

---

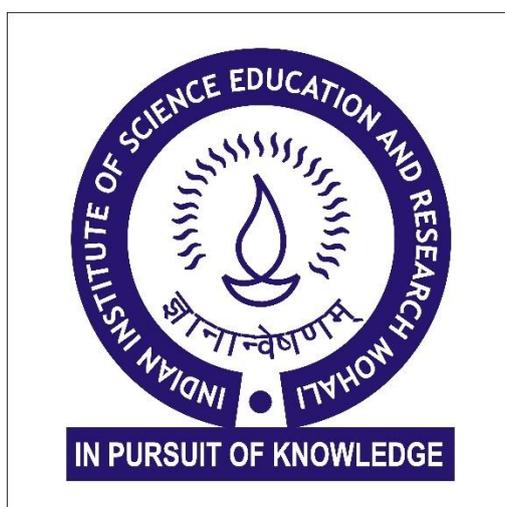
# EXPLORING ALTERNATE METABOLIC PATHWAYS IN SINGLE GENE DELETION STRAIN OF *ESCHERICHIA COLI* USING FLUX BALANCE ANALYSIS

---

Mridul

[MS13044]

*A dissertation submitted for the partial fulfilment of BS-MS dual degree in Science.*



Indian Institute of Science Education and Research Mohali

May 2018



## **Certificate of Examination**

This is to certify that the dissertation titled “Exploring alternate metabolic pathways in single gene deletion strain of *Escherichia coli* using Flux Balance Analysis” submitted by Mr. Mridul (Reg. No. MS13044) for the partial fulfilment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Kuljeet Singh Sandhu

Dr. Rajesh Ramachandran

Dr. Shashi Bhushan Pandit

(Supervisor)

Dated: April 20<sup>th</sup>, 2018



## **Declaration**

The work presented in this dissertation has been carried out by me under the guidance of Dr. Shashi Bhushan Pandit at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort has been made to indicate this clearly, with due acknowledgment of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Mridul

(Candidate)

Dated: April 20<sup>th</sup>, 2018

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Shashi Bhushan Pandit

(Supervisor)



## **Acknowledgement**

I cannot be grateful enough to my supervisor Dr. Shashi Bhushan Pandit who supported me in every way possible to help me complete my thesis. He always guided me out of all troubling situations and always motivated me to give my best. I would also like to thank all my lab members especially Deeksha Thakur for their invaluable support and insights both on and off the field.

I would like to express my gratitude to my parents and close friends who have been dependable pillars of strength for me all through my time here. Lastly I would like to acknowledge the support provided by department Biology and IISER Mohali and thank them for providing me an opportunity to pursue my desired project.

# List of figures

<b>Figure 1.1: Schematic representation showing .....</b>	<b>2</b>
<b>    a) metabolic pathway b) metabolic network.</b>	
<b>Figure 1.2: Showing various steps in flux balance .....</b>	<b>5</b>
<b>Figure 3.1: Comparison of chemical structures of carbon sources .....</b>	<b>13</b>
<b>Figure 3.2: Figure showing maltose (a) and glucosamine.....</b>	<b>14</b>
<b>    and N-acetyl glucosamine (b) utilization pathways.</b>	
<b>Figure 3.3: Scatter plot of fitness score of single gene deletion.....</b>	<b>16</b>
<b>    strain on glucose versus maltose.</b>	
<b>Figure 3.4: Scatter plot of fitness score of single gene deletion.....</b>	<b>18</b>
<b>    strain on glucose versus glucosamine.</b>	
<b>Figure 3.5: Scatter plot of fitness score of single gene deletion.....</b>	<b>18</b>
<b>    strain on glucose versus N-acetyl glucosamine.</b>	

# List of Tables

<b>Table 3.1: FBA computed biomass of wild type <i>E. coli</i> .....</b>	<b>20</b>
<b>Table 3.2: FBA computed single gene deletions under Glucose.....</b>	<b>21</b>
<b>Table 3.3: FBA computed single gene deletions under Maltose.....</b>	<b>22</b>
<b>Table 3.4: FBA computed single gene deletions under Glucosamine.....</b>	<b>23</b>
<b>Table 3.5: FBA computed single gene deletions under N-acetyl Glucosamine.....</b>	<b>24</b>

## Abstract

Microorganisms exhibit diverse metabolic capability and show physiological adaptation to genotype/environment perturbations. Metabolic robustness is attributed mostly to complex interplay and dynamic interactions among metabolic network components. In the event of disruption of metabolic gene function, it has been shown that fluxes are routed through alternate pathways to maintain constant flow of metabolites in order to sustain cellular growth. To understand resilient nature of *Escherichia coli* metabolic networks, the study of flux rerouting in single gene deletion strains can be studied using constraint based methods such as Flux Balance Analysis (FBA), which facilitates computation of *in silico* fluxes. Due to limitation of experimental growth rate in continuous culture condition for every single gene deletion strain, in the present study, we explore the possibility of using experimental large-scale single gene deletion in *E. coli* (fitness score data of generated from growth on solid media) to understand metabolic flux distribution in genetic perturbation. In the present study, we have used fitness scores of single deletion strains only on fermentable carbon sources *viz.* glucose, maltose, glucosamine, and N-acetyl glucosamine and used FBA with biomass function optimization to analyze flow of fluxes in alternate pathways.

# Contents

List of figures.....	i
List of Tables .....	ii
Abstract.....	iii
Chapter 1.....	1
Introduction.....	1
1.1 Metabolic network.....	1
1.2 Flux Balance Analysis (FBA) .....	3
1.3 Application of FBA on phenotype prediction in <i>Escherichia coli</i> .....	6
1.4 Enzyme promiscuity.....	7
1.5 Objective .....	8
Chapter 2.....	9
Methodology.....	9
2.1 <i>E. coli</i> Fitness score data .....	9
2.2 FBA modeling using COBRA ToolBox .....	9
2.3 FBA modeling of single gene deletion strains using COBRA ToolBox .....	10
Chapter 3.....	12
Results and Discussions.....	12
3.1 Comparative analysis of fitness scores data of <i>E. coli</i> single gene deletion mutants grown in fermentable carbon source .....	12
3.2 Metabolic flux analysis and comparison it with fitness score obtained from single gene deletion mutants.....	19
Conclusions.....	26
Bibliography .....	27

# Chapter 1

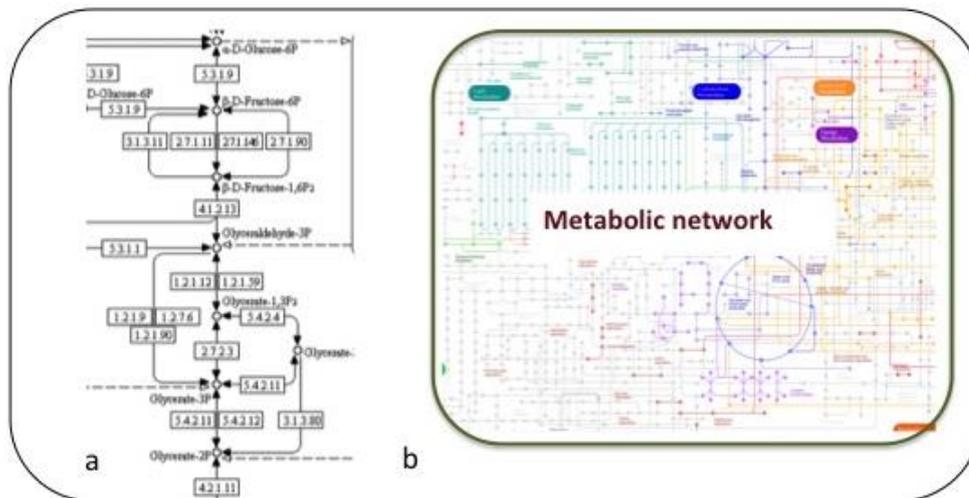
## Introduction

### 1.1 Metabolic network

Cellular metabolism consists of all chemical transformations catalyzed by enzymes or spontaneous reactions within a cell and forms the basis of any biological process. Traditionally, metabolism is described as organization of metabolic pathways, which usually is a linear arrangement of metabolites connected by enzymes responsible for chemical conversion of input substrate/s to output product/s (Figure 1.1a). In fact, enzyme function description also includes its role in a given metabolic pathway. The pathways can be organized into metabolism of a specific chemical compound such as ‘Lysine metabolism’ or ‘glucose metabolism’ that includes both catabolism (degradation) and anabolism (biosynthesis). In many instances, pathways can also be organized for give a generalized metabolic context such as ‘energy metabolism’, ‘lipid metabolism’, and ‘carbohydrate metabolism’ (Tanabe and Kanehisa, 2012; Kanehisa et al., 2017). Since metabolite/s in a given pathway can be present in multiple pathways or enzyme from a pathway can be involved in multiple pathways, it is impossible to study effect of perturbation of a metabolite or an enzyme on the cellular phenotype by using isolated metabolic pathways. This necessitated the description of metabolism as interconnected metabolites through enzymes as edges in metabolic networks (Figure 1.1b). Biochemical networks is one of the earliest mathematical application of graph based description and analysis (Jeong et al., 2000).

The system level approach to study intracellular metabolism has revolutionized understanding as well as prediction of phenotypic behavior of cell under a given environmental condition or genotypic perturbations (Dunphy and Papin, 2017; Yilmaz and

Walhout, 2017). Metabolic networks have been extensively used in metabolic engineering of cells to biosynthetically produce large quantities of a desired product, especially when chemical synthesis of such compounds is challenging (Guo et al., 2017).



**Figure 1.1:** Schematic representation showing a) metabolic pathway b) metabolic network. Images have been derived from KEGG database.

In a simplest description, metabolic networks are composed of nodes (vertices), as metabolites, and edges are enzyme/s, which catalyzes conversion of one metabolite to other. Further, depending on the reversibility of enzymatic reaction edge can be directional or undirected edges. Many studies have used graph theoretical approaches to derive global network properties such as topological properties and robustness (Samal and Martin, 2011) to determine the importance of individual enzyme or metabolites. However, limitation of this representation has been that one edge (enzyme) can link multiple input nodes and many output nodes (metabolites), of which only one or two metabolites are crucial in network description. Recent approaches have suggested hypergraphs to represent metabolic network (Carbonell et al., 2012).

Usually, metabolic network is reconstructed using information from annotated genes and known enzymatic reactions and their role in metabolic pathways. However, in the post-genomic era, metabolic networks are reconstructed using genomic, proteomic, and

phenotypic data obtained from various sources (Fiest et al., 2009; Orth et al., 2014). Such genomic scale metabolic networks are also referred as genome scale metabolic network models (GSMNM). One of the limitations in abstracting metabolism as metabolic networks is that regulation of cell metabolism such as variations in enzyme expressions, protein levels via translation or post translational modifications and feedback regulation of enzymes are not taken into account. Since phenotype is net result of both regulation of enzymes and metabolic pathways, the prediction of phenotype from genotype or varying environmental conditions becomes challenging. Attempts are made in recent studies (Guo and Feng 2016) to include regulation on static metabolic network to improve genotype-phenotype correlations. Despite deficiencies in description of metabolic network, it has been used extensively in predict range of cellular functions such as cellular growth capabilities on various substrates, predict phenotype under single/multiple gene knockouts on genomic scale, tracing carbon in radio active labeling studies (O'Brien and Palsson, 2014; O'Brien et al., 2015).

In metabolic networks, flux through a reaction that is rate of substrate conversion to product can be useful in prediction of phenotypes and validating of genome scale metabolic networks (Gianchandani et al., 2010). Usually, this is performed using Flux Balance Analysis (FBA), which is mathematical approach to analyze flow of metabolites in a metabolic network (Orth et al., 2010).

## **1.2 Flux Balance Analysis (FBA)**

Briefly, FBA involves describing metabolic fluxes using a set of linear equations where all metabolic reactions are balanced and concentrations of metabolites do not change over time also referred as steady state condition. Further, these linear algebraic equations of fluxes are solved under biological or chemical constraints to obtain optimal solution to generate flux distribution with optimizing an objective function (Orth et al., 2010). Hence,

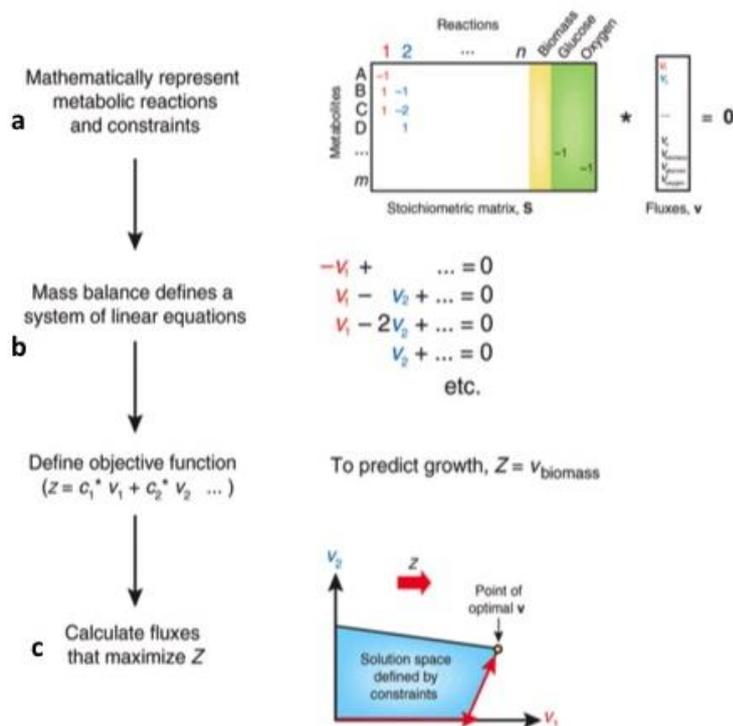
FBA facilitates computational prediction of systemic phenotypes in the form of fluxes in a biochemical network.

As described before, genome scale constructed metabolic models consist of stoichiometric balanced reactions, which are mathematically represented in stoichiometric matrix  $\mathbf{S}$  (Figure 1.2a). In matrix  $\mathbf{S}$ , the row is set of all metabolites or network components and the column corresponds to reactions or description of interaction among network components. Thus, stoichiometric coefficient of a reaction lies in the cell of column and row of the reaction and metabolite respectively. By definition for a given reaction, stoichiometric coefficient of the input metabolite or substrate is given a negative coefficient, while product metabolite is assigned a positive coefficient. A coefficient of zero is assigned to metabolite that does not participate in a particular reaction. Hence, stoichiometric matrix  $\mathbf{S}$  puts constraints on the flow of metabolites through the network as well as capture quantitatively and chemically consistent accounting of biochemical network. The biochemical network as described by  $\mathbf{S}$  can be multiplied by a column vector  $\mathbf{v}$  (contains fluxes through reactions), such that (Figure 1.2b)

$$\mathbf{S} \cdot \mathbf{v} = d\mathbf{x}/dt,$$

,where  $\mathbf{x}$  is the concentration of metabolites, and  $d\mathbf{x}/dt$  is the rate of change of metabolite. In order to solve for  $\mathbf{v}$ , to obtain fluxes through metabolic network steady state condition is assumed i.e.  $d\mathbf{x}/dt = 0$ , which essentially means there is no net mass consumed or produced in the system. This leads to equation:

$$\mathbf{S} \cdot \mathbf{v} = 0,$$



**Figure 1.2:** Showing various steps in flux balance analysis a) Generation of Stoichiometric matrix b) Linear equation  $S \cdot v = 0$  and c) description of objective function; linear programming is used find optimal flux to maximize/minimize objective function (Image adopted from Orth et al., 2010)

The linear programming can be used to determine solution of this equation and any  $v$  flux vector that satisfies above equation is within the solution space of  $S$  or is said to be in null space of  $S$ . The solution of this linear equation for most metabolic networks is underdetermined, *i.e.* there are more unknown variables than equations involving these. In other words, there are fewer metabolites than reactions whose fluxes are to be predicted and there are no unique solutions to this system of equations. FBA overcomes this by using Linear Programming (LP) to strategy by optimizing for a particular flux while ensuring that constraints on mass is maintained in steady state conditions. Here, FBA seeks to identify single point in solution space by maximizing or minimizing an objective function (Figure 1.2c) given by

$$Z = c^T \cdot v$$

,where fluxes are given by  $\mathbf{v}$  and  $\mathbf{c}$  is a vector weights indicating how much each reaction contributes to the objective function. The objective function is a linear combination of  $\mathbf{c}$  and  $\mathbf{v}$ . Examples of objective function in FBA are maximizing growth rate, maximizing (minimizing) ATP production or a particular metabolite. In order to obtain optimal solution of  $\mathbf{S} \cdot \mathbf{v}=0$ , and objective function, FBA employs LP methods to solve for this system of linear equations. Hence, FBA result in a solution vector  $\mathbf{v}$ , the flux distribution, which is either maximizes or minimizes the objective function. In addition to previously described constraints, FBA solution space can be narrowed by specifying additional constraints such as maximum and minimum allowable fluxes through reactions (Orth et al., 2010).

A very common objective for FBA is biomass production, which consists of metabolites conversion into biomass constituents such as nucleic acids, proteins, and lipids. This biomass is represented mathematically as ‘biomass reaction’, which consumes precursor metabolites at stoichiometry in biomass production. Since biomass is represented in the model, maximal growth can be accomplished by finding conditions that result in maximal flux through biomass reaction (Orth et al., 2010).

### **1.3 Application of FBA on phenotype prediction in *Escherichia coli***

FBA has demonstrated reasonable agreement with experimental data such as in *E. coli* for gene essentiality or metabolic gene knockouts (Long and Antoniewicz, 2014; Long et al., 2018). Extensive work on computational prediction of flux response in *E. coli* under genetic perturbations such as gene knockouts shows reliable prediction overall. Despite development of both experimental and computational tools to study fluxes in *E. coli* metabolic network, it is still challenging to predict perturbation responses to gene deletions (Long and Antoniewicz, 2014). Such studies are also hampered by limitation of experimental data on

continuous culture, which is mostly used in studying metabolic fluxes. To study metabolic response to gene deletion on a large scale, we attempted to utilize the effect on colony sizes (on solid media) measured as fitness score upon gene deletion (Nichols et al., 2010). We used this data because it is comprehensive *E. coli* single gene deletion data.

## 1.4 Enzyme promiscuity

In general, cells show diverse metabolic capability and show physiological adaptation to genotype/environment perturbations. This robust behavior has been mostly attributed to the complex interplay and dynamic interactions of components in the metabolic network (Ari and Casadesús, 1998). Moreover, experimental studies have shown that in the event of metabolic gene function disruption, apart from transcriptional, translational, and post-translational regulation of enzymes, fluxes are routed through alternate pathway/s to maintain constant flow of essential metabolites (Ishii et al., 2007; Nakahigashi, et al., 2009). These alternate metabolic routes often involve enzymes, which catalyze analogous substrates/reactions. For example, *Pseudomonas diminuta* uses phosphotriesterase to release phosphate from organophosphate insecticide, which is not its native substrate (Raushel and Holden, 2000).

Even though enzymes are usually described as efficient and specific catalysts to its substrate and reaction it catalyzes, many enzymes harbor capabilities to catalyze other reactions and/or substrates apart from the ones for which they are physiologically specialized or evolved (Khersonsky and Tawfik, 2010). These adventitious secondary (promiscuous) reactions are generally orders of magnitude less efficient than their evolved activities. Such low level of promiscuous activity, usually undetectable, can become important if substrate/enzyme concentration changes due to some factors (Ari and Casadesús, 1998). Promiscuous enzymes could confer fitness benefit to the organism under new selective pressures with these enzymes serving as starting point in the emergence of new enzyme functions and/or divergence of enzyme families (Nobeli et al., 2009).

## 1.5 Objective

Given limitations in generating suitable experimental data for every single metabolic gene deletion in *E. coli* for flux analysis, we explored possibility of using *E. coli* growth data on solid media (fitness scores (Nichols et al., 2010)) to understand the effect of gene deletion and distribution of fluxes. To facilitate understanding of metabolic gene defect we formulated following hypothesis:

A. If biomass from FBA models is unchanged or reduced marginally compared to wild type biomass, however fitness score predicts lower growth then we analyze (here FBA models are unable to predict *in vivo* conditions):

1. Could this be a result of pleotropic effect of deleted gene?
2. The alternate pathway evoked because of gene deletion may lead to harmful intermediates, which can affect growth of bacteria, such as reactive superoxide radicals.

B. If gene deletion predicts zero biomass from FBA models, however, fitness score shows growth defect:

1. Could alternate pathways missing in FBA model?
2. In such conditions, promiscuous enzymes can assist in survival of bacteria. However, does not compensate full activity of the deleted gene.

In the present study, we have used fitness scores of *E. coli* single deletion strains only on fermentable carbon sources *viz.* glucose, maltose, glucosamine, and N-acetyl glucosamine and used FBA to analyze flow of fluxes in alternate pathways.

# Chapter 2

## Methodology

### 2.1 *E. coli* Fitness score data

In work of Nichols et al., 2010, effect of gene deletion in *E. coli* is measured using fitness score, which is relative size of colony size of strain having single gene deletion with respect to wild type colony size. Here, wild type colony size is represented by average colony sizes of all single deletion strains. We obtained fitness scores of 3979 genes on 324 conditions from URL <http://ecoliwiki.net/tools/chemgen/>. The growth data of *E. coli* on fermentable carbon sources glucose, maltose, glucosamine and N-acetyl glucosamine are retrieved for analysis.

### 2.2 FBA modeling using COBRA ToolBox

In order to perform FBA, we have used constraint-based reconstruction and analysis (COBRA) method using freely available MATLAB toolbox referred as COBRA Toolbox (Becker et al., 2007). The genome scale metabolic model of *E. coli* iJO1366 is obtained from BIGG models database (King et al., 2016). We have optimized biomass under various fermentable carbon sources. Further, we performed single gene deletion and again optimized the biomass to understand predicted effect of gene deletions on *E. coli* growth. We have used unlimited oxygen uptake ( $-1000 \text{ mmol gDW}^{-1} \text{ Hr}^{-1}$ ), which is mimicking aerobic growth condition and set a constant rate of glucose uptake at  $-20 \text{ mmol gDW}^{-1} \text{ Hr}^{-1}$ . In order to model biomass growth in alternate carbon source, we have maintained same uptake rate of substrate in aerobic conditions.

List of commands:

```
initCobraToolbox() //To initialize CobraToolbox
```

```

load('iJO1366.mat ') //load E. coli model

model = changeRxnBounds(model,'EX_glc__D_e',-20,'l'); //set lower bound of the glucose

model=changeRxnBounds(model,'EX_o2_e',-1000,'l'); //set lower bound for oxygen

model=changeObjective(model,'BIOMASS_Ec_iJO1366_WT_53p95M');//set objective
function

FBAsolution= optimizeCbModel(model,'max',0,0) // optimization step

printFluxVector(model, FBAsolution.x, true); //print flux values

```

## 2.3 FBA modeling of single gene deletion strains using COBRA ToolBox

For each metabolic gene, we find list of its associate enzymatic reactions in BIGG models database. To model the effect of gene deletion, we used a simple approach of making the flux of zero through this reaction by setting both upper and lower bound of reaction to 0. Rest all exchange reactions of carbon source and oxygen is maintained as in FBA of wild type. Apart from this, we also used function single reaction deletion (*'singleRxnDeletion'*) to model effect of gene deletion on biomass.

### Additional list of commands:

```

model=changeRxnBounds(model,'REACTION_NAME',0,'b'); // makes reaction bound to zero

model=changeObjective(model,'BIOMASS_Ec_iJO1366_WT_53p95M');//set objective
function

FBAsolution= optimizeCbModel(model,'max',0,0) //optimization step

printFluxVector(model, FBAsolution.x, true); //print flux values

```

*[grRatio, grRateKO, grRateWT, hasEffect, delRxns, fluxSolution] = singleRxnDeletion(model,'FBA')*

We have compared biomass in wild type and gene knockout strains and performed analysis of gene knockout effects.

# Chapter 3

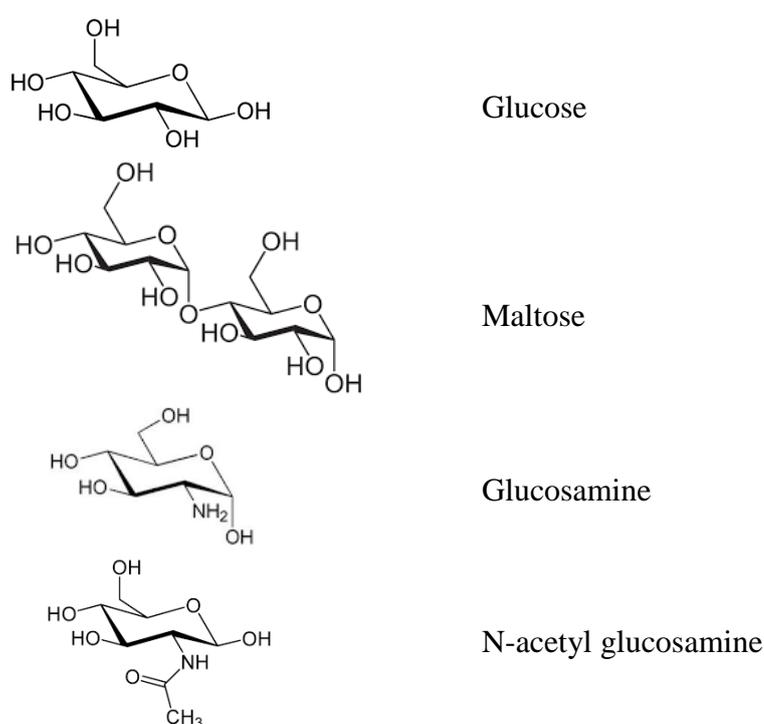
## Results and Discussions

### 3.1 Comparative analysis of fitness scores data of *E. coli* single gene deletion mutants grown in fermentable carbon source

In the seminal work of Nichols et al., phenotypic responses of single gene deletions strains under various physiological and drug stresses was studied with an aim to relate genotype to phenotype, function annotation of uncharacterized proteins and generate conditionally essential genes (Nichols et al., 2010). In this study, they performed high throughput phenotypic analysis and provided quantitative phenotypic responses of all single gene deletion mutants of *E. coli* under 324 conditions. The quantitative response is given by fitness score, which has been previously used for assessing genetic interactions (Typas et al., 2008; Bochner, 2009). Essentially, fitness score provides a measure of change in colony size of a single gene deletion strain with respect to the average colony size of all single gene deletions on all conditions, which is considered as representative for wild type *E. coli* colony size.. Hence, positive and negative fitness score represents increased and decreased colony sizes in comparison to average colony size respectively. Thus, a positive and negative fitness score of a single gene deletion strain grown in a given condition is interpreted as improved and reduced growth respectively. In the present work, we re-analyzed *E. coli* fitness score data growth on the fermentable carbon source (glucose, maltose, glucosamine and N-acetyl glucosamine) with a sole objective to explore possibility of using such data in metabolic network analysis and Flux Balance Analysis (FBA). Mostly, biomass (representative of bacterial growth) optimized in FBA is correlated with bacteria grown in liquid culture. Here, the optimized biomass in FBA is compared to fitness scores, which is a statistical relative

measure of colony sizes. Such studies will extend usefulness of FBA for bacteria grown on solid media as well as facilitate analysis of large fitness growth data.

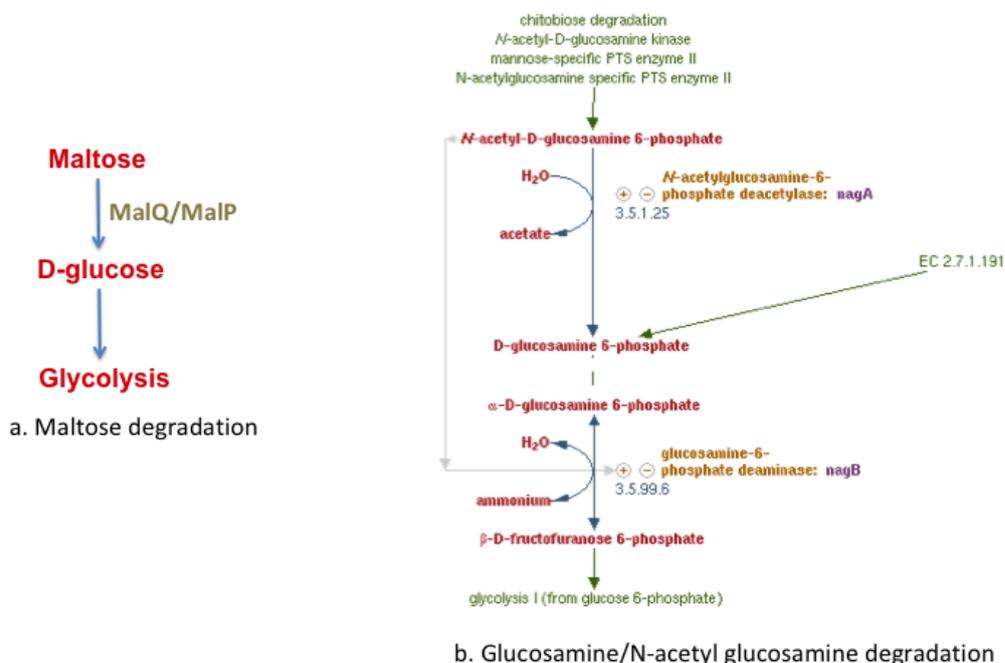
First, we compared single gene deletion strains fitness score on maltose, glucosamine, and N-acetyl glucosamine with respect to glucose to identify and verify respective carbon source specific pathways. Since glucose is the simplest carbohydrate among set of fermentable carbon sources in this study (Figure 3.1), we expected a common set of essential genes for growth on these carbon sources and glucose. Furthermore, for growth on maltose, glucosamine, and N-acetyl glucosamine there will be conditionally essential genes, which are involved in transport of sugars/amino sugars or metabolic pathway to degrade sugars/amino sugars to glucose and genes required for regulation of genes in such metabolic pathways. To facilitate this analysis, *E. coli* metabolic pathway involved in maltose, glucosamine and N-acetyl glucosamine from EcoCyc/literature (Keseler et al., 2016) is shown in Figure 3.2a and 3.2b.



**Figure 3.1:** Comparison of chemical structures of carbon sources showing similarities in their chemical nature.

For maltose utilization, in *E. coli*, MalQ is a key enzyme with  $\alpha$ -1-4 glucanotransferase activity in maltose degradation that recognizes maltose and a longer maltodextrins and preferentially removes glucose from reducing ends of maltose that can enter glycolysis (Figure 3.2a) (Park et al., 2011). Additionally, enzyme MalP recognize maltotetraose and longer maltodextrins removes glucose from non-reducing ends to generate glucose-1-phosphate (Park et al., 2011), which can enter glycolysis. Maltose is transported by high affinity maltose transport system encoded by gene malEFGK (Dippel and Boos, 2005).

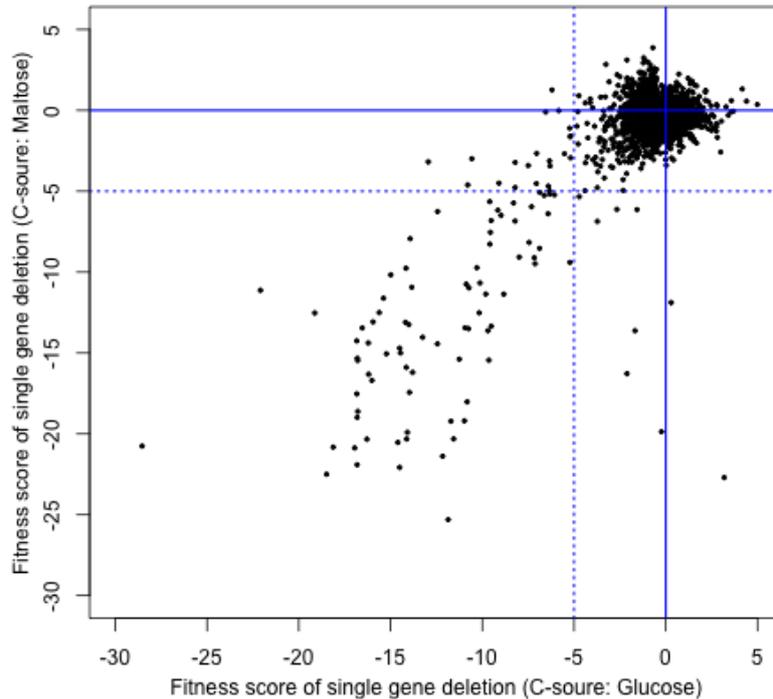
The degradation pathways of glucosamine and N-acetyl glucosamine involve common enzyme NagB, which catalyzes conversion of D-glucosamine 6 phosphate to D-fructofuranose 6-phosphate. NagA enzyme removes acetate group from N-acetyl glucosamine to generate D-glucosamine 6-phosphate, which can be utilized by NagB (Figure 2b). Additionally, gene NagE is involved in transport of glucosamine and N-acetyl glucosamine that also phosphorylates these substrates (Alvarez-Añorve et al., 2009).



**Figure 3.2:** Figure showing maltose (a) and glucosamine and N-acetyl glucosamine (b) utilization pathways.

The scatter plot (Figure 3.3) shows comparison of fitness scores of single gene deletion strains grown in maltose and glucose as sole carbon sources. As is evident in Figure 3, fitness scores  $\leq 5$  are better correlated in comparison to fitness score of all genes. In general, fitness scores on glucose for same gene deletion is lesser compared to fitness score on maltose. Next, we focused on single gene deletions having negative fitness scores in maltose with no apparent loss of fitness in glucose. Genes having fitness score of  $\geq 0$  or  $\geq -5$  are considered to have no effect on growth in respective carbon sugars. The gene *malT* has fitness scores of -22.7 and 3.2 in maltose and glucose respectively. As expected, *malT* is involved in maltose utilization. It is a transcriptional activator of maltose regulon, which are responsible for uptake and catabolism of malto-oligosaccharide. Similarly, as expected gene *malQ*, which encoded enzyme for catabolism of maltose (Figure 2a), shows poor growth in maltose in comparison to glucose. Among components of maltose transporter *malG*, *malK*, *malF* shows varying degree of defect in growth in maltose. This shows that transport complex can probably assemble in absence of either of these components or maltose is transported in cell using other transport system. Interestingly, *MalE*, an important maltose binding protein exported in periplasmic space is responsible for activation of maltose transport does not show extensive fitness defect in maltose (-6.1) relative to glucose (-1.6).

Further, we analyzed genes with fitness defect in glucose relative to maltose. As mentioned before, gene fitness scores are correlated between maltose and glucose. However, genes such as *pyrB*, *pyrC*, *pyrD*, *trpA*, and *cysA* exhibited relative poor growth in glucose compared to maltose. Of these, *pyrB*, *pyrC* and *pyrD* are involved in *de novo* biosynthesis of uridine-5'-phosphate. superpathway of pyrimidine, purine, and histidine biosynthesis pathway (EcoCyc).

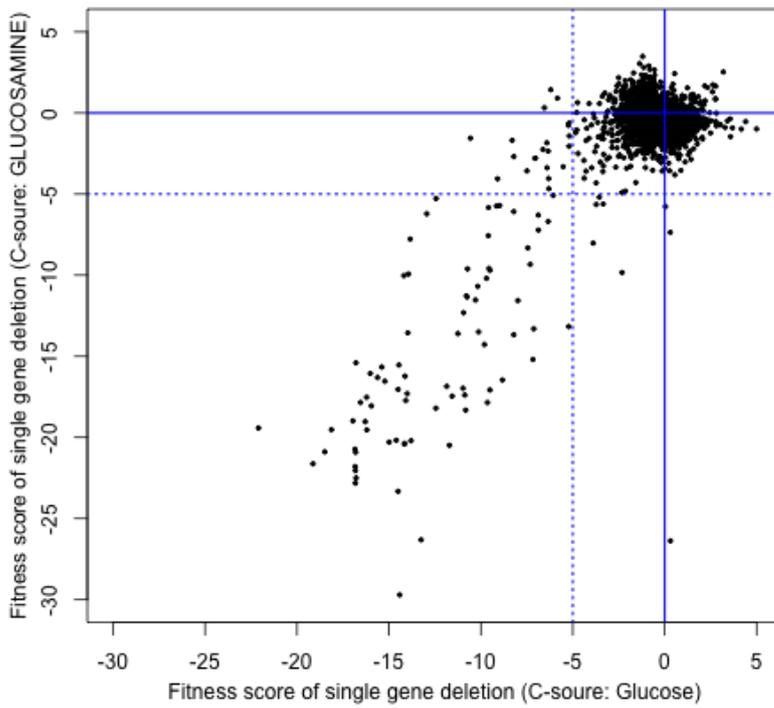


**Figure 3.3:** Scatter plot of fitness score of single gene deletion strain on glucose versus maltose.

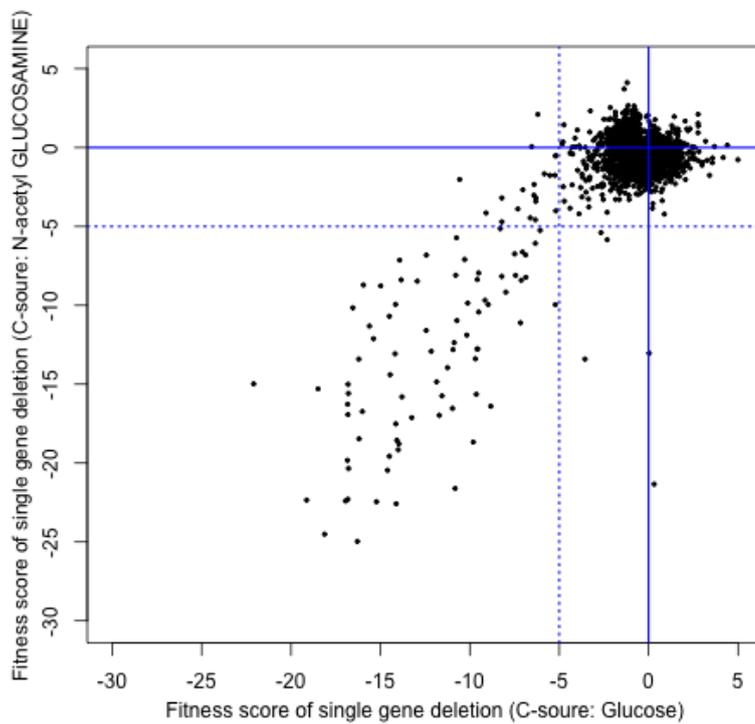
Similarly, genes fitness scores on sole carbon source as glucose are compared with amino sugars glucosamine and N-acetyl glucosamine as shown in Figure 3.4 and 3.5 respectively. As observed before, higher correlation in fitness score is evident in scores are  $\leq -5$ . Moreover, relatively a greater affect in growth on glucosamine and N-acetyl glucosamine is observed with respect to glucose as shown by higher magnitude of negative fitness score in amino sugars for a given deleted gene. Since NagB is an essential step in utilization of glucosamine, nagB deleted strain shows a low fitness score on glucosamine/N-acetyl glucosamine in comparison to glucose. A recent study has shown nagB that it is an essential gene for growth on glucosamine/N-acetyl glucosamine in *Streptococcus pneumoniae* (Afzal et al., 2016). However, *E. coli*  $\Delta$ nagB strain is able to utilize amino sugars as sole carbon source. It is possible that alternate activity or promiscuous activity of enzyme GlnS, which converts fructose-6-phosphate to glucosamine-6-phosphate in the biosynthesis of glucosamine compensates for deaminase activity in  $\Delta$ nagB strains. In growth on N-acetyl glucosamine, apart from NagB,  $\Delta$ nagA strain also shows lower fitness score in comparison to glucose. A

protein of unknown function encoded by *yceB* has relatively more defective growth on N-acetyl glucosamine (fitness score: -13.4) in comparison to glucose (fitness score: -3.6). This gene may be involved in specific function in utilization of amino sugars that can be explored experimentally.

Interestingly, deletion of *nagE*, which encodes phosphoenolpyruvate:sugar phosphotransferase system (PTS) permeases involved in transport of glucosamine/N-acetyl glucosamine does not show any growth defect on these carbon sources. It is possible that in  $\Delta$ *nagE* strain other PTS permeases, such as for mannose/glucose may be responsible for transport of these amino sugars. In a recent study, it has been shown that components of mannose PTS *viz.* ManL, ManM and ManN plays dominant role in transport of amino sugars in *S. mutans* (Moye et al., 2014). Importantly, deletion of ManX, ManY and ManZ, which are *E. coli* homologues of ManL, ManM and ManN respectively shows slight lower (poor) fitness score on amino sugars.



**Figure 3.4:** Scatter plot of fitness score of single gene deletion strain on glucose versus glucosamine.



**Figure 3.5:** Scatter plot of fitness score of single gene deletion strain on glucose versus N-acetyl glucosamine.

### **3.2 Metabolic flux analysis and comparison it with fitness score obtained from single gene deletion mutants**

In this study, we explore the possibility of using fitness score data in conjunction with theoretical metabolic flux models to understand the effect of gene deletions on metabolic networks and associated fluxes. Moreover, this will allow improving metabolic reconstruction models.

Overall, we have optimized biomass in both full metabolic flux model and for a given metabolic deleted gene. Then, relative biomass from FBA is compared with fitness score. Given that we are using known stoichiometric balanced reactions from *E. coli*, we may not observe concurrence between biomass optimized FBA models and fitness scores of single gene deletions strains. One of the reasons could be that a cell under various genotypic perturbations may not be involved in higher growth (improving biomass). However, such studies will provide limits of prediction from metabolic networks and give an opportunity to improve these theoretical models. To analyze and understand effect of gene deletion on metabolic networks, we hypothesized following in two situations:

If biomass from FBA models is unchanged or reduced marginally compared to wild type biomass, however fitness score predicts lower growth then we analyze (here FBA models are unable to predict *in vivo* conditions):

1. Could this be a result of pleotropic effect of deleted gene?
2. The alternate pathway in gene deletion may lead to harmful intermediates, which can affect growth of bacteria, such as reactive superoxide radicals.

If gene deletion predicts zero biomass from FBA models, however, fitness score shows growth defect. This could be

1. Could alternate pathways missing in FBA model?

2. Promiscuous enzymes can assist in survival of bacteria. However, does not compensate full activity of the deleted gene.

For our study, we have used recent full metabolic reconstructed *E. coli* model iJO1366 and used ‘cobraToolbox’ to perform flux balance analysis under various genotypic perturbation and change in carbon source for computation of bacterial growth (see methods). First, we optimized biomass production without any gene deletion where carbon source uptake rate is set at 20 mmol gDW<sup>-1</sup> Hr<sup>-1</sup> and aerobic condition is assumed with maximal allowed flux through exchange rate of oxygen of 1000 mmol gDW<sup>-1</sup> Hr<sup>-1</sup>. Next, we used CobraToolBox function to delete gene and compute biomass under gene deletion condition.

The biomass computed using FBA models for wild type *E. coli* under various carbon sources (Glucose, Maltose, Glucosamine, and N-acetyl glucosamine). As expected maximum biomass is for growth on maltose as it is made up of 2 glucose units. This is followed by N-acetyl glucosamine, which will provide additional 2 carbons from acetate moiety.

**Table 3.1: FBA computed biomass of wild type *E. coli***

	<b>Name of Carbon Source</b>	<b>Uptake rate</b>	<b>Biomass</b>
1	Glucose	-20	1.986
2	Maltose	-20	3.985
3	Glucosamine	-20	1.986
4	N-acetyl glucosamine	-20	2.546

Subsequently, we deleted one gene at a time and re-computed biomass in FBA models. To begin with, genes are sorted based on their fitness score and we took top 30 genes for further analysis. Tables 3.2, 3.3, 3.4 and 3.5 summarizes biomass of single gene deletion under various sole carbon source. As is evident, most of the genes having lower fitness score are common in fermentable carbon sources.

**Table 3.2: FBA computed single gene deletions under Glucose**

Sr. No	Gene Deleted	Growth rate	Growth rate relative to WT	Fitness Score
1	ECK0075-LEUB	0	-1.9861	-28.542293
2	ECK0002-THRA	1.9861	0	-22.093327
3	ECK0033-CARA	1.9861	0	-19.127788
4	ECK3161-ARGG	0	-1.9861	-18.4895
5	ECK2019-HISA	0	-1.9861	-18.126139
6	ECK2017-HISB	0	-1.9861	-16.956621
7	ECK1255-TRPB	0	-1.9861	-16.844067
8	ECK2021-HISI	0	-1.9861	-16.837168
9	ECK2836-LYSA	0	-1.9861	-16.822236
10	ECK3823-METE	1.9861	0	-16.820471
11	ECK2814-ARGA	0	-1.9861	-16.809406
12	ECK0244-PROA	1.9844	-0.0017	-16.792507
13	ECK2909-SERA	1.9441	-0.042	-16.78734
14	ECK2016-HISC	0	-1.9861	-16.543581
15	ECK3951-ARGH	0	-1.9861	-16.291227
16	ECK2020-HISF	0	-1.9861	-16.211824
17	ECK3933-METF	0	-1.9861	-16.19766
18	ECK3764-ILVA	1.9861	0	-16.014049
19	ECK3822-METR		-1.9861	-15.951693
20	ECK2597-TYRA	0	-1.9861	-15.620196
21	ECK2014-HISG	0	-1.9861	-15.38656
22	ECK2555-PURL	0	-1.9861	-15.224041
23	ECK2323-AROC	0	-1.9861	-14.988352
24	ECK3376-AROB	0	-1.9861	-14.601909
25	ECK3763-ILVD	0	-1.9861	-14.501324
26	ECK3947-PPC	1.9801	-0.006	-14.500779
27	ECK3762-ILVE	0	-1.9861	-14.459306
28	ECK3998-PURH	0	-1.9861	-14.410078
29	ECK4210-CYSQ	0	-1.9861	-14.184101
30	ECK2747-CYSD	0	-1.9861	-14.156461

**Table 3.3: FBA computed single gene deletions under Maltose**

Sr. No	Gene Deleted	Growth rate	Growth rate realted to WT	Fitness Score
1	ECK0075-LEUB	0	-3.9854	-20.755706
2	ECK0002-THRA	3.9854	0	-11.123402
3	ECK0033-CARA	3.9854	0	-12.527022
4	ECK3161-ARGG	0	-3.9854	-22.496025
5	ECK2019-HISA	0	-3.9854	-20.821367
6	ECK2017-HISB	0	-3.9854	-20.871902
7	ECK1255-TRPB	0	-3.9854	-14.25015
8	ECK2021-HISI	0	-3.9854	-17.529904
9	ECK2836-LYSA	0	-3.9854	-21.912441
10	ECK3823-METE	3.9854	0	-18.986523
11	ECK2814-ARGA	0	-3.9854	-15.33372
12	ECK0244-PROA	3.9819	-0.0035	-15.463079
13	ECK2909-SERA	3.901	-0.0844	-18.615858
14	ECK2016-HISC	0	-3.9854	-13.452438
15	ECK3951-ARGH	0	-3.9854	-20.33496
16	ECK2020-HISF	0	-3.9854	-14.379435
17	ECK3933-METF	0	-3.9854	-16.319253
18	ECK3764-ILVA	3.9854	0	-16.695228
19	ECK3822-METR		-3.9854	-13.078199
20	ECK2597-TYRA	0	-3.9854	-12.505323
21	ECK2014-HISG	0	-3.9854	-11.611893
22	ECK2555-PURL	0	-3.9854	-15.050241
23	ECK2323-AROC	0	-3.9854	-10.171083
24	ECK3376-AROB	0	-3.9854	-20.531727
25	ECK3763-ILVD	0	-3.9854	-14.706607
26	ECK3947-PPC	3.9733	-0.0121	-22.0784
27	ECK3762-ILVE	0	-3.9854	-15.001737
28	ECK3998-PURH	0	-3.9854	-32.465045
29	ECK4210-CYSQ	0	-3.9854	-13.098526
30	ECK2747-CYSD	0	-3.9854	-9.757666

**Table 3.4: FBA computed single gene deletions under Glucosamine**

Sr. No	Gene Deleted	Growth rate	Growth rate related to WT	Fitness Score
1	ECK0075-LEUB	0	-1.9867	-31.952112
2	ECK0002-THRA	1.9867	0	-19.419673
3	ECK0033-CARA	1.9867	0	-21.628953
4	ECK3161-ARGG	0	-1.9867	-20.889105
5	ECK2019-HISA	0	-1.9867	-19.522231
6	ECK2017-HISB	0	-1.9867	-18.979149
7	ECK1255-TRPB	0	-1.9867	-20.725781
8	ECK2021-HISI	0	-1.9867	-21.803142
9	ECK2836-LYSA	0	-1.9867	-22.816676
10	ECK3823-METE	1.9867	0	-22.044893
11	ECK2814-ARGA	0	-1.9867	-20.909752
12	ECK0244-PROA	1.985	-0.0017	-15.397838
13	ECK2909-SERA	1.9447	-0.042	-22.501464
14	ECK2016-HISC	0	-1.9867	-17.847272
15	ECK3951-ARGH	0	-1.9867	-19.029234
16	ECK2020-HISF	0	-1.9867	-17.524683
17	ECK3933-METF	0	-1.9867	-19.53524
18	ECK3764-ILVA	1.9867	0	-16.058728
19	ECK3822-METR		-1.9867	-18.059164
20	ECK2597-TYRA	0	-1.9867	-16.309859
21	ECK2014-HISG	0	-1.9867	-15.663783
22	ECK2555-PURL	0	-1.9867	-16.536849
23	ECK2323-AROC	0	-1.9867	-20.293052
24	ECK3376-AROB	0	-1.9867	-20.176259
25	ECK3763-ILVD	0	-1.9867	-17.041873
26	ECK3947-PPC	1.9807	-0.006	-23.327251
27	ECK3762-ILVE	0	-1.9867	-15.540219
28	ECK3998-PURH	0	-1.9867	-29.70611
29	ECK4210-CYSQ	0	-1.9867	-10.025838
30	ECK2747-CYSD	0	-1.9867	-20.380946

**Table 3.5: FBA computed single gene deletions under N-acetyl Glucosamine**

Sr. No	Gene Delted	Growth rate	Growth rate relative to WT	Fitness Score
1	ECK0075-LEUB	0	-2.546	-34.445652
2	ECK0002-THRA	2.546	0	-14.98107
3	ECK0033-CARA	2.546	0	-22.345368
4	ECK3161-ARGG	0	-2.546	-15.297851
5	ECK2019-HISA	0	-2.546	-24.50751
6	ECK2017-HISB	0	-2.546	-22.400293
7	ECK1255-TRPB	0	-2.546	-19.829989
8	ECK2021-HISI	0	-2.546	-16.273763
9	ECK2836-LYSA	0	-2.546	-22.295987
10	ECK3823-METE	2.546	0	-16.930991
11	ECK2814-ARGA	0	-2.546	-15.01158
12	ECK0244-PROA	2.5438	-0.0022	-15.585281
13	ECK2909-SERA	2.4912	-0.0548	-20.343845
14	ECK2016-HISC	0	-2.546	-10.159378
15	ECK3951-ARGH	0	-2.546	-24.964691
16	ECK2020-HISF	0	-2.546	-13.410851
17	ECK3933-METF	0	-2.546	-18.466317
18	ECK3764-ILVA	2.546	0	-16.736738
19	ECK3822-METR		-2.546	-8.713386
20	ECK2597-TYRA	0	-2.546	-11.317083
21	ECK2014-HISG	0	-2.546	-12.117833
22	ECK2555-PURL	0	-2.546	-22.452039
23	ECK2323-AROC	0	-2.546	-8.772376
24	ECK3376-AROB	0	-2.546	-20.462213
25	ECK3763-ILVD	0	-2.546	-19.57331
26	ECK3947-PPC	2.5415	-0.0045	-10.697374
27	ECK3762-ILVE	0	-2.546	-14.398676
28	ECK3998-PURH	0	-2.546	-33.386079
29	ECK4210-CYSQ	0	-2.546	-13.072453
30	ECK2747-CYSD	0	-2.546	-9.9423

The predicted biomass of  $\Delta$ leuB strain under various carbon sources is 0. However, fitness score suggests  $\Delta$ leuB strain shows growth defect. LeuB is involved in leucine biosynthesis, which is important for bacterial survival. A recent study has shown that D-malate dehydrogenase is a generalist enzyme, which can perform same enzymatic reaction as

LeuB (Vorobieva et al., 2014). This may be the reason of low fitness on fermentable carbon sources. We are analyzing other gene deletion strains to understand its effect of *in silico* metabolic models.

The preliminary analysis of comparing single gene deletion strains fitness scores with FBA predicted biomass on various fermentable carbon sources showed that not all gene deletion predicted no or poor growth phenotype in FBA models. In fact, in FBA we obtained either same biomass as without gene deletion or zero biomass. Further, detailed analysis of metabolic models and FBA will be required to fully explore reliability of fitness score prediction using flux distribution in metabolic models.

## Conclusions

In the first section, as expected comparison of fitness scores of single gene deletion strains between maltose/glucosamine/N-acetyl glucosamine to glucose identified fermentable carbon sources specific enzymes. However, we identified transporter of amino sugars NagE is dispensable because mannose transporter, ManXYZ can actively transport amino sugars in absence of NagE. Another gene yceB, without known function, showed growth fitness defect on amino sugars that suggest condition specific gene function of which can be explored experimentally.

The preliminary analysis of comparing single gene deletion strains fitness scores with FBA predicted biomass on various fermentable carbon sources showed that not all gene deletion predicted no or poor growth phenotype in FBA models. In fact, in FBA we obtained either same biomass as without gene deletion or zero biomass. Further, detailed analysis of metabolic models and FBA will be required to fully explore reliability of fitness score prediction using flux distribution in metabolic models.

## Bibliography

Afzal, M., Shafeeq, S., Manzoor, I., Henriques-Normark, B., & Kuipers, O. P. (2016). N-acetylglucosamine-Mediated Expression of nagA and nagB in *Streptococcus pneumoniae*. *Frontiers in cellular and infection microbiology*, *6*, 158.

Alvarez-Anorve, L. I., Bustos-Jaimes, I., Calcagno, M. L., & Plumbridge, J. (2009). Allosteric regulation of glucosamine-6-phosphate deaminase (NagB) and growth of *Escherichia coli* on glucosamine. *Journal of bacteriology*, *191*(20), 6401-6407.

Bochner, B. R. (2008). Global phenotypic characterization of bacteria. *FEMS microbiology reviews*, *33*(1), 191-205.

Carbonell, P., Fichera, D., Pandit, S. B., & Faulon, J. L. (2012). Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms. *BMC systems biology*, *6*(1), 10.

D'Ari, R., & Casadesús, J. (1998). Underground metabolism. *Bioessays*, *20*(2), 181-186.

Dippel, R., & Boos, W. (2005). The maltodextrin system of *Escherichia coli*: metabolism and transport. *Journal of bacteriology*, *187*(24), 8322-8331.

Dunphy, L. J., & Papin, J. A. (2018). Biomedical applications of genome-scale metabolic network reconstructions of human pathogens. *Current opinion in biotechnology*, *51*, 70-79.

Edwards, J. S., & Palsson, B. O. (2000). Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions. *BMC bioinformatics*, *1*(1), 1.

Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., & Palsson, B. Ø. (2009). Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, *7*(2), 129.

Gianchandani, E. P., Chavali, A. K., & Papin, J. A. (2010). The application of flux balance analysis in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *2*(3), 372-382.

Guo, W., & Feng, X. (2016). OM-FBA: integrate transcriptomics data with flux balance analysis to decipher the cell metabolism. *PloS one*, *11*(4), e0154188.

Guo, W., Sheng, J., & Feng, X. (2017). Mini-review: in vitro metabolic engineering for biomanufacturing of high-value products. *Computational and structural biotechnology journal*, *15*, 161-167.

Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A. and Ho, P.Y. (2007). Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science*, 316(5824), 593-597.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, 407(6804), 651.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2016). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, 45(D1), D353-D361.

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M. and Latendresse, M. (2012). EcoCyc: fusing model organism databases with systems biology. *Nucleic acids research*, 41(D1), D605-D612.

King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O. and Lewis, N.E. (2015). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic acids research*, 44(D1), D515-D522.

Long, C. P., & Antoniewicz, M. R. (2014). Metabolic flux analysis of *Escherichia coli* knockouts: lessons from the Keio collection and future outlook. *Current opinion in biotechnology*, 28, 127-133.

Long, C. P., Gonzalez, J. E., Feist, A. M., Palsson, B. O., & Antoniewicz, M. R. (2017). Dissecting the genetic and metabolic mechanisms of adaptation to the knockout of a major metabolic enzyme in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 201716056.

Moye, Z. D., Burne, R. A., & Zeng, L. (2014). Uptake and metabolism of N-acetylglucosamine and glucosamine by *Streptococcus mutans*. *Applied and environmental microbiology*, 80(16), 5053-5067.

Nakahigashi, K., Toya, Y., Ishii, N., Soga, T., Hasegawa, M., Watanabe, H., Takai, Y., Honma, M., Mori, H. and Tomita, M. (2009). Systematic phenome analysis of *Escherichia coli* multiple knockout mutants reveals hidden reactions in central carbon metabolism. *Molecular systems biology*, 5(1), 306.

- Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., Wong, A. and Shales, M. (2011). Phenotypic landscape of a bacterial cell. *Cell*, *144*(1), 143-156.
- Nobeli, I., Favia, A. D., & Thornton, J. M. (2009). Protein promiscuity and its implications for biotechnology. *Nature biotechnology*, *27*(2), 157.
- O'Brien, E. J., & Palsson, B. O. (2015). Computing the functional proteome: recent progress and future prospects for genome-scale models. *Current opinion in biotechnology*, *34*, 125-134.
- O'Brien, E. J., Monk, J. M., & Palsson, B. O. (2015). Using genome-scale models to predict biological capabilities. *Cell*, *161*(5), 971-987.
- Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., & Palsson, B. Ø. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular systems biology*, *7*(1), 535.
- Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis?. *Nature biotechnology*, *28*(3), 245.
- Park, J.T., Shim, J.H., Tran, P.L., Hong, I.H., Yong, H.U., Oktavina, E.F., Nguyen, H.D., Kim, J.W., Lee, T.S., Park, S.H. and Boos, W. (2011). Role of maltose enzymes in glycogen synthesis by *Escherichia coli*. *Journal of bacteriology*, *193*(10), 2517-2526.
- Raushel, F. M., & Holden, H. M. (2006). Phosphotriesterase: an enzyme in search of its natural substrate. *Advances in Enzymology and Related Areas of Molecular Biology, Volume 74*, 51-93.
- Samal, A., & Martin, O. C. (2011). Randomizing genome-scale metabolic networks. *Plos One*, *6*(7), e22295.
- Schellenberger, J., Que, R., Fleming, R.M., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S. and Kang, J. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. *Nature protocols*, *6*(9), 1290.
- Tawfik, O. K. A. D. S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual review of biochemistry*, *79*, 471-505.

Typas, A., Nichols, R.J., Siegele, D.A., Shales, M., Collins, S.R., Lim, B., Braberg, H., Yamamoto, N., Takeuchi, R., Wanner, B.L. and Mori, H. (2008). High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nature methods*, 5(9), 781.

Yilmaz, L. S., & Walhout, A. J. (2017). Metabolic network modeling with model organisms. *Current opinion in chemical biology*, 36, 32-39.