target productivity, one should also make sure that the pathways that are essential for cell maintenance gets adequate resources.
- The constraint-based analysis is not always applicable/appropriate for predicting a modification in strain design inorder to enhance target production. FBA has a shortfall in predicting side reactions that can carry flux *in silico*. This is because any such flux to the side reactions could suppress the necessary resources needed for optimal growth. This could have broken down their entire research work as it primarily revolved around FBA and FVA. However, since the objective is to find a viable strain design strategy, this defect of FBA can be slightly overlooked but still poses as a limitation of this work wherein desired product production pathways which are not highly tied with growth could not be identified.
- There does not exist any one-to-one relationship between a gene and a reaction. Therefore, if at all FVA identified a reaction as a potential knockout, identifying the corresponding set of knockout genes is tricky. There could exist some genes which are mapped to a knockout reaction and several other essential reactions. Hence knocking out such genes could reduce the optimal target production, which is not desirable. Hence their research work should have found it hard to identify the correct set of gene modification strategies given all these complications.

2.a)

As the first step, the BiGG IDs corresponding to formate, glucose, galactose and maltose are to be found out. This is done with the help of the keyword search in BiGG database (http://bigg.ucsd.edu/).

BiGG ID	Name
EX_for_e	Formate exchange
EX_glc__D_e	D-Glucose exchange
EX_gal_e	D-Galactose exchange
EX_malt_e	Maltose exchange

Their presence in the model is verified and stored using the below commands:

```
formate_bigg_id = 'EX_for_e';
glucose_bigg_id = 'EX_glc__D_e';
galactose_bigg_id = 'EX_gal_e';
maltose_bigg_id = 'EX_malt_e';

formate_id = find(ismember(model.rxns,formate_bigg_id));
glucose_id = find(ismember(model.rxns,glucose_bigg_id));
galactose_id = find(ismember(model.rxns,galactose_bigg_id));
maltose_id = find(ismember(model.rxns,maltose_bigg_id));
```

As the next step, we compute the maximum flux of formate exchange using *fluxVariability* function in CobraToolbox. We set *optPercentage* argument as 10 since we want the maximum flux such that biomass should be atleast 10% of wild-type.

Carbon source	Max formate production ($\mathrm{mmolgDW^{-1}h^{-1}}$)
Glucose	84.0164
Galactose	83.6872
Maltose	156.8031

It is aerobic condition because there is oxygen uptake in each of the different conditions. This is verified by checking the lower bound of 'EX_o2_e' in each of these conditions.

For finding the lethal genes, we use *singleGeneDeletion* which is in-built function in CobraToolbox. The genes corresponding to a grRatio of 0 are considered to be lethal. The following code snippet does this job.

```
[grRatio, grRateKO, grRateWT, hasEffect, delRxn, fluxSolution] = singleGeneDeletion(model);
lethal_genes = model.genes(find(grRatio == 0))
```

The output for all the three carbon sources is like:

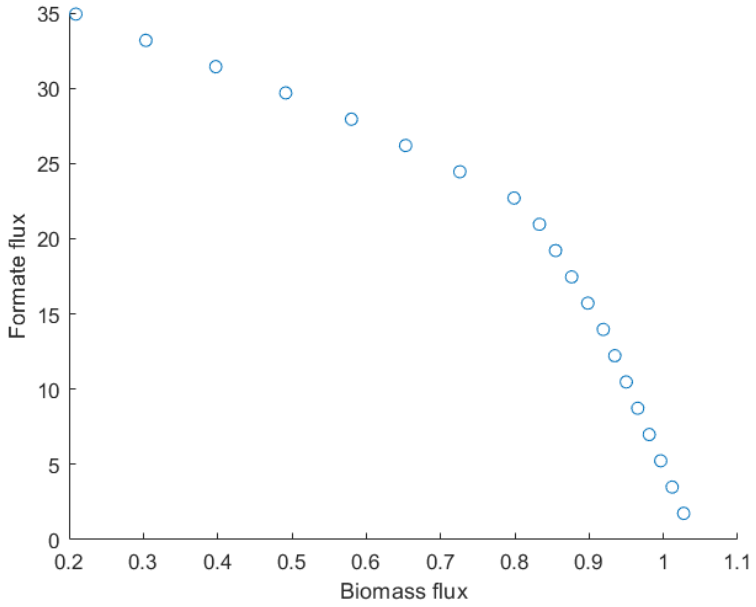
```
Single gene deletion analysis in progress ...
100%    [.....................................]

lethal_genes =

  0×1 empty cell array
```

This implies that no lethal genes were found in all the three different conditions. It was quite surprising but all I could find was lethal reactions but not lethal genes. Since more than one gene can correspond to a reaction in general, this might indicate that there could be lethal gene-pairs (or triplets or other combinations) without lethal genes.

2.b)

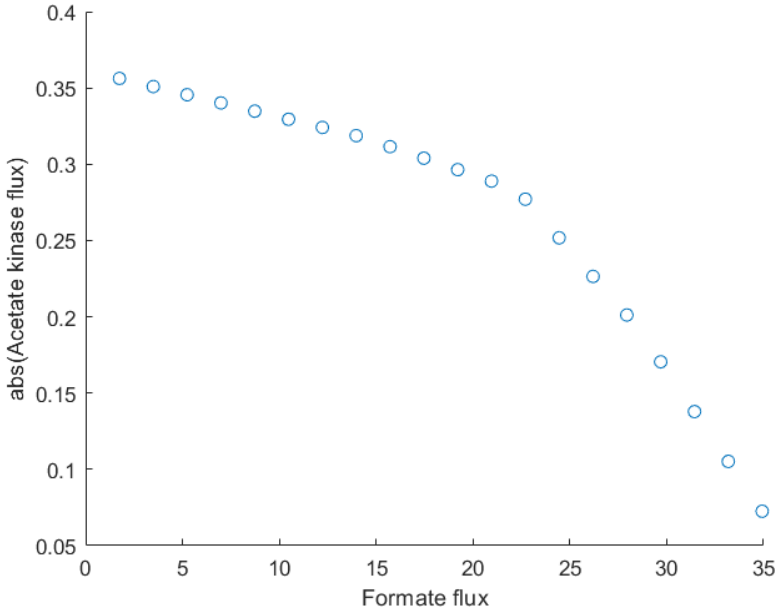
According to FSEOF, potential knockouts have target flux decreasing with increasing product flux. Over-expression targets will have flux increasing with increasing product flux. The source condition is set to default. The maximum product flux (corresponding to biomass being atleast 0% of wild-type) is 38.8065 $\mathrm{mmolgDW^{-1}h^{-1}}$. It is varied from 90% of the maximum value to its minimum value (corresponding to maximum biomass production) with 20 steps.

The formate flux vs biomass flux is shown below:

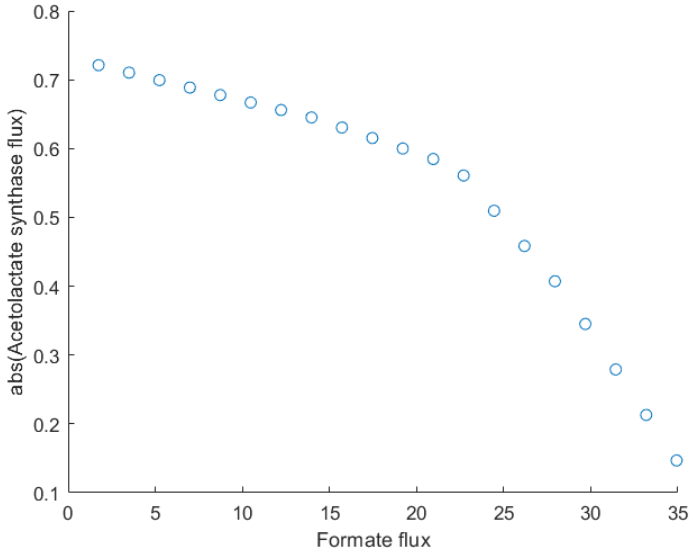


<u>Potential knockouts</u>

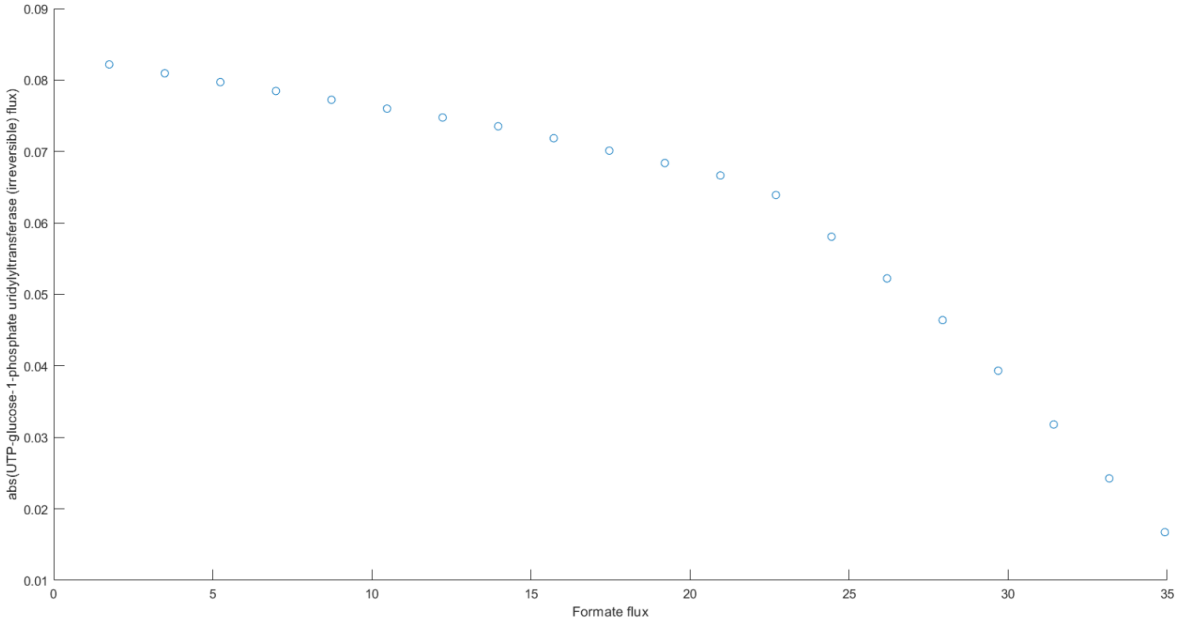
1. Acetate kinase



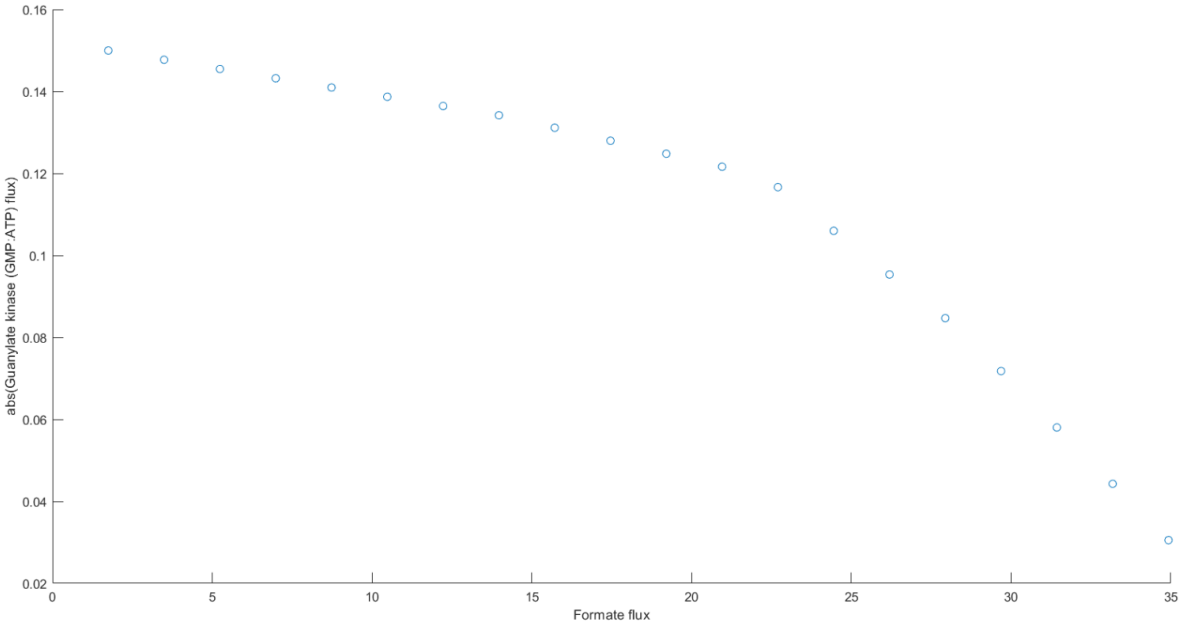
2. Acetolactate synthase



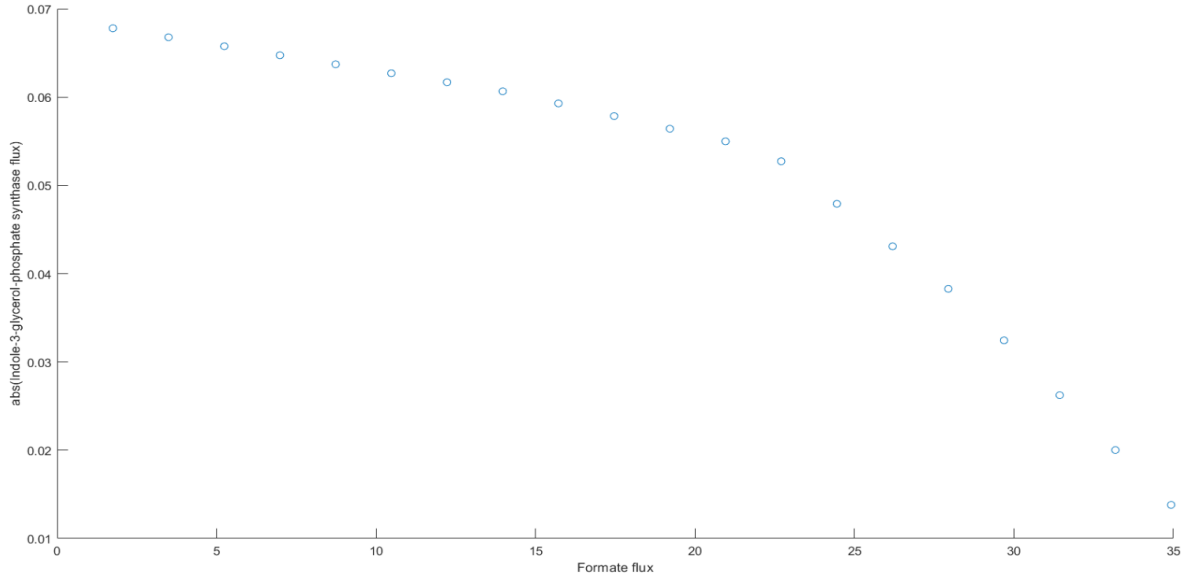
3. UTP-glucose-1-phosphate uridylyltransferase (irreversible)



4. Guanylate kinase (GMP:ATP)

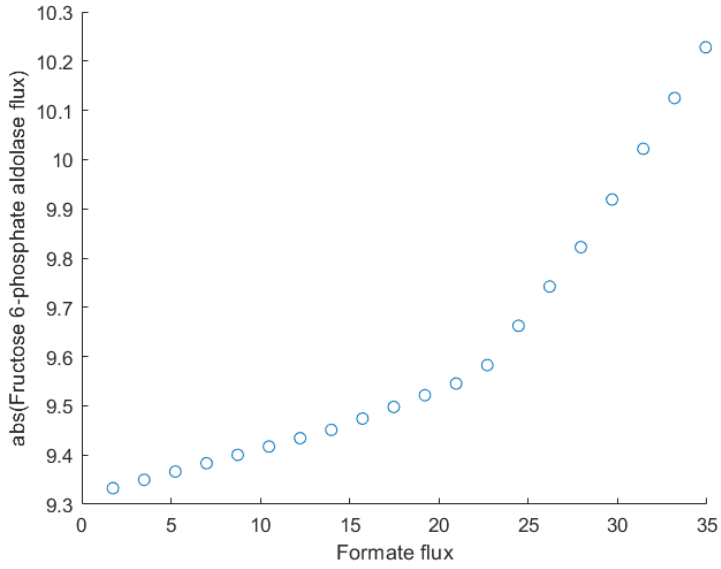


5. Indole-3-glycerol-phosphate synthase

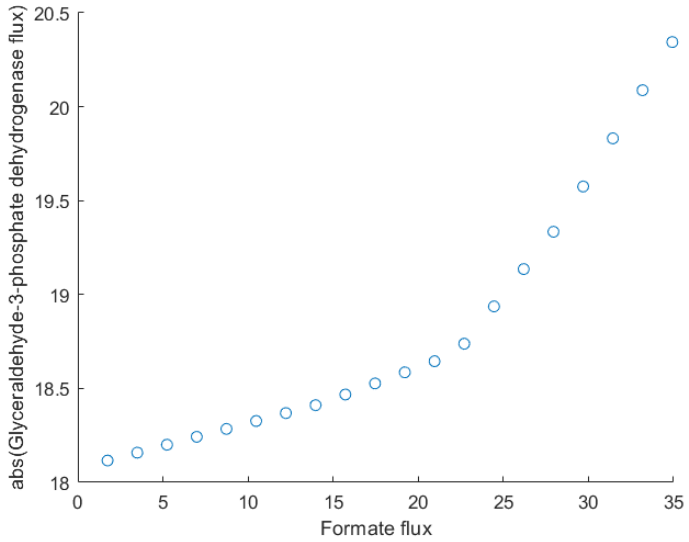


Over-expression

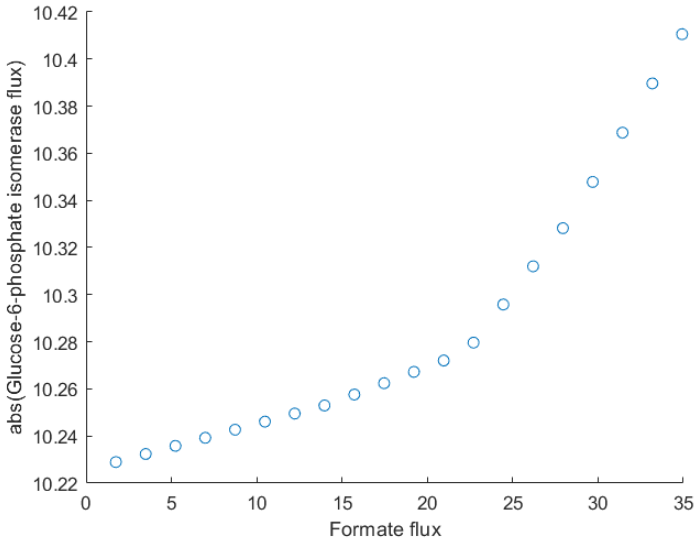
1. Fructose 6-phosphate aldolase



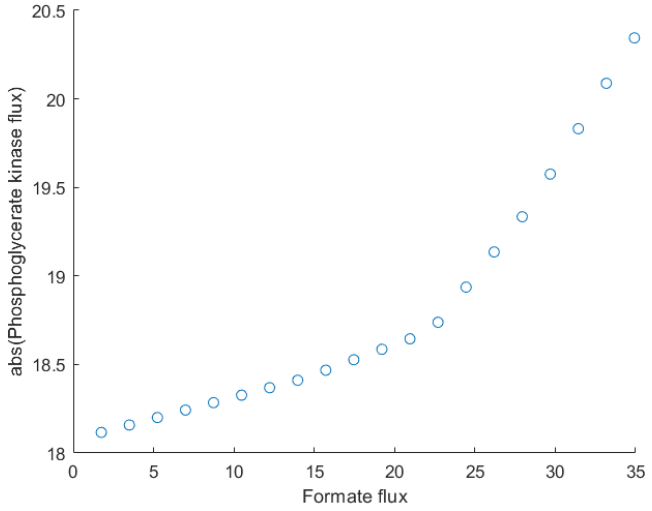
2. Glyceraldehyde-3-phosphate dehydrogenase



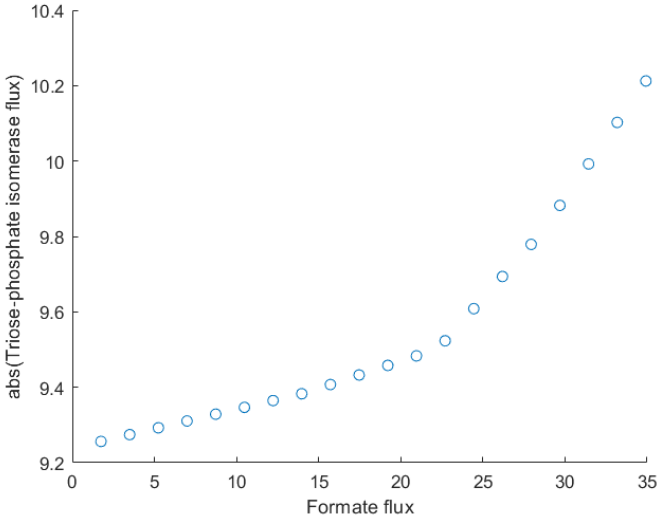
3. Glucose-6-phosphate isomerase



4. Phosphoglycerate kinase



5. Triose-phosphate isomerase



References

[1] Stanford Natalie J., Millard Pierre, Swainston Neil, *RobOKoD: microbial strain design for (over)production of target compounds*, Frontiers in Cell and Developmental Biology, Vol. 3, doi: 10.3389/fcell.2015.00017