Drug Target Identification Using Flux Balance Analysis

Ayush Deva 201501098
Pranay Gupta 20161088

1. Introduction

A drug target is a key molecule, specific gene or protein, involved in a particular metabolic or signaling pathway specific to a disease or for the survival of a microbial pathogen, whose activity is aimed to be modified through a drug resulting in desirable therapeutic effect. [4.5]

Essentially, in the process of discovering the drug for a particular disease or against a particular pathogen, scientists look for molecules like certain proteins, enzymes that are essential to the survival of the pathogens which their drugs can act upon without much side-effects on other metabolisms. To understand which molecules are essential, apart from experimental observations, insights can be drawn from utilizing our knowledge of the biological networks, for example, metabolic networks and protein interaction networks.

Flux Balance Analysis (FBA), a stoichiometric analysis technique, is one such class of methods for drug target identification that provides us with flux distribution between various reactions in a metabolic network which can then be used to **a**. identify essential metabolites and reactions that can serve as potential drug targets and **b**. study the response to perturbations such as gene deletions [7,8].

2. Background

In this project, we studied the comprehensive model presented by Raman et al. [3] of metabolic pathways of mycolic acid essential to the survival of Mycobacterium Tuberculosis, the bacteria responsible for TB. They perform FBA on this model to gain insights into the metabolic capabilities of the pathway and analyze proteins essential to it, thereby identifying potential drug targets.

2.1 **Why are Mycolic Acid Pathways (MAP) being studied?**

The bacteria responsible for Mycobacterium tuberculosis has a distinctive thick cell wall. This wall prevents antibiotics and the immune system to fight with the bacterium. Mycolic acid is a major constituent of this protective layer and has been shown to be critical for the survival of the bacterium [1]. In fact, many of the front-line anti-TB drugs such as isoniazid and ethionamide have targeted the proteins and enzymes involved in the synthesis of mycolic acid.

3. Methodology

The MAP model involves 28 proteins and 197 metabolites participating in 219 reactions. Each reaction is associated with a flux value (the rate of turnover of molecules). In addition to that, we have 28 exchange fluxes originating from external metabolites such as ATP, NADP, ADP etc. indicating their entry into or exit from the MAP system.

3.1 **Objective Functions**

The cell wall contains varying proportions of different types of mycolates, the optimal production of which can be modeled in two different ways:

a. **Obj1**: Mycolates can be produced and contribute to the objective independently but with different priorities (so the production of the most important mycolate is preferred).

$$< c_1 \,.\, v > \;=\; -0.4926v \;-\; 0.2334v_{cis-methoxy} \;-0.0327v_{trans-methoxy} \;-\; 0.2117v_{cis-keto} \;-0.0297v_{trans-keto}$$

b. **Obj2**: The ratios of different mycolates are fixed, so all of them contribute to the objective (in particular ratios) together and not independently.

$$< c_2 \,.\, v > \;=\; -1.00v_{mycolates}$$

where $v_{mycolates}$ represents the flux of the hypothetical reaction:

$$0.4926\ \alpha mycolate \;+\; 0.2334\ cis-methoxy-mycolate \;+\; 0.0327\ trans-methoxy-mycolate \;+0.2117\ cis-keto-mycolate \;+\; 0.0297\ trans-keto-mycolate$$

3.2 **Stoichiometric Matrix**

Using the data about the 197 metabolites and 219 reactions [1] (available here), we construct a stoichiometric matrix **(S)** corresponding to this model of size 197 x 247, each row corresponding to a metabolite and each column to each of the 219 reactions and 28 exchange fluxes.

For the analysis corresponding to the second objective function, we add another metabolite called *mycolates* (#198), one reaction corresponding to the one mentioned above and an exchange flux corresponding to *mycolates* which is what we want to maximize. Thus, in this case, our stoichiometric matrix is of size 198 x 249.

3.3 **Integer Linear Program (ILP)**

The ILP corresponding to our Flux Balance Analysis model would be:

Decision Variable: $v = [v_1\ v_2\ ....\ v_{219}\ b_1\ b_2\ ....\ b_{28}]$

where v_i denotes the reaction fluxes and b_i denotes exchange fluxes

Note: For Obj2, we add v_{220} & b_{29} as described above

Objective: $min\ c^T v$

Constraints:

$Sv = 0$ Net flux for any metabolite is 0 at steady state

$0 < v_i < 1$ Upper and Lower Bounds on Reaction Fluxes

$-1 < b_i < 1$ Upper and Lower Bounds on Exchange Fluxes

4. Results

The ILP was solved using *linprog* utility in MATLAB which uses the Interior-Point Algorithm to solve for the same.

4.1 **Without any intervention**

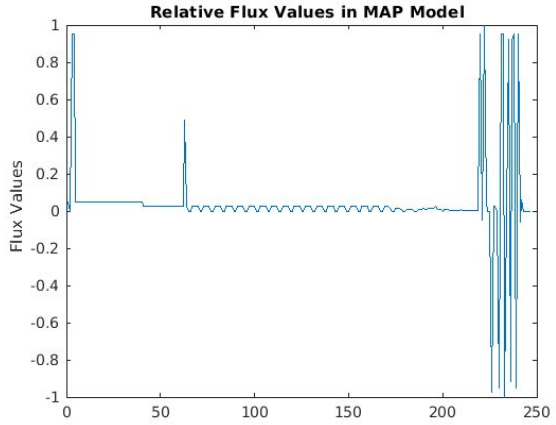


Fig. 1a: For Obj1

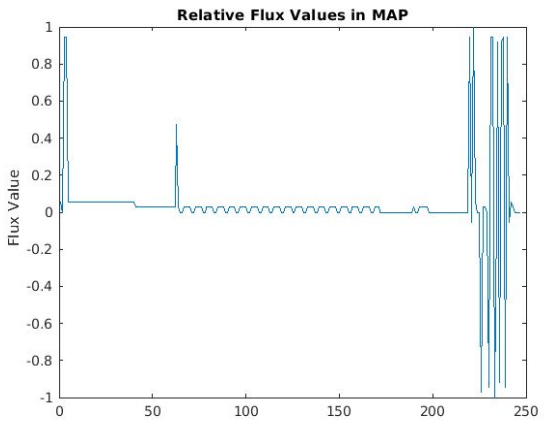


Fig. 1b: For Obj2

(Note: Click on the link for enlarged images)

4.2 **Gene Deletions**

In the attempt to identify a potential drug target, [3] studies the effect of in silico gene detection for each of the 28 genes. If a gene is deleted, all the reactions corresponding to it are stopped, i.e., flux is constrained to be 0 for such reactions. We can then analyze the change in flux (for each of the two objective functions) and depending on that, classify whether or not a gene is essential to the metabolic pathway.

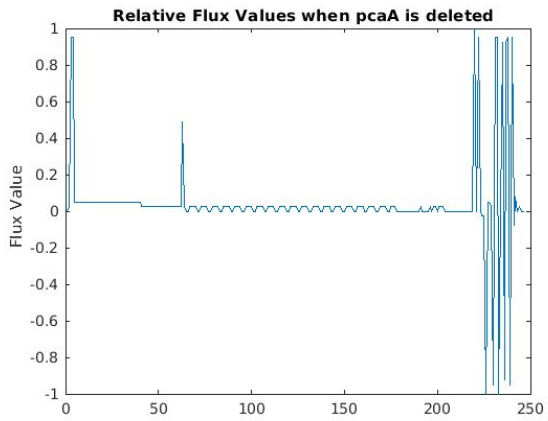
In this report, we provide results for two specific genes (the ones also analyzed in detail by [3]), pcaA, and InhA, due to their interesting results as we shall see.

The results for the following experiments are performed using the same ILP as described above with an additional constraint:

$v_j = 0 \quad \forall\, j \in X$ where X is the set of reactions inhibited by the deletion of the gene.

4.3 **Inhibiting pcaA**

pcaA is a protein essential for one of the reactions (#194) that is essential for the production of α-mycolate. Upon deleting pcaA, this reaction is inhibited, thus there is no production of α-mycolate. Now, since the second objective function (Obj2) requires fixed ratios of all the mycolates, the biomass produced and the overall flux distribution goes to zero (Fig. 2a), while in the first objective function, even though the production of α-mycolate is inhibited, the other mycolates are produced and the overall flux distribution remains relatively unchanged (Fig. 2b).



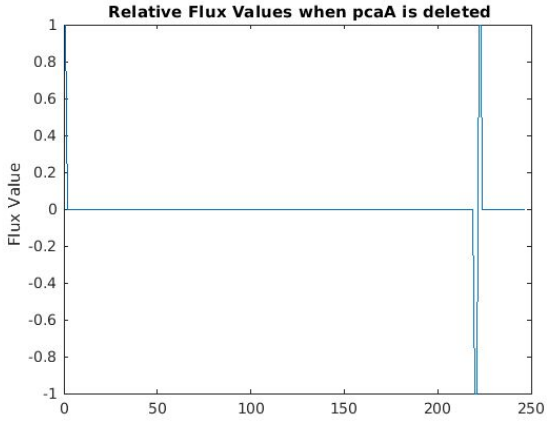


Fig. 1a: For Obj1

Fig. 1b: For Obj2

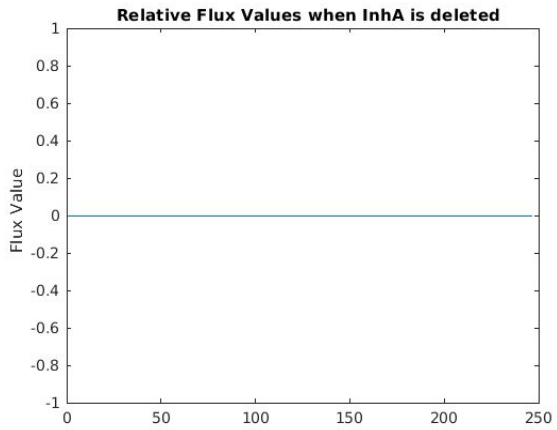
Experimental evidence for pcaA deletion shows that the bacilli can grow and survive for limited periods by producing a significant excess of keto-mycolate, to compensate for the absence of α-mycolate, in line, with our output corresponding to Obj1. Thus, for our case, Obj1 seems to be a better representation of the objective function that Obj2. Hence, from henceforth, we shall focus our discussions through results obtained from Obj1.

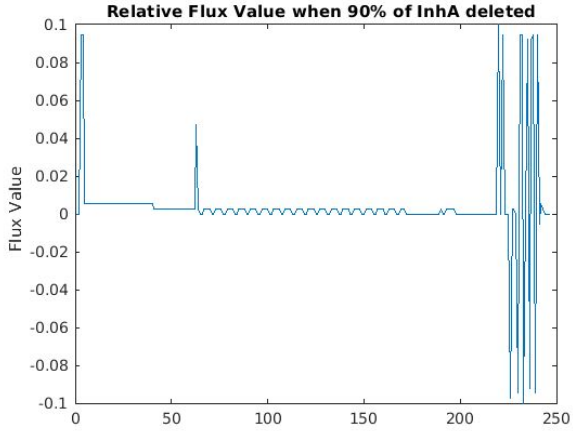
Point to Note

Fig. 1b doesn't completely match with the one in [3]. The reason for this is that the fluxes for reactions 1-2 and the exchange fluxes for their corresponding external metabolites are not affected by this inhibition. However, the flux values for these exchange fluxes do not appear in the objective function, hence, the relative flux values can take multiple values as long as it satisfies the constraint (i.e., net flux is 0). Here, we see that these fluxes take the extreme values of either 1 or 0. We shall see such discrepancies later as well (when inhibiting InhA for example) and the reason for them would be the same (unless stated otherwise).

4.3 **Inhibiting inhA**

Next, we study the effect of inhibiting inhA which catalyzes every sixth reaction from 69 to 189. Inhibiting inhA stops the production of C_n-acyl-ACP, a metabolite essential for the production of all types of mycolates. Hence, by inhibiting inhA, there are no mycolates produced and the relative fluxes go to zero for almost all reactions except reactions 1-2 and their corresponding external metabolites. In fact, inhA is a well-known target for drugs such as isoniazid and ethionamide.





[Fig. 3a:](#) For Obj1

[Fig. 3b:](#) For Obj1 (90% of InhA inhibited)

Again, the graph is not exactly similar to [3] for reasons mentioned above in section 4.3 -> Point to Note. For Fig. 3b, notice the change in the relative scale of the y-axis. The overall flux distribution looks very similar but the values are much inhibited (almost 90% for all the fluxes).

5. Side Effects *(Our contribution)*

As we saw in the example of inhibiting InhA, it not only inhibits the production of the undesirable mycolates but also affects other metabolites which may be useful for other metabolisms. Inhibition of such metabolites would lead to unintended side-effects due to the intake of drug targeting InhA. Hence, it may not be in our best interest to completely inhibit InhA and we must find the correct amount of dosage that leads to partial inhibition of InhA which maintains the production of certain essential metabolites above a certain threshold.

5.1 **Background** (talk about the other paper)

The idea was inspired by a paper by Li et al. [2] in which the authors study what they call the mass flow of metabolites in the metabolic network. Mass flow of a metabolite is defined as the sum of all the fluxes that lead to the production of that metabolite (see Appendix 1).

In the pathologic state, the mass flows of some metabolites are out of healthy ranges which directly result in the disease symptoms. As a result of drug intake, and thereby inhibition of certain enzymes and reactions, the mass flow of such metabolites can be brought back to the normal range but might lead to side-effects, i.e., pushing other metabolites outside their healthy range. While this may not be completely avoidable, we can try and minimize this side-effect through certain control mechanisms.

In our metabolic network of Mycolic Acid Pathways, we do not have a well-defined healthy range for the metabolites, rather, we wish to minimize the production of these mycolates (by inhibiting the InhA protein to a certain extent) such that the mass flow of other essential mycolates doesn't drop below a certain threshold.

5.2 **How much of InhA to inhibit to ensure a particular metabolite in a healthy range**

To define the problem formally, as we saw the results for inhibiting InhA by 90%, we need to

come up with the maximum value of $\alpha \in [0, 1]$ such that by inhibiting InhA by a factor of α, the mass flow of essential metabolites does not drop below a certain threshold.
Let X be the set of all the reactions inhibited by inhibiting a particular enzyme (here, InhA), Y be the set of all the essential metabolites whose mass flow needs to be controlled, $\Lambda = \{\lambda_i\}_{i \in Y}$ be the threshold values (lower limit) of the mass flow for these metabolites and v_i^{ss} be the steady-state flux value for the reactions inhibited. Then, we have,

$max.$	α	
$s.t$	$Sv = 0$	(Flux Balance)
	$v_i <= (1-\alpha)v_i^{ss} \quad \forall\ i \in X$	(Fluxes are inhibited by a factor α)
	$x_i = mass\ flow\ (see\ Appendix\ 1)$	(Mass Flow)
	$x_i >= \lambda_i \quad \forall\ i \in Y$	(Mass Flow Above Threshold)
	$0 <= v_i, x_i <= 1$	(Upper & Lower Bounds)
	$0 <= \alpha <= 1$	(Upper & Lower Bounds)

5.3 **Experiments and Results**
While our model is generalizable for any enzyme (to inhibit) and any set of essential metabolites whose mass flow needs to be regulated, here, we choose InhA as our drug target and the Malonyl CoA as our essential metabolite. Malonyl CoA plays a key role in chain elongation in fatty acid biosynthesis and polyketide biosynthesis [6]. Malonyl CoA is produced by a single reaction (#4) and the relative steady-state flux value corresponding to that when no enzymes were inhibited was 0.9474, when InhA was 90% inhibited it was 0.0947 and 0.0 when InhA was completely inhibited.

The following were the obtained values of α for different threshold levels of Malonyl CoA.

$\lambda_{Malonyl-CoA}$	α
0.0947	91%
0.9474	8%
0.3	71%
0.5	51%
0.7	31%

Note: The deviation of α from the expected values of 90% and 0% in the first two cases is because even when InhA is inhibited, the first few reactions that don't depend on it for their production can occur to some extent as long as they can be utilized in other reactions/exchange flux (similar to our Point to Note in section 4.3).

The relative flux distribution when $\lambda_{Malonyl-CoA} = 0.3$ is:

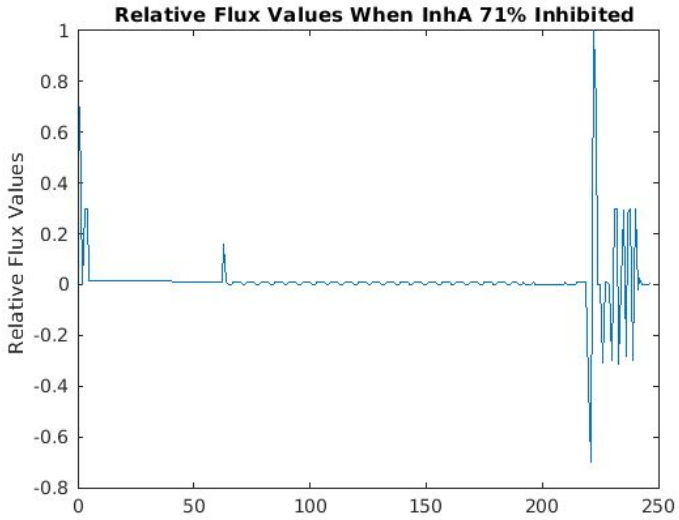


Fig. 4a: $\lambda_{Malonyl-CoA} = 0.3$

6. Conclusion

In this report, we study how we can apply Flux Balance Analysis (FBA) on metabolic networks of important compounds in a pathogen to identify potential drug targets that, if inhibited, stops the growth of the pathogen. In particular, we study the paper by Raman et al. [3], in which the authors analyze the Mycolic Acid Pathways, a compound present in the cell wall of the pathogen, Mycobacterium bacilli. We formulate the FBA problem as an ILP and consider two possible objective functions and show that performing FBA on one of them shows results much closer to experimental findings.

Later, we study the effect on the relative flux distribution after inhibiting certain enzymes to identify drug targets. It is noteworthy that our results match very closely, if not exactly, to the ones presented in the paper with certain deviations that were properly explained. We identify InhA as a potential drug target, which when inhibited, brings down the entire flux distribution to close to zero.

As part of our contribution to the findings in the paper by Raman et al., we analyze the side-effects of inhibiting the InhA enzyme. We show that by inhibiting one protein, we can inhibit the mass flow of other essential metabolites which is undesirable. So, we re-formulate our ILP to find the maximum inhibition to the InhA protein such that the mass flow of other essential metabolites stays above a certain threshold value. We provide the mathematical formulation of the ILP and also provide the results for our experiments on Malonyl-CoA as the essential metabolite for various threshold values.

7. References

[1] Draper P, Daffé M (2005) The cell envelope of Mycobacterium tuberculosis with special reference to the capsule and outer permeability barrier. In: Cole ST, editor. Tuberculosis and the tubercle bacillus. Washington, D.C.: American Society of Microbiology Press. pp. 261–273

[2] Li, Zhenping, Rui-Sheng Wang, and Xiang-Sun Zhang. "Two-stage flux balance analysis of metabolic networks for drug target identification." *BMC systems biology* 5.S1 (2011): S11.

[3] Raman, Karthik, Preethi Rajagopalan, and Nagasuma Chandra. "Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs." *PLoS computational biology* 1.5 (2005).

[4] Smith C: Drug target validation: Hitting the target. Nature 2003, 422:341-347.

[5] Smith C: Drug target identification: A question of biology. Nature 2004, 428:225-231.

[6] https://en.wikipedia.org/wiki/Malonyl-CoA

[7] Edwards JS, Palsson BO (2000) The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. Proc Natl Acad Sci USA 97: 5528–5533

[8] Becker SA, Palsson BO (2005) Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: An initial draft to the two-dimensional annotation. BMC Microbiology 5: 8.

Appendix 1

Suppose that there are *m* metabolites $C_1, C_2, ..., C_m$ and *n* reactions $R_1, R_2, ..., R_n$ in a metabolic network. Consider two coefficient matrices T and U such that the kth column of matrix T denotes the coefficients of reactants in reaction R_k , while the kth row of matrix U denotes the coefficients of metabolites produced by reaction R_k. Thus, the stoichiometric matrix defined in our project, $S = -T + U^T$.

Mass flow of a metabolite is defined as the total flow produced by all the reactions in the metabolic network. In the stead state, it is equal to the total flow consumed by all the reactions in the network. Mathematically, the mass flow for metabolite C_j is defined as:

$$x_i = \sum_{j=1}^{n} t_{ij} v_j = \sum_{j=1}^{n} u_{ji} v_j$$

Codes and Plots

All the MATLAB codes used for generating the above plots can be found here.