# Predicting the carbon source for *Bacillus subtilis* by integrating gene expression profiles into a constraint-based metabolic model

von Kulwadee Thanamit, Pharm. D,

geboren am 11.09.1989 in Trang, Thailand

Gutachter

1. Prof. Dr. Rainer König (Jena)

2. Prof. Dr. Stefan Schuster (Jena)

3. Prof. Dr. Christoph Kaleta (Kiel)

Tag der öffentlichen Verteidigung: 05.07.2022

# Summary

Elucidating cellular metabolism led to many breakthroughs in biotechnology, synthetic biology, and health sciences. To date, deriving metabolic fluxes by $^{13}$C tracer experiments is the most prominent approach for studying metabolic fluxes quantitatively with high accuracy and precision. However, the technique has a high demand for experimental resources. Alternatively, flux balance analysis (FBA) has been employed to estimate metabolic fluxes without labeling experiments. It is less informative but can benefit from the low costs and low experimental efforts and gain flux estimates in experimentally difficult conditions. Methods to integrate relevant experimental data have been emerged to improve FBA flux estimations. For this, data from transcription profiling is often used since it is easy to generate at the genome scale, typically embedded by a binarization of differentially and non-differentially expressed genes coding for the respective enzymes. However, employing defined thresholds can result in disregarding the fine-grained regulation of metabolism. Besides this, thermodynamically infeasible loops (TIL) are a well-known complication in constraint-based modeling, leading to unrealistic flux distributions.

In order to integrate transcriptomic data more efficiently and improve a context-specific model extraction method, the novel method named Linear Programming based Gene Expression Model (LPM-GEM) was established in this thesis. LPM-GEM linearly embeds gene expression into FBA constraints. It avoids binarization and preserves the natural characteristic of gene expression profiles as continuous data. Additionally, three strategies were implemented to reduce thermodynamically infeasible loops, which is necessary for such an omics-based model building. A model of *Bacillus subtilis* (*B. subtilis)* grown in eight different carbon sources was built as a case study to demonstrate the concept of LPM-GEM. The method obtained good flux predictions based on the respective transcription profiles when validating with $^{13}$C-tracer based metabolic flux data of the same conditions. LPM-GEM could well predict the specific carbon sources. When testing the model on another unseen dataset that was not used during training, good prediction performance was also observed.

Furthermore, LPM-GEM outperformed a well-established model building method. Employing LPM-GEM integrates gene expression data efficiently. The method supports gene expression-based FBA models and can be applied as an alternative to estimate metabolic fluxes when tracer experiments are inappropriate.

# Zusammenfassung

Die Aufklärung des zellulären Stoffwechsels hat zu vielen Durchbrüchen in der Biotechnologie, der synthetischen Biologie und den Gesundheitswissenschaften geführt. Bis heute ist die Rekonstruktion von Stoffwechselflüssen durch $^{13}$C-Tracerexperimente der bekannteste Ansatz zur quantitativen Untersuchung von Stoffwechselflüssen. Er ist sehr genau und präzise. Allerdings hat diese Technik einen hohen Bedarf an experimentellen Ressourcen. Alternativ wurde die Flux Balance Analysis (FBA) entwickelt, um metabolische Flüsse ohne Markierungsexperimente abzuschätzen. Sie ist weniger informativ, kann aber durch den geringen Kosten und dem geringen experimentellen Aufwand vorteilhaft sein, auch wenn Flussabschätzungen unter experimentell schwierigen Bedingungen benötigt werden. Es wurden verschiedene Methoden zur Integration relevanter experimenteller Daten entwickelt, um die FBA-Fluxschätzungen zu verbessern. Daten aus Transkriptions-Profiling werden oft ausgewählt, da sie auf genomweit einfach zu generieren sind, typischerweise eingebettet durch eine Binarisierung von differentiell und nicht-differentiell exprimierten Genen, die für die jeweiligen Enzyme kodieren. Die Verwendung von definierten Schwellenwerten kann jedoch dazu führen, dass die feinkörnige Regulation des Stoffwechsels außer Acht gelassen wird. Zusätzlich sind bei der Einbettung derartiger Omics Daten thermodynamisch unmögliche Kreise (thermodynamically infeasible loops, TIL) eine bekannte Komplikation in derartigen Modellierungen, die zu unrealistischen Flussverteilungen führen.

Um Transkriptomdaten effizienter zu integrieren und eine kontextspezifische Modellextraktionsmethode zu verbessern, wurde in dieser Arbeit eine neue Methode namens Linear Programming based Gene Expression Model (LPM-GEM) etabliert. LPM-GEM bettet die Genexpression linear in FBA Zwangsbedingungen ein. Es vermeidet die Binarisierung und bewahrt die natürliche Charakteristik von Genexpressionsprofilen als kontinuierliche Daten. Zusätzlich wurden drei Strategien implementiert, um TIL zu reduzieren, was für eine solche omics-basierte Modellbildung notwendig ist. Ein Modell von *Bacillus subtilis* (*B. subtilis*), das in acht verschiedenen

Kohlenstoffquellen gewachsen ist, wurde als Fallstudie erstellt, um das Konzept von LPM-GEM zu demonstrieren. Die Methode erzielte gute Flussvorhersagen basierend auf den jeweiligen Transkriptionsprofilen bei der Validierung mit $^{13}$C-Tracer-basierten metabolischen Flussdaten der gleichen Bedingungen. LPM-GEM konnte die spezifischen Kohlenstoffquellen gut vorhersagen. Beim Testen des Modells auf einem weiteren ungesehenen Datensatz, der beim Training nicht verwendet wurde, wurde ebenfalls eine gute Vorhersageleistung beobachtet. Darüber hinaus übertraf LPM-GEM eine andere, gut etablierte Methode zur Modellbildung. Der Einsatz von LPM-GEM integriert Genexpressionsdaten effizient. Die Methode unterstützt Genexpressions-basierte FBA-Modelle und kann als Alternative zur Schätzung metabolischer Flüsse eingesetzt werden, wenn Tracer-Experimente ungeeignet sind.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| $^{13}$C-MFA | $^{13}$C metabolic flux analysis |
| AOS | Alternative optimal solution |
| AR | Associated reaction |
| *B. subtilis* | *Bacillus subtilis* |
| CBM | Constraint-based modeling |
| CcpA | Catabolite control protein A |
| CR | Core reaction |
| DNA | Deoxyribonucleic acid |
| *E. coli* | *Escherichia coli* |
| EFM | Elementary flux mode analysis |
| EPA | Extreme pathway analysis |
| FBA | Flux balance analysis |
| FVA | Flux variability analysis |
| GC-MS | Gas chromatography-mass spectrometry |
| GIMME | Gene Inactivity Moderated by Metabolism and Expression |
| GPR | Gene-protein-reaction |
| GRN | Gene regulatory network |
| IFFPR | Iterative Feasible Flux Space Reduction |
| iMAT | Integrative Metabolic Analysis Tool |
| LC-MS | Liquid chromatography-mass spectrometry |
| ll-COBRA | loopless-COBRA |
| LP | Linear programming |
| LPM-GEM | Linear Programming based Gene Expression Model |
| *M. tuberculosis* | *Mycobacterium tuberculosis* |
| mCADRE | metabolic Context-specificity Assessed by Deterministic Reaction Evaluation |
| MILP | Mixed-integer linear programming |
| MIQP | Mixed-integer quadratic programming |
| mRNA | Messenger ribonucleic acid |

| | |
|---|---|
| NGS | Next generation sequencing |
| NLP | Non-linear programming |
| *P. falciparum* | *Plasmodium falciparum* |
| PPI | Protein-protein interaction network |
| PPP | Pentose phosphate pathway |
| PROM | Probabilistic Regulation of Metabolism |
| QP | Quadratic programming |
| RED-TIL | REDucing the number of Thermodynamically Infeasible Loops |
| *S. aureaus* | *Staphylococcus aureus* |
| TCA | Tricarboxylic acid |
| TIL | Thermodynamic infeasible loop |

# Chapter 1

## Introduction

### 1.1    Metabolism

Metabolism is a term referred to as the set of chemical reactions inside an organism or cell. These reactions are involved in sustaining the living state of cells and can be divided into two categories: catabolism and anabolism. Catabolism is a process of breaking down large molecules (for example, carbohydrates, proteins, lipids) into smaller molecules (such as monosaccharides, amino acids, fatty acids). The process usually involves releasing energy. Opposite to catabolism, anabolism is a process of building macromolecules and requires energy. Catabolism and anabolism work together and keep the balance to maintain the normal functions of cells. Several control mechanisms at different levels, e.g., transcriptional regulation, translational regulation, enzyme-metabolite interactions, are employed and collaborated to achieve this balance. Gaining insight into metabolism is important since it can guide researchers to discover components and mechanisms behind these biochemical events, leading to numerous breakthroughs in many fields such as synthetic biology, metabolic engineering, biotechnology, or health sciences (Bideaux et al. 2016, Chenard et al. 2017, Chiewchankaset et al. 2019, Dang et al. 2017, DeWaal et al. 2018, Gatto et al. 2015, Lu et al. 2015, Shan et al. 2018, Sung et al. 2017, Veras et al. 2019, Yao et al. 2019, Zhong et al. 2014).

### 1.2    Network reconstruction and metabolic network

Due to high throughput technology and next generation sequencing (NGS), many biological components, e.g., genes, enzymes, proteins, or metabolites, have been discovered. The demand to systematically pool, categorize, and link these components together to understand how they interact with each other and understand the general mechanisms behind these components is incredibly high. It led to a change in biology research from reductionism to holistic approaches and emerged a new interdisciplinary field of study called systems biology (Chuang et al. 2010, Palsson 2015).

System biology combines computational and mathematical methods to study biology mechanistically. A vast number of various types of networks such as gene regulatory networks (GRN), protein-protein interaction networks (PPI), or metabolic networks have been constructed as well as many theories have been brought up to explain biological phenomena within the same organism or across organisms (Barabasi and Oltvai 2004, Palsson 2015, Koutrouli et al. 2020, Liu et al. 2020). A constraint-based metabolic network, a network combining stoichiometric data of biochemical reactions inside an organism (Figure 1.1), is used to systematically observe the functional state of metabolism (Palsson 2015, Liu et al. 2020). The network can be used to study metabolism when combined with laboratory experiments as in $^{13}$C metabolic flux analysis (Dai and Locasale 2017, Long and Antoniewicz 2019) or when employing purely computational approaches (Palsson 2015) (details in the following sections). Due to more data availability, a large number of metabolic models has been generated, and the number will continue to grow and cover more and more organisms, as well as the size of the models growths as more components and reactions have been discovered and annotated.
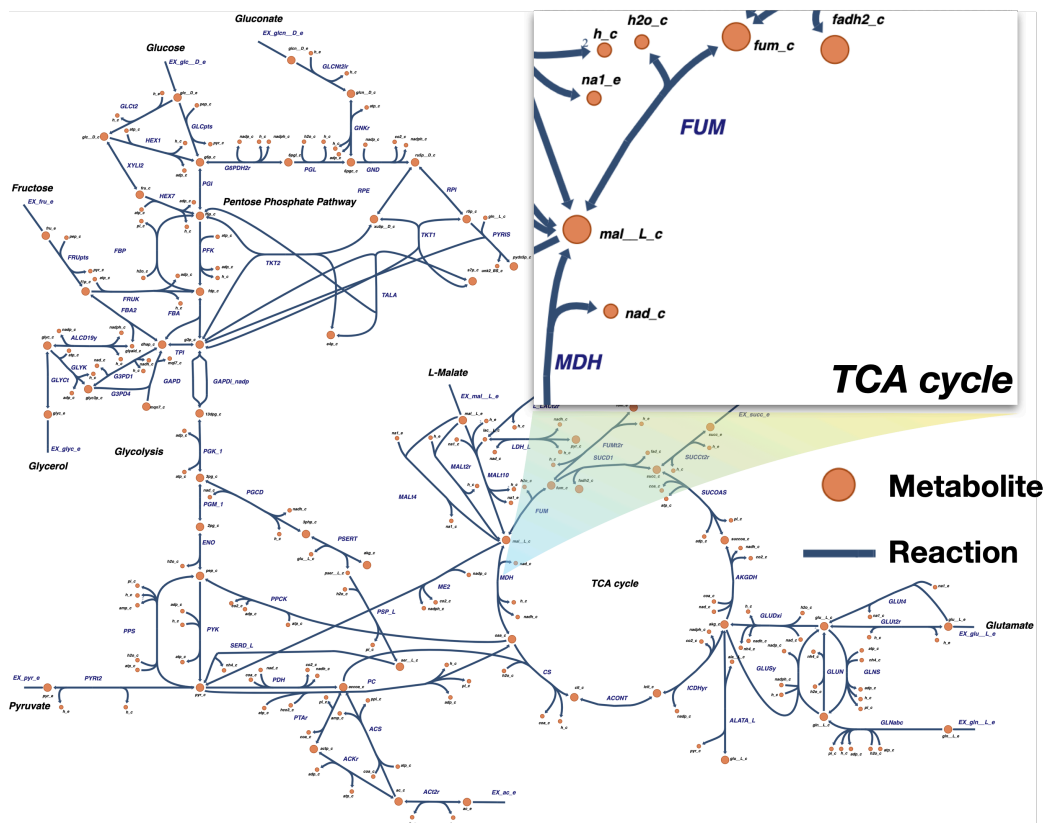


(caption on the next page)

**Figure 1.1. Central energy metabolism inside the metabolic network of _Bacillus subtilis_ generated from a web application, Escher (King et al. 2015).** Central energy metabolism is commonly selected as a starting point to investigate the metabolism since it is where energy is generated. As shown in the zoom-in box, each edge of the network represents a reaction, while each node represents a metabolite. These reactions are connected via different metabolites and form the network. The flow of metabolites of each reaction is generally observed and used to assess the metabolic state. TCA cycle: tricarboxylic acid cycle; FUM: fumarase; MDH: malate dehydrogenase.

## 1.3    $^{13}$C metabolic flux analysis

To access the state of metabolism, we generally observe the flow of metabolites (metabolic flux). By far, $^{13}$C metabolic flux analysis ($^{13}$C-MFA) is the most prominent approach to determine fluxes in metabolic pathways (Buescher et al. 2015, Long and Antoniewicz 2019). Typically, major metabolic pathways of central energy metabolism such as glycolysis, pentose phosphate pathway (PPP), and the tricarboxylic acid (TCA) cycle are included in $^{13}$C-MFA models to observe metabolic fluxes. Metabolites can be traced via $^{13}$C isotope labeling and measured employing mass spectrometry, either gas chromatography-mass spectrometry (GC-MS) or liquid chromatography-mass spectrometry (LC-MS). The metabolite data are then processed and fitted to a metabolic network of the studied organism to estimate the metabolic fluxes. Because of its high accuracy and precision for flux visibility, it has been used as a gold standard and successfully applied in many different fields (Buescher et al. 2012, Chubukov et al. 2013, Zhong et al. 2014, Bideaux et al. 2016, Hauslein et al. 2016, DeWaal et al. 2018, Veras et al. 2019, Yao et al. 2019). Though the method can provide precise fluxes in high resolution, it highly demands experimental and computational resources (Antoniewicz 2015, Buescher et al. 2015, Dai and Locasale 2017, Antoniewicz 2018), shown in Figure 1.2.
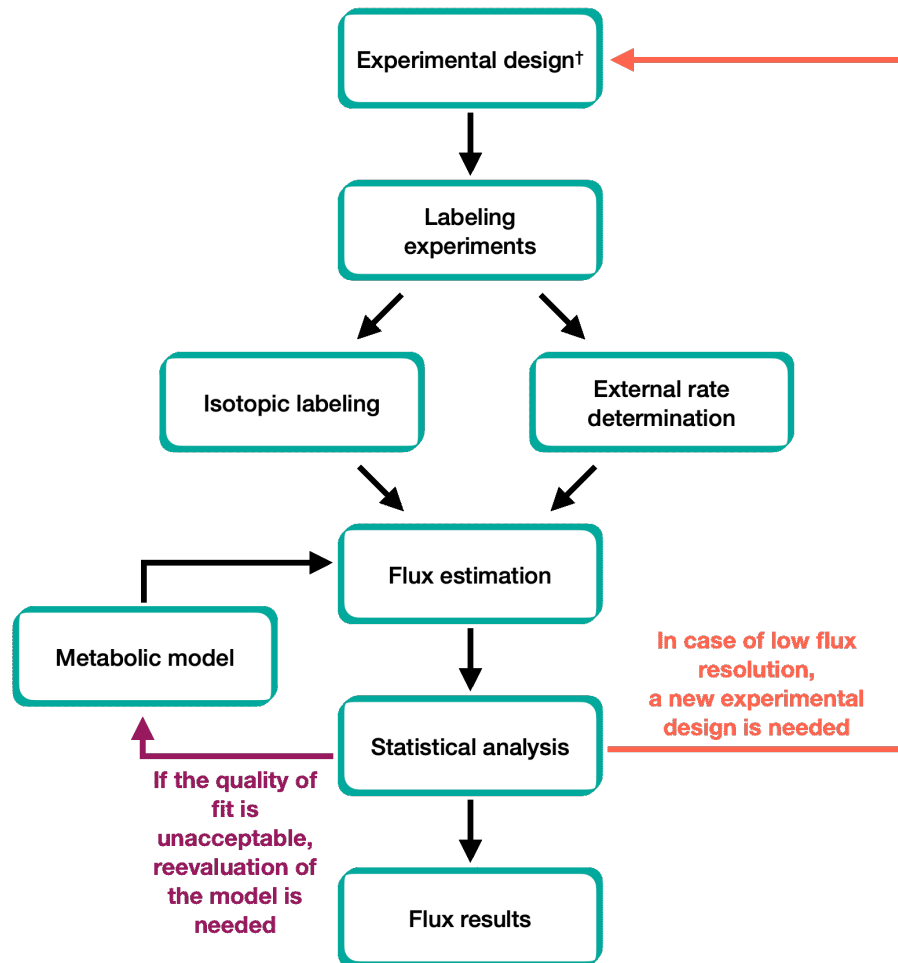
† Design for single, mixed or parallel labeling experiment

**Figure 1.2. Overview of ¹³C-MFA procedures (Adapted from Long and Antoniewicz 2019).** The experimental design starts by selecting suitable $^{13}C$ tracers to perform labeling experiments to investigate the pathway of interest. During labeling experiments, samples are collected to perform external rate determination (i.e., substrate uptake rate, product secretion rate, and growth rate) and isotopic labeling measurement via GC-MS or LC-MS. The external rates and quantified metabolite labeling are fed into the metabolic model to estimate fluxes by minimizing the differences between the observed and simulated isotope labeling patterns. Later, statistical analysis is performed to assess the fitness of fluxes and find the best global fit. When the quality of the fit is unacceptable, the metabolic model may need to be modified, e.g., by adding or removing reactions. If the fit is acceptable, the resolution needs to be assessed. In case of low flux resolution, a new experimental design is generally

conducted. Finally, flux results are obtained when the fit is good, and a high flux resolution is achieved.

One important factor, which greatly determines the quality of flux visibility, is $^{13}$C tracer selection. Poorly selected $^{13}$C tracers can result in inadequate flux resolution, which leads to misinterpretation or uninterpretable results. In general, *in silico* simulations are recommended to find the optimal tracer with a narrow flux confidence interval among all pre-selected tracers. As a starting point, [1,2-$^{13}$C]glucose and [1,6-$^{13}$C]glucose tracers applied in parallel labeling experiments are recommended as the global default for high-resolution $^{13}$C-MFA to study central energy metabolism (Long and Antoniewicz 2019). Further modifications can be added to suit specific needs in interested pathways or non-modeled organisms. By far, the most challenging limitation for this approach is to determine metabolic fluxes in complex environments in which it is difficult to derive the origin of certain metabolites, for example, metabolites in different compartments, e.g., the nucleus or cytosol (Antoniewicz 2018), or in scenarios where pathogens are in host cells, and host cells and pathogens consume or produce the same metabolites.

## 1.4    Constraint-based modeling

### 1.4.1  Overview of constraint-based modeling and optimization methods

Constraint-based modeling (CBM), a mathematical approach commonly used in mathematics and engineering fields to solve optimization problems, has been successfully adapted to biology and emerged as an alternative to estimate fluxes in cell metabolism without conducting labeling experiments (Bordbar et al. 2014, Palsson 2015). CBM solves a problem of interest and obtains an optimal feasible solution by optimizing an objective function (goal of the problem) within an allowable solution space. The optimization methods, which are usually applied, can be put into five major classes: linear programming (LP), mixed-integer linear programming (MILP), quadratic programming (QP), mixed-integer quadratic programming (MIQP), and non-linear programming (NLP). LP and MILP are generally the most used and preferable optimization methods among all five classes (Heirendt et al. 2019, Palsson 2015). Particularly, LP has become a core optimization method (most CBM techniques were developed

using LP (Heirendt et al. 2019)) since it has the great advantage of having the simplest mathematical form and requires the least computational resources.

### 1.4.1.1 Linear programming

LP is an optimization technique using a linear objective function and a linear set of inequality constraints to find an optimal solution within a defined, convex solution space (Figure 1.3). LP is widely known as a special subclass of convex optimization (Boyd and Vandenberghe 2004). The approach can be expressed in the standard form below.

$$\text{Maximize} \qquad c^T x \qquad (1.1)$$

subject to

$$Ax \leq b \qquad (1.2)$$

where $x$ is a vector of variables to be solved ($x \in \mathbb{R}^n$). $c^T$ is a vector of coefficients ($c \in \mathbb{R}^n$), which contribute to the objective function of the problem. $A$ is a given matrix ($A \in \mathbb{R}^{mxn}$). $b$ is a vector which determines the constraints ($b \in \mathbb{R}^m$). By optimizing equation (1.1) while satisfying all constraints from equation (1.2), the optimal solution which maximizes equation (1.1) is found (Figure 1.3).
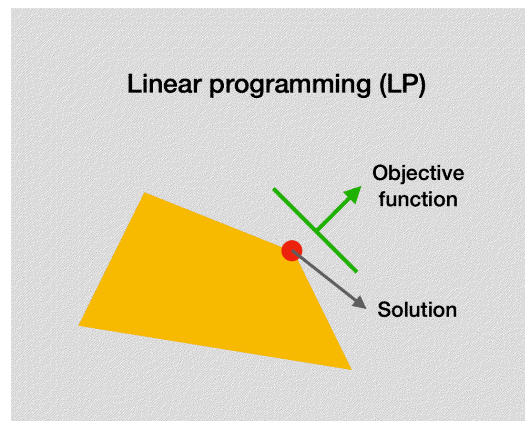


**Figure 1.3. The graphical representation of LP.** The orange quadrilateral is the convex solution space. After optimizing the objective function, the optimal solution is found on the red spot.

The following problem of oil blending demonstrates an easy example for an LP application. A fossil fuel company produces two types of oil blends, $A$ and $B$. Each oil blend contains similar $i$ types of oil sources but differs in amount.

The difference in the amount of each type $i$ of oil sources brings different costs and revenue to oil blend $A$ and $B$. To maximize a total profit for each production, the company must find the optimal amount between oil blends $A$ and $B$ production that creates high revenue but requires low cost to increase profit (profit = revenue – cost). The problem can be formulated as follows.

$$\text{Maximize} \quad \left(\sum_{i=1}^{n} r_A x_{A,i} + \sum_{i=1}^{n} r_B x_{B,i}\right) - \left(\sum_{i=1}^{n} c_A x_{A,i} + \sum_{i=1}^{n} c_B x_{B,i}\right) \quad (1.3)$$

subject to

$$\frac{x_{A,i}}{\sum_{i=1}^{n} x_{A,i}} \geq a_{A,i} \quad (1.4)$$

$$\frac{x_{B,i}}{\sum_{i=1}^{n} x_{B,i}} \geq a_{B,i} \quad (1.5)$$

$$x_{A,i} + x_{B,i} \leq b_i \quad (1.6)$$

where $x_{A,i}$ is a vector of oil source variables in oil blend $A$, and $x_{B,i}$ is a vector of oil source variables in oil blend $B$. $r_A$ is a vector of coefficients from revenue of oil blend $A$, while $r_B$ is a vector of coefficients from revenue of oil blend $B$. $r_A$, $c_A$, $r_B$, and $c_B$ are representing the coefficients from revenue of oil blend $A$, cost of oil blend $A$, revenue of oil blend $B$, and cost of oil blend $B$, respectively. $a_{A,i}$ and $a_{B,i}$ are constraints for a minimum ratio of each oil source $i$ in oil blend $A$ and $B$. $b_i$ is a constraint for an available stock of each oil source $i$. After maximizing the objective function (shown in equation (1.3)), which is the function of revenue subtracted with cost, if the problem is feasible under the confined solution space, the optimal amount of blend $A$ and $B$ production that maximizes the profit is found.

Besides the oil blending problem, LP has been frequently used to solved other similar optimization problems. Another example for LP application is shown below (see section 1.4.2.1 Flux balance analysis).

### 1.4.1.2 Mixed-integer linear programming

Besides using linear constraints to determine the feasible solution space as in LP, MILP also employs integers to form the optimization problem. Introducing integers into the system splits the continuous solution space into fractional sub-spaces, and the optimal solution can be found on the integer point (Figure 1.4). The standard form of MILP is similar to LP shown in the following:

$$\text{Maximize} \quad c^T x \quad (1.7)$$

subject to

$$Ax \leq b \tag{1.8}$$

$$\exists x_i \in \mathbb{Z} \tag{1.9}$$

where $x$ is a vector of variables to be solved ($x \in \mathbb{R}^n$). $c^T$ is a vector of coefficients ($c \in \mathbb{R}^n$), which contribute to the objective function of the problem. $A$ is a given matrix ($A \in \mathbb{R}^{mxn}$). $b$ is a vector which determines the constraints ($b \in \mathbb{R}^m$). In MILP, some $x$ are restricted to be integers, as shown in equation (1.9).
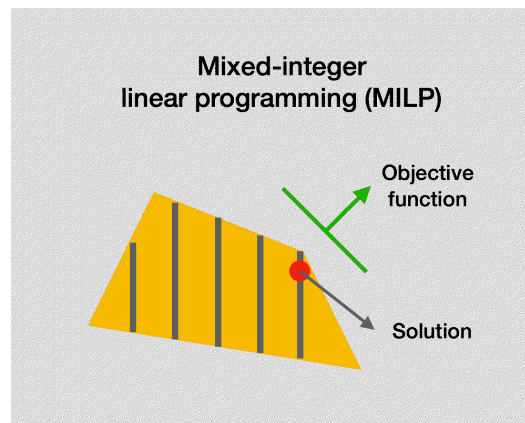


**Figure 1.4. The graphical representation of MILP.** The orange quadrilateral is the convex solution space. By introducing integers, the solution space is separated into several fractional subspaces. Each gray line represents feasible solutions in the system. The optimal solution can be found on the red spot after optimizing the objective function.

By including discrete values in the system, MILP can solve optimization problems that LP cannot solve, such as special problems that involve decision making (e.g., yes/no, 1/0) or integer quantity problems (e.g., number of cars, houses, cells, persons, reactions). The example showing how MILP is applied to solve the problem containing integers and continuous variables, which distinguishes MILP from LP, can be explained with a variant version of a well-known knapsack problem.

In this problem, there are two types of items, $x_i$ and $y_i$. $x_i$ is inseparable, but $y_i$ can be split into different fractions. Both $x_i$ and $y_i$ come with different values ($v_{x,i}$ and $v_{y,i}$) and weights ($w_{x,i}$ and $w_{y,i}$). Although a goal is fitting items into a bag as many as possible while maximizing a total value, the optimal

solution of $x_i$ and $y_i$ must be found under a maximum weight capacity of the bag of $W$ kilograms. The problem can be formulated below.

$$\text{Maximize} \quad \sum_{i=1}^{n}(v_{x,i}x_i + v_{y,i}y_i) \qquad (1.10)$$

subject to

$$\sum_{i=1}^{n}(w_{x,i}x_i + w_{y,i}y_i) \leq W \qquad (1.11)$$

$$\sum_{i=1}^{n} x_i \geq 1 \qquad (1.12)$$

where $x_i$ is an integer variable as the item $x_i$ must be counted integrally ($x_i \in \mathbb{Z}$). Meanwhile, $y_i$ is a continuous variable since the item $y_i$ can be fractional. In equation (1.12), it is also required that at least one $x_i$ be placed inside the bag. If all conditions are met, the algorithm finds the optimal solution for a combination of items from each item type that provides the highest total value. Another example to demonstrate the MILP optimization problem is shown in the newly developed method from this thesis (see Chapter 3, Materials and Methods).

### 1.4.2 Fundamental methods in constraint-based modeling

Being the purely computational approach, CBM circumvents limitations of $^{13}$C-MFA and provides researchers opportunities to easily reconstruct metabolic networks at the genome scale as well as explore metabolic fluxes under different conditions ranging from simple (laboratory conditions) to complex environments (clinical settings). Many techniques, e.g., flux balance analysis (FBA), flux variability analysis (FVA), elementary flux mode analysis (EFM), or extreme pathway analysis (EPA), were developed and deployed to provide good predictive models and solve arisen questions such as increasing production yields in metabolic engineering (Bideaux et al. 2016, Chiewchankaset et al. 2019, Dang et al. 2017, Veras et al. 2019), improving bacteria strains by knocking out genes/enzymes in biotechnology (Lu et al. 2015, Yao et al. 2019, Zhong et al. 2014), or discovering new drug targets in health sciences (Chenard et al. 2017, DeWaal et al. 2018, Gatto et al. 2015, Shan et al. 2018, Sung et al. 2017). The most used methods among other CBM approaches, which are considered fundamental techniques, are FBA and its variation, FVA (Orth et al. 2010, Bordbar et al. 2014, Palsson 2015).

### 1.4.2.1       Flux balance analysis

FBA, known as the basis of flux estimation in CBM, employs LP for the optimization method to predict metabolic fluxes inside metabolic networks (Orth et al. 2010). To implement FBA for flux simulation (Figure 1.5), first, the models are assumed to be at a steady state. The steady-state condition means there is no accumulation of mass in the system resulting in constant metabolite concentrations over time. After selecting and applying the objective function $c^T v_i$ of the model to optimize (e.g., biomass production in bacteria or cancer cells), the metabolic flux vector $v_i$ is derived within a confined solution space based on physiochemical constraints from a stoichiometric matrix $S_i$ of metabolites and reactions inside the metabolic network. Each $v_i$ is set to fall in between a lower bound $lb_i$ and an upper bound $ub_i$ of each reaction $i$. The mathematical formulation is shown in the following.

$$\text{Maximize} \qquad\qquad c^T v_i \qquad\qquad\qquad (1.13)$$

subject to

$$S_i \cdot v_i = 0 \qquad\qquad (1.14)$$

$$lb_i \leq v_i \leq ub_i \qquad\qquad (1.15)$$

where $c^T$ is a vector of weights, which accounts to how much each reaction contributes to the objective function. If there is only one reaction aimed for minimization or maximization (for example, biomass production reaction), $c^T$ is set to be the vector of zeros with a one at the position of the reaction of interest. To note, as FBA requires no experiment, choosing the rational objective function or modeling assumption is critical to final prediction results (Bordbar et al. 2014). Maximizing biomass production is commonly selected as the objective function for model optimization, as shown in bacterial or cancer models (Orth et al. 2010, Bordbar et al. 2014, Palsson 2015). The formation of biomass accumulation is based on the hypothesis that the primary objective of cells is to grow, which is reasonable for bacteria and cancer cells. However, in other circumstances, setting the biomass production reaction as the objective function can be inappropriate. Alternative objective functions such as minimizing ATP production to find conditions that show optimal metabolic energy efficiency or maximizing a particular metabolite production reaction to

determine production capability of a cell of interest can be a reasonable choice for the mentioned purposes (Raman and Chandra 2009, Bordbar et al. 2014).
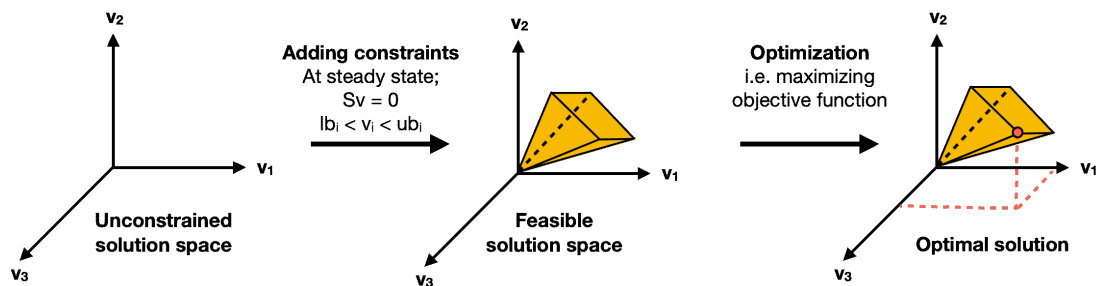


**Figure 1.5. The basic concept of FBA.** With no constraint, any solution is possible within the solution space. To estimate fluxes with FBA, the system is in a steady state. Stoichiometric constraints are introduced together with capacity constraints (upper and lower bounds). The solution space is shrunk to form the feasible solution space. After setting the objective function, e.g., maximize a biomass production reaction, the optimal solution is found inside the feasible solution space.

Besides stoichiometric constraints, additional constraints, e.g., thermodynamic constraints from reaction directionality or maximal flux (maximal enzymatic velocity, $V_{max}$) from enzyme kinetic data, can also be added to limit the solution space further. However, this information is often not available as it can be experimentally demanding (Schellenberger et al. 2011, Bordbar et al. 2014). Overall, by assuming a steady-state condition, FBA allows generating simplified models and requires no reaction kinetic parameters to construct the models at a genome scale, which can be a great benefit of constraint-based modeling compared to not only ^{13}C-MFA but also other computational modeling methods, for example, kinetic models (Bordbar et al. 2014).

### 1.4.2.2 Flux variability analysis

At different steady-state conditions, feasible minimal and maximal fluxes for each reaction in the metabolic network can vary between environments and differ from their lower and upper bounds. Under the same framework as FBA, FVA is a well-known technique in CBM to determine flux ranges (Mahadevan and Schilling 2003). FVA is explained in the following.

For each reaction $i$ in the metabolic network, it is assumed that the metabolic network is in steady state, and the stoichiometry is fulfilled. At a fixed objective function value, e.g., growth rate, FVA solves two LP problems, i.e., minimization and maximization of flux $v_i$ to identify the flux span for reaction $i$ while satisfying the FBA constraints, i.e.

$$\text{Minimize} \qquad v_i \qquad\qquad (1.16)$$

subject to

$$S_i \cdot v_i = 0 \qquad\qquad (1.17)$$

$$lb_i \leq v_i \leq ub_i \qquad\qquad (1.18)$$

For maximization, the objective function (equation (1.16)) switches from minimization to maximization, and the same procedure applies. After performing FVA for all reactions in the network, the flux span is obtained. The flux span is a useful measure to determine flux characteristics and assess flexibility of the network in the specific environment (Palsson 2015). Furthermore, as these metabolic network models generally allow a large range of solutions, it is possible that there are multiple optimal solutions existing for the same objective function. The alternative solution is called alternative optimal solution (AOS). FVA can also be combined with other CBM methods such as EPA and randomized sampling to identify AOS (Palsson 2015).

### 1.4.3 Limitation of constraint-based modeling

### 1.4.3.1 Physiochemical constraints alone may fail to reflect the actual flux state

Although the CBM approach is beneficial for predicting metabolic fluxes in the metabolic network without conducting $^{13}C$ isotope labeling experiments, it has its limitation. Myriad biological components from different regulatory levels, e.g., genomics, transcriptomics, proteomics, and metabolomics, are linked together in a fine-grained network to regulate metabolism to ensure survival and growth. The physiochemical data as the only layer of data in the model was shown to be inadequate to provide a good estimation of metabolic fluxes. This led to many attempts to integrate other relevant experimental data, in particular, transcriptomic data into the metabolic model to improve flux predictions (Becker and Palsson 2008, Shlomi et al. 2008, Chandrasekaran and Price 2010, Zur et al. 2010, Wang et al. 2012, Heirendt et al. 2019). Although the data is indirect

compared to $^{13}C$ labeling data, it is less labor-intensive to generate (Lowe et al. 2017, Uygun et al. 2016, van den Esker and Koets 2019) and was shown to deliver good predictions while preserving the simplicity of the model (Becker and Palsson 2008, Chandrasekaran and Price 2010, Shlomi et al. 2008, Wang et al. 2012, Zur et al. 2010).

In the past years, various techniques were developed to integrate gene expression data and construct context-specific metabolic models. Most of the methods (e.g., Integrative Metabolic Analysis Tool (iMAT) (Shlomi et al. 2008, Zur et al. 2010), Gene Inactivity Moderated by Metabolism and Expression (GIMME) (Becker and Palsson 2008), Probabilistic Regulation of Metabolism (PROM) (Chandrasekaran and Price 2010) or the metabolic Context-specificity Assessed by Deterministic Reaction Evaluation (mCADRE) (Wang et al. 2012)) integrate gene expression data to the metabolic network by setting a threshold to define qualitatively binary expressed/non-expressed reactions.

### 1.4.3.1.1    Gene Inactivity Moderated by Metabolism and Expression

To build context-specific models from transcriptomic data, GIMME (Becker and Palsson 2008) introduces a threshold to define an expression state of a gene, which refers to an activity of a reaction. If a gene expression level is higher than the threshold, the reaction is active and remains in the model. If not, the reaction is first removed from the model, and it may be re-inserted to the reduced model by the algorithm if the particular reaction is required for the model to achieve its functional goal (e.g., biomass production or ATP production). However, the algorithm is set to minimize the number of re-inserted reactions as these reactions are shown to be inconsistent with the gene expression data (Becker and Palsson 2008). Finally, GIMME creates the context-specific model containing reactions with minimal disagreement with the transcription profiles under the selected threshold.

### 1.4.3.1.2    Integrative Metabolic Analysis Tool

In order to process and integrate transcriptomic data into the metabolic network, iMAT (Shlomi et al. 2008, Zur et al. 2010) first categorizes reactions into three groups: highly expressed (active) reaction, moderately expressed reaction, and lowly expressed (inactive) reaction. The categorization is based on gene

expression levels of corresponding genes related to each reaction, defined by two cutoffs, i.e., the upper and lower thresholds. By this, gene expression profiles are binarized, i.e., they are converted from a continuous to a discrete value. Then, iMAT forms a MILP optimization problem aiming to maximize the number of highly and lowly expressed reaction groups since their activity is consistent with their expression state (Shlomi et al. 2008, Zur et al. 2010). However, the moderately expressed reaction group is not part of the optimization goal since the situation is unclear whether it is controlled by transcriptional regulation or not, as the gene expression level falls between the upper and lower thresholds. If the model predicts the moderately expressed reaction having a non-zero flux, the reaction is interpreted as a post-transcriptional regulated reaction. With this concept, iMAT can integrate gene expression data into the metabolic models and create various context-specific metabolic models based on different conditions.

### 1.4.3.1.3    Probabilistic Regulation of Metabolism

Like GIMME, PROM (Chandrasekaran and Price 2010) also uses a fixed threshold to define the state of a gene. When the threshold is higher than the expression level, the gene is set to be off. In contrast, if the threshold is lower than the expression level, the gene is set to be on. Despite using a similar approach as GIMME to binarize gene expression data, PROM introduces probability to further process the binarized gene expression data in order to prevent the algorithm from completely turning the corresponding reaction on or off. By calculating the probability of the gene being expressed to an activity of its transcription factor, the method predicts how often the specific gene is being on or off across all samples in each condition. The computed probability ranges from zero (completely off) to one (completely on). PROM uses the calculated probability to adjust a flux range for each corresponding reaction. This allows PROM to predict more than two states of the reaction making flux predictions become less binary (Chandrasekaran and Price 2010). Still, the threshold is implemented to decide the expression state of each gene, which can be problematic when the inappropriate threshold value is selected. Moreover, compared to other approaches which require only gene expression data, PROM needs prior knowledge of transcription factor-gene interactions. The method

also requires a large number of transcriptomic datasets of the same condition to correctly calculate the probability (Chandrasekaran and Price 2010), while the other approaches (e.g., GIMME, iMAT) can use a single transcriptomic dataset per condition (Becker and Palsson 2008, Shlomi et al. 2008, Zur et al. 2010).

### 1.4.3.1.4 The metabolic Context-specificity Assessed by Deterministic Reaction Evaluation

In mCADRE (Wang et al. 2012), the method first transforms gene expression levels in continuous values to binary values. This separates genes into two groups: present call and absent call. While the present call is assigned to one, the absent call (a marginal call also counts in this case) is assigned to zero. With this binarized data, mCADRE calculates expression-based evidence scores to quantify how often a gene is expressed across all samples in each condition. The expression-based evidence score is compared to a certain threshold to identify two sets of corresponding reactions: a high-confidence core set (the expression-based evidence score is above the threshold) and a non-core set (the expression-based evidence score is below the threshold). Notably, different data binarization techniques may require different values for the threshold (Wang et al. 2012). Then, mCADRE builds context-specific models by keeping high-confidence core reactions and removing unnecessary non-core reactions if removing these reactions does not prevent the model from achieving the functional goal.

Although these methods succeeded in improving flux prediction compared to relying on physiochemical data alone, employing defined thresholds to decide the activity state of genes may disregard the fine-grained regulation of metabolism. Also, since appropriate values for the thresholds can vary between different genes, organisms, or conditions, finding the suitable threshold can be challenging. A novel technique to efficiently integrate transcriptomic data while preserving the characteristic of the data is still in need.

### 1.4.3.2 Thermodynamic infeasible loops

Many CBM methods, including FBA, ignore the imposition of the loop law. The loop law is similar to Kirchhoff's second law for electrical circuits (Price et al. 2002). It states that there must not be any closed cycle or loop in the metabolic network with a non-zero net flux at a steady state. The non-zero net flux loop is also recognized as a type 3 extreme pathway in EPA (Figure 1.6, right). Such loops would disregard the second law of thermodynamics and are hence thermodynamically infeasible.
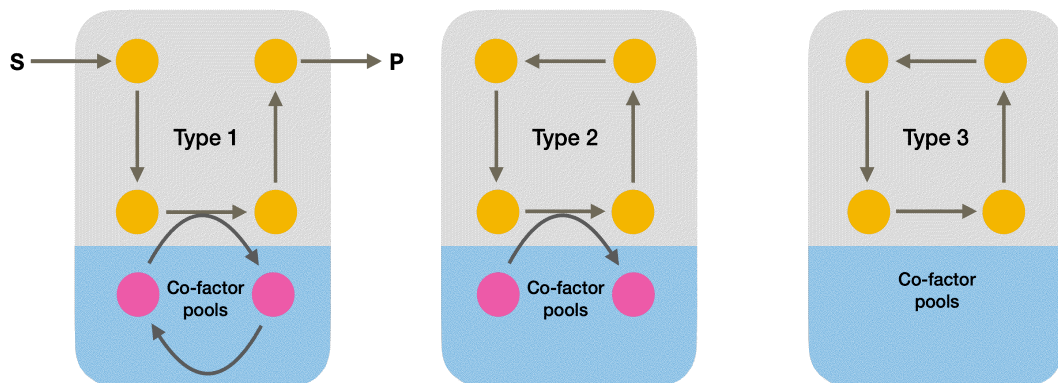


**Figure 1.6. Three basic categories of extreme pathways.** At steady state, sets of fluxes can be defined into three types via EPA: type 1, type 2, and type 3. Type 1 has exchange fluxes connecting between the system and the external environment. The substrate S is used to drive other reactions in the network as well as used in co-factor pools and produce the product P. For type 2, although fluxes show no connection with external exchange fluxes, reactions are connected to reactions in co-factor pools (e.g., ATP-ADP, NADPH-NADP+) which drive the reactions internally. However, in type 3, fluxes do not connect to any exchange reactions that can act as a driving force for the cycle. This situation is considered thermodynamically infeasible and should be removed. S; substrate and P; product.

Neglecting the loop law allows thermodynamically infeasible loops (TIL) to be used by the optimization algorithms resulting in an unrealistic flux distribution (Figure 1.7). The problem of avoiding thermodynamically infeasible loops can be solved by imposing thermodynamic constraints such as standard-state free energy of reactions into the optimization. However, it is very challenging to acquire this information for the whole metabolic network as well as to implement it computationally (Schellenberger et al. 2011).
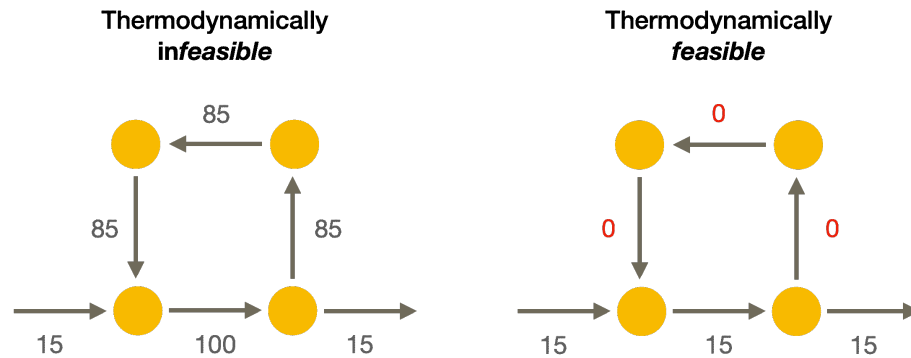
**Figure 1.7. The illustration shows a thermodynamically infeasible solution (left) and a thermodynamically feasible solution (right).** The application of the loop law does not change the optimal value for the objective function of the optimization; however, it indeed restricts the allowable flux distribution inside the metabolic network.

In order to impose the loop law into the optimization and simplify the computation complexity, Schellenberger *et al.* introduced a method called loopless-COBRA (ll-COBRA) (Schellenberger et al. 2011). Instead of enumerating all possible loops in the metabolic network using EPA or EFM, ll-COBRA introduces a new binary variable derived from reaction directionality to force the optimization to occur in the thermodynamically feasible region (loopless) region and obtain loopless flux distribution. By utilizing reaction directionality that is the readily existing information in the metabolic network, the method bypasses the need for standard-state free energy and only creates a MILP problem. As a result of this, the approach takes less time to calculate the thermodynamically feasible flux distribution for the whole network than other methods to solve a similar problem (Schellenberger et al. 2011). However, although ll-COBRA reduces the computational time significantly, it takes a significant amount of time to find the loopless optimal solution; particularly, when the computation requires many iterations. As a result of this, the novel method, which can give a comparable result but consume less runtime, is still in need.

## 1.5    *Bacillus subtilis*

### 1.5.1   Overview

*Bacillus subtilis* (*B. subtilis*), also known as hay bacillus or grass bacillus, is a Gram-positive, rod-shaped bacterium. It can be found in soil and in the

gastrointestinal tract of ruminants and humans. Due to its high competence of extracellular DNA uptake for genetic manipulation, *B. subtilis* gained its popularity in research and has been studied extensively in many aspects (Figure 1.8) such as bacterial cell division, surface motility, protein secretion, biofilm formation, and secondary metabolite production (Kovács 2019, Errington and Aart 2020, Su et al. 2020, Harwood et al. 2018). Also, *B. subtilis* has been used to study sporulation in bacteria as the bacterium undergoes sporulation processes and forms an endospore to survive in stressful environments, e.g., due to change of pH, radiation, starvation, or an extreme temperature (Paredes-Sabja et al. 2011, McKenney et al. 2013). Equivalent to *Escherichia coli (E. coli)* as a model organism for Gram-negative bacteria, *B. subtilis* is the model organism for Gram-positive bacteria (Kovács 2019, Errington and Aart 2020, Su et al. 2020).
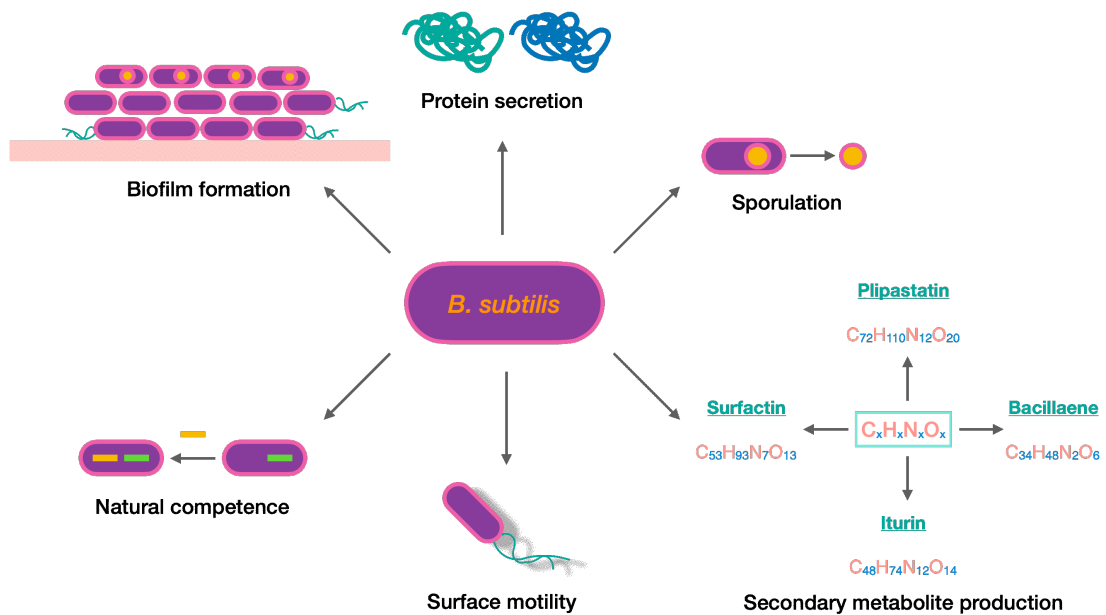


**Figure 1.8. Major usages of *B. subtilis* in different aspects.**

### 1.5.2 Carbon source utilization in *Bacillus subtilis*

*B. subtilis* can utilize many different carbon sources such as monosaccharides, disaccharides, amino acids, or large molecules like cellulose to generate building blocks and energy for growth promotion (Stulke and Hillen 2000). However, like other bacteria, glucose is the preferred carbon source for *B. subtilis* (Kleijn et al. 2010, Stulke and Hillen 2000). When glucose is available in an environment, the bacteria exert catabolite repression to quickly uptake

glucose into the system and utilize it. Catabolite repression is a global regulatory system to allow micro-organisms to switch and utilize the preferred carbon sources (considered to be more beneficial for the cell) even though the other carbon sources are still present. The adaptation can be achieved by inhibiting the synthesis of enzymes involved in the catabolism of other carbon sources than the preferred ones (Deutscher 2008, Stulke and Hillen 2000). In the case of *B. subtilis,* this process is controlled by catabolite control protein A (CcpA), which is a transcription factor regulating carbon utilization in Gram-positive bacteria (Stulke and Hillen 2000, Lorca et al. 2005).

Besides glucose, malate is recognized as another highly preferred carbon source of *B. subtilis* (Kleijn et al. 2010, Meyer and Stulke 2013). Kleijn *et al.* discovered that while *B. subtilis* suppressed the utilization of other carbon sources (gluconate, glycerol, succinate, fructose, arabinose, or pyruvate) when glucose was available, the bacteria co-utilized malate and glucose and achieved a higher growth rate compared to growing the bacteria on malate or glucose as the only carbon source (Kleijn et al. 2010).

# Chapter 2

## Objectives

Gaining insight into metabolism is of importance to understand how cells behave or adapt to changes in the environment. Studying metabolic fluxes provides access to this valuable information. Although $^{13}$C-MFA is by far the most prominent approach providing flux estimates with high accuracy and precision, it is experimentally and computationally expensive. Alternatively, FBA has been employed to estimate metabolic fluxes without labeling experiments. It is less informative but can benefit from the low costs and low experimental efforts and gain flux estimates in experimentally difficult conditions. Methods to integrate relevant experimental data have been established to improve FBA flux estimations. Transcriptomic profiling is often selected as the data of choice since it is easy to generate at the genome scale. Most methods typically embed gene expression data by binarizing differential and non-differential expressed genes coding for the respective enzymes. However, employing defined thresholds can result in disregarding the fine-grained regulation of metabolism. Thus, the first aim of this study was to develop a novel context-specific model extraction method employing transcriptomic data in a continuous way to improve flux prediction.

Besides this, TIL is a well-known problem in constraint-based modeling, leading to unrealistic flux distributions, and particularly when studying such omics-based models. Though the problem is well-addressed theoretically, the practice of TIL removal is still challenging due to computational complexity. This led to the second goal of the study, which was to develop a new method to improve TIL reduction.

Nutrition is essential for any cell to guarantee its survival. Having a model that can identify correct carbon sources of a cell is very beneficial to investigate how the cell utilizes its carbon sources under different environments. In this study, the Gram-positive bacterium *B. subtilis,* which is frequently used as the model organism in research and has good experimental data available, was

used to develop the context-specific model that can predict the correct carbon sources (this served as the final purpose of the study).

# Chapter 3

# Materials and Methods

## 3.1 Data assembly

### 3.1.1 Experimental data of the eight-carbon-source study (first dataset)

Published microarray gene expression data of the *B. subtilis* strain BSB1 was used. BSB1 is a tryptophan prototrophic derivative of strain 168. Tiling arrays covering the whole genome of *B. subtilis* 168 were used to measure the expression level (Nicolas et al. 2012). In this dataset, *B. subtilis* was grown in minimal medium in eight different carbon source conditions (glucose, fructose, gluconate, glutamate/succinate, glycerol, malate, malate/glucose, pyruvate). The data was taken from the original publication (Table S2 from (Nicolas et al. 2012)). In order to validate the model, metabolic flux data from $^{13}C$ isotope labeling experiments of the same eight carbon source conditions (Table S4 from (Chubukov et al. 2013)) was used. This data is denoted as the first dataset in the following.

### 3.1.2 Experimental data of the nutritional-shift study (second dataset)

In order to validate the trained model with a separate, unseen dataset, publicly available gene expression data and $^{13}C$-tracer based metabolic flux data from a time-series experiment of two nutritional shifts, i.e., the shift from glucose to glucose plus malate and the shift from malate to malate plus glucose, were used (Buescher et al. 2012). Gene expression data and $^{13}C$ metabolic flux data were generated using the same experimental protocol described in the first dataset. *B. subtilis* was grown in minimal medium on a single carbon substrate until an $OD_{600}$ of 0.5 was achieved. Then, the other substrate (glucose or malate) was added to the culture to assess the bacterial behavior at 0 (before the addition of the other substrate), 5, 10, 15, 25, 45, 60, and 90 minutes after the other substrate was added. Both gene expression data and $^{13}C$ metabolic flux data were taken from the BaSysBio database (https://basysbio.ethz.ch/openbis/basysbio_openbis.html). This data is denoted as the second dataset in the following.

## 3.2    Data pre-processing

The gene expression data of the first and second dataset had been pre-processed by computing the median of the estimated transcription signal of all probes assigned to one corresponding gene (Buescher et al. 2012, Nicolas et al. 2012). The gene expression data in the second dataset had been further processed by quantile normalization (Buescher et al. 2012). All gene expression levels were provided after log2 transformation (Buescher et al. 2012, Nicolas et al. 2012). In order to obtain the gene symbols, BSU identifiers were matched with gene symbols using bioDBnet, version 2.1 (Mudunuri et al. 2009). In the first dataset, each condition contained three biological replicates.  For most of the time points of the second dataset, three biological replicates were available. The rest had two biological replicates. For each condition or time point, gene expression levels across the available replicates were averaged.

Gene-protein-reaction (GPR) mapping was performed to map gene expression values to proteins and reactions. The GPR data was taken from the original publication of Chubukov *et al.* (Chubukov et al. 2013) and the metabolic network of *B. subtilis* 168 from the BiGG Models database (BiGG ID iYO844) (King et al. 2016). The data was compared with the information from UniProt (UniProt Consortium 2018) and KEGG (Kanehisa et al. 2017, Kanehisa and Goto 2000, Kanehisa et al. 2016) and corrected if these databases contained different information. Additionally, literature about two more genes (lrgA, lrgB) coding for a pyruvate transporter was found (van den Esker et al. 2017). Therefore, these two genes were added and linked to the corresponding reaction in the GPR data. [13]C metabolic flux data from Chubukov *et al.* (Chubukov et al. 2013) and Buescher *et al.* (Buescher et al. 2012) were used as published without further processing.

## 3.3    Model building

### 3.3.1  Building the metabolic model

The iYO844 model of *B. subtilis* was transferred from Matlab to R (stoichiometric matrix, lower and upper bounds, reversibility, metabolite, and reaction names) to develop a MILP based model in R. The ^{13}C metabolic flux data was assessed if it fitted into the metabolic model in R. The solution from the ^{13}C model needed to be a feasible solution complying with all set

constraints. Checking the feasibility of $^{13}$C metabolic flux data to the metabolic model was done to ensure that the prediction results from the approach developed within this study (named as Linear Programming based Gene Expression Model, LPM-GEM) and $^{13}$C metabolic flux data could be efficiently compared. However, initial trials showed that no solution was found in the solution space when flux values from the $^{13}$C metabolic flux data were used allowing only one exchange reaction flux to be non-zero, i.e., from the specific transporter of the corresponding carbon source. In turn, it was possible to find a feasible solution when fluxes from other exchange reactions besides the exchange reaction of the corresponding carbon source were allowed to be non-zero for the influx. Although the solution was found, the flux values from other exchange reactions were substantially high, which was unrealistic. Hence, to find a reasonable boundary for each of these exchange reactions, an optimization problem was set up by letting the solution deviate from $^{13}$C metabolic flux data by maximal 0.1 mmol h$^{-1}$ gcdw$^{-1}$. After optimization, a sum of fluxes from other exchange reactions for each different condition was obtained and compared amongst all conditions. The lowest possible value (sum of fluxes = 0.688 mmol h$^{-1}$ gcdw$^{-1}$) restricting the influx of all other metabolites (not the metabolite of the corresponding condition) into the cell was applied to the model. The list of all exchange reactions besides the designated carbon sources of the corresponding minimal medium is provided in the appendix in Table A1.

### 3.3.2 Defining the set of reactions for the optimization criterion

As described below, the generated models were validated with a well-defined gold standard, i.e., flux values based on the $^{13}$C labeling data from the original publication. This gold standard data was available for 40 reactions, mostly covering central energy metabolism (Chubukov et al. 2013). Hence, these reactions were used in the optimization process when gene expression values were mapped to predicted fluxes in the metabolic model (explained in the next section). These reactions are called core reactions in the following. A selection of further reactions was added to the optimization function of the model to improve the model predictions, called associated reactions in the following. Associated reactions were added following three criteria, (1) they needed to be

reactions that were directly connected (via an exchanging metabolite) to the core reactions in central energy metabolism or amino acid biosynthesis, (2) important metabolites in glycolysis or tricarboxylic acid (TCA) cycle (e.g., glyceraldehyde 3-phosphate, pyruvate, oxaloacetate, α-ketoglutarate) which are substrates or products of these reactions, and (3) at least one of the associated genes to the reactions needed to be differentially expressed in at least one out of the eight carbon sources of the first dataset when compared to the expression of *B. subtilis* in the control medium (LB medium) (Buescher et al. 2012). In order to identify differentially expressed genes, T-tests were performed comparing the expression value of the corresponding gene in each specific carbon source condition *versus* its expression in the control medium. The Benjamini-Hochberg method was used to correct for multiple testing across all genes (Benjamini and Hochberg 1995). The p-value cutoff was 0.05. By this, 119 genes and 138 reactions were assembled (Table A1).

## 3.4 Formulating the optimization criterion

Assuming that the metabolic flux correlates linearly with the expression of the gene coding for the responsible enzyme of the corresponding reaction (Figure A1), gene expression values were linearly mapped to predicted fluxes formulated within the following optimization problem.

Let $v_{ri,c}^{fit}$ represent a gene expression-based flux for reaction $ri$ ( $ri$ is a reaction that is part of core or associated reactions) in condition $c$. $v_{ri,c}^{fit}$ is based on information from gene expression data and the flux range,

$$v_{ri,c}^{fit} = V_{ri}^{min} + \left( \bar{g}_{ri,c} - g_{ri}^{min} \right) \left[ \frac{\left( V_{ri}^{max} - V_{ri}^{min} \right)}{\left( g_{ri}^{max} - g_{ri}^{min} \right)} \right] \tag{3.1}$$

where $\bar{g}_{ri,c}$ is the averaged gene expression value of the gene associated with reaction $ri$ in condition $c$. $g_{ri}^{min}$ is the minimum gene expression value across all conditions of the gene associated with reaction $ri$, $g_{ri}^{max}$ is the maximum gene expression value. $V_{ri}^{min}$ is the minimum possible flux and $V_{ri}^{max}$ is the maximum possible flux across all conditions obtained from flux variability analysis (FVA, see below) for core reactions (CR) and associated reactions (AR).

Under the FBA framework, the metabolism is assumed to be at steady state. Hence, there is no accumulation of mass, and specifically, no change of metabolite concentrations over time. $S_r$ is the stoichiometric matrix of the metabolic network, $v_{r,c}$ represents the predicted flux in the metabolic network for reaction $r$ ($r$ is any reaction in the network) in condition $c$. The variable $v_{r,c}$ must satisfy the constraints from the stoichiometry, as well as lower $lb_r$ and upper bounds $ub_r$, i.e.,

$$S_r \cdot v_{r,c} = 0 \tag{3.2}$$

$$lb_r \leq v_{r,c} \leq ub_r \tag{3.3}$$

Subject to constraints (3.2) to (3.3), the optimization problem was formulated by

$$\text{Minimize} \quad \sum_{ri,c} w_{ri} \cdot \left| v_{ri,c} - v_{ri,c}^{fit} \right| + \alpha \sum_{ro,c} v_{ro,c} \tag{3.4}$$

$$w_{ri} = \begin{cases} \dfrac{1}{V_{ri}^{weight}}, & \forall ri \in \text{CR} \\[2ex] \dfrac{1}{V_{ri}^{weight}+100}, & \forall ri \in \text{AR} \end{cases} \tag{3.5}$$

The formulated objective function is a trade-off between two optimization criteria. The first term, $\sum_{ri,c} w_{ri} \cdot \left| v_{ri,c} - v_{ri,c}^{fit} \right|$, minimizes the error between the predicted flux $v_{ri,c}$ and the gene expression-based flux $v_{ri,c}^{fit}$. The weight $w_{ri}$ is introduced to adjust the term through equation (3.5). The predicted flux $v_{ri,c}$ was adjusted by averaged gene expression values using the weight $w_{ri}$ for each gene encoding the reaction $ri$. $V_{ri}^{weight}$ was obtained by selecting the maximum of absolute values of the maximum or minimum flux from the FVA derived maximal flux values. The weight was set as the reciprocal of this value to make reactions with small and high variances of fluxes equally important to the objective function. The associated reactions (AR) were down-weighted by adding the constant +100 in the denominator. Moreover, reactions which $V_{ri}^{min}$ and $V_{ri}^{max}$ were zero, were discarded. This resulted in lower numbers of reactions leading to 98 reactions basing on 116 genes.

The second term in formula (3.4), $\alpha \sum_{ro,c} v_{ro,c}$, was set to minimize the sum of all predicted fluxes $v_{ro,c}$ from reactions being not CR nor AR coping for the problem of obtaining thermodynamically infeasible loops. To obtain an appropriate $\alpha$ value, the sum of sums of absolute values of fluxes $v_{ro,c}$ across

all conditions, $\sum_{i=1}^{c}\sum_{ro,c}|v_{ro,c}|$, and a total model mapping discrepancy were assessed for each $\alpha$ variation (see Chapter 4, Results). The total model mapping discrepancy $d$ is a coefficient used to measure an overall deviated distance between $v_{ri,c}$ and $v_{ri,c}^{fit}$ from all reactions $ri$ across all conditions. It reflects how good $v_{ri,c}$ resembles $v_{ri,c}^{fit}$ and was derived by

$$d = \sum_{i=1}^{c}\sum_{ri,c}|v_{ri,c} - v_{ri,c}^{fit}| \qquad (3.6)$$

After comparing the sum of sums of absolute values of fluxes $v_{ro,c}$ and the total model mapping discrepancy from different $\alpha$ values, the value of 0.01 was set and selected since the sum of sums of absolute values of fluxes $v_{ro,c}$ was considerably reduced while the total model mapping discrepancy was only moderately increased (see Chapter 4, Results).

The biomass constraint was set for each different condition $c$ based on the publications of Chubukov *et al.* and Buescher *et al.* (Chubukov et al. 2013, Buescher et al. 2012). During learning of the model based on the corresponding gene expression profiles, the lower bounds for all eight carbon source exchange reactions were opened to allow influxes of any possible carbon source. The maximum substrate rate (negative lower bounds) reported in Chubukov *et al.* (Chubukov et al. 2013) across all conditions for each carbon source was taken. These lower bounds of all eight carbon source transporter reactions were set to the minimum values for all conditions. This setting was done to ensure that the method predicts carbon sources without prior knowledge of the carbon source in a certain carbon source condition.

### 3.4.1 Reducing the search space employing Iterative Feasible Flux Space Reduction

A realistic estimate of the lower and upper bounds for the reactions in the model was needed to correctly map the expression data to the metabolic flux. At different steady-state conditions, the feasible minimum and maximum flux within the solution space can differ from the initially set lower and upper bounds of each reaction. FVA is a well-known technique to determine flux ranges (see Chapter 1, Introduction). FVA was applied to determine the minimum and maximum possible fluxes as follows. For each reaction $r$ in condition $c$, in FVA, it is assumed that the metabolic network is in a steady state and the

stoichiometry is fulfilled, as referred to equation (3.2) and (3.3). FVA minimizes and maximizes the flux $v_{r,c}$ to find an upper and lower bound for the respective reaction.

In general, doing this for every reaction should narrow down the flux range of each reaction. However, the observed boundaries did not differ substantially after performing FVA for every core and associated reaction. To further reduce the solution space and limit its flexibility, an iterative approach was developed (Figure 3.1). The approach was part of the training scheme to narrow down the flux ranges; hence, this was applied only to the training data. The method was termed Iterative Feasible Flux Space Reduction (IFFPR).
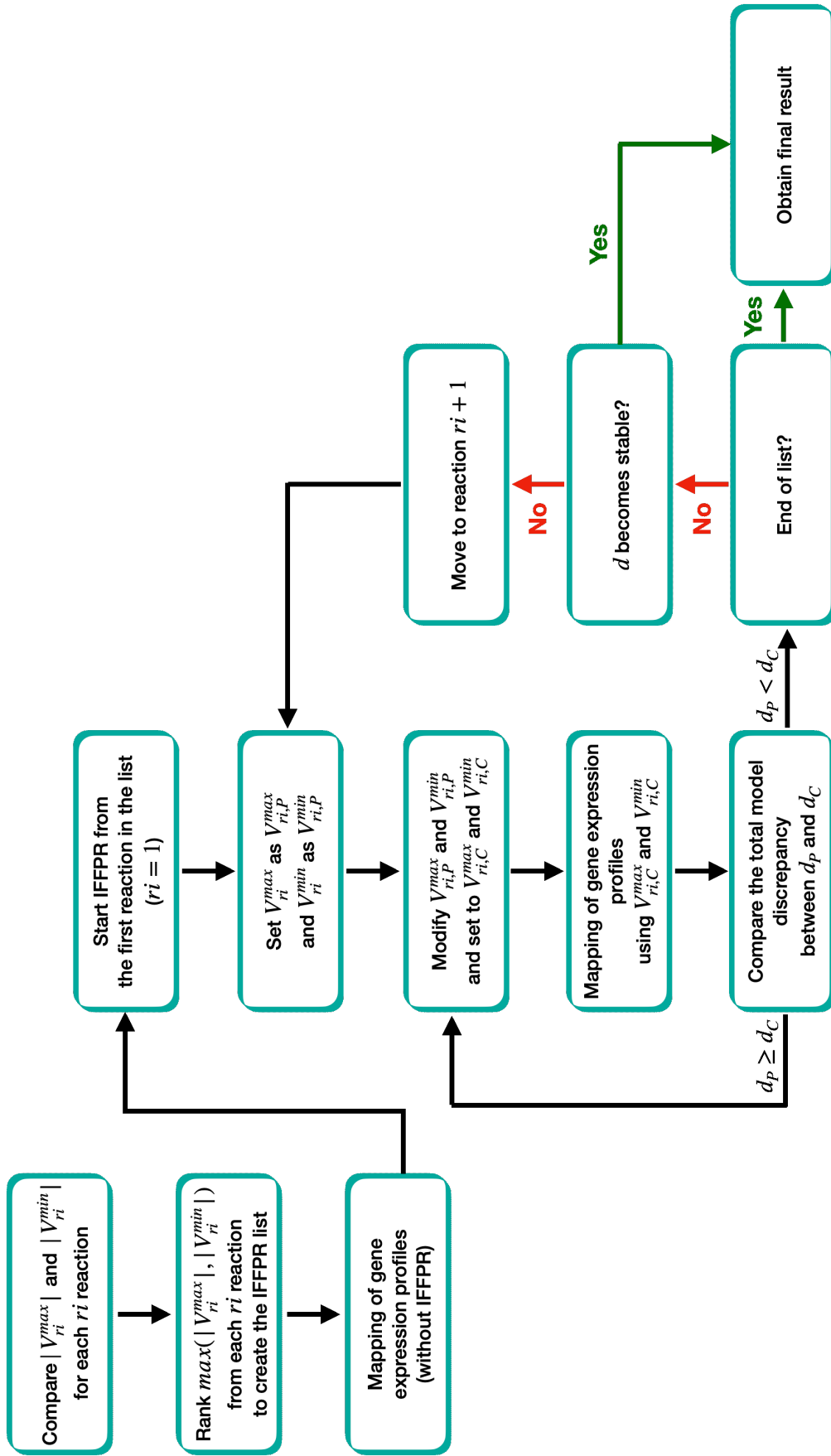
**Figure 3.1. The diagram of IFFPR.** The algorithm is explicitly explained in the text below.

As depicted in Figure 3.1, the approach iteratively adjusts flux ranges based on the total model discrepancy and can be explained in detail as follows:

a)    $V_{ri}^{max}$ and $V_{ri}^{min}$ are acquired by the above-described FVA for each reaction.

b)    Absolute values of $V_{ri}^{max}$ and $V_{ri}^{min}$ from each reaction are compared and the maximum of these values used as the representative maximal bound for this reaction. Representative maximal bounds from all reactions are used to rank the reactions.  The reaction with the highest value is placed at the top position ($ri = 1$).

c)    The first reaction $ri$, with $ri = 1$ is selected.

d)    $V_{ri,P}^{max} = V_{ri}^{max}$ and $V_{ri,P}^{min}= V_{ri}^{min}$ is set.

e)
$$V_{ri,C}^{max} = 0.5 * V_{ri,P}^{max} \tag{3.7}$$

$$V_{ri,C}^{min} = 0.5 * V_{ri,P}^{min} \tag{3.8}$$

in which $V_{ri,C}^{max}$ is the new maximal possible flux for the current iteration, $V_{ri,C}^{min}$ is the new minimal possible flux for the current iteration. $V_{ri,P}^{max}$ and $V_{ri,P}^{min}$ are reduced by half every iteration. Since the reaction can be unidirectional or bi-directional, $V_{ri,P}^{max}$ and $V_{ri,P}^{min}$ can have similar or different signs. Equation (3.7) and (3.8) are applied to reduce $V_{ri,P}^{max}$ and $V_{ri,P}^{min}$. If $V_{ri,P}^{max}$ and $V_{ri,P}^{min}$ have the same sign, either equation (3.7) or (3.8) is used depending on the sign aiming to reduce the flux range (Figure 3.2).
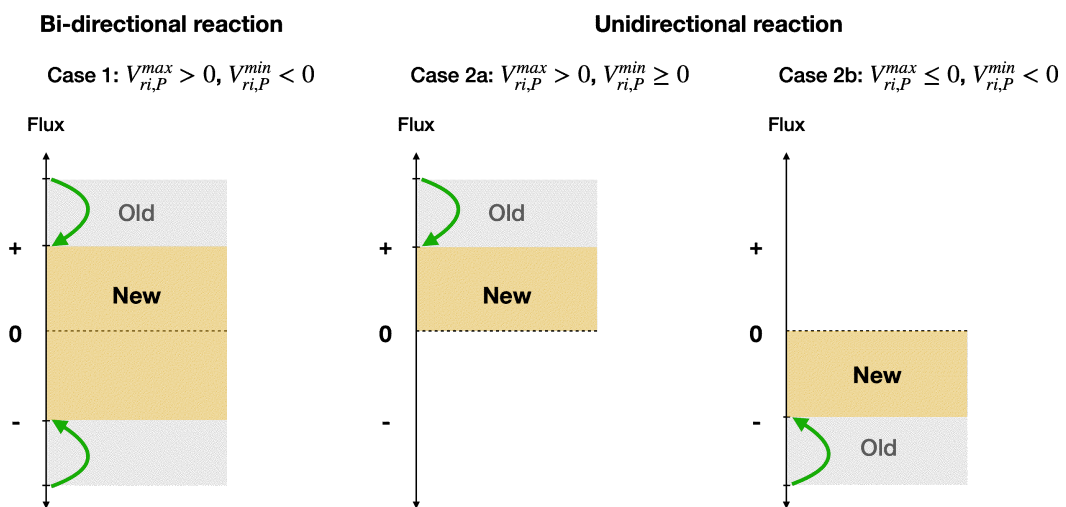


**Figure 3.2. The graphical presentation of how IFFPR modifies a solution space regarding reaction directions.** In the case of bi-directional reactions

(case 1), both equations (3.7) and (3.8) are applied to restrict the solution space. However, in the case of unidirectional reactions (case 2a and 2b), only equation (3.7) or (3.8) is applied to case 2a or 2b, respectively.

f)   $V_{ri,C}^{max}$ and $V_{ri,C}^{min}$ are applied as $V_{ri}^{max}$ and $V_{ri}^{min}$ in formula (3.1) mapping gene expression values to flux.

g)   After optimizing the objective function in formula (3.4), the total model mapping discrepancies are compared between the previous $d_P$ and the new iteration $d_C$. If the total model mapping discrepancy from the previous run is greater or equal ($d_P \geq d_C$), the algorithm proceeds with the next iteration and proceeds with step e).

h)   The inner iterative process terminates for reaction $ri$. The next reaction in the list is selected by setting $ri = ri + 1$, and the algorithm proceeds with step d).

i)   The algorithm terminates if the total model mapping discrepancy becomes stable, or all reactions are processed.

In this study, the process was terminated before the algorithm reached the end of the list. As the algorithm processed around 80% of the reactions in the list, the total model mapping discrepancy became stable (see Chapter 4, Results). This was explainable as the rest of the reactions (~ 20%) already showed narrow flux ranges (Table 4.1). Reducing the flux ranges for these reactions could not influence the total model mapping discrepancy any further but only cost more computational time (see Chapter 4, Results). Hence, the algorithm was stopped, and the sets of $V_{ri,C}^{max}$ and $V_{ri,C}^{min}$ were obtained.

## 3.5    Reducing the number of thermodynamically infeasible loops

As mentioned in the previous chapter (see Chapter 1, Introduction), thermodynamically infeasible loops can cause unrealistic fluxes predicted from constraint-based modeling methods (e.g., FBA, FVA). Many attempts have been made to solve the problem. By far, the most successful method which is considered as a gold standard method is ll-COBRA. The method successfully reduces computational complexity by introducing reaction directionality as a binary variable into the optimization function and generates only a MILP level problem to inhibit TIL at the genome scale.  Although ll-COBRA requires less

runtime compared to other methods (Schellenberger et al. 2011), it is still computationally intensive. In order to speed up the process to remove TIL, in this study, a novel iterative procedure was developed to detect and remove TIL called REDucing the number of Thermodynamically Infeasible Loops (RED-TIL). After obtaining flux prediction results from the mapping procedure (see section 3.4 Formulating the optimization criterion), the results were used as an input for a MILP problem to identify TILs and to exclude them.

External reactions are not regarded. Applying a maximal flux value threshold (threshold = 0.01) for TIL to be allowed, the set of reactions $supp(v)$ known as the support of $v$ is assembled. $supp(v)$ contains a subset of the internal reactions ($v \geq 0.01$). The value of 0.01 was applied as a trade-off between CPU time and reasonable results. Next, an optimization problem is put up to determine the length of a minimum-containing TIL in the solution by

minimize $\qquad \sum_r \lambda_r$ $\hfill$ (3.9)

subject to

$$\sum_r S_r \cdot \lambda_r = 0 \qquad (3.10)$$

$$\lambda_r \geq inFC_r \qquad (3.11)$$

$$\sum_r inFC_r \geq 2 \qquad (3.12)$$

$$inFC_r \in \{0, 1\} \qquad (3.13)$$

where $\lambda_r$ is the flux of reaction $r$ ($\forall r \in supp(v)$), $S_r$ is a stoichiometric matrix of the metabolic network with metabolites and reactions, $inFC_r$ is a binary variable which equals to 1 for a reaction that is involved in the potential TIL. In a system that contains a TIL, there must be at least two reactions involved enforced by equation (3.12). If a solution of the problem put up by equation (3.9) – (3.13) is found, a TIL (of length k) is detected. A constraint is added not allowing this TIL by

$$\sum_{i=1}^{k} inFC_{r_i} \leq k - 1 \qquad (3.14)$$

Equation (3.14) forces the algorithm to search for a solution that puts at least one of these variables $inFC_{r_1}, inFC_{r_2}, \ldots, inFC_{r_k}$ to 0 which leads to the TIL to be discarded from the solution. In the next optimization iteration, the mapping procedure is re-optimized using equation (3.1) to (3.5) together with the newly added constraint from equation (3.14), followed by finding new TIL employing

the MIP problem described by equation (3.9) – (3.13). The algorithm stops when no TIL above the threshold can be found.

## 3.6    Workflow for validating the model

The overview of the entire process is illustrated in Figure 3.3. The process started by allowing the model to learn based on gene expression profiles as explained above. The model was trained iteratively to reduce the total model mapping discrepancy and improve the model fitting. The total model mapping discrepancy was used to measure how well the estimated fluxes fitted to gene expression data. At the end of the model training, the best parameter setting was obtained at the lowest possible total model mapping discrepancy (total model mapping discrepancy becomes stable or the end of the IFFPR list is reached) (see section 3.4.1 Reducing the search space employing Iterative Feasible Flux Space Reduction). Then, the primary carbon source for each of the eight carbon source conditions was predicted by selecting the transporter (one out of eight potential transporters) with the highest flux in the corresponding condition. The prediction was validated by comparing it to the known carbon source of the according condition. Additionally, the flux predictions of the 40 core reactions were compared with the fluxes of the original publication (Chubukov et al. 2013) derived by $^{13}C$ tracer analysis, and the similarity was quantified by calculating Pearson's correlation of the predicted fluxes and $^{13}C$ derived fluxes of the core reactions. The Benjamini-Hochberg method was used to correct for multiple testing across all reactions (Benjamini and Hochberg 1995). Furthermore, the model was applied to an unknown dataset, i.e., data from the time series of *B. subtilis* grown on glucose spiked with malate and malate spiked with glucose as described above.
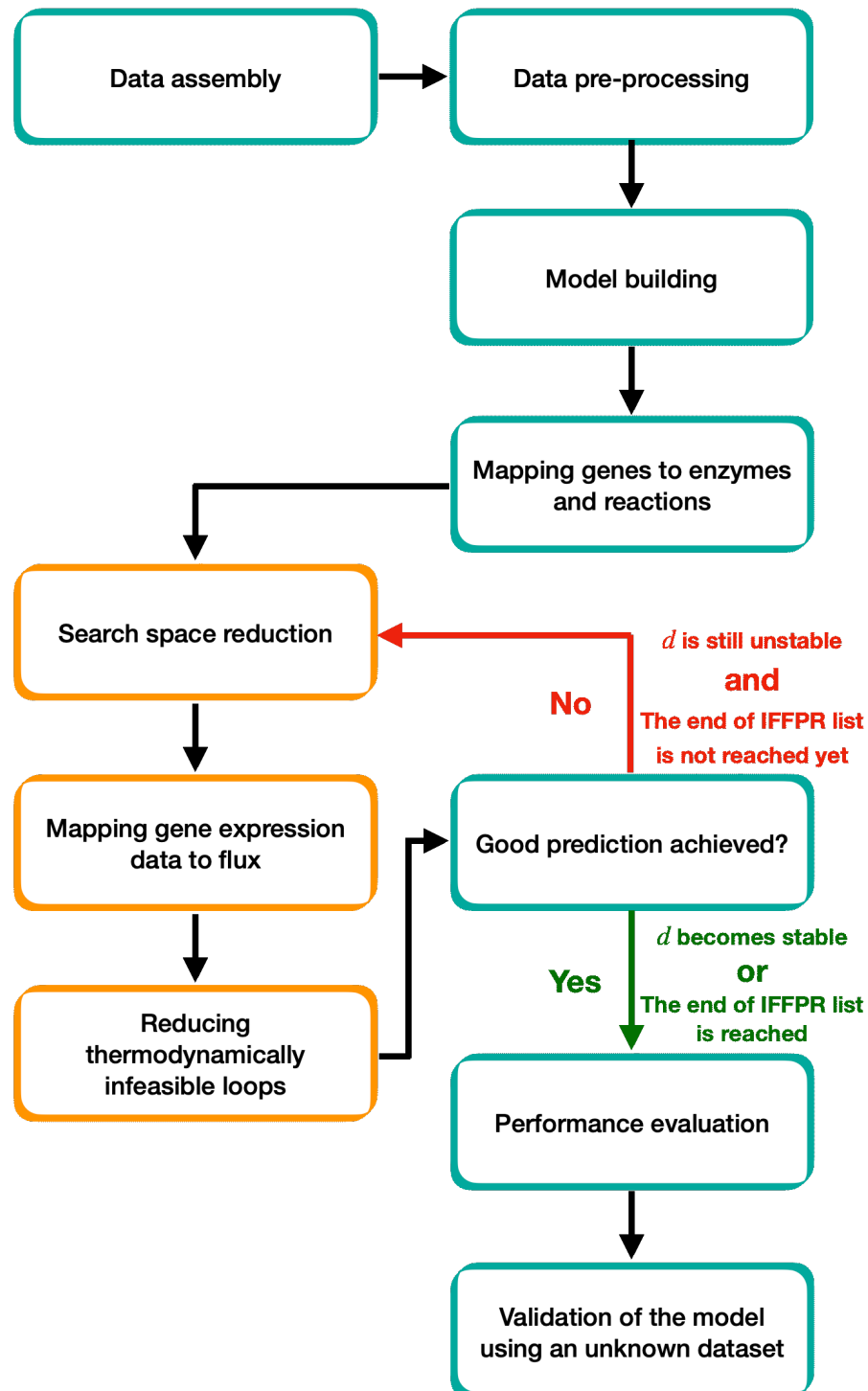
**Figure 3.3. The workflow of LPM-GEM.** The boxes in orange color are all the processes involving in model training. $d$: total model mapping discrepancy. (details, see main text).

## 3.7    Implementation of the integrative Metabolic Analysis Tool

To benchmark LPM-GEM, its prediction performance was compared with one of the well-known existing tools for generating context-specific metabolic

models, i.e., the integrative Metabolic Analysis Tool (iMAT) (Shlomi et al. 2008, Zur et al. 2010). As iMAT is implemented within the Cobra toolbox (Heirendt et al. 2019), both tutorials, i.e., from the toolbox (Heirendt et al. 2019) and from the iMAT protocol (Shlomi et al. 2008, Zur et al. 2010), was followed to ensure following the correct implementation of the method. The same metabolic model (as described above), gene expression data for the eight different carbon source conditions (Nicolas et al. 2012), and GPR mapping as in LPM-GEM were employed. Minimal lower bounds for all eight carbon source exchange reactions were also taken from Chubukov *et al.* (Chubukov et al. 2013). Upper (a cutoff for non-zero flux reactions) and lower thresholds (a cutoff for zero flux reactions) were set equal to +/- 0.3 SD from gene expression values following the suggested iMAT discretization process (Stempler et al. 2014, Shlomi et al. 2008, Zur et al. 2010).

## 3.8    Implementation environment

All analyses were performed using R version 3.3.3 (www.r-project.org). The Cobra toolbox version 3.0 and Matlab version R2019a (www.mathworks.com) were used to obtain the initial stoichiometric matrix, lower and upper bounds, reversibility information, and metabolite and reaction names from the initial metabolic network and prediction results from iMAT models for benchmarking. All further analysis was performed using R. The Gurobi optimizer version 9.0.2 (www.gurobi.com) was used to solve mixed-integer linear programming problems.

# Chapter 4

## Results

### 4.1 Employing three strategies to construct thermodynamically feasible context-specific models

In order to efficiently integrate gene expression profiles into the metabolic models and obtain realistic flux distribution, LPM-GEM was developed based on solutions that are thermodynamically feasible. Three different strategies (i.e., IFFPR, RED-TIL, and adding a penalty for the sum of fluxes) were employed to reduce TIL. All strategies were able to improve the constructed models as described in the following.

#### 4.1.1 Iterative Feasible Flux Space Reduction

Under various environments such as temperature, pH, or different carbon sources, a cell or an organism can adapt itself to changes and achieve different steady-state conditions of fluxes. A feasible flux range for each reaction at the specific steady-state condition can differ from its original upper and lower bounds in each particular condition. This feasible flux range generally represents the actual minimal and maximal flux boundaries better than the original upper and lower bounds from the metabolic network. To assess flux characteristics under different circumstances, FVA is typically used to find the allowable flux range for each reaction in the metabolic network (see Chapter 1, Introduction).

In this thesis, FVA was first performed to reduce the maximal and minimal flux boundaries for all fitted reactions (core and associated reactions) of the model. However, it was observed that the resulting flux ranges did not substantially differ from the original upper and lower bounds for many reactions. These high flux ranges were expected not to reflect realistic situations and may facilitate using TILs by the optimization procedure when fitting the model to the transcription profiles. Hence, IFFPR was developed to iteratively reduce the flux boundaries from FVA by comparing the discrepancy between the fluxes derived from the expression values and the optimal feasible flux when obtaining newly

adjusted maximal ($V_{ri}^{max}$) and minimal ($V_{ri}^{min}$) possible fluxes for every fitted reaction. This discrepancy was used as an indicator to measure how well the predicted fluxes resemble gene expression values, and it is denoted as the "total model mapping discrepancy" in the following (see section 3.4 Formulating the optimization criterion). Employing IFFPR led to new, considerably reduced flux bounds (Table 4.1).

**Table 4.1.** Initial and final bounds after implementing IFFPR.

| Reaction | Minimum flux (Before) | Maximum flux (Before) | Minimum flux (After) | Maximum flux (After) |
|---|---|---|---|---|
| 2S6HCCi | 4.52E-05 | 0.00028 | 4.52E-05 | 0.00028 |
| ACKr | -210 | 210 | -0.411 | 0.41 |
| ACONT | -6.02 | 45.2 | -3.01 | 22.6 |
| ACOTA | -136 | 0 | -0.0665 | 0 |
| AKGDH | 0 | 44.6 | 0 | 22.3 |
| AKGt2r | 0 | 0.688 | 0 | 0.688 |
| ALCD19y | -6.22E+00 | 5.72 | -6.22E+00 | 5.72 |
| ALDD31_1 | 0 | 0.688 | 0 | 0.688 |
| ARGSL | -0.655 | 42.1 | -0.00128 | 0.0822 |
| ARGabc | 0 | 0.688 | 0 | 0.688 |
| ARGt2r | 0 | 0.688 | 0 | 0.688 |
| ASPO1 | 0 | 119 | 0 | 29.7 |
| ASPT | 0 | 185 | 0 | 0.00565 |
| ASPTA | -186 | 0.362 | -46.6 | 0.0905 |
| CDPDSP_BS | 0.00952 | 0.059 | 0.00952 | 0.059 |
| CITt10 | 0 | 0.688 | 0 | 0.688 |
| CITt14 | 0 | 840 | 0 | 0.000801 |
| CITt15 | 0 | 840 | 0 | 0.103 |
| CITt2r | -840 | 0.688 | -0.103 | 8.40E-05 |
| CLPNS2_BS | 8.50E-05 | 0.000527 | 8.50E-05 | 0.000527 |
| CS | 0 | 45.2 | 0 | 22.6 |
| CYSS_2 | 0 | 78.9 | 0 | 0.0385 |
| CYSTGL_1 | 0 | 25.1 | 0 | 0.0122 |
| CYSTS_2 | 0 | 118 | 0 | 0.0018 |
| ENO | -116 | 28.3 | -29.1 | 7.08 |
| FRUK | 0 | 6.22 | 0 | 6.22 |
| FBA | -23.4 | 9.13 | -11.7 | 4.57 |
| FBA2 | -6.22 | 5.72 | -6.22 | 5.72 |
| FBP | 0 | 23.4 | 0.00E+00 | 0.0228 |
| FEDCabc | 0 | 0.00364 | 0 | 0.00364 |
| FRUpts | 0 | 5.72 | 0 | 5.72 |
| FUM | -15.4 | 209 | -1.93 | 26.1 |
| FUMt2r | -17.1 | 0.688 | -0.134 | 0.00537 |
| G6PDH2r | 0 | 72.1 | 0 | 0.0352 |

| Reaction | Minimum flux (Before) | Maximum flux (Before) | Minimum flux (After) | Maximum flux (After) |
|---|---|---|---|---|
| GAPD | 0 | 17.7 | 0 | 17.7 |
| GAPDi_nadp | 0 | 30.6 | 0 | 0.0299 |
| GHMT2r | 6.24E-05 | 48.4 | 1.22E-07 | 0.0945 |
| GLCNt2ir | 0 | 5.13 | 0 | 5.13 |
| GLCpts | 0 | 9.01 | 0 | 9.01 |
| GLUDxi | 0 | 133 | 0 | 0.0081 |
| GLUSy | 0 | 136 | 0 | 0.00104 |
| GLUt2r | 0 | 3.59 | 0 | 3.59 |
| GLYCt | -2.93 | 6.22 | -2.93 | 6.22 |
| GLYK | 0 | 6.22 | 0 | 6.22 |
| GLYO1 | 0 | 49.7 | 0 | 3.11 |
| GNKr | 0 | 5.13 | 0 | 5.13 |
| HISTD | 0 | 17.3 | 0 | 1.65E-05 |
| HISt2r | 0 | 0.688 | 0 | 0.688 |
| HSTPTr | 0 | 17.3 | 0 | 0.0169 |
| ICDHyr | -6.02 | 45.2 | -3.01 | 22.6 |
| ICITt10 | 0 | 0.688 | 0 | 0.688 |
| ICITt2 | -11.3 | 0.688 | -5.63 | 0.344 |
| LCADi | 0 | 0.688 | 0 | 0.688 |
| LDH_L | -17.5 | 0.688 | -4.37 | 0.172 |
| L_LACt2r | -17.5 | 27.2 | -4.37 | 6.8 |
| MALt10 | 0 | 26.5 | 0 | 1.66 |
| MALt2r | 0 | 26.5 | 0 | 26.5 |
| MALt4 | 0 | 26.5 | 0 | 0.0259 |
| MDH | -185 | 139 | -23.1 | 17.3 |
| ME2 | 0 | 235 | 0 | 0.000448 |
| MCITL2 | 0 | 24.9 | 0 | 0.0121 |
| OXGDC | 0 | 44.5 | 0 | 0.0217 |
| PC | 0 | 210 | 0 | 0.0032 |
| PDH | 0 | 24.9 | 0 | 24.9 |
| PFK | 0 | 9.13 | 0 | 9.13 |
| PGCD | -0.555 | 119 | -0.00108 | 0.232 |
| GND | -6.16 | 72.1 | -0.385 | 4.5 |
| PGI | -72.2 | 9.55 | -9.03 | 1.19 |
| PGK_1 | -30.6 | 17.7 | -30.6 | 17.7 |
| PGM_1 | -116 | 28.3 | -29.1 | 7.08 |
| PHETA1 | -3.97 | 0 | -3.97 | 0 |
| PPCK | 0 | 235 | 0 | 3.67 |
| PPS | 0 | 210 | 0 | 0.0032 |
| PRAGSr | 0 | 18.8 | 0 | 0.00229 |
| PTAr | -210 | 210 | -0.411 | 0.41 |
| PYK | 0 | 236 | 0 | 14.8 |
| PYRt2 | -20.8 | 8.26 | -10.4 | 4.13 |
| RPE | -25.1 | 47.8 | -0.784 | 1.49 |
| RPI | -25.1 | 1.85 | -1.57 | 0.116 |
| SERAT | 0 | 78.9 | 0 | 0.0385 |

| Reaction | Minimum flux (Before) | Maximum flux (Before) | Minimum flux (After) | Maximum flux (After) |
|---|---|---|---|---|
| SERD_L | 0 | 118 | 0 | 0.000902 |
| SHSL1_1 | 0 | 25.1 | 0 | 0.00614 |
| SHSL2 | 0 | 25.1 | 0 | 0.0982 |
| SHSL4r | 0 | 25.1 | 0 | 0.0245 |
| SUCCt2r | -13.9 | 3.35 | -13.9 | 3.35 |
| SUCD1 | -0.688 | 48 | -0.344 | 24 |
| SUCOAS | -187 | 44.5 | -23.4 | 5.57 |
| TALA | -12.5 | 24 | -0.389 | 0.75 |
| THRD | 0 | 43.9 | 0 | 11 |
| THRD_L | 0 | 25.1 | 0 | 0.098 |
| THRS | 0 | 44.2 | 0 | 5.53 |
| TKT1 | -12.5 | 24 | -0.778 | 1.5 |
| TKT2 | -12.6 | 23.8 | -0.791 | 1.49 |
| TPI | -23.7 | 12.7 | -11.9 | 6.34 |
| TRPAS1 | 0 | 78.9 | 0 | 0.077 |
| TRPS1 | 0 | 0.0573 | 0 | 0.0573 |
| TYRTA | -3.94 | 0 | -3.94 | 0 |
| UNK5 | 0 | 1.38 | 0 | 1.38 |

*For full reaction name, see the publication of *B. subtilis* 168 (iYO844) model (King et al. 2016).

From all 98 fitted reactions, there were 50 reactions that either minimal or maximal bound (unidirectional reaction) or both bounds (bidirectional reaction) were reduced by at least 80% from the original value (Table 4.1). The reduced bounds greatly reduced the search space. The result of the search space reduction was observed by the declining total model mapping discrepancy. Notably, the total model mapping discrepancy decreased considerably by 95.65% from the original discrepancy (Figure 4.1). The lower total model mapping discrepancy indicated that the fitting of gene expression profiles to the metabolic model was improved after the solution space was forced to be more restricted by the newly adjusted bounds. All flux predictions with and without employing IFFPR are listed in Table A3 and Table A4. In summary, IFFPR improved the experimental data integration more efficiently, observed by a decreased total model mapping discrepancy.
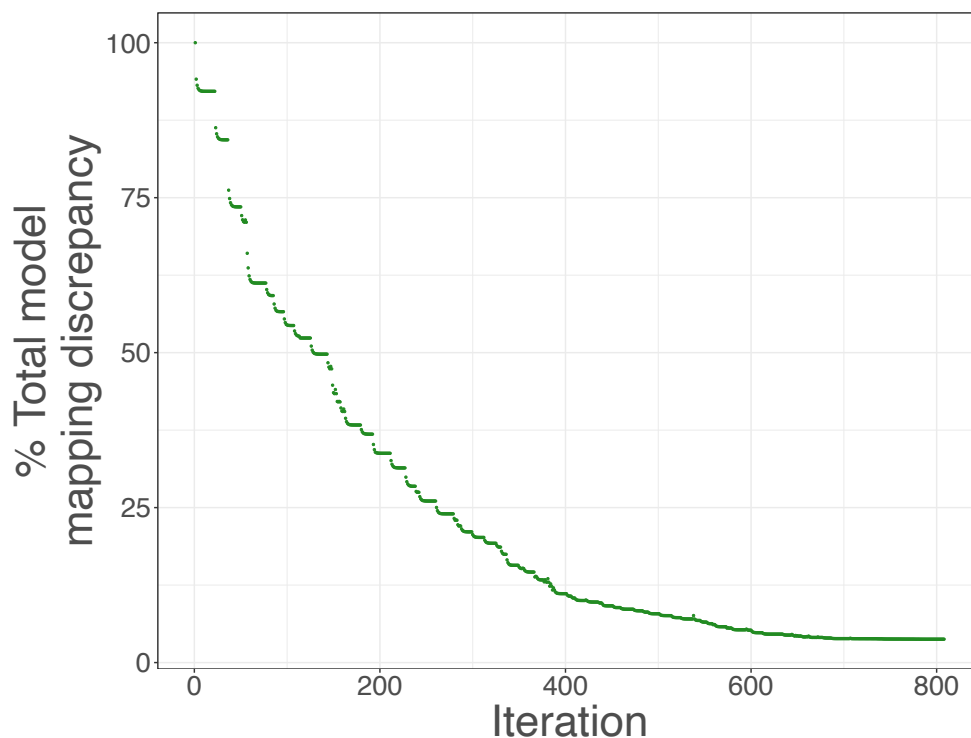
**Figure 4.1. The total model mapping discrepancy calculated over all eight conditions at $\alpha$ = 0.01 is shown with respect to the number of iterations of IFFPR algorithm.** As the algorithm proceeds near the end of the list of reactions (>600 iterations), the total model mapping discrepancy does not further decrease.

### 4.1.2  REDucing the number of Thermodynamically Infeasible Loops

ll-COBRA is a well-established and efficient method to remove TILs in constraint-based modeling. The method generates one large MILP problem finding an optimal solution while enforcing fluxes from internal reactions participating in all detected cycles to be zero (Schellenberger et al. 2011). Although the approach is very powerful, it is computationally demanding. Hence, RED-TIL was developed based on MILP to solve the same problem while improving the computation speed. The novel iterative approach required considerably less running time. To provide the thermodynamically feasible flux distribution, RED-TIL solves an FBA problem, identifies and removes TIL in the solution space iteratively by a bottom-up approach. The process is repeated until no TIL above a certain threshold (threshold = 0.01) is detected (see Chapter 3, 3.5 Reducing the number of thermodynamically infeasible loops).

In order to compare these approaches, both methods were implemented using the same R programming environment and the same numerical solver, yielding very similar solutions (Pearson's correlation coefficient r = 0.96, Figure 4.2). The predicted fluxes from RED-TIL and ll-COBRA are provided in Table A4 and Table A5.
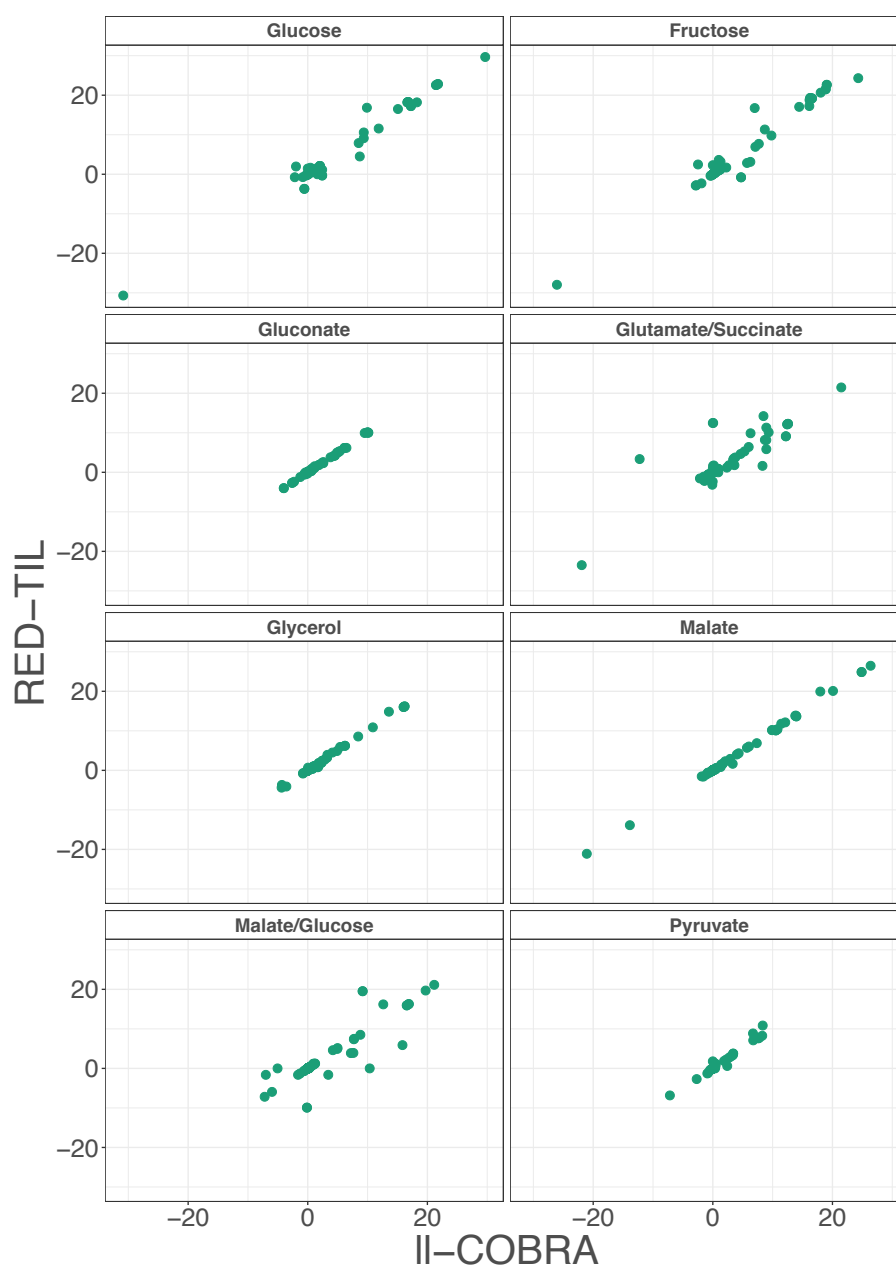


**Figure 4.2. Flux distribution between RED-TIL and ll-COBRA.** Scatterplots display the predicted fluxes of the core and associated reactions (98 reactions) from RED-TIL *versus* the predicted fluxes from ll-COBRA from all eight different carbon source conditions.

Even though similar results from ll-COBRA and RED-TIL were observed, the runtimes from both methods were different. Explicitly, when many iterations were needed, i.e., performing FVA for the whole metabolic network, the benefit of employing RED-TIL was noticeable. In this study, FVA was performed for every reaction in the network (1,250 reactions equal to 2,500 iterations per condition). While ll-COBRA required 42.83 hours to perform this task for all eight conditions, RED-TIL only needed 12.89 hours (Figure 4.3).
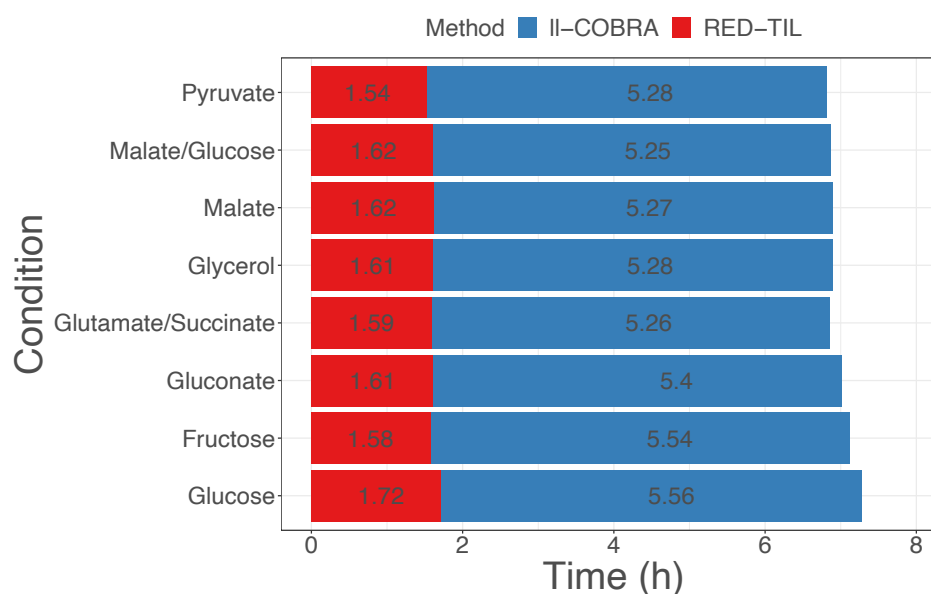


**Figure 4.3. Computational running time (hours) of ll-COBRA and RED-TIL for each carbon source condition after performing FVA.**

In summary, for removing thermodynamically infeasible loops, the new approach RED-TIL led to similar models in much faster running time when compared to a well-established, commonly used method.

### 4.1.3  Adding a penalty for the sum of fluxes

In addition to optimizing the fitting of gene expression to the fluxes of the reactions, minimizing the sum of all predicted fluxes from reactions which were either core reactions or associated reactions was also considered to restrict the optimizer employing thermodynamically infeasible loops (see Chapter 3, 3.4 Formulating the optimization criterion). Gauging between low total model mapping discrepancy and a low total sum of fluxes $v_{ro,c}$ across all conditions

before applying IFFPR and RED-TIL, different strengths of the penalty (ranging from $\alpha$ = 0 (no penalty) to α = 10 (high penalty)) were tested (Figure 4.4). In the end, $\alpha$ = 0.01 was selected as a suitable trade-off parameter for preventing high fluxes. The considerable decrease from the total sum of fluxes (compare to $\alpha$ = 0) was observed while leaving the total model mapping discrepancy moderate.
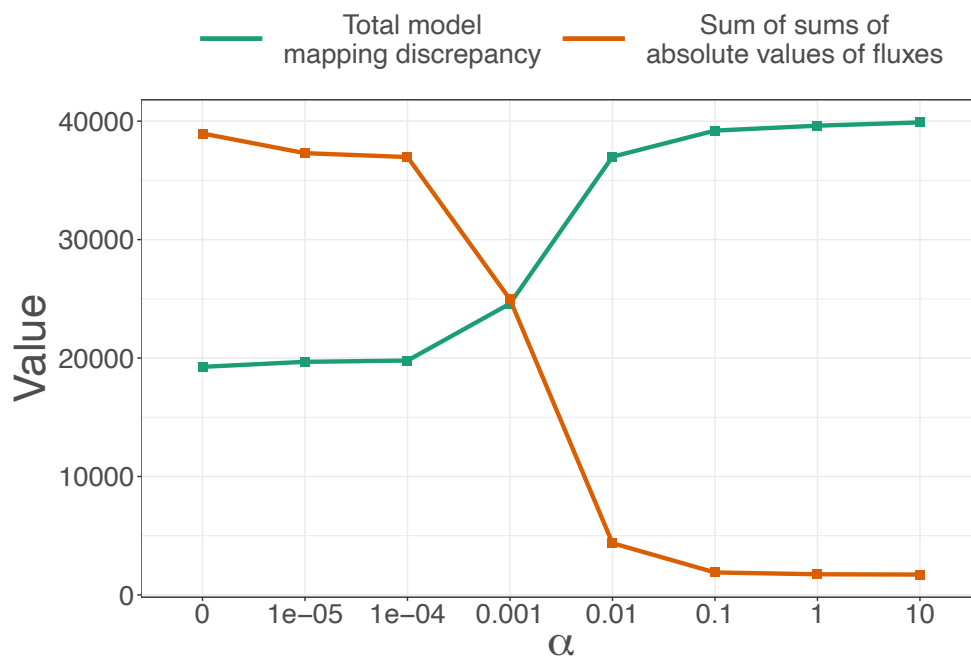


**Figure 4.4. Trade-off between the sum of sums of absolute values of fluxes from non-core and non-associated reactions and the total model mapping discrepancy calculated across all eight conditions at different values of parameter $\alpha$ before applying IFFPR and RED-TIL.** At $\alpha$ = 0.01, the sum of sums of absolute values of fluxes drops considerably while the total model mapping discrepancy only moderately increases.

By employing all three different strategies, i.e., IFFPR, RED-TIL, and adding a penalty for the sum of fluxes, a novel mechanistic approach to extract context-specific metabolic models from gene expression profiles was established to provide thermodynamically feasible flux estimations, and we investigated this implementation described in the following sections.

## 4.2    LPM-GEM identifies the correct carbon sources

LPM-GEM mapped gene expression levels onto the reactions by a regression approach (see Chapter 3, 3.4 Formulating the optimization criterion) and

generated context-specific metabolic models employing RED-TIL and IFFPR. This was done across each of the eight carbon sources (glucose, fructose, gluconate, glutamate, succinate, glycerol, malate, pyruvate). Besides the core reactions, for which $^{13}C$ metabolic flux data was available, "associated reactions" were considered for fitting the solution to the transcription profiles of the corresponding coding genes. "Associated reactions" were neighbors of the core reactions and were assumed to be important for the carbon source prediction (see Chapter 3, 3.3.2 Defining the set of reactions for the optimization criterion). A major objective of the thesis was to construct models enabling to predict the carbon source. For this, after optimization, the prediction results of the eight transporter reactions (glucose, fructose, gluconate, glutamate, succinate, glycerol, malate, pyruvate) were assessed to identify the major carbon source for each condition. The z-scores of the corresponding carbon source transporters in each condition were compared to predict the primary (highest z-score) and secondary (second highest z-score) carbon sources (Figure 4.5 a and b).

a

**b**



Condition

|  | Glucose | Fructose | Gluconate | Glutamate/<br>Succinate | Glycerol | Malate | Malate/<br>Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| Glucose | 1 | 2 |  |  |  | 2 |  |  |
| Fructose | 2 | 1 |  |  |  |  |  |  |
| Gluconate |  |  | 1 |  |  |  |  |  |
| Glutamate |  |  |  | 1 |  |  |  | 2 |
| Succinate |  |  | 2 | 2 | 2 |  |  |  |
| Glycerol |  |  |  |  | 1 |  |  |  |
| Malate |  |  |  |  |  | 1 | 1 |  |
| Pyruvate |  |  |  |  |  |  | 2 | 1 |

(Prediction — vertical axis label)
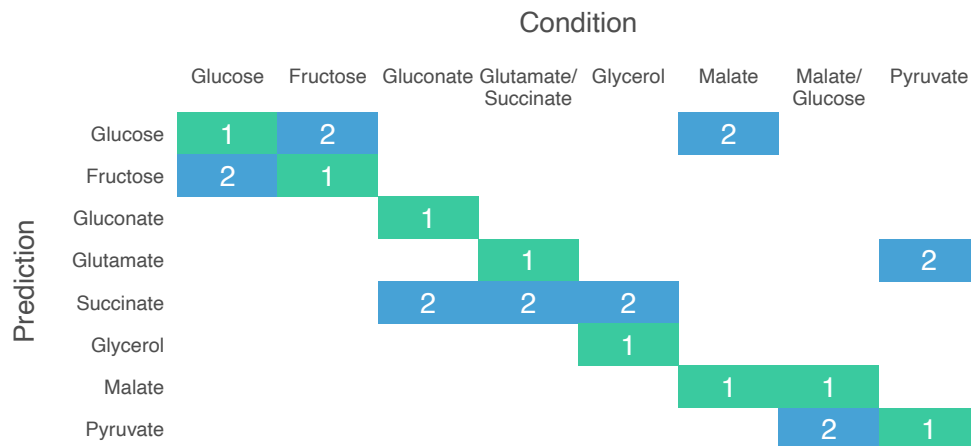
**Figure 4.5. a** Z-scores of the predictions of the carbon source transporters of the eight carbon sources study. A higher z-score indicates a higher probability for a specific carbon source; **b** Carbon source predictions of the eight-carbon-sources study, 1: first prediction, 2: second prediction.

For carbon source conditions with only one carbon source, all predictions were correct (n = 6, glucose, fructose, gluconate, glycerol, malate, pyruvate). For the two carbon sources (glutamate/succinate, malate/glucose), the primary carbon source was also correctly predicted. The secondary carbon source (succinate) was predicted correctly for the carbon sources glutamate and succinate. However, for the carbon sources malate and glucose, pyruvate was predicted as a second carbon source instead of glucose. Notably, the z-score of the glucose transporter was only slightly below the z-score of the pyruvate transporter (Figure 4.5 a).

Overall, the method could well predict the carbon sources based on gene expression profiles of the respective conditions.

## 4.3    Benchmarking result between LPM-GEM and iMAT

For benchmarking, LPM-GEM was compared with the well-known method iMAT (Shlomi et al. 2008, Zur et al. 2010). The iMAT models were constructed using the same metabolic network of *B. subtilis,* the same gene expression profiles from the eight carbon source conditions (glucose, fructose, gluconate, glutamate/succinate, glycerol, malate, malate/glucose, pyruvate), and the same set of GPR mapping as LPM-GEM. The iMAT standard protocol was followed

(Shlomi et al. 2008, Stempler et al. 2014, Zur et al. 2010) (see Chapter 3, 3.7 Implementation of the Integrative Metabolic Analysis Tool). The flux predictions from both methods were compared to the gold standard ($^{13}$C tracer derived fluxes) for all 40 reactions (Chubukov et al. 2013) for which $^{13}$C metabolic flux data was available. All flux predictions are listed in Table A4 and Table A6. Pearson's correlation coefficients for these reactions between the predictions and the gold standard are shown in Figure 4.6 (and listed in Table A7 and Table A8).
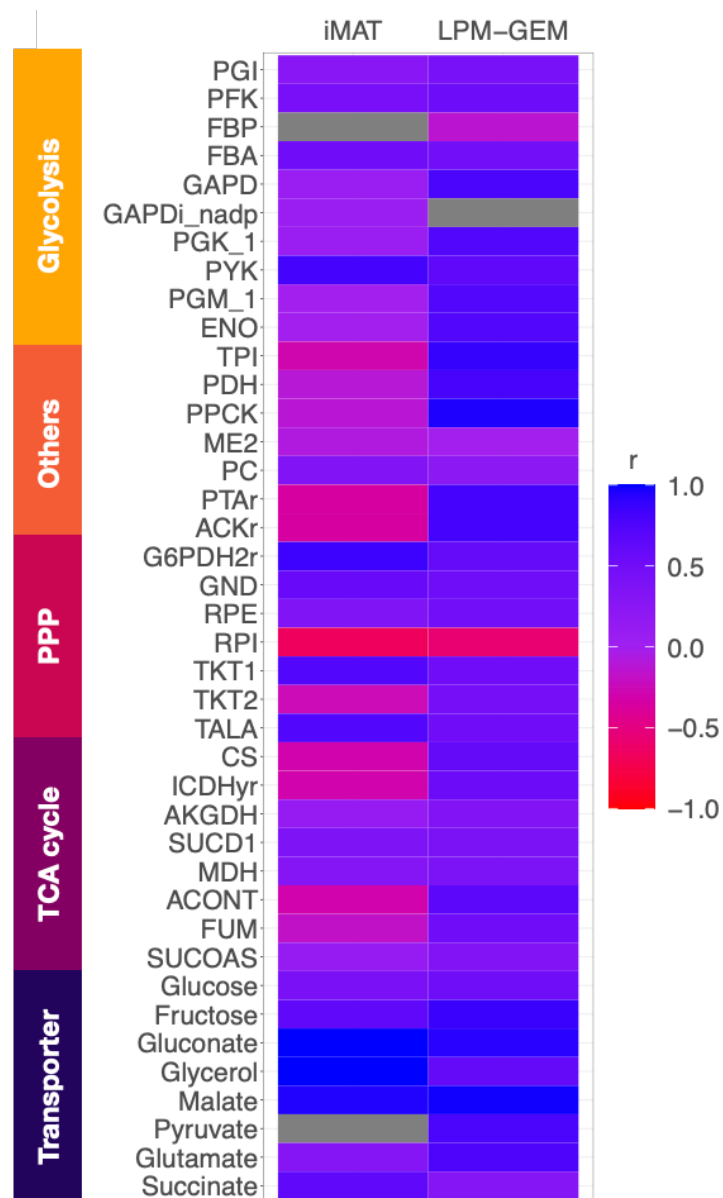


**Figure 4.6. Prediction performance of LPM-GEM and iMAT.** For all 40 core reactions for which gold standard data (from $^{13}$C tracer analysis) was available, Pearson's correlation coefficients between the predicted fluxes and the fluxes

from the gold standard are shown (grey: fluxes are predicted to be zero in every condition); PPP: pentose phosphate pathway, TCA: tricarboxylic acid.

Overall, the number of reactions that showed a good positive correlation when compared flux predictions to $^{13}$C metabolic flux data from LPM-GEM was higher than iMAT, showing the better performance of the model. For most reactions in substrate uptakes, glycolysis, TCA cycle, and others, it was clear that LPM-GEM performed better. However, in PPP, iMAT gave moderately better flux prediction (Figure 4.6). On average, LPM-GEM outperformed iMAT (averaged Pearson's correlation coefficient, LPM-GEM: r = 0.55, iMAT: r = 0.22).

Next, the carbon source predictions were compared between LPM-GEM and iMAT. Although iMAT correctly predicted all two conditions of two-carbon-source combinations (LPM-GEM: one condition was correct, while the other condition, only the primary source was correctly predicted), it correctly predicted only three out of six single carbon sources (LPM-GEM: all six out of six were correctly predicted) (Figure 4.7 a and b). Interestingly, iMAT completely failed to predict the pyruvate transporter at all as shown in Figure 4.7 a. It was because the flux from the transporter was predicted to be zero in all conditions (Table A6). In summary, LPM-GEM led to better flux predictions when compared to iMAT.

**a**
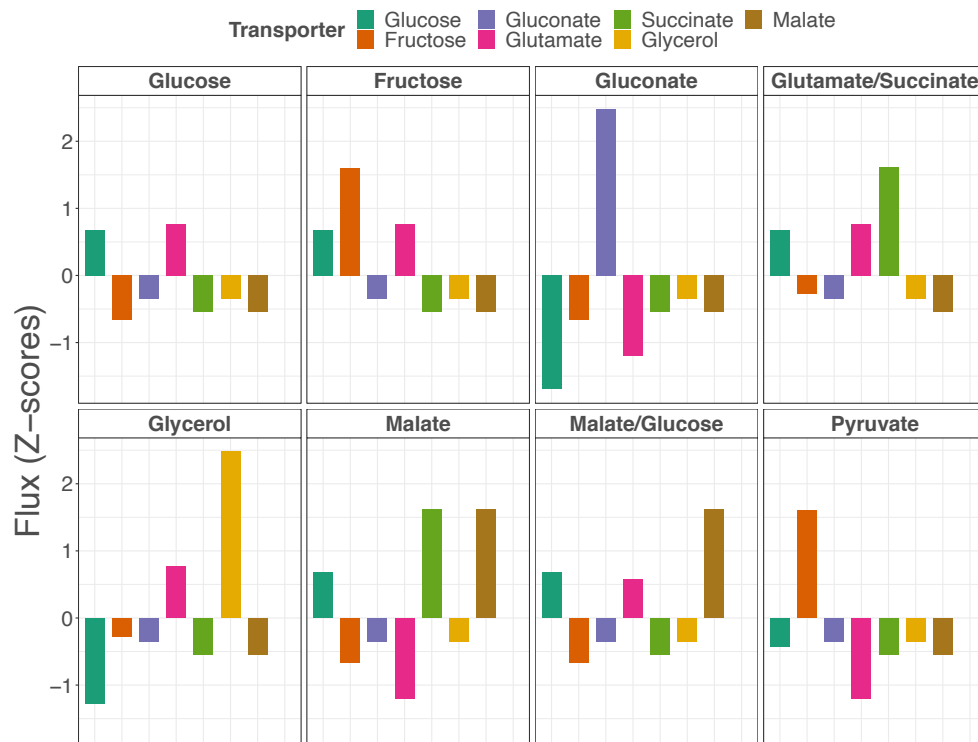


**b**
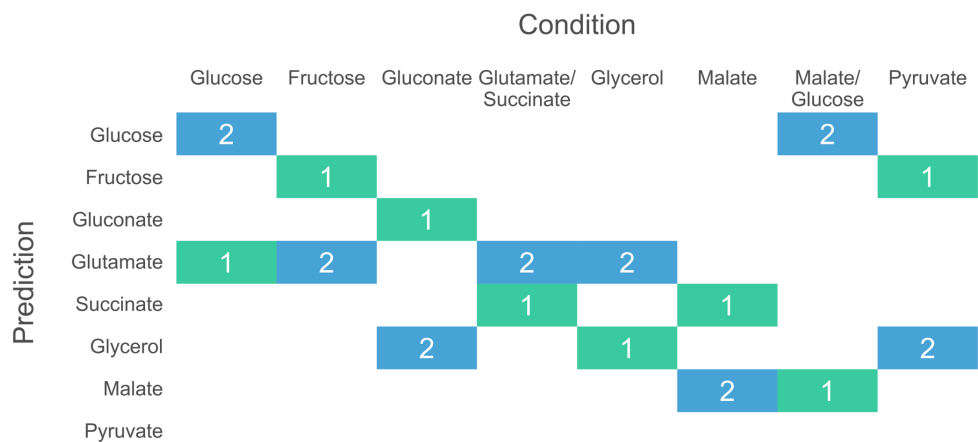


**Figure 4.7. a** Predictions of the carbon source of the first dataset using iMAT. A higher z-score indicates a higher probability for a specific carbon source; **b** Prediction of the carbon source for the eight carbon sources study from iMAT, 1: first prediction, 2: second prediction.

## 4.4    LPM-GEM identifies the carbon sources of the validation set but shows limitations when predicting time-lapse fluxes

So far, LPM-GEM successfully delivered good flux predictions based on transcriptomic data of *B. subtilis* in eight different steady-state conditions.

However, in a natural environment, the bacteria may need to switch from one carbon source to another. Particularly, glucose and malate are preferred carbon sources for which such a switch may occur (Kleijn et al. 2010, Meyer and Stulke 2013). To further test LPM-GEM, the method was applied to a publicly available time-series dataset consisting of two nutritional shifts, i.e., from glucose to glucose plus malate (malate was added as an additional carbon source) and from malate to malate plus glucose (glucose was added as an additional carbon source). In these assays, *B. subtilis* was grown on a single substrate leading to a steady-state-like initial condition. Then, the other substrate was added. Transcription profiles and $^{13}$C flux data were generated in a time series until the shift was performed (at an endpoint at 90 min) and the new steady state was reached, according to the authors of the original study (Buescher et al. 2012). The model which had been trained on the first dataset was applied to the data. The predictions were investigated of the two major carbon sources for the two initial conditions (before adding the other carbon source - only glucose and only malate at steady state) and for the two endpoint conditions (90 min after adding malate to glucose (glucose to glucose plus malate), and 90 min after adding glucose to malate (malate to malate plus glucose)). To predict the carbon sources, the z-scores of the transporters across these steady-state conditions were computed in a similar way to the study with the eight carbon sources (described above). Then, the predicted results were compared with the gold standard. All four out of four steady-state conditions were predicted correctly (Figure 4.8 a and b).

**a**



**b**
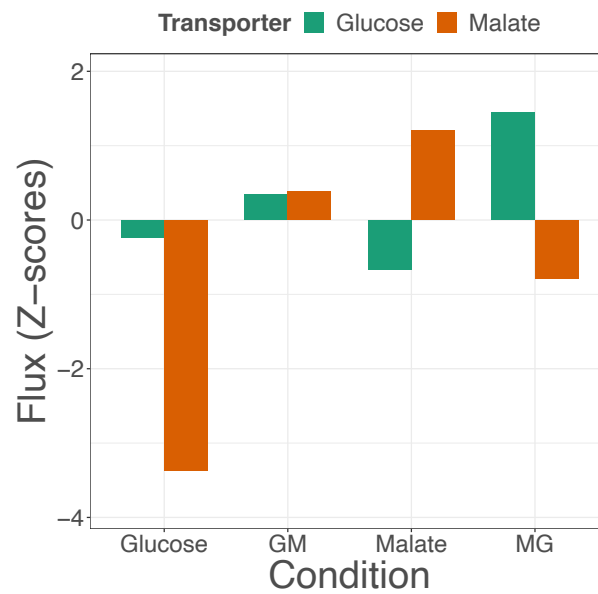


**Figure 4.8. a** Z-scores of the predictions of the carbon source for the nutritional shift. A higher z-score indicates a higher probability for a specific carbon source; **b** Prediction of the nutritional shift for the initial and the endpoint conditions. For **a** and **b**, GM: glucose to glucose plus malate shift, 90 min after adding malate; MG: malate to malate plus glucose shift, 90 min after adding glucose.

Next, to investigate how the LPM-GEM model predicted the time-series of the shifts, the predicted fluxes of the malate and glucose transporters were compared with the fluxes from the gold standard across all time points (Figure 4.9).
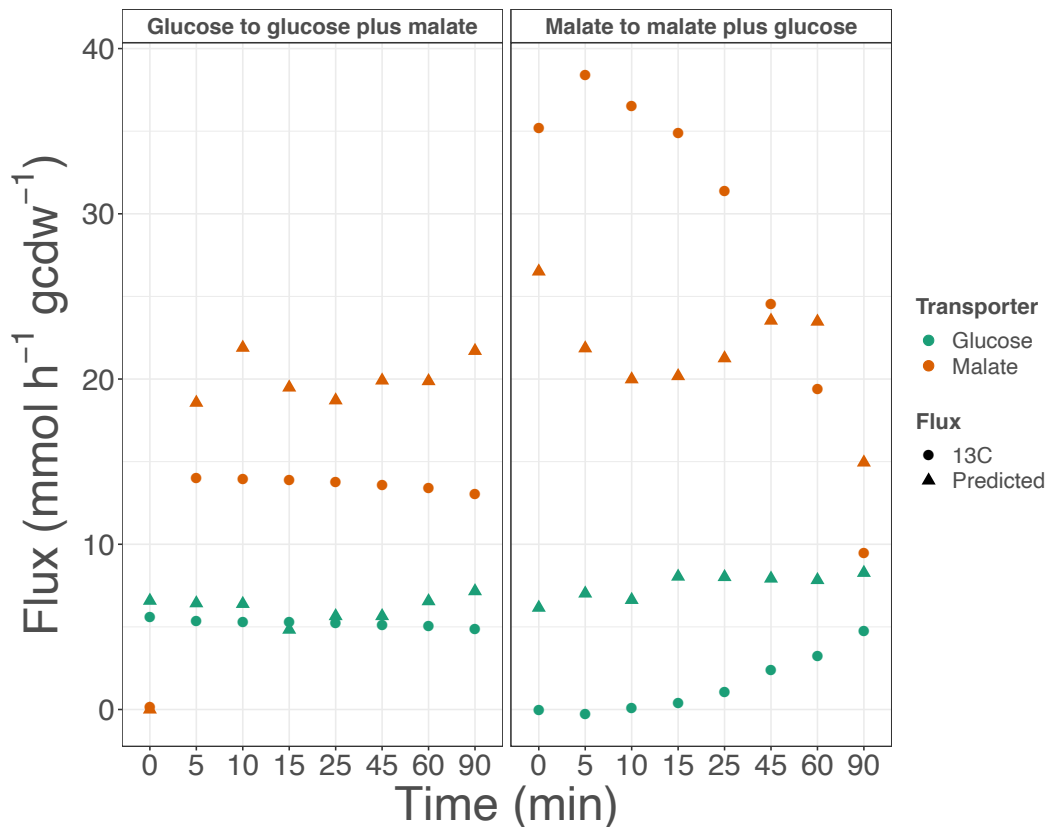
**Figure 4.9. Carbon source shifts between glucose and malate.** After adding the second substrate, the malate transporter quickly adjusts the uptake rate as shown in both shifts. In contrast, the glucose transporter delays its response.

When the carbon source was shifted from malate to malate plus glucose, our predicted fluxes from glucose and malate transporters correlated quite well with the gold standard (r = 0.68 for the glucose transporter, r = 0.48 for the malate transporter) (Table 4.2). Good prediction results for the shift from glucose to glucose plus malate were also expected, as although the order was changed, the conditions were based on the same carbon sources. While the prediction for the malate transporter was very good (r = 0.98 for the shift of glucose to glucose plus malate), the prediction for the glucose transporter was very poor, i.e., the flux prediction was even negatively correlated to the gold standard (r = -0.21). This discrepancy will be discussed later. The flux predictions from both shifts are provided in Table A9 and Table A10.

**Table 4.2.** List of Pearson's correlation coefficients (r) between predicted fluxes from glucose and malate transporters and $^{13}C$ metabolic flux data from the corresponding shifts (glucose to glucose plus malate and malate to malate plus glucose).

| Transporter | Shift | r | P-value (adjusted for multiple testing) |
|---|---|---|---|
| Glucose | Glucose to glucose plus malate | -0.21 | 0.69 |
| Malate | Glucose to glucose plus malate | 0.98 | 5.28E-05 |
| Glucose | Malate to malate plus glucose | 0.68 | 6.35E-02 |
| Malate | Malate to malate plus glucose | 0.48 | 0.16 |

In summary, applying the model trained with the data from the first study (eight carbon sources) to the unknown data from the validation set (glucose/malate carbon source shift), the model correctly identified the carbon sources at baseline and at the endpoints of the carbon source shifts. For the prediction of the time lapse, the model predicted the time-lapse behavior of the main nutrients for the shift from malate to malate plus glucose correctly. However, the model had major difficulties for the shift from glucose to glucose plus malate, and this will be discussed in the next chapter.

# Chapter 5

# Discussion

## 5.1 Technical aspects of LPM-GEM as a novel transcriptomics-based context-specific model extraction method

Most of existing context-specific extraction methods employ binarization to integrate gene expression data into the metabolic models. Generally, the methods binarize transcriptomic data using a threshold or thresholds (upper and lower thresholds in iMAT (Shlomi et al. 2008, Zur et al. 2010)) to define an expression state of a gene being on or off. Then, the algorithms use the binarized data to allow or restrict fluxes in the metabolic network (see details in Chapter 1, 1.4.3.1 Physiochemical constraints alone may fail to reflect the actual flux state). Though the methods successfully improved metabolic flux estimation compared to the normal FBA method, the binarization of transcriptomic data may disregard the fine-grained regulation of metabolism and overlook interesting biological events. Also, finding the optimal value of the thresholds can be troublesome since these values can vary between genes, organisms, conditions, or measurement techniques. To solve the arisen question and find the algorithm to efficiently integrate transcriptomic data and construct context-specific models, LPM-GEM was established based on a linear regression approach to estimate metabolic fluxes in a systems view. By assuming that the metabolic flux and the expression of the gene coding for the responsible enzyme of the corresponding reaction correlate linearly, the method formulated the linear optimization problem and avoided binarization of gene expression data by mapping gene expression values directly to fluxes through the gene expression-based flux constraint, $v_{ri,c}^{fit}$. Doing this, LPM-GEM circumvented the need of setting thresholds to decide whether the reaction of interest would be expressed or non-expressed and let the expression of genes guide the model, which also preserved the continuous flux predictions compared to the method employing binarization shown in Table A4 and A6. Moreover, LPM-GEM does not require a large number of datasets per condition as in PROM (Chandrasekaran and Price 2010). Similar to GIMME (Becker and

Palsson 2008) or iMAT (Shlomi et al. 2008, Zur et al. 2010), it can work well with a single dataset per condition. However, as the expression level is compared relatively, it needs at least two different conditions to linearly map gene expression values to fluxes.

Besides, TIL is a well-known issue in CBM when thermodynamic feasibility is neglected during the optimization process, resulting in unrealistic flux distribution. In order to obtain thermodynamically feasible flux predictions and increase the prediction performance, three different strategies (i.e., IFFPR, RED-TIL, and adding a penalty for the sum of fluxes) were developed to address this issue, and all of them were employed to support LPM-GEM. IFFPR was implemented to reduce the solution space by iteratively reducing the flux ranges of each considered reaction. By reducing the upper and lower bounds, the solution space became more constrained, which automatically forced the optimization algorithm to avoid fitting unreasonably high flux values to gene expression values. It indirectly prevented the algorithm from involving TIL in the system and improved the data integration process to be more efficient (total model mapping discrepancy decreased considerably). ll-COBRA is an existing well-established method to reduce TIL directly. IL-COBRA is very powerful, but also very CPU intensive. To speed up the process of TIL removal, a new method RED-TIL was developed. ll-COBRA formulates one large problem and searches for an optimal solution in a predefined-thermodynamic feasible region. In turn, employing a bottom-up design, RED-TIL splits the overall problem into smaller problems by detecting a TIL in the optimal solution, excluding it from the solution space, and re-optimizing the solution iteratively until no TIL (above a certain flux limit) is detected. While the results were comparable, the computational speed for RED-TIL was considerably faster. On average, RED-TIL removed TIL from the relevant solution space three times faster than ll-COBRA. Such a speed-up was relevant for this work as many computations were needed to generate solutions for a larger range of different parameter settings, particularly for optimizing the upper and lower bounds for each reaction when running IFFPR. Lastly, the third mean to reduce TIL was penalizing the sum of fluxes of non-core and non-associated reactions. Selecting a suitable tradeoff value ($\alpha$ = 0.01 in this thesis), these reactions were limited to reduced fluxes since high flux values can be a source for TIL, while

still addressing the optimization of good flux – gene expression value fits for the core and associated reactions. By this, an FBA model was set up based on a linear fit between the expression of the encoding genes for an enzyme and its predicted flux with considerably reduced TIL.

## 5.2 LPM-GEM shows improved prediction performances when compared to an existing method

LPM-GEM generated the context-specific models to predict the main carbon sources for the model organism *B. subtilis* in eight different conditions. LPM-GEM mostly identified the correct major carbon sources for all eight conditions based on the corresponding gene expression profiles. Seven out of eight conditions were predicted correctly – except malate/glucose condition. Although LPM-GEM could predict the correct carbon source for the primary carbon source, it predicted pyruvate as the secondary carbon source instead of glucose. A hypothesis behind this wrong prediction can be explained in the following. In central energy metabolism, pyruvate and malate are closely related. Kleijn *et al.* observed that when *B. subtilis* was fed with glucose and malate, there was an overflow metabolites secreted from consumed malate and pyruvate was the main overflow product on malate (Kleijn et al. 2010). Although there were only malate and glucose in malate/glucose condition in the eight-carbon-source dataset (Chubukov et al. 2013), it was possible that in this case pyruvate was also generated by the consumed malate, and the bacteria might sense this, which caused coding genes for the pyruvate transporter (lrgA, lrgB) to be expressed. It explained why LPM-GEM predicted pyruvate as the secondary carbon source instead of glucose in malate/glucose condition. Moreover, the flux prediction results from LPM-GEM were also validated with $^{13}$C metabolic flux data. For most reactions in substrate uptakes, glycolysis, TCA cycle, and others (the rest of reactions which could not be put in the specific pathway), the flux prediction results correlated well with $^{13}$C metabolic flux data.

Then, the approach was benchmarked with the well-known existing method, iMAT. Overall, LPM-GEM yielded better flux predictions and, on average, better predictions of the carbon source. A reason for this may be that our method requires no binarization of the transcription profiles. iMAT and other

approaches (Becker and Palsson 2008, Chandrasekaran and Price 2010, Shlomi et al. 2008, Wang et al. 2012, Zur et al. 2010) need expression level thresholds to decide whether a reaction needs to be active (constrained or part of the optimization to have a non-zero flux) or not. The drawback of employing binarization to integrate gene expression profiles was observed in the case of pyruvate transporter. It was clearly seen that iMAT failed to predict the pyruvate transporter having the highest flux at the condition in which pyruvate was the carbon source, as the expression values of corresponding genes of the pyruvate transporter (lrgA, lrgB) in every condition were lower compared to expression values of other genes. Even though the expression values of these two genes in the pyruvate condition were much higher than in any other condition, they were still lower than the lower threshold and were hence regarded as non-expressed by the iMAT algorithm. Thus, the method could not detect this expression and falsely categorized the corresponding reaction as non-expressed, resulting in an estimated zero flux in every condition for the pyruvate transporter (Table A6). In turn, LPM-GEM avoided binarization for data integration and employed the linear regression approach to estimate the metabolic fluxes. LPM-GEM predicted fluxes for the pyruvate transporter in every condition, and the predicted fluxes showed a high positive correlation when validated with $^{13}$C metabolic flux data (r = 0.78) (Table A7). In summary, LPM-GEM yielded improved flux predictions compared to iMAT, a well-established context-specific model extraction method based on transcriptomic data.

## 5.3 LPM-GEM provides good flux predictions for metabolic states mediated by transcriptional regulation but is not sensitive enough to capture fluxes controlled by other mechanisms

Using the model trained with the dataset of the eight carbon sources, LPM-GEM was validated with a second publicly available dataset in which shifts in the carbon source from glucose to glucose plus malate and from malate to malate plus glucose were investigated (Buescher et al. 2012). The model correctly predicted the carbon sources of the initial setting and the endpoints. Besides, the model predictions of the shift from malate to malate plus glucose correlated well with the gold standard ($^{13}$C tracer derived flux from the original study).

Although gene expression data is scalable and easy to obtain compared to $^{13}$C metabolic flux data (Lowe et al. 2017, Uygun et al. 2016, van den Esker and Koets 2019), it provides only indirect information for the estimation of metabolic fluxes. For some conditions or settings, the metabolic flux may not be controlled by transcription of the corresponding enzyme coding genes. Other flux control mechanisms, e.g., related to limitations due to substrate or product concentrations, translational regulation, covalent modification of the enzymes, or allosteric regulation, can influence the metabolic flux (Wegner et al. 2015). This limitation was observed in the nutrient shift from glucose to glucose plus malate. The very poor prediction results suggested that this shift might not be controlled by transcriptional regulation. The observation from the modeling is in line with the observations reported in the original study by Buescher *et al.* (Buescher et al. 2012). They assumed that these shifts are mediated by fundamentally different control mechanisms. In order to confirm their assumption, Buescher *et al.* performed a multi-omics analysis of time-lapse profiles from promoter activity, mRNA, and protein abundance to identify post-transcriptional events (Buescher et al. 2012). After correlating the gene expression levels with the protein levels, they observed high positive correlations in gene-protein pairs related to glycolysis such as phosphoglycerate mutase (r = 0.96), PTS glucose transporter (r = 0.88), and glyceraldehyde 3-phosphate dehydrogenase (r = 0.96) for the shift from malate to malate plus glucose. However, they could not find correlations in gene-protein pairs related to glycolysis in the glucose to glucose plus malate shift. From this, they concluded that the shift from glucose to glucose plus malate was dominantly controlled by post-transcriptional mechanisms (in contrast to the malate to malate plus glucose shift), or proteins for glycolysis are constitutively expressed. The latter is reasonable. As the benefit of glucose consumption is very high compared to malate, it may be beneficial to keep proteins for glycolysis constitutively expressed under the malate condition. This observation serves as a good example of a limitation of LPM-GEM. The method relies on gene expression profiles to predict metabolic fluxes. Hence, it requires a basic understanding of the cellular biology of the regulatory mechanism beforehand to avoid studying mechanisms that are likely to depend on other regulation mechanisms than transcriptional regulation.

# Chapter 6

## Conclusion and Perspectives

This work introduced a novel computational approach to efficiently integrate gene expression profiles into a metabolic network called Linear Programming based Gene Expression Model (LPM-GEM). Unlike many previous methods, LPM-GEM linearly embeds gene expression values into FBA constraints and adjusts flux levels continuously. It avoids defining thresholds and binarization of the expression data, preventing loss of information and improving the data integration process. Also, to reduce solutions containing thermodynamically infeasible fluxes, LPM-GEM was supported by the implementation of three different strategies (i.e., IFFPR, RED-TIL, and adding a penalty for the sum of fluxes). IFFPR and adding a penalty for the sum of fluxes limits the FBA algorithm to use TIL during the optimization process, and RED-TIL is a direct method which detects and discards TIL from the solution space. RED-TIL successfully improved the computational speed compared to ll-COBRA, which is a well-established and commonly used method for removing TIL. The speed-up was crucial for the success of our method, especially as many iterations were required. With all components, LPM-GEM successfully predicted carbon sources for *B. subtilis* from both datasets (the eight-carbon-source dataset and the time-series nutrient shift dataset). By this, LPM-GEP yielded improved prediction performance (validated by $^{13}$C metabolic flux data) compared to the well-known existing method, iMAT. This study supports the idea that FBA analysis based on gene expression profiles can serve as an alternative to $^{13}$C tracer analysis when conducting $^{13}$C labeling experiments can be hard to achieve. For example, it can be used to investigate intracellular pathogens as they can consume or produce the same metabolite as the host. In a publication of Zimmerman *et al.*, iMAT was used to create context-specific models to help identifying the carbon sources of *Mycobacterium tuberculosis* (*M. tuberculosis*) inside the host cell since metabolome analysis alone could not distinguish the carbon sources used by the host or pathogen. Using transcriptomics-based context-specific models, they identified 33 carbon sources of *M. tuberculosis*,

and 31 predicted carbon sources were also detected by metabolomic measurements (Zimmermann et al. 2017). A similar thing can be applied to other pathogens (i.e., *S. aureaus* in osteocytes or *P. falciparum* in erythrocytes) or to track a metabolite in a cell that has different compartments. With the new method developed in this research, LPM-GEM should be assumed to deliver better flux prediction performance if performing the same task. Although gene-expression-based FBA methods cannot provide the same flux resolution as $^{13}$C-MFA, the methods require much less experimental and computational efforts while providing decent flux visibility.

# References

Antoniewicz MR. 2015. Methods and advances in metabolic flux analysis: a mini-review. J Ind Microbiol Biotechnol, 42 (3):317-325.

Antoniewicz MR. 2018. A guide to (13)C metabolic flux analysis for the cancer biologist. Exp Mol Med, 50 (4):19.

Barabasi AL, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. Nat Rev Genet, 5 (2):101-113.

Becker SA, Palsson BO. 2008. Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol, 4 (5):e1000082.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological), 57 (1):289-300.

Bideaux C, Montheard J, Cameleyre X, Molina-Jouve C, Alfenore S. 2016. Metabolic flux analysis model for optimizing xylose conversion into ethanol by the natural C5-fermenting yeast Candida shehatae. Appl Microbiol Biotechnol, 100 (3):1489-1499.

Bordbar A, Monk JM, King ZA, Palsson BO. 2014. Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet, 15 (2):107-120.

Boyd S, Vandenberghe L. 2004. Convex Optimization. Cambridge: Cambridge University Press.

Buescher JM, Antoniewicz MR, Boros LG, Burgess SC, Brunengraber H, Clish CB, DeBerardinis RJ, Feron O, Frezza C, Ghesquiere B, Gottlieb E, Hiller K, Jones RG, Kamphorst JJ, Kibbey RG, Kimmelman AC, Locasale JW, Lunt SY,

Maddocks OD, Malloy C, Metallo CM, Meuillet EJ, Munger J, Noh K, Rabinowitz JD, Ralser M, Sauer U, Stephanopoulos G, St-Pierre J, Tennant DA, Wittmann C, Vander Heiden MG, Vazquez A, Vousden K, Young JD, Zamboni N, Fendt SM. 2015. A roadmap for interpreting (13)C metabolite labeling patterns from cells. Curr Opin Biotechnol, 34:189-201.

Buescher JM, Liebermeister W, Jules M, Uhr M, Muntel J, Botella E, Hessling B, Kleijn RJ, Le Chat L, Lecointe F, Mader U, Nicolas P, Piersma S, Rugheimer F, Becher D, Bessieres P, Bidnenko E, Denham EL, Dervyn E, Devine KM, Doherty G, Drulhe S, Felicori L, Fogg MJ, Goelzer A, Hansen A, Harwood CR, Hecker M, Hubner S, Hultschig C, Jarmer H, Klipp E, Leduc A, Lewis P, Molina F, Noirot P, Peres S, Pigeonneau N, Pohl S, Rasmussen S, Rinn B, Schaffer M, Schnidder J, Schwikowski B, Van Dijl JM, Veiga P, Walsh S, Wilkinson AJ, Stelling J, Aymerich S, Sauer U. 2012. Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism. Science, 335 (6072):1099-1103.

Chandrasekaran S, Price ND. 2010. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. Proc Natl Acad Sci U S A, 107 (41):17845-17850.

Chenard T, Guenard F, Vohl MC, Carpentier A, Tchernof A, Najmanovich RJ. 2017. Remodeling adipose tissue through in silico modulation of fat storage for the prevention of type 2 diabetes. BMC Syst Biol, 11 (1):60.

Chiewchankaset P, Siriwat W, Suksangpanomrung M, Boonseng O, Meechai A, Tanticharoen M, Kalapanulak S, Saithong T. 2019. Understanding carbon utilization routes between high and low starch-producing cultivars of cassava through Flux Balance Analysis. Sci Rep, 9 (1):2964.

Chuang HY, Hofree M, Ideker T. 2010. A decade of systems biology. Annu Rev Cell Dev Biol, 26:721-744.

Chubukov V, Uhr M, Le Chat L, Kleijn RJ, Jules M, Link H, Aymerich S, Stelling J, Sauer U. 2013. Transcriptional regulation is insufficient to explain substrate-induced flux changes in Bacillus subtilis. Mol Syst Biol, 9:709.

Dai Z, Locasale JW. 2017. Understanding metabolism with flux analysis: From theory to application. Metab Eng, 43 (Pt B):94-102.

Dang L, Liu J, Wang C, Liu H, Wen J. 2017. Enhancement of rapamycin production by metabolic engineering in Streptomyces hygroscopicus based on genome-scale metabolic model. J Ind Microbiol Biotechnol, 44 (2):259-270.

Deutscher J. 2008. The mechanisms of carbon catabolite repression in bacteria. Curr Opin Microbiol, 11 (2):87-93.

DeWaal D, Nogueira V, Terry AR, Patra KC, Jeon SM, Guzman G, Au J, Long CP, Antoniewicz MR, Hay N. 2018. Hexokinase-2 depletion inhibits glycolysis and induces oxidative phosphorylation in hepatocellular carcinoma and sensitizes to metformin. Nat Commun, 9 (1):446.

Errington J, Aart LTV. 2020. Microbe Profile: Bacillus subtilis: model organism for cellular development, and industrial workhorse. Microbiology (Reading), 166 (5):425-427.

Gatto F, Miess H, Schulze A, Nielsen J. 2015. Flux balance analysis predicts essential genes in clear cell renal cell carcinoma metabolism. Sci Rep, 5:10738.

Harwood CR, Mouillon J-M, Pohl S, Arnau J. 2018. Secondary metabolite production and the safety of industrially important members of the Bacillus subtilis group. FEMS Microbiology Reviews, 42 (6):721-738.

Hauslein I, Manske C, Goebel W, Eisenreich W, Hilbi H. 2016. Pathway analysis using (13) C-glycerol and other carbon tracers reveals a bipartite metabolism of Legionella pneumophila. Mol Microbiol, 100 (2):229-246.

Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdottir HS, Wachowiak J, Keating SM, Vlasov V, Magnusdottir S, Ng CY, Preciat G, Zagare A, Chan SHJ, Aurich MK, Clancy CM, Modamio J, Sauls JT, Noronha A, Bordbar A, Cousins B, El Assal DC, Valcarcel LV, Apaolaza I, Ghaderi S, Ahookhosh M, Ben Guebila M, Kostromins A, Sompairac N, Le HM, Ma D, Sun Y, Wang L, Yurkovich JT, Oliveira MAP, Vuong PT, El Assal LP, Kuperstein I, Zinovyev A, Hinton HS, Bryant WA, Aragon Artacho FJ, Planes FJ, Stalidzans E, Maass A, Vempala S, Hucka M, Saunders MA, Maranas CD, Lewis NE, Sauter T, Palsson BO, Thiele I, Fleming RMT. 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat Protoc, 14 (3):639-702.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res, 28 (1):27-30.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res, 44 (D1):D457-462.

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res, 45 (D1):D353-D361.

King ZA, Drager A, Ebrahim A, Sonnenschein N, Lewis NE, Palsson BO. 2015. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. PLoS Comput Biol, 11 (8):e1004321.

King ZA, Lu J, Drager A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE. 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res, 44 (D1):D515-522.

Kleijn RJ, Buescher JM, Le Chat L, Jules M, Aymerich S, Sauer U. 2010. Metabolic fluxes during strong carbon catabolite repression by malate in Bacillus subtilis. J Biol Chem, 285 (3):1587-1596.

Koutrouli M, Karatzas E, Paez-Espino D, Pavlopoulos GA. 2020. A Guide to Conquer the Biological Network Era Using Graph Theory. Frontiers in Bioengineering and Biotechnology, 8 (34).

Kovács ÁT. 2019. Bacillus subtilis. Trends in Microbiology, 27 (8):724-725.

Liu C, Ma Y, Zhao J, Nussinov R, Zhang Y-C, Cheng F, Zhang Z-K. 2020. Computational network biology: Data, models, and applications. Physics Reports, 846:1-66.

Long CP, Antoniewicz MR. 2019. High-resolution (13)C metabolic flux analysis. Nat Protoc, 14 (10):2856-2877.

Lorca GL, Chung YJ, Barabote RD, Weyler W, Schilling CH, Saier MH, Jr. 2005. Catabolite repression and activation in Bacillus subtilis: dependency on CcpA, HPr, and HprK. J Bacteriol, 187 (22):7826-7839.

Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. 2017. Transcriptomics technologies. PLoS Comput Biol, 13 (5):e1005457.

Lu H, Liu X, Huang M, Xia J, Chu J, Zhuang Y, Zhang S, Noorman H. 2015. Integrated isotope-assisted metabolomics and (13)C metabolic flux analysis reveals metabolic flux redistribution for high glucoamylase production by Aspergillus niger. Microb Cell Fact, 14:147.

Mahadevan R, Schilling CH. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng, 5 (4):264-276.

McKenney PT, Driks A, Eichenberger P. 2013. The Bacillus subtilis endospore: assembly and functions of the multilayered coat. Nat Rev Microbiol, 11 (1):33-44.

Meyer FM, Stulke J. 2013. Malate metabolism in Bacillus subtilis: distinct roles for three classes of malate-oxidizing enzymes. FEMS Microbiol Lett, 339 (1):17-22.

Mudunuri U, Che A, Yi M, Stephens RM. 2009. bioDBnet: the biological database network. Bioinformatics, 25 (4):555-556.

Nicolas P, Mader U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, Bidnenko E, Marchadier E, Hoebeke M, Aymerich S, Becher D, Bisicchia P, Botella E, Delumeau O, Doherty G, Denham EL, Fogg MJ, Fromion V, Goelzer A, Hansen A, Hartig E, Harwood CR, Homuth G, Jarmer H, Jules M, Klipp E, Le Chat L, Lecointe F, Lewis P, Liebermeister W, March A, Mars RA, Nannapaneni P, Noone D, Pohl S, Rinn B, Rugheimer F, Sappa PK, Samson F, Schaffer M, Schwikowski B, Steil L, Stulke J, Wiegert T, Devine KM, Wilkinson AJ, van Dijl JM, Hecker M, Volker U, Bessieres P, Noirot P. 2012. Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. Science, 335 (6072):1103-1106.

Orth JD, Thiele I, Palsson BO. 2010. What is flux balance analysis? Nat Biotechnol, 28 (3):245-248.

Palsson BØ. 2015. Systems Biology: Constraint-based Reconstruction and Analysis. Cambridge: Cambridge University Press.

Paredes-Sabja D, Setlow P, Sarker MR. 2011. Germination of spores of Bacillales and Clostridiales species: mechanisms and proteins involved. Trends Microbiol, 19 (2):85-94.

Price ND, Famili I, Beard DA, Palsson BO. 2002. Extreme pathways and Kirchhoff's second law. Biophys J, 83 (5):2879-2882.

Raman K, Chandra N. 2009. Flux balance analysis of biological systems: applications and challenges. Brief Bioinform, 10 (4):435-449.

Schellenberger J, Lewis NE, Palsson BO. 2011. Elimination of thermodynamically infeasible loops in steady-state metabolic models. Biophys J, 100 (3):544-553.

Shan M, Dai D, Vudem A, Varner JD, Stroock AD. 2018. Multi-scale computational study of the Warburg effect, reverse Warburg effect and glutamine addiction in solid tumors. PLoS Comput Biol, 14 (12):e1006584.

Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E. 2008. Network-based prediction of human tissue-specific metabolism. Nat Biotechnol, 26 (9):1003-1010.

Stempler S, Yizhak K, Ruppin E. 2014. Integrating transcriptomics with metabolic modeling predicts biomarkers and drug targets for Alzheimer's disease. PLoS One, 9 (8):e105383.

Stulke J, Hillen W. 2000. Regulation of carbon catabolism in Bacillus species. Annu Rev Microbiol, 54:849-880.

Su Y, Liu C, Fang H, Zhang D. 2020. Bacillus subtilis: a universal cell factory for industry, agriculture, biomaterials and medicine. Microb Cell Fact, 19 (1):173.

Sung J, Kim S, Cabatbat JJT, Jang S, Jin YS, Jung GY, Chia N, Kim PJ. 2017. Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. Nat Commun, 8:15393.

UniProt Consortium T. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res, 46 (5):2699.

Uygun S, Peng C, Lehti-Shiu MD, Last RL, Shiu SH. 2016. Utility and Limitations of Using Gene Expression Data to Identify Functional Associations. PLoS Comput Biol, 12 (12):e1005244.

van den Esker MH, Koets AP. 2019. Application of Transcriptomics to Enhance Early Diagnostics of Mycobacterial Infections, with an Emphasis on Mycobacterium avium ssp. paratuberculosis. Vet Sci, 6 (3).

van den Esker MH, Kovacs AT, Kuipers OP. 2017. YsbA and LytST are essential for pyruvate utilization in Bacillus subtilis. Environ Microbiol, 19 (1):83-94.

Veras HCT, Campos CG, Nascimento IF, Abdelnur PV, Almeida JRM, Parachin NS. 2019. Metabolic flux analysis for metabolome data validation of naturally xylose-fermenting yeasts. BMC Biotechnol, 19 (1):58.

Wang Y, Eddy JA, Price ND. 2012. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC Syst Biol, 6:153.

Wegner A, Meiser J, Weindl D, Hiller K. 2015. How metabolites modulate metabolic flux. Curr Opin Biotechnol, 34:16-22.

Yao R, Li J, Feng L, Zhang X, Hu H. 2019. (13)C metabolic flux analysis-guided metabolic engineering of Escherichia coli for improved acetol production from glycerol. Biotechnol Biofuels, 12:29.

Zhong C, Li F, Liu M, Yang XN, Zhu HX, Jia YY, Jia SR, Piergiovanni L. 2014. Revealing differences in metabolic flux distributions between a mutant strain and its parent strain Gluconacetobacter xylinus CGMCC 2955. PLoS One, 9 (6):e98772.

Zimmermann M, Kogadeeva M, Gengenbacher M, McEwen G, Mollenkopf HJ, Zamboni N, Kaufmann SHE, Sauer U. 2017. Integration of Metabolomics and Transcriptomics Reveals a Complex Diet of Mycobacterium tuberculosis during Early Macrophage Infection. mSystems, 2 (4).

Zur H, Ruppin E, Shlomi T. 2010. iMAT: an integrative metabolic analysis tool. Bioinformatics, 26 (24):3140-3142.

# Appendix

**Table A1:** List of GPR associations from central energy metabolism of *B. subtilis*.

| Gene symbol | Gene ID | Flux name |
|---|---|---|
| ackA | BSU29470 | ACKr |
| citA | BSU09440 | CS |
| citB | BSU18000 | ACONT |
| citZ | BSU29140 | CS |
| dctP | BSU04470 | SUCCt2r |
| eno | BSU33900 | ENO |
| fbaA | BSU37120 | FBA |
| fbp | BSU40190 | FBP |
| fruA | BSU14400 | FRUpts |
| fumC | BSU33040 | FUM |
| gapA | BSU33940 | GAPD |
| gapB | BSU29020 | GAPDi_nadp |
| glpF | BSU09280 | GLYCt |
| gltP | BSU02340 | GLUt2r |
| gndA | BSU23860 | GND |
| gntP | BSU40070 | GLCNt2ir |
| gntZ | BSU40080 | GND |
| icd | BSU29130 | ICDHyr |
| levD | BSU27070 | FRUpts |
| levE | BSU27060 | FRUpts |
| levF | BSU27050 | FRUpts |
| levG | BSU27040 | FRUpts |
| ptsH | BSU13900 | FRUpts |
| ptsI | BSU13910 | FRUpts |
| lrgA | BSU28910 | PYRt2 |
| lrgB | BSU28900 | PYRt2 |
| maeA | BSU37050 | ME2 |
| maeN | BSU31580 | MALt4 |
| yflS | BSU07570 | MALt4 |
| maeN | BSU31580 | MALt10 |
| yflS | BSU07570 | MALt10 |
| maeN | BSU31580 | MALt2r |
| yflS | BSU07570 | MALt2r |
| malS | BSU29880 | ME2 |
| mdh | BSU29120 | MDH |
| mleA | BSU23550 | ME2 |
| odhA | BSU19370 | AKGDH |
| odhB | BSU19360 | AKGDH |
| pckA | BSU30560 | PPCK |
| pdhA | BSU14580 | PDH |
| pdhB | BSU14590 | PDH |
| pdhC | BSU14600 | PDH |
| pdhD | BSU14610 | PDH |
| pfkA | BSU29190 | PFK |
| pgi | BSU31350 | PGI |
| pgk | BSU33930 | PGK_1 |
| pgm | BSU33910 | PGM_1 |
| pps | BSU18830 | PPS |
| pta | BSU37660 | PTAr |

| Gene symbol | Gene ID | Flux name |
|---|---|---|
| ptsG | BSU13890 | GLCpts |
| ptsH | BSU13900 | GLCpts |
| ptsI | BSU13910 | GLCpts |
| pycA | BSU14860 | PC |
| pyk | BSU29180 | PYK |
| rpe | BSU15790 | RPE |
| sdhA | BSU28440 | SUCD1 |
| sdhB | BSU28430 | SUCD1 |
| sdhC | BSU28450 | SUCD1 |
| sucC | BSU16090 | SUCOAS |
| sucD | BSU16100 | SUCOAS |
| tkt | BSU17890 | TKT1 |
| tkt | BSU17890 | TKT2 |
| tpiA | BSU33920 | TPI |
| ykgB | BSU13010 | G6PDH2r |
| ytsJ | BSU29220 | ME2 |
| ywjH | BSU37110 | TALA |
| ywlF | BSU36920 | RPI |
| zwf | BSU23850 | G6PDH2r |
| fruK | BSU14390 | FRUK |
| glpK | BSU09290 | GLYK |
| serA | BSU23070 | PGCD |
| serC | BSU10020 | PSERT |
| yoaD | BSU18560 | PGCD |
| rsbX | BSU04740 | PSP_L |
| sdaAB | BSU15850 | SERD_L |
| sdaAA | BSU15860 | SERD_L |
| fbaA | BSU37120 | FBA2 |
| gntK | BSU40060 | GNKr |
| glxK | BSU40040 | GLYCK |
| nadB | BSU27870 | ASPO1 |
| aspB | BSU22370 | ASPTA |
| mtnE | BSU13580 | UNK5 |
| citM | BSU07610 | CITt10 |
| citM | BSU07610 | CITt11 |
| citM | BSU07610 | CITt12 |
| citM | BSU07610 | CITt13 |
| citM | BSU07610 | CITt14 |
| citH | BSU39060 | CITt14 |
| citM | BSU07610 | CITt15 |
| cimH | BSU38770 | CITt2r |
| yraO | BSU26860 | CITt2r |
| yfiY | BSU08440 | FEDCabc |
| yfhA | BSU08460 | FEDCabc |
| yfiZ | BSU08450 | FEDCabc |
| citM | BSU07610 | ICITt10 |
| citM | BSU07610 | ICITt2 |
| metA | BSU21910 | HSST |
| scoA | BSU38990 | OCOAT1 |
| scoB | BSU38980 | OCOAT1 |
| prpB | BSU24120 | MCITL2 |
| yodQ | BSU19710 | SDPDS |
| metI | BSU11870 | SHSL1_1 |
| metI | BSU11870 | SHSL2 |
| metI | BSU11870 | SHSL3 |
| metI | BSU11870 | SHSL4r |
| gabD | BSU03910 | SSALy |
| purB | BSU06440 | ADSL1r |
| purB | BSU06440 | ADSL2r |
| argH | BSU29440 | ARGSL |

| Gene symbol | Gene ID | Flux name |
|---|---|---|
| ansB | BSU23570 | ASPT |
| yflS | BSU07570 | FUMt2r |
| dctP | BSU04470 | FUMt2r |
| menD | BSU30820 | 2S6HCCi |
| gabT | BSU03900 | ABTA |
| argD | BSU11220 | ACOTA |
| yflS | BSU07570 | AKGt2r |
| yoaB | BSU18540 | AKGt2r |
| yugH | BSU31400 | ALATA_L |
| patA | BSU14000 | APTA1i |
| gudB | BSU22960 | GLUDxi |
| rocG | BSU37790 | GLUDxi |
| gltA | BSU18450 | GLUSy |
| gltB | BSU18440 | GLUSy |
| hisH | BSU34890 | HSTPTr |
| hisC | BSU22620 | HSTPTr |
| rocD | BSU40340 | ORNTA_1 |
| menD | BSU30820 | OXGDC |
| hisC | BSU22620 | PHETA1 |
| hisH | BSU34890 | PHETA1 |
| hisC | BSU22620 | TYRTA |
| hisH | BSU34890 | TYRTA |
| ywiE | BSU37240 | CLPNS2_BS |
| clsB | BSU37190 | CLPNS2_BS |
| clsA | BSU36590 | CLPNS2_BS |
| yhdN | BSU09530 | ALCD19y |
| iolS | BSU39780 | ALCD19y |
| yhdN | BSU09530 | ALCD19_L |
| iolS | BSU39780 | ALCD19_L |
| glpQ | BSU02130 | GPDDA4 |
| bacD | BSU37710 | LAAL24 |
| bacD | BSU37710 | LAAL25 |
| bacD | BSU37710 | LAAL27 |
| bacD | BSU37710 | LAAL26 |
| bacD | BSU37710 | LAAL28 |
| speA | BSU14630 | ARGDC |
| argI | BSU40320 | ARGN_1 |
| artP | BSU23980 | ARGabc |
| artQ | BSU23970 | ARGabc |
| artR | BSU23960 | ARGabc |
| lysP | BSU33330 | ARGt2r |
| rocE | BSU40330 | ARGt2r |
| rocC | BSU37760 | ARGt2r |
| aldX | BSU39860 | LCADi |
| aldY | BSU38830 | LCADi |
| dhaS | BSU19310 | LCADi |
| ldh | BSU03050 | LDH_L |
| lctP | BSU03060 | L_LACt2r |
| yvfH | BSU34190 | L_LACt2r |
| bacD | BSU37710 | LAAL33 |
| bacD | BSU37710 | LAAL34 |
| trpB | BSU22640 | TRPS1 |
| trpA | BSU22630 | TRPS1 |
| trpP | BSU10010 | TRPt2r |
| cysK | BSU00730 | AHSERL4 |
| cysK | BSU00730 | CYSS_2 |
| mccB | BSU27250 | CYSTGL_1 |
| bacD | BSU37710 | LAAL9 |
| coaBC | BSU15700 | PPNCL |
| mccB | BSU27250 | TRPAS1 |

| Gene symbol | Gene ID | Flux name |
|---|---|---|
| hutH | BSU39350 | HISDr |
| hisD | BSU34910 | HISTD |
| hutM | BSU39390 | HISt2r |
| ybgF | BSU02400 | HISt2r |
| bacD | BSU37710 | LAAL40 |
| bacD | BSU37710 | LAAL41 |
| bacD | BSU37710 | LAAL42 |
| bacD | BSU37710 | LAAL10 |
| bacD | BSU37710 | LAAL16 |
| bacD | BSU37710 | LAAL22 |
| bacD | BSU37710 | LAAL4 |
| tdh | BSU16990 | THRD |
| ilvA | BSU21770 | THRD_L |
| thrC | BSU32250 | THRS |
| pssA | BSU02270 | CDPDSP_BS |
| mccA | BSU27260 | CYSTS_2 |
| glyA | BSU36900 | GHMT2r |
| bacD | BSU37710 | LAAL12 |
| bacD | BSU37710 | LAAL14 |
| bacD | BSU37710 | LAAL18 |
| bacD | BSU37710 | LAAL21 |
| bacD | BSU37710 | LAAL3 |
| bacD | BSU37710 | LAAL31 |
| bacD | BSU37710 | LAAL37 |
| bacD | BSU37710 | LAAL6 |
| bacD | BSU37710 | LAAL7 |
| bacD | BSU37710 | LAAL8 |
| cysE | BSU00930 | SERAT |
| aldX | BSU39860 | ALDD31_1 |
| aldY | BSU38830 | ALDD31_1 |
| dhaS | BSU19310 | ALDD31_1 |
| kbl | BSU17000 | GLYAT |
| thiO | BSU11670 | GLYO1 |
| bacD | BSU37710 | LAAL1 |
| bacD | BSU37710 | LAAL13 |
| bacD | BSU37710 | LAAL19 |
| bacD | BSU37710 | LAAL2 |
| bacD | BSU37710 | LAAL29 |
| bacD | BSU37710 | LAAL35 |
| purD | BSU06530 | PRAGSr |
| thiO | BSU11670 | SARCOX |

*After applying a P-value cutoff of 0.05 cutoff (after correction for multiple testing), reactions of the according genes which were differentially expressed in at least one of all eight conditions are highlighted in yellow.

**For full reaction name, see the publication of *B. subtilis* 168 (iYO844) model (King et al. 2016).

**Table A2:** List of other exchange reactions in the metabolic model of *B. subtilis* (excluding carbon sources used in this study).

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| EX_2ddglcn(e) | 0 | 0 | 0.00E+00 | 0.00E+00 | 0 | 0 | 0.00E+00 | 0.00E+00 |
| EX_2hxmp(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_2pg(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_2pglyc(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_3amba(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_3amp(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_3cmp(e) | 0.00E+00 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_3gmp(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_3pg(e) | -0.209 | 0 | -0.149 | 0 | 0 | 0 | 0.00E+00 | 0.00E+00 |
| EX_3ump(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_4abut(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_5mtr(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_6pgc(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_Larab(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_Lcyst(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_abt__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ac(e) | 3.43 | 0.682 | 0.0428 | 0.0349 | 0.281 | 10.6 | 7.71 | 0.0553 |
| EX_acac(e) | 0.679 | 0.597 | 0.503 | 0.212 | 0.0376 | 0.467 | 0.6 | 1.02 |
| EX_acgam(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_acmana(e) | 0 | 0 | 0 | 0.00E+00 | 0.00E+00 | 0.00E+00 | 0.00E+00 | 0.00E+00 |
| EX_acnam(e) | -0.167 | -0.13 | -0.00676 | 0 | 0 | 0 | -0.245 | 0 |
| EX_actn__R(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ade(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_adn(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_akg(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ala_B(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ala__D(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ala_L_Thr__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ala_L_asp__L(e) | 0.00E+00 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ala__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ala_L_gln__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| EX_ala_L_glu__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_L_alagly(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ala_L_his__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ala_L_leu__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_alaala(e) | -0.219 | -0.413 | -0.243 | 0 | 0 | 0 | -1.91 | -6.13E-02 |
| EX_alltn(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00E+00 |
| EX_amp(e) | -0.108 | -1.12 | -0.0768 | -0.0402 | -0.0731 | -0.213 | -0.137 | -0.0311 |
| EX_amylase(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_antim(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_arab__D(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_arab__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_arbt(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_arg__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_argp(e) | 0 | 0 | 0 | 0.0613 | -0.00673 | 0 | 0.158 | 0.0458 |
| EX_arsenb(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_arsna(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_arsni2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_asn__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_asp__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_bilea(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_btd_RR(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_buts(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_cbl2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_cd2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_cellb(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_cgly(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_chitob(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_chol(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_chols(e) | 0 | 0 | 0.00E+00 | 0.00E+00 | 0 | 0 | 0 | 0 |
| EX_chor(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_cit(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_citr__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_cmp(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_cobalt2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00E+00 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| EX_crn(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_cro4(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_csn(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ctbt(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_cu2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_cys__D(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_cys__L(e) | -0.0336 | -0.0302 | -0.0239 | 0 | 0 | 0 | 0 | 0 |
| EX_cyst__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00E+00 |
| EX_cytd(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_dad_2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_dcyt(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_dextrin(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_dha(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_diact(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_djenk(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_drib(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_dtmp(e) | 1.11 | 0.985 | 0.828 | 0.428 | 0.532 | 1.31 | 1.73 | 0.28 |
| EX_ectoine(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_etha(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_eths(e) | 0 | 0 | 0 | -0.0375 | 0 | 0 | 0 | 0 |
| EX_etoh(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_fe2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ferrich(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ferxa(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_fol(e) | 0.139 | 0.108 | 0.0997 | 0.0418 | 0 | 0.119 | 0.197 | 0.0231 |
| EX_for(e) | 0 | 0 | 0 | 0.0419 | 0 | 0 | 0 | 0 |
| EX_fum(e) | 0.114 | 0.101 | 0.0811 | 0.104 | 0.0705 | 0.257 | 0.303 | 0.0786 |
| EX_g1p(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_g6p(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gal(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_galctr__D(e) | 0 | 0 | 0 | -0.216 | 0 | 0 | 0 | -0.161 |
| EX_galt(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_galur(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gam6p(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| EX_gam(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gbbtn(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_glcr(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_glcur(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gln__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_glu__D(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_glx(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gly_asn__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gly_asp__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gly(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gly_gln__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gly_glu__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gly_met__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gly_pro__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_glyb(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_glyc3p(e) | 0 | 0 | 0 | 0 | 0 | -0.322 | 0 | 0 |
| EX_glyclt(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_glycogen(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gmp(e) | -1.26 | -1.2 | -0.864 | 0 | -0.00025 | 0 | -0.197 | -0.023 |
| EX_gsn(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gthox(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gthrd(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_gua(e) | 0 | 0 | 0 | -0.00383 | 0 | -0.00992 | 0 | 0 |
| EX_h2o2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_hexs(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_hg2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_his__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_hqn(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_hxan(e) | 0 | 0 | 0 | -0.0378 | 0 | 0 | 0 | 0 |
| EX_icit(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ile__L(e) | -0.224 | -0.201 | -0.159 | -0.0835 | -0.129 | 0 | -0.285 | -0.0645 |
| EX_inost(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ins(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_istnt(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| EX_lac__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_lanth(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_lcts(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_leu__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_lipt(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_lys__L(e) | -0.192 | -0.172 | -0.137 | -0.0716 | 0 | 0 | -0.244 | -0.0553 |
| EX_madg(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_mal__D(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_malt(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_malttr(e) | 0 | 0 | -0.0157 | -0.169 | -0.0862 | -0.0399 | -0.0039 | -0.0254 |
| EX_man1p(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.255 |
| EX_man6p(e) | 0 | 0 | 0 | 0.714 | 0.171 | 0 | 0 | 0 |
| EX_man(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_mbdg(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_melib(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_meoh(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_met__D(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_met_L_ala__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_met__L(e) | -0.0669 | -0.0601 | -0.0476 | 0 | -0.0681 | 0 | -0.085 | -0.0193 |
| EX_metox__R(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_metox(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_mn2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_mnl(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_mobd(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_mops(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_mso3(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_nac(e) | 0 | 0 | 0 | -0.00381 | 0 | -0.00987 | 0 | -0.00294 |
| EX_ni2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_no2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_no3(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_orn__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_pala(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_pep(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_phe__L(e) | -0.0442 | -0.0316 | -0.0424 | 0 | 0 | 0 | 0 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| EX_pnto__R(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ppa(e) | 0 | 0 | 0 | 0 | 0.346 | 0.974 | 0 | 0 |
| EX_ppi(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_pro__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_prolb(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_pser__D(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_pser__L(e) | 0 | -0.132 | 0 | 0 | 0 | -0.211 | 0 | 0 |
| EX_pur(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_raffin(e) | 0 | 0 | 0 | -0.0554 | 0 | 0 | 0 | -0.0428 |
| EX_rib__D(e) | 1.42 | 2.36 | 0.951 | 0.0557 | 0.123 | 0.206 | 0.543 | 0.0632 |
| EX_ribflv(e) | 0 | 0 | 0 | 0 | 5.00E-04 | 0 | 0 | 0 |
| EX_rmn(e) | -0.283 | -0.322 | -0.238 | -0.0325 | -1.96 | 0 | -0.355 | -0.00719 |
| EX_salcn(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_sbt__D(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ser__D(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ser__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_spmd(e) | 0 | 0 | 0 | 0 | 0 | 0.147 | 0 | 0 |
| EX_srb__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_starch(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_subtilisin(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_sucr(e) | 0 | 0 | 0 | -0.0046 | 0 | 0 | 0 | -0.00969 |
| EX_sula(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_taur(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_thiog(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_thr__L(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_thym(e) | -1.12 | -0.997 | -0.838 | -0.236 | -0.44 | -1.32 | -0.843 | -0.134 |
| EX_thymd(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_tmp(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_tre(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_trp__L(e) | -0.0321 | -0.0288 | -0.0228 | -0.012 | -0.0217 | -0.031 | -0.0408 | -0.00924 |
| EX_tyr__L(e) | 0 | 0 | 0 | 0 | 0.022 | 0 | 0 | 0 |
| EX_ump(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_ura(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_urate(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| EX_urea(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_uri(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_val__L(e) | -0.427 | -0.163 | -0.304 | -0.113 | 0 | 0 | 0 | -0.0415 |
| EX_xan(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_xtsn(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_xyl__D(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_zn2(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EX_f6p(e) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*Common exchange reactions used by the model are highlighted in yellow.

**For full reaction name, see the publication of *B. subtilis* 168 (iYO844) model (King et al. 2016).

**Table A3:** Predicted flux from RED-TIL before implementing IFFPR.

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| 2S6HCCi | 0.000157 | 0.000141 | 1.12E-04 | 5.85E-05 | 0.000106 | 0.000152 | 2.00E-04 | 4.52E-05 |
| ACKr | -0.286 | -0.257 | -2.27 | -0.107 | 2.47 | -0.276 | 19.6 | -0.0726 |
| ACONT | 17.3 | 19.1 | 3.88 | 15.7 | 2.52 | 27.3 | -1.62 | 7.63 |
| ACOTA | -0.114 | 0 | 0 | -0.0425 | 0 | 0 | 0 | 0 |
| AKGDH | 18.9 | 20.9 | 2.93 | 16.4 | 4.05 | 27.8 | 0 | 7.4 |
| AKGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ALCD19y | -3.11E-01 | 4.72 | -0.544 | -0.363 | 0 | -0.826 | -0.636 | -0.49 |
| ALDD31_1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ARGSL | 0.114 | 0 | 0 | 0.0425 | 0 | 0 | 0 | 0 |
| ARGabc | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ARGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ASPO1 | 46.8 | 54.1 | 0 | 45.2 | 0 | 39.5 | 0 | 2.41 |
| ASPT | 0 | 0 | 1.62 | 0 | 0.206 | 0 | 0 | 0 |
| ASPTA | -48 | -57.6 | -2.6 | -51.2 | -0.884 | -43.8 | -1.29 | -2.7 |
| CDPDSP_BS | 0.033 | 0.0297 | 0.0235 | 0.0123 | 0.0224 | 0.0319 | 0.042 | 0.00952 |
| CITt10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| CITt14 | 0.00189 | 0.0017 | 0.00135 | 0.000705 | 0.00128 | 0.00183 | 0.0024 | 0.000545 |
| CITt15 | 0.417 | 0.374 | 0 | 0.155 | 0 | 0.403 | 0 | 0.12 |
| CITt2r | -0.423 | -0.38 | -0.00424 | -0.158 | -0.654 | -0.405 | -1.63 | -0.122 |
| CLPNS2_BS | 0.000295 | 0.000265 | 0.00021 | 1.10E-04 | 2.00E-04 | 2.85E-04 | 3.75E-04 | 8.50E-05 |
| CS | 17.3 | 19.1 | 3.88 | 15.7 | 3.18 | 27.3 | 0 | 7.63 |
| CYSS_2 | 0.0336 | 0.0302 | 0 | 0.0125 | 0 | 0.0325 | 0 | 0 |
| CYSTGL_1 | 0 | 0 | 0.159 | 0 | 0.0228 | 0 | 0.0427 | 0 |
| CYSTS_2 | 0 | 0 | 0.0239 | 0 | 0.0228 | 0 | 0.0427 | 0 |
| ENO | 25.2 | 25.8 | 11.7 | 18.2 | 16.4 | 13.9 | 16.3 | -2.75 |
| FRUK | 1.19 | 0.998 | 0.544 | 2.64 | 0.787 | 1.31 | 1.05 | 0.525 |
| FBA | 10.3 | 8.69 | 4.56 | 7.53 | 5.09 | 6.41 | 8.25 | -1.66 |
| FBA2 | 1.09 | 4.72 | -0.544 | -0.298 | 0 | -0.826 | -0.636 | -0.49 |
| FBP | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.19 |
| FEDCabc | 0.00204 | 0.00183 | 0.00145 | 0.000759 | 0 | 0.000345 | 0.00123 | 0.000586 |
| FRUpts | 2.28 | 5.72 | 0 | 2.34 | 0.787 | 0.488 | 0.415 | 0.0354 |
| FUM | 19.1 | 21.3 | 6.48 | 19.8 | 5.06 | 14 | -11.1 | 6.98 |
| FUMt2r | 0 | 0 | 0 | 0 | 0 | 0 | -11.1 | -3.77 |
| G6PDH2r | 1.65 | 1.06 | 0 | 0 | 0 | 0 | 2.84 | 0 |
| GAPD | 24.1 | 24.8 | 11.9 | 15.5 | 16.6 | 12.5 | 16.7 | 0 |
| GAPDi_nadp | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.70E+00 |
| GHMT2r | 0.129 | 0.0603 | 0.0478 | 0.0481 | 0.0455 | 0.0648 | 0.0853 | 6.24E-05 |
| GLCNt2ir | 2.11 | 1.14 | 5.13 | 0.889 | 1.05 | 1.39 | 1.18 | 0 |
| GLCpts | 8.79 | 6.27 | 0.6 | 3.61 | 1.76 | 3.32 | 7.84 | 0 |
| GLUDxi | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GLUSy | 19.2 | 11.7 | 1.39 | 0 | 0 | 21.6 | 0.963 | 0 |
| GLUt2r | 0.567 | 1.17 | 1.19 | 3.51 | 0.865 | 0.601 | 0 | 3.06 |
| GLYCt | 1.97 | -2.48 | 2.39 | 3.76 | 6.22 | 2.91 | 0.635 | 2.76 |
| GLYK | 1.66 | 2.25 | 1.84 | 3.4 | 6.22 | 2.08 | 0 | 2.27 |
| GLYO1 | 0 | 2.54 | 0 | 5.52 | 0 | 3.28 | 0 | 0 |
| GNKr | 2.11 | 1.14 | 5.13 | 0.889 | 1.05 | 1.39 | 1.18 | 0 |
| HISTD | 0.0482 | 0 | 0 | 0.018 | 0 | 0 | 0 | 0 |
| HISt2r | 0 | 0.0433 | 0.0343 | 0 | 0.0327 | 0.0466 | 0.0613 | 0.0139 |
| HSTPTr | 0.0482 | 0 | 0 | 0.018 | 0 | 0 | 0 | 0 |
| ICDHyr | 17.3 | 19.1 | 1.05 | 13.1 | 2.52 | 26.1 | -1.62 | 4.47 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| ICITt10 | 0.06 | 0.0539 | 0.0427 | 0.0224 | 0.0407 | 0.058 | 0 | 0.0173 |
| ICITt2 | -0.06 | -0.0539 | -2.87 | -2.64 | -0.0407 | -1.3 | 0 | -3.18 |
| LCADi | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LDH_L | -7.6 | -12.4 | -6.4 | -9.87 | -13.7 | -4.38 | -15 | -2.01 |
| L_LACt2r | -7.6 | -11.1 | -6.4 | -8.24 | -12.9 | 17.2 | -10.5 | -0.126 |
| MALt10 | 0 | 1.36 | 0 | 1.63 | 0.807 | 21.6 | 4.51 | 1.88 |
| MALt2r | 0 | 1.36 | 0 | 1.63 | 0.807 | 0 | 19.7 | 1.88 |
| MALt4 | 0 | 1.36 | 0 | 1.63 | 0.807 | 4.95 | 2.3 | 1.88 |
| MDH | 18.5 | 22.6 | 6.48 | 21.7 | 7.48 | 31.6 | 14.7 | 10.7 |
| ME2 | 0.612 | 2.78 | 0 | 3.02E+00 | 0 | 8.9 | 0.62 | 1.9 |
| MCITL2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| OXGDC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PDH | 18.3 | 17.3 | 0 | 10.6 | 6.36 | 24.9 | 21.1 | 7.82 |
| PFK | 9.13 | 7.69 | 4.02 | 4.89 | 4.31 | 5.1 | 7.2 | 0 |
| PGCD | 0.355 | 0.235 | 0.186 | 0.132 | 0.177 | 0.252 | 0.332 | 0.0463 |
| GND | 3.76 | 2.2 | 5.13 | -3.81 | 1.05 | -3.43 | 4.01 | -3.14 |
| PGI | 6.99 | 6.4 | 0.494 | 7.57 | 1.66 | 7.49 | 4.81 | -0.0428 |
| PGK_1 | 24.1 | 24.8 | 11.9 | 15.5 | 16.6 | 12.5 | 16.7 | -2.7 |
| PGM_1 | 25.2 | 25.8 | 11.7 | 18.2 | 16.4 | 13.9 | 16.3 | -2.75 |
| PHETA1 | -0.104 | 0 | 0 | -0.0387 | 0 | -0.1 | -0.107 | 0 |
| PPCK | 0 | 0 | 0 | 0 | 3.42 | 0 | 13.5 | 2.79 |
| PPS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00E+00 |
| PRAGSr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PTAr | -0.286 | -0.257 | -2.27 | -0.107 | 2.47 | -0.276 | 19.6 | -0.0726 |
| PYK | 13.6 | 13.7 | 11.2 | 12.1 | 17.4 | 9.97 | 21.3 | 0 |
| PYRt2 | 2.32 | 2.35 | 0 | 0.098 | 1.38 | 8.26 | 8.26 | 8.26 |
| RPE | 2.33 | 1.46 | 3.52 | -2.61 | 0.817 | -2.33 | 2.63 | -2.1 |
| RPI | -1.43 | -0.741 | -1.61 | 1.21 | -0.233 | 1.1 | -1.38 | 1.04 |
| SERAT | 0.0336 | 0.0302 | 0 | 0.0125 | 0 | 0.0325 | 0 | 0 |
| SERD_L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SHSL1_1 | 0 | 0 | 0.136 | 0 | 0 | 0 | 0 | 0 |
| SHSL2 | 0.0669 | 0.0601 | 0.0715 | 0.0249 | 0.0681 | 0.0646 | 0.128 | 0 |
| SHSL4r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| SUCCt2r | 0 | 0.351 | 1.93 | 3.35 | 0.809 | -13.9 | 0 | 3.35 |
| SUCD1 | 18.9 | 21.3 | 4.86 | 19.8 | 4.85 | 14 | 0 | 10.8 |
| SUCOAS | 18.8 | 20.8 | 2.73 | 16.4 | 3.98 | 27.8 | -0.128 | 7.4 |
| TALA | 1.27 | 0.732 | 1.76 | -1.27 | 0.408 | -1.11 | 1.37 | -1.05 |
| THRD | 0.112 | 2.69 | 0.391 | 5.56 | 0.118 | 3.45 | 0.221 | 0.0693 |
| THRD_L | 0.224 | 0.201 | 0 | 0.0835 | 0.129 | 0.216 | 0.242 | 0.0645 |
| THRS | 0.446 | 2.99 | 0.469 | 5.69 | 0.321 | 3.77 | 0.603 | 0.166 |
| TKT1 | 1.27 | 0.732 | 1.76 | -1.27 | 0.408 | -1.11 | 1.37 | -1.05 |
| TKT2 | 1.07 | 0.732 | 1.76 | -1.34 | 0.408 | -1.21 | 1.26 | -1.05 |
| TPI | 12.7 | 15.3 | 5.62 | 9.29 | 11.1 | 7.32 | 7.16 | 0.0106 |
| TRPAS1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TRPS1 | 0.0321 | 0 | 0 | 0.012 | 0 | 0 | 0 | 0 |
| TYRTA | -0.0654 | 0 | 0 | -0.0244 | 0 | 0 | 0 | 0 |
| UNK5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*For full reaction name, see the publication of *B. subtilis* 168 (iYO844) model (King et al. 2016).

**Table A4:** Predicted flux from RED-TIL after implementing IFFPR.

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| 2S6HCCi | 0.000157 | 0.000141 | 1.12E-04 | 5.85E-05 | 0.000106 | 0.000152 | 2.00E-04 | 4.52E-05 |
| ACKr | -0.284 | -0.255 | -2.66 | -0.106 | 1.76 | 9.91 | 19 | -0.0726 |
| ACONT | 17.2 | 19.3 | 2.6 | 8.16 | 1.84 | 14 | -1.62 | 7.63 |
| ACOTA | -0.114 | 0 | 0 | -0.0425 | 0 | 0 | 0 | 0 |
| AKGDH | 18 | 19.2 | 4.25 | 6.77 | 3.24 | 13.9 | 0 | 3.78 |
| AKGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ALCD19y | 2.41E+00 | 4.72 | -0.544 | -0.363 | 0 | -0.826 | -0.636 | -0.49 |
| ALDD31_1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ARGSL | 0.114 | 0 | 0 | 0.0425 | 0 | 0 | -1.39E-17 | 0 |
| ARGabc | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ARGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ASPO1 | 29.7 | 24.3 | 0 | 21.5 | 0 | 20.1 | 0 | 0.65 |
| ASPT | 0 | 0 | 0.000916 | 0 | 0.000728 | 0 | 0 | 1.76 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| ASPTA | -30.8 | -28 | -1.19 | -21.9 | -0.68 | -21.1 | -1.29 | -2.7 |
| CDPDSP_BS | 0.033 | 0.0297 | 0.0235 | 0.0123 | 0.0224 | 0.0319 | 0.042 | 0.00952 |
| CITt10 | 0.06 | 0 | 0 | 0.0224 | 0 | 0.058 | 0 | 0.0173 |
| CITt14 | 0.00189 | 0.0017 | 0.00135 | 0.000705 | 0.00128 | 0.00183 | 0.0024 | 0.000545 |
| CITt15 | 0.0136 | 0.0838 | 0 | 0 | 0 | 0.0219 | 0 | 0 |
| CITt2r | -0.0796 | -0.0891 | -2.37 | -2.09 | -4.16 | -0.0824 | -2.2 | -0.019 |
| CLPNS2_BS | 0.000295 | 0.000265 | 0.00021 | 1.10E-04 | 2.00E-04 | 2.85E-04 | 3.75E-04 | 8.50E-05 |
| CS | 17.2 | 19.3 | 4.96 | 10.2 | 6 | 14 | 0.575 | 7.63 |
| CYSS_2 | 0.0324 | 0.0289 | 0 | 0.0116 | 0.0187 | 0.0267 | 0.0333 | 0 |
| CYSTGL_1 | 0.00668 | 0.00598 | 0.0255 | 0.00506 | 0.00405 | 0.00909 | 0.00941 | 0 |
| CYSTS_2 | 0.00124 | 0.00126 | 0.0239 | 0.000953 | 0.00405 | 0.00582 | 0.00941 | 0 |
| ENO | 21.5 | 22.1 | 10.1 | 12.4 | 16 | 10.5 | 16.6 | 0.0696 |
| FRUK | 0 | 0.998 | 0.544 | 2.7 | 0.787 | 1.31 | 1.05 | 0.409 |
| FBA | 9.36 | 8.69 | 4.35 | 8.7 | 4.9 | 7.35 | 8.79 | 0.106 |
| FBA2 | 2.41 | 4.72 | -0.544 | -0.363 | 0 | -0.826 | -0.636 | -0.374 |
| FBP | 1.81E-13 | 0 | 0 | 0 | 0 | 0 | 0 | 0.303 |
| FEDCabc | 0.00204 | 0.00183 | 0.00138 | 0.000759 | 0 | 0.000345 | 0.00123 | 0.000586 |
| FRUpts | 2.41 | 5.72 | 0 | 2.34 | 0.787 | 0.488 | 0.415 | 0.0354 |
| FUM | 18.1 | 19.6 | 6.15 | 10.1 | 2.43 | -0.134 | -9.92 | 8.85 |
| FUMt2r | 0 | 0 | -0.0186 | -0.0455 | 0 | -0.134 | -9.92 | -0.0357 |
| G6PDH2r | 0.0352 | 0 | 0 | 0 | 0 | 0 | 0.233 | 0 |
| GAPD | 21.8 | 20.9 | 10.2 | 12.6 | 16.2 | 10.8 | 16.9 | 0 |
| GAPDi_nadp | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00E+00 |
| GHMT2r | 0.0807 | 0.0603 | 0.0017 | 0.0393 | 0.0455 | 0.0648 | 0.0853 | 6.24E-05 |
| GLCNt2ir | 2.01 | 1.14 | 5.13 | 0.889 | 1.05 | 1.39 | 0.953 | 0 |
| GLCpts | 8.69 | 6.27 | 0.6 | 3.61 | 1.76 | 3.32 | 7.63 | 0 |
| GLUDxi | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GLUSy | 0.00104 | 0.000863 | 0 | 0 | 0 | 0.000617 | 0 | 0 |
| GLUt2r | 0.567 | 1.17 | 1.19 | 3.49 | 0.865 | 0.601 | 0 | 3.06 |
| GLYCt | -1.97 | -2.48 | 2.39 | 3.76 | 6.22 | 2.91 | 0.635 | 2.76 |
| GLYK | 0.443 | 2.25 | 1.84 | 3.4 | 6.22 | 2.08 | 0 | 2.27 |
| GLYO1 | 0 | 2.73 | 0 | 0 | 0 | 0 | 0 | 0 |
| GNKr | 2.01 | 1.14 | 5.13 | 0.889 | 1.05 | 1.39 | 0.953 | 0 |
| HISTD | 0 | 0 | 0 | 0.00919 | 0 | 0 | 0 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| HISt2r | 0.0482 | 0.0433 | 0.0343 | 0.00879 | 0.0327 | 0.0466 | 0.0613 | 0.0139 |
| HSTPTr | 0 | 0 | 0 | 0.00919 | 0 | 0 | 0 | 0 |
| ICDHyr | 16.4 | 17.5 | 2.3 | 3.5 | 1.84 | 12.1 | -1.62 | 0.848 |
| ICITt10 | 0 | 0.0539 | 0.0427 | 0 | 0.0407 | 0 | 0 | 0 |
| ICITt2 | -0.847 | -1.9 | -0.34 | -4.66 | -0.0407 | -1.85 | 0 | -6.79 |
| LCADi | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LDH_L | -1.89 | -2.85 | -4.01 | -2.26 | -4.37 | 0 | -7.21 | -1.29 |
| L_LACt2r | -1.89 | -2.77 | -4.01 | -2.15 | -3.68 | 1.62 | -5.98 | -1.18 |
| MALt10 | 0 | 0.085 | 0 | 0.102 | 0.692 | 1.62 | 1.23 | 0.118 |
| MALt2r | 0 | 1.36 | 0 | 1.63 | 0.807 | 24.9 | 19.7 | 1.88 |
| MALt4 | 0 | 0.00133 | 0 | 0.0016 | 0.000788 | 0.00971 | 0.0192 | 0.00184 |
| MDH | 18.1 | 21 | 6.15 | 11.9 | 3.93 | 26.4 | 6.06 | 10.8 |
| ME2 | 0 | 0 | 2.49E-05 | 9.73E-06 | 0 | 0.000385 | 4.98 | 0.000298 |
| MCITL2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| OXGDC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PC | 0.283 | 1.9 | 0 | 0 | 2.75 | 0 | 0 | 0 |
| PDH | 18.2 | 17.3 | 0 | 10.6 | 8.49 | 24.9 | 21.1 | 7.82 |
| PFK | 9.36 | 7.69 | 3.8 | 6 | 4.11 | 6.03 | 7.74 | 0 |
| PGCD | 0.307 | 0.235 | 0.14 | 0.124 | 0.177 | 0.252 | 0.332 | 0.0463 |
| GND | 2.05 | 1.14 | 1.5 | -2.17 | 1.05 | -1.58 | 1.19 | -0.314 |
| PGI | 8.51 | 7.11 | 1 | 5.3 | 1.66 | 5.7 | 7.21 | -0.0428 |
| PGK_1 | 21.8 | 20.9 | 10.2 | 12.6 | 16.2 | 10.8 | 16.9 | 0 |
| PGM_1 | 21.5 | 22.1 | 10.1 | 12.4 | 16 | 10.5 | 16.6 | 0.0696 |
| PHETA1 | -0.104 | 0 | 0 | -0.0387 | 0 | -0.0671 | -0.00211 | 0 |
| PPCK | 0 | 0 | 0 | 1.19 | 0 | 11.4 | 4.19 | 1.17 |
| PPS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00E+00 |
| PRAGSr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PTAr | -0.284 | -0.255 | -2.66 | -0.106 | 1.76 | 9.91 | 19 | -0.0726 |
| PYK | 9.89 | 10 | 9.63 | 7.51 | 13.6 | 18 | 12.6 | 1.19 |
| PYRt2 | 1.16 | 1.17 | 0 | 0.049 | 0.689 | 4.32 | 4.93 | 8.26 |
| RPE | 1.22 | 0.756 | 0.998 | -1.5 | 0.817 | -1.08 | 0.782 | -0.212 |
| RPI | -0.826 | -0.387 | -0.499 | 0.662 | -0.233 | 0.498 | -0.405 | 0.103 |
| SERAT | 0.0324 | 0.0289 | 0 | 0.0116 | 0.0187 | 0.0267 | 0.0333 | 0 |
| SERD_L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| SHSL1_1 | 0.00544 | 0.00472 | 0.00159 | 0.0041 | 0 | 0.00327 | 0 | 0 |
| SHSL2 | 0.0681 | 0.0613 | 0.0255 | 0.0259 | 0.0494 | 0.0704 | 0.0944 | 0 |
| SHSL4r | 0 | 0 | 0.00636 | 0 | 0 | 0.0131 | 0 | 0 |
| SUCCt2r | 0 | 0.351 | 1.93 | 3.35 | -0.809 | -13.9 | 0 | 3.35 |
| SUCD1 | 18 | 19.6 | 6.17 | 10.1 | 2.43 | 0 | 0 | 7.13 |
| SUCOAS | 18 | 19.2 | 4.21 | 6.74 | 3.19 | 13.8 | -0.0944 | 3.78 |
| TALA | 0.711 | 0.378 | 0.499 | -0.714 | 0.408 | -0.508 | 0.392 | -0.106 |
| THRD | 0.16 | 2.89 | 0.639 | 0.0505 | 0.118 | 0.168 | 0.221 | 0.0693 |
| THRD_L | 0.217 | 0.195 | 0.128 | 0.0785 | 0.148 | 0.194 | 0.275 | 0.0645 |
| THRS | 0.487 | 3.18 | 0.845 | 0.17 | 0.34 | 0.468 | 0.636 | 0.166 |
| TKT1 | 0.711 | 0.378 | 0.499 | -0.714 | 0.408 | -0.508 | 0.392 | -0.106 |
| TKT2 | 0.51 | 0.378 | 0.499 | -0.789 | 0.408 | -0.575 | 0.39 | -0.106 |
| TPI | 11.9 | 11.9 | 5.39 | 4.64 | 10.9 | 4.01 | 7.7 | 0 |
| TRPAS1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TRPS1 | 0.0321 | 0 | 0 | 0.012 | 0 | 0 | 0 | 0 |
| TYRTA | -0.0654 | 0 | 0 | -0.0244 | 0 | 0 | 0 | 0 |
| UNK5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*For full reaction name, see the publication of *B. subtilis* 168 (iYO844) model (King et al. 2016).

**Table A5:** Predicted flux from ll-COBRA.

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| 2S6HCCi | 0.000157 | 0.000141 | 1.12E-04 | 5.85E-05 | 0.000106 | 0.000152 | 2.00E-04 | 4.52E-05 |
| ACKr | -0.284 | -0.255 | -2.61 | -0.106 | 2.29 | 9.91 | 9.17 | -0.0726 |
| ACONT | 17.2 | 16.3 | 2.6 | 8.94 | 1.84 | 14 | 3.43 | 7.63 |
| ACOTA | -0.114 | 0 | 0 | -0.0425 | 0 | 0 | 0 | 0 |
| AKGDH | 16.7 | 16.2 | 4.54 | 12.2 | 3.24 | 13.9 | 0 | 3.4 |
| AKGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ALCD19y | 2.41E+00 | 4.72 | -0.544 | -0.363 | 0 | -0.826 | -0.636 | -0.49 |
| ALDD31_1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ARGSL | 0.114 | 0 | 0 | 0.0425 | 0 | 0 | 0.00E+00 | 8.67E-19 |
| ARGabc | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| ARGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ASPO1 | 29.7 | 24.3 | 0 | 21.5 | 0.083 | 20.1 | 0 | 2.41 |
| ASPT | 0 | 0 | 0.0291 | 0 | 0 | 0 | 0 | 0 |
| ASPTA | -30.8 | -26.1 | -1.25 | -21.9 | -0.762 | -21.1 | -1.29 | -2.7 |
| CDPDSP_BS | 0.033 | 0.0297 | 0.0235 | 0.0123 | 0.0224 | 0.0319 | 0.042 | 0.00952 |
| CITt10 | 0.06 | 0.00654 | 0 | 0 | 0 | 0.058 | 0.0763 | 0 |
| CITt14 | 0.00189 | 0.0017 | 0.00135 | 0.000705 | 0.00128 | 0.00183 | 0.0024 | 0.000545 |
| CITt15 | 0.0136 | 0.0838 | 0 | 0.0849 | 0 | 0.0119 | 0 | 0 |
| CITt2r | -0.0796 | -0.0957 | -2.29 | -0.0872 | -3.58 | -0.0724 | -7.01 | -0.00172 |
| CLPNS2_BS | 0.000295 | 0.000265 | 0.00021 | 1.10E-04 | 2.00E-04 | 2.85E-04 | 3.75E-04 | 8.50E-05 |
| CS | 17.2 | 16.3 | 4.88 | 8.94 | 5.41 | 14 | 10.4 | 7.63 |
| CYSS_2 | 0.0324 | 0.0289 | 0 | 0.0116 | 0.0187 | 0.0267 | 0.0333 | 0 |
| CYSTGL_1 | 0.00668 | 0.00598 | 0.0255 | 0.00506 | 0.00405 | 0.00909 | 0.00941 | 0 |
| CYSTS_2 | 0.00124 | 0.00126 | 0.0239 | 0.000953 | 0.00405 | 0.00582 | 0.00941 | 0 |
| ENO | 21.5 | 19 | 9.98 | 12.4 | 16 | 10.5 | 16.6 | 0.0696 |
| FRUK | 0 | 0.998 | 0.544 | 2.7 | 0.787 | 1.31 | 1.05 | 0.409 |
| FBA | 9.36 | 8.69 | 4.35 | 8.7 | 4.9 | 7.35 | 8.79 | 0.106 |
| FBA2 | 2.41 | 4.72 | -0.544 | -0.363 | 0 | -0.826 | -0.636 | -0.374 |
| FBP | 4.08E-13 | 0 | 0 | 0 | 0 | 0 | 0 | 0.303 |
| FEDCabc | 0.00204 | 0.00183 | 0.00138 | 0.000759 | 0 | 0.000345 | 0.00123 | 0.000586 |
| FRUpts | 2.41 | 5.72 | 0 | 2.34 | 0.787 | 0.488 | 0.415 | 0.0354 |
| FUM | 16.8 | 16.6 | 6.13 | 0.0604 | 2.43 | -0.131 | -0.134 | 6.72 |
| FUMt2r | 0 | 0 | -0.368 | 0 | 0 | -0.131 | -0.134 | -0.0357 |
| G6PDH2r | 0.0352 | 0 | 0 | 0 | 0 | 0 | 0.233 | 0 |
| GAPD | 21.8 | 18.9 | 10.1 | 12.6 | 16.2 | 10.8 | 16.9 | 0 |
| GAPDi_nadp | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00E+00 |
| GHMT2r | 0.0807 | 0.0603 | 0.00182 | 0.0482 | 0.0455 | 0.0648 | 0.0853 | 6.24E-05 |
| GLCNt2ir | 2.01 | 1.14 | 5.13 | 0.889 | 1.05 | 1.39 | 0.953 | 0 |
| GLCpts | 8.69 | 6.27 | 0.6 | 3.61 | 1.76 | 3.32 | 7.63 | 0 |
| GLUDxi | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GLUSy | 0.00104 | 0.000863 | 0 | 0 | 0 | 0.000617 | 0 | 0 |
| GLUt2r | 0.567 | 1.17 | 1.19 | 3.51 | 0.865 | 0.601 | 0 | 3.06 |
| GLYCt | -1.97 | -2.48 | 2.39 | 3.76 | 6.22 | 2.91 | 0.635 | 2.76 |
| GLYK | 0.443 | 2.25 | 1.84 | 3.4 | 6.22 | 2.08 | 0 | 2.27 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| GLYO1 | 0 | 0.839 | 0 | 0 | 0 | 0 | 0 | 0 |
| GNKr | 2.01 | 1.14 | 5.13 | 0.889 | 1.05 | 1.39 | 0.953 | 0 |
| HISTD | 0 | 0 | 0 | 0.018 | 0 | 0 | 0 | 0 |
| HISt2r | 0.0482 | 0.0433 | 0.0343 | 0 | 0.0327 | 0.0466 | 0.0613 | 0.0139 |
| HSTPTr | 0 | 0 | 0 | 0.018 | 0 | 0 | 0 | 0 |
| ICDHyr | 15.1 | 14.4 | 2.6 | 8.94 | 1.84 | 12.1 | -1.62 | 0.472 |
| ICITt10 | 0 | 0.0474 | 0.0427 | 0.0224 | 0.0407 | 0 | 0 | 0.0173 |
| ICITt2 | -2.17 | -1.89 | -0.0427 | -0.0224 | -0.0407 | -1.85 | -5.05 | -7.18 |
| LCADi | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LDH_L | -0.574 | -2.85 | -4.01 | -1.42 | -4.37 | 0 | -7.21 | -0.919 |
| L_LACt2r | -0.574 | -2.77 | -4.01 | -1.32 | -4.32 | 1.62 | -5.98 | -0.801 |
| MALt10 | 0 | 0.085 | 0 | 0.102 | 0.0505 | 1.62 | 1.23 | 0.118 |
| MALt2r | 0 | 1.36 | 0 | 8.3 | 0.807 | 24.9 | 19.7 | 1.88 |
| MALt4 | 0 | 0.00133 | 0 | 0.0016 | 0.000788 | 0.00971 | 0.0192 | 0.00184 |
| MDH | 16.8 | 18 | 6.13 | 8.47 | 3.29 | 26.4 | 15.8 | 8.34 |
| ME2 | 0 | 0 | 0.00E+00 | 0.00E+00 | 0 | 0.000356 | 4.98 | 0.376 |
| MCITL2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| OXGDC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PC | 1.6 | 0.00266 | 0 | 0.915 | 2.81 | 0 | 0 | 0 |
| PDH | 18.2 | 16.1 | 0 | 9.33 | 8.43 | 24.9 | 21.1 | 7.82 |
| PFK | 9.36 | 7.69 | 3.8 | 6 | 4.11 | 6.03 | 7.74 | 1.60E-12 |
| PGCD | 0.307 | 0.235 | 0.14 | 0.133 | 0.177 | 0.252 | 0.332 | 0.0463 |
| GND | 2.05 | 1.14 | 1.15 | -2.16 | 1.05 | -1.58 | 1.19 | -0.314 |
| PGI | 8.51 | 7.11 | 1 | 5.3 | 1.66 | 5.7 | 7.21 | -0.0428 |
| PGK_1 | 21.8 | 18.9 | 10.1 | 12.6 | 16.2 | 10.8 | 16.9 | 0 |
| PGM_1 | 21.5 | 19 | 9.98 | 12.4 | 16 | 10.5 | 16.6 | 0.0696 |
| PHETA1 | -0.104 | 0 | 0 | -0.0387 | 0 | -0.0671 | -0.00211 | 0 |
| PPCK | 0 | 0 | 0 | 0 | 0 | 11.4 | 4.19 | 0.416 |
| PPS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00E+00 |
| PRAGSr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PTAr | -0.284 | -0.255 | -2.61 | -0.106 | 2.29 | 9.91 | 9.17 | -0.0726 |
| PYK | 9.89 | 7 | 9.52 | 6.31 | 13.6 | 18 | 12.6 | 0.44 |
| PYRt2 | 1.16 | 1.17 | 0 | 0.049 | 0.689 | 4.32 | 4.93 | 8.26 |
| RPE | 1.22 | 0.756 | 0.768 | -1.5 | 0.817 | -1.08 | 0.782 | -0.212 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| RPI | -0.826 | -0.387 | -0.384 | 0.654 | -0.233 | 0.498 | -0.405 | 0.103 |
| SERAT | 0.0324 | 0.0289 | 0 | 0.0116 | 0.0187 | 0.0267 | 0.0333 | 0 |
| SERD_L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SHSL1_1 | 0.00544 | 0.00472 | 0.00159 | 0.0041 | 0 | 0.00327 | 0 | 0 |
| SHSL2 | 0.0681 | 0.0613 | 0.0255 | 0.026 | 0.0494 | 0.0704 | 0.0944 | 0 |
| SHSL4r | 0 | 0 | 0.00636 | 0 | 0 | 0.0131 | 0 | 0 |
| SUCCt2r | 0 | 0.351 | 1.93 | -12.2 | -0.809 | -13.9 | 0 | 3.35 |
| SUCD1 | 16.7 | 16.6 | 6.47 | 0 | 2.43 | 0 | 0 | 6.75 |
| SUCOAS | 16.6 | 16.2 | 4.51 | 12.2 | 3.19 | 13.8 | -0.0944 | 3.4 |
| TALA | 0.711 | 0.378 | 0.384 | -0.714 | 0.408 | -0.508 | 0.392 | -0.106 |
| THRD | 0.16 | 0.995 | 0.669 | 0.0416 | 0.118 | 0.168 | 0.221 | 0.0693 |
| THRD_L | 0.217 | 0.195 | 0.128 | 0.0785 | 0.148 | 0.194 | 0.275 | 0.0645 |
| THRS | 0.487 | 1.29 | 0.875 | 0.161 | 0.34 | 0.468 | 0.636 | 0.166 |
| TKT1 | 0.711 | 0.378 | 0.384 | -0.714 | 0.408 | -0.508 | 0.392 | -0.106 |
| TKT2 | 0.51 | 0.378 | 0.384 | -0.789 | 0.408 | -0.575 | 0.39 | -0.106 |
| TPI | 11.9 | 9.8 | 5.39 | 4.64 | 10.9 | 4.01 | 7.7 | 0 |
| TRPAS1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TRPS1 | 0.0321 | 0 | 0 | 0.012 | 0 | 0 | 0 | 0 |
| TYRTA | -0.0654 | 0 | 0 | -0.0244 | 0 | 0 | 0 | 0 |
| UNK5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*For full reaction name, see the publication of *B. subtilis* 168 (iYO844) model (King et al. 2016).

**Table A6:** Predicted flux from iMAT.

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| 2S6HCCi | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ACKr | 1000 | 1000 | 0 | 1000 | 0 | 1 | -1 | 0 |
| ACONT | 1 | 1.1 | 1 | 174 | 31 | 237 | 1 | 1 |
| ACOTA | -1 | -1 | 0 | -1 | 0 | -273 | -1 | -1 |
| AKGDH | 5 | 7.8 | 2 | 177 | 34.2 | 238 | 4 | 1 |
| AKGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ALCD19y | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| ALDD31_1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ARGSL | 1 | 1 | 1 | 1 | 217 | 1 | 1 | 1 |
| ARGabc | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ARGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ASPO1 | 642 | 1 | 1 | 590 | 0 | 149 | 504 | 1 |
| ASPT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ASPTA | -988 | -11.9 | -4 | -597 | -224 | -154 | -706 | -5 |
| CDPDSP_BS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CITt10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CITt14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CITt15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CITt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CLPNS2_BS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CS | 1 | 1.1 | 1 | 174 | 31 | 237 | 1 | 1 |
| CYSS_2 | 4 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| CYSTGL_1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CYSTS_2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ENO | 221 | 195 | 26.5 | 210 | 144 | 398 | 472 | 1 |
| FRUK | 0 | 4.72 | 0 | 0 | 0 | 0 | 0 | 0 |
| FBA | 1 | 5.72 | 1 | 1 | 1 | 1 | 2.32 | 1 |
| FBA2 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 5.72 |
| FBP | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| FEDCabc | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| FRUpts | 0 | 5.72 | 0 | 1 | 1 | 0 | 0 | 5.72 |
| FUM | 344 | 14.7 | 5 | 183 | 253 | 245 | 199 | 4 |
| FUMt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G6PDH2r | 8.63 | 6.63 | 1 | 3.63 | 2 | 11.3 | 8.63 | 1 |
| GAPD | 229 | 203 | 137 | 323 | 146 | 409 | 977 | 129 |
| GAPDi_nadp | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 59.2 |
| GHMT2r | 7.79 | 7.94 | 110 | 113 | 7 | 4 | 999 | 73.2 |
| GLCNt2ir | 0 | 0 | 5.13 | 0 | 0 | 0 | 0 | 0 |
| GLCpts | 7.63 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| GLUDxi | 1000 | 1 | 885 | 1 | 1 | 1000 | 1 | 1 |
| GLUSy | 1 | 1 | 1 | 264 | 1 | 1 | 1 | 1 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| GLUt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GLYCt | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| GLYK | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| GLYO1 | 6.58 | 6.94 | 106 | 112 | 6 | 1 | 998 | 73.2 |
| GNKr | 0 | 0 | 5.13 | 0 | 0 | 0 | 0 | 0 |
| HISTD | 1.79 | 4.94 | 1 | 1 | 1 | 1 | 1 | 1 |
| HISt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| HSTPTr | 1.79 | 4.94 | 1 | 1 | 1 | 1 | 1 | 1 |
| ICDHyr | 1 | 1.1 | 1 | 174 | 31 | 237 | 1 | 1 |
| ICITt10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ICITt2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LCADi | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LDH_L | 0 | 0 | 0 | 0 | -1 | -344 | -1000 | -1 |
| L_LACt2r | 0 | 0 | 0 | 0 | -1 | -343 | -975 | -1 |
| MALt10 | 0 | 0 | 0 | 0 | 0 | 1 | 24.5 | 0 |
| MALt2r | 0 | 0 | 0 | 0 | 0 | 24.5 | 1 | 0 |
| MALt4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| MDH | -1 | 11 | 3 | 181 | 253 | 242 | -1 | 1 |
| ME2 | 345 | 3.69 | 2 | 2.56 | 0 | 28.9 | 227 | 3 |
| MCITL2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| OXGDC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PC | 348 | 1 | 1 | 1 | 1.79 | 1 | 204 | 24.7 |
| PDH | 978 | 392 | 1 | 527 | 284 | 1 | 456 | 49.8 |
| PFK | 1 | 1 | 1 | 1 | 1 | 1 | 2.32 | 1 |
| PGCD | 11.8 | 8.94 | 110 | 114 | 7 | 5 | 1000 | 74.2 |
| GND | 6.79 | 6.63 | 6.13 | -1 | 2 | 11.3 | 8.63 | 1 |
| PGI | -1 | 1 | -1 | 4 | 1 | -11.3 | -1 | -1 |
| PGK_1 | 229 | 203 | 137 | 323 | 146 | 408 | 977 | 69.5 |
| PGM_1 | 221 | 195 | 26.5 | 210 | 144 | 398 | 472 | 1 |
| PHETA1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| PPCK | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 19.7 |
| PPS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PRAGSr | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2.84E-14 |
| PTAr | 1000 | 1000 | 0 | 1000 | 0 | 1 | -1 | 0 |

| Reaction | Glucose | Fructose | Gluconate | Glutamate/ Succinate | Glycerol | Malate | Malate/ Glucose | Pyruvate |
|---|---|---|---|---|---|---|---|---|
| PYK | 206 | 183 | 1 | 200 | 139 | 385 | 460 | 1 |
| PYRt2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| RPE | 2 | 0 | 0 | -3 | 0 | 4.63 | 3.31 | -3 |
| RPI | -4.79 | -6.63 | -6.13 | -2 | -2 | -6.63 | -5.31 | -4 |
| SERAT | 4 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| SERD_L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SHSL1_1 | 4 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| SHSL2 | 1 | 1 | 0 | 1 | 5 | 0 | 4 | 0 |
| SHSL4r | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| SUCCt2r | 0 | 0 | 0 | 3.35 | 0 | 3.35 | 0 | 0 |
| SUCD1 | 5 | 7.8 | 2 | 180 | 34.2 | 242 | 4 | 2 |
| SUCOAS | -1 | 4.8 | 221 | 174 | 1 | 237 | -1 | 1 |
| TALA | 3 | 1.23 | 5.13 | 1 | 1 | 5.63 | 4.31 | 2 |
| THRD | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| THRD_L | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| THRS | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| TKT1 | 3 | 1.23 | 5.13 | 1 | 1 | 5.63 | 4.31 | 2 |
| TKT2 | -1 | -1.23 | -5.13 | -4 | -1 | -1 | -1 | -4.99 |
| TPI | -536 | -1 | -736 | -1 | -1 | 1 | 2.32 | 6.72 |
| TRPAS1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TRPS1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TYRTA | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| UNK5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*For full reaction name, see the publication of *B. subtilis* 168 (iYO844) model (King et al. 2016).

**Table A7:** List of Pearson's correlation coefficients (r) between LPM-GEM flux prediction results and ^{13}C metabolic flux data.

| Reaction | r | P-value (adjusted for multiple testing) |
|---|---|---|
| Glucose | 0.51 | 0.16 |
| Fructose | 0.87 | 1.89E-02 |
| Gluconate | 0.93 | 4.94E-03 |
| Glycerol | 0.61 | 0.12 |
| Malate | 0.98 | 3.36E-04 |
| Pyruvate | 0.78 | 4.75E-02 |
| Glutamate | 0.76 | 5.16E-02 |
| Succinate | 0.3 | 0.27 |
| PGI | 0.43 | 0.19 |
| PFK | 0.54 | 0.16 |
| FBP | -0.14 | 0.66 |
| FBA | 0.49 | 0.16 |
| TPI | 0.89 | 1.62E-02 |
| GAPD | 0.78 | 4.75E-02 |
| GAPDi_nadp | NA | NA |
| PGK_1 | 0.73 | 5.72E-02 |
| PYK | 0.64 | 0.11 |
| G6PDH2r | 0.59 | 0.13 |
| GND | 0.52 | 0.16 |
| RPE | 0.48 | 0.16 |
| RPI | -0.57 | 0.95 |
| TKT1 | 0.5 | 0.16 |
| TKT2 | 0.45 | 0.18 |
| TALA | 0.5 | 0.16 |
| PDH | 0.8 | 4.04E-02 |
| CS | 0.6 | 0.13 |
| ICDHyr | 0.55 | 0.16 |
| AKGDH | 0.34 | 0.25 |
| SUCD1 | 0.4 | 0.21 |
| MDH | 0.39 | 0.21 |
| PPCK | 0.96 | 1.90E-03 |
| ME2 | -0.01 | 0.55 |
| PC | 0.23 | 0.33 |
| PTAr | 0.81 | 4.01E-02 |
| PGM_1 | 0.73 | 5.72E-02 |
| ENO | 0.73 | 5.72E-02 |
| ACONT | 0.67 | 9.09E-02 |
| FUM | 0.5 | 0.16 |
| ACKr | 0.81 | 4.01E-02 |
| SUCOAS | 0.34 | 0.25 |

*For full reaction name, see the publication of *B. subtilis* 168 (iYO844) model (King et al. 2016).

**Table A8:** List of Pearson's correlation coefficients (r) between flux prediction results from integrative metabolic analysis tool (iMAT) and $^{13}$C metabolic flux data.

| Reaction | r | P-value (adjusted for multiple testing) |
|---|---|---|
| Glucose | 0.42 | 0.47 |
| Fructose | 0.64 | 0.19 |
| Gluconate | 1 | 0 |
| Glycerol | 1 | 0 |
| Malate | 0.95 | 2.21E-03 |
| Pyruvate | NA | NA |
| Glutamate | 0.31 | 0.51 |
| Succinate | 0.65 | 0.19 |
| PGI | 0.26 | 0.56 |
| PFK | 0.44 | 0.47 |
| FBP | NA | NA |
| FBA | 0.5 | 0.38 |
| TPI | -0.29 | 0.86 |
| GAPD | 0.07 | 0.73 |
| GAPDi_nadp | 0.07 | 0.73 |
| PGK_1 | 0.07 | 0.73 |
| PYK | 0.81 | 5.57E-02 |
| G6PDH2r | 0.85 | 3.67E-02 |
| GND | 0.58 | 0.27 |
| RPE | 0.36 | 0.49 |
| RPI | -0.67 | 1 |
| TKT1 | 0.73 | 0.11 |
| TKT2 | -0.25 | 0.86 |
| TALA | 0.73 | 0.11 |
| PDH | -0.11 | 0.84 |
| CS | -0.3 | 0.86 |
| ICDHyr | -0.3 | 0.86 |
| AKGDH | 0.13 | 0.73 |
| SUCD1 | 0.38 | 0.49 |
| MDH | 0.31 | 0.51 |
| PPCK | -0.12 | 0.84 |
| ME2 | -0.08 | 0.84 |
| PC | 0.35 | 0.49 |
| PTAr | -0.35 | 0.86 |
| PGM_1 | -0.01 | 0.79 |
| ENO | -0.01 | 0.79 |
| ACONT | -0.3 | 0.86 |
| FUM | -0.18 | 0.86 |
| ACKr | -0.35 | 0.86 |
| SUCOAS | 0.13 | 0.73 |

*For full reaction name, see the publication of *B. subtilis* 168 (iYO844) model (King et al. 2016).

**Table A9:** Predicted flux from glucose to glucose plus malate shift.

| Reaction | 0 min | 5 min | 10 min | 15 min | 25 min | 45 min | 60 min | 90 min |
|---|---|---|---|---|---|---|---|---|
| 2S6HCCi | 0.000149 | 0.000154 | 0.000157 | 0.00016 | 0.000162 | 0.00017 | 0.000176 | 0.000186 |
| ACKr | -0.624 | -0.25 | -0.252 | -3.68 | 6.34 | 2.79 | 4.29 | 6.49 |
| ACONT | 6.96 | 9.67 | 8.47 | 6.15 | 2.1 | 5.19 | 7.24 | 6.93 |
| ACOTA | -0.0157 | -0.112 | -0.0363 | 0 | 0 | 0 | 0 | 0 |
| AKGDH | 9.46 | 12.1 | 10.9 | 6.89 | 3.91 | 4.82 | 5.53 | 7.08 |
| AKGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ALCD19y | 4.38 | 4.61 | 3.48 | 1.24 | 0.948 | 0.868 | -0.00033 | 1.43 |
| ALDD31_1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ARGSL | 0.0157 | 0.112 | 0.0363 | -0.0732 | -0.0641 | -0.000344 | 0 | 0 |
| ARGabc | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ARGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ASPO1 | 0 | 0 | 0.22 | 0 | 0 | 0 | 0 | 0 |
| ASPT | 0.433 | 0 | 0.00085 | 0.172 | 0 | 0 | 0 | 0 |
| ASPTA | -1.43 | -1.47 | -1.27 | -1.19 | -0.986 | -1.1 | -1.13 | -1.2 |
| CDPDSP_BS | 0.0314 | 0.0325 | 0.033 | 0.0336 | 0.0342 | 0.0358 | 0.037 | 0.0392 |
| CITt10 | 0 | 0 | 0 | 0 | 0.062 | 0 | 0.0199 | 0 |
| CITt14 | 0.00179 | 0.00186 | 0.00189 | 0.00192 | 0.00196 | 0.00205 | 0.00212 | 0.00224 |
| CITt15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CITt2r | -0.00566 | -0.00586 | -0.00596 | -0.00606 | -0.0682 | -0.00647 | -0.0266 | -0.00388 |
| CLPNS2_BS | 0.00028 | 0.00029 | 0.000295 | 3.00E-04 | 0.000305 | 0.00032 | 0.00033 | 0.00035 |
| CS | 6.96 | 9.67 | 8.47 | 6.15 | 2.1 | 5.19 | 7.24 | 6.93 |
| CYSS_2 | 0 | 0.00193 | 0 | 0 | 0 | 0 | 0.00315 | 0.00348 |
| CYSTGL_1 | 0.0319 | 0.0311 | 0.0354 | 0.0362 | 0.00936 | 0.0118 | 0.00527 | 0.00784 |
| CYSTS_2 | 0.0319 | 0.0311 | 0.0336 | 0.0342 | 0.0065 | 0.00991 | 0.000868 | 0.00666 |
| ENO | 17.4 | 17.2 | 17.2 | 16.6 | 17.7 | 15.3 | 15.9 | 15.7 |
| FRUK | 1.09 | 1.11 | 1.4 | 3.25 | 1.98 | 1.66 | 1.61 | 0.848 |
| FBA | 6.11 | 5.95 | 6.47 | 7.72 | 6.95 | 6.65 | 7.39 | 7.28 |
| FBA2 | 4.38 | 4.61 | 3.48 | 1.24 | 2.31 | 1.32 | 0.962 | 1.43 |
| FBP | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| FEDCabc | 0.00193 | 0.002 | 0.00204 | 0.00207 | 0.0021 | 0.00221 | 0.00228 | 0.00082 |
| FRUpts | 5.48 | 5.72 | 4.88 | 4.49 | 4.29 | 2.98 | 2.57 | 2.28 |

| Reaction | 0 min | 5 min | 10 min | 15 min | 25 min | 45 min | 60 min | 90 min |
|---|---|---|---|---|---|---|---|---|
| FUM | 9.4 | -5.91 | -10.7 | -11.9 | -13.9 | -11.4 | -9.75 | -12.3 |
| FUMt2r | -0.506 | -9.64 | -15 | -13.9 | -13.9 | -12.8 | -12.1 | -15.8 |
| G6PDH2r | 0.0182 | 0 | 0.0178 | 0.0142 | 0.226 | 0.857 | 0 | 0.0101 |
| GAPD | 17.7 | 17.4 | 17.5 | 16.8 | 16.6 | 15.1 | 15.1 | 16 |
| GAPDi_nadp | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GHMT2r | 0.0637 | 0.0794 | 0.0671 | 0.0141 | 0.0601 | 0.000235 | 0.0396 | 0.0656 |
| GLCNt2ir | 4.55 | 4.19 | 1.54 | 0.393 | 0.769 | 0 | 0 | 2.46 |
| GLCpts | 6.58 | 6.43 | 6.39 | 4.83 | 5.65 | 5.65 | 6.56 | 7.16 |
| GLUDxi | 0.000729 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GLUSy | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GLUt2r | 2.95 | 3.28 | 2.91 | 2.48 | 1.82 | 1.01 | 0.32 | 0.194 |
| GLYCt | -2.49 | -2.87 | -1.56 | -0.684 | -0.416 | -0.317 | 0 | -1.23 |
| GLYK | 1.9 | 1.74 | 1.92 | 0.561 | 0.532 | 0.552 | 0 | 0.196 |
| GLYO1 | 0 | 0.304 | 0 | 0 | 0 | 0 | 0 | 0 |
| GNKr | 4.55 | 4.19 | 1.54 | 0.393 | 0.769 | 0 | 0 | 2.46 |
| HISTD | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| HISt2r | 0.0458 | 0.0474 | 0.0482 | 0.049 | 0.0499 | 0.0523 | 0.0539 | 0.0572 |
| HSTPTr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ICDHyr | 6.96 | 9.36 | 8.47 | 4.81 | 2.1 | 3.12 | 4.57 | 5.42 |
| ICITt10 | 0.057 | 0.059 | 0.06 | 0.061 | 0 | 0.0651 | 0 | 0.0712 |
| ICITt2 | -0.057 | -0.367 | -0.06 | -1.41 | 0 | -2.13 | -2.67 | -1.58 |
| LCADi | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LDH_L | -8.82 | -6.27 | -8.09 | -7.39 | -5.65 | -5.04 | -4.8 | -7.61 |
| L_LACt2r | -8.82 | -5.18 | -6.8 | -6.25 | -4.55 | -3.87 | -3.63 | -6.34 |
| MALt10 | 0 | 1.09 | 1.29 | 1.15 | 1.1 | 1.17 | 1.17 | 1.28 |
| MALt2r | 0 | 17.5 | 20.6 | 18.3 | 17.6 | 18.7 | 18.7 | 20.4 |
| MALt4 | 0 | 0.0171 | 0.0201 | 0.0179 | 0.0172 | 0.0183 | 0.0183 | 0.0199 |
| MDH | 8.43 | 12.7 | 10.6 | 7.63 | 3.08 | 8.34 | 9.66 | 9.37 |
| ME2 | 0.973 | 0.000121 | 0.62 | 7.31E-05 | 1.71 | 0.17 | 0.468 | 0.000266 |
| MCITL2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| OXGDC | 0 | 0 | 0 | 0 | 0.0118 | 0.0125 | 0 | 0 |
| PC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PDH | 7.13 | 10.3 | 9.46 | 0 | 9.72 | 9.25 | 12.9 | 14.9 |
| PFK | 5.01 | 4.84 | 5.07 | 4.47 | 4.98 | 4.99 | 5.78 | 6.44 |

| Reaction | 0 min | 5 min | 10 min | 15 min | 25 min | 45 min | 60 min | 90 min |
|---|---|---|---|---|---|---|---|---|
| PGCD | 0.248 | 0.302 | 0.261 | 0.212 | 0.233 | 0.184 | 0.223 | 0.266 |
| GND | 0.652 | 0.355 | -1.16 | 0.407 | 0.62 | 0.857 | -0.587 | 0.86 |
| PGI | 6.42 | 6.28 | 6.23 | 4.67 | 5.27 | 4.64 | 6.39 | 6.97 |
| PGK_1 | 17.7 | 17.4 | 17.5 | 16.8 | 16.6 | 15.1 | 15.1 | 16 |
| PGM_1 | 17.4 | 17.2 | 17.2 | 16.6 | 17.7 | 15.3 | 15.9 | 15.7 |
| PHETA1 | -0.0985 | -0.102 | -0.104 | 0 | 0 | 0 | 0 | 0 |
| PPCK | 0.0376 | 1.53 | 1.06 | 0.289 | 0 | 2.06 | 1.29 | 1.24 |
| PPS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00E+00 |
| PRAGSr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PTAr | -0.624 | -0.25 | -0.252 | -3.68 | 6.34 | 2.79 | 4.29 | 6.49 |
| PYK | 5.04 | 6.16 | 6.61 | 7.41 | 7.69 | 8.69 | 7.98 | 7.47 |
| PYRt2 | 0.313 | 0.00184 | 0.79 | -0.198 | -2.15 | -1.29 | 2.07 | 7.67 |
| RPE | -0.769 | -0.879 | -0.834 | 0 | 0.363 | 0.563 | -0.391 | 0.474 |
| RPI | -1.42 | -1.23 | 0.322 | -0.407 | -0.256 | -0.293 | 0.196 | -0.386 |
| SERAT | 0 | 0.00193 | 0 | 0 | 0 | 0 | 0.00315 | 0.00348 |
| SERD_L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SHSL1_1 | 0 | 0 | 0.00173 | 0.002 | 0.00286 | 0.00186 | 0.0044 | 0.00118 |
| SHSL2 | 0.0954 | 0.0969 | 0.1 | 0.0342 | 0.0065 | 0.00991 | 0.0402 | 0.072 |
| SHSL4r | 0 | 0 | 0.0069 | 0.00799 | 0.222 | 0.148 | 0.00505 | 0.0047 |
| SUCCt2r | 0 | -8.45 | -6.66 | -4.92 | -3.92 | -3.42 | -3.18 | -3.6 |
| SUCD1 | 9.46 | 3.62 | 4.22 | 1.96 | 0 | 1.41 | 2.35 | 3.49 |
| SUCOAS | 9.36 | 12 | 10.8 | 6.84 | 3.68 | 4.66 | 5.48 | 7.01 |
| TALA | -0.304 | -0.34 | -0.332 | 0.0327 | 0.182 | 0.282 | -0.196 | 0.237 |
| THRD | 0.165 | 0.461 | 0.174 | 0.231 | 0.189 | 0.261 | 0.23 | 0.22 |
| THRD_L | 0.181 | 0.189 | 0.182 | 0.184 | 0 | 0.0834 | 0.24 | 0.253 |
| THRS | 0.45 | 0.758 | 0.465 | 0.526 | 0.303 | 0.464 | 0.593 | 0.604 |
| TKT1 | -0.304 | -0.34 | -0.332 | 0.0327 | 0.182 | 0.282 | -0.196 | 0.237 |
| TKT2 | -0.465 | -0.538 | -0.502 | -0.0327 | 0.182 | 0.282 | -0.196 | 0.237 |
| TPI | 12 | 11.9 | 11.5 | 9.16 | 9.43 | 8.13 | 7.95 | 8.48 |
| TRPAS1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TRPS1 | 0 | 0.0315 | 0 | 0 | 0 | 0 | 0 | 0 |
| TYRTA | -0.0621 | -0.0643 | -0.0654 | -0.0653 | 0 | 0 | 0 | 0 |
| UNK5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*For full reaction name, see the publication of *B. subtilis* 168 (iYO844) model (King et al. 2016).

**Table A10:** Predicted flux from malate to malate plus glucose shift.

| Reaction | 0 min | 5 min | 10 min | 15 min | 25 min | 45 min | 60 min | 90 min |
|---|---|---|---|---|---|---|---|---|
| 2S6HCCi | 0.000181 | 0.000197 | 2.00E-04 | 0.000202 | 0.000207 | 0.000213 | 0.000213 | 0.000197 |
| ACKr | 3.68 | 8.66 | 9.96 | 4.91 | 3.63 | 4.44 | 5.8 | 1.98 |
| ACONT | 19.8 | 9.48 | 5.01 | 3.54 | 5.35 | 6.3 | 4.24 | 6.45 |
| ACOTA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.0471 |
| AKGDH | 17.3 | 8.32 | 4.5 | 4.41 | 2.72 | 6.4 | 5.72 | 8.03 |
| AKGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ALCD19y | -9.66E-05 | -2.43 | -1.25 | -3.54 | 0.345 | -0.0993 | 1.94 | 2.73 |
| ALDD31_1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ARGSL | 0 | 0 | 0 | -0.0396 | 0 | 0 | 0 | 0.0471 |
| ARGabc | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ARGt2r | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ASPO1 | 9.42 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ASPT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ASPTA | -10.6 | -1.27 | -1.29 | -1.29 | -1.34 | -1.37 | -1.37 | -1.37 |
| CDPDSP_BS | 0.0381 | 0.0414 | 0.042 | 0.0426 | 0.0437 | 0.0448 | 0.0448 | 0.0414 |
| CITt10 | 0 | 0.0753 | 0.0436 | 0 | 0.0584 | 0 | 0 | 0 |
| CITt14 | 0.00218 | 0.00237 | 0.0024 | 0.00244 | 0.0025 | 0.00256 | 0.00256 | 0.00237 |
| CITt15 | 0.0731 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CITt2r | -0.078 | -0.0796 | -0.912 | -7.7 | -6.2 | -5.49 | -8.27 | -8.8 |
| CLPNS2_BS | 0.00034 | 0.00037 | 0.000375 | 3.80E-04 | 0.00039 | 4.00E-04 | 4.00E-04 | 0.00037 |
| CS | 19.8 | 9.48 | 5.88 | 11.2 | 11.5 | 11.8 | 12.5 | 15.2 |
| CYSS_2 | 0.0335 | 0.0374 | 0 | 0 | 0.0203 | 0.0367 | 0.0357 | 0.0356 |
| CYSTGL_1 | 0.00756 | 0.0103 | 0.0489 | 0.0469 | 0.026 | 0.00888 | 0.00989 | 0.00662 |
| CYSTS_2 | 0.00521 | 0.00479 | 0.0427 | 0.0433 | 0.0242 | 0.00888 | 0.00989 | 0.00662 |
| ENO | 9.59 | 6.63 | 8.97 | 9.91 | 15.1 | 22.1 | 20.9 | 20.5 |
| FRUK | 0 | 3.89 | 2.03 | 3.56 | 0 | 0.343 | 0 | 0 |
| FBA | 7.38 | 8.13 | 6.87 | 9.56 | 7.98 | 9.43 | 9.23 | 8.64 |
| FBA2 | 2.29 | -2.43 | -1.25 | -3.54 | 0.345 | 0.794 | 1.94 | 2.73 |
| FBP | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| FEDCabc | 0.00133 | 0.000983 | 0 | 0 | 0.000438 | 0.00209 | 0.00248 | 0.00198 |
| FRUpts | 2.29 | 1.46 | 0.78 | 0.0109 | 0.345 | 1.14 | 1.94 | 2.73 |

| Reaction | 0 min | 5 min | 10 min | 15 min | 25 min | 45 min | 60 min | 90 min |
|---|---|---|---|---|---|---|---|---|
| FUM | 3.41 | 1.72 | -2.42 | -0.168 | -4.28 | -9.71 | -7.22 | 1.94 |
| FUMt2r | 0 | -0.13 | -2.42 | -0.129 | -4.28 | -9.71 | -7.22 | -0.071 |
| G6PDH2r | 0 | 0 | 0.0138 | 0.0157 | 0.0153 | 0 | 0.0262 | 0.0224 |
| GAPD | 7.6 | 6.96 | 9.31 | 10.3 | 15.5 | 21.6 | 21.2 | 20.9 |
| GAPDi_nadp | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.58E-13 |
| GHMT2r | 0.0773 | 0.0841 | 0.0853 | 0.0945 | 0.0887 | 0.091 | 0.091 | 0.0841 |
| GLCNt2ir | 5.13 | 0.0319 | 0.0388 | 0.531 | 0.747 | 3 | 3.47 | 5.08 |
| GLCpts | 0 | 7.02 | 6.63 | 8.04 | 8.02 | 7.93 | 7.84 | 8.27 |
| GLUDxi | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GLUSy | 3.91 | 0.000213 | 0 | 0 | 0 | 0.000813 | 0 | 0 |
| GLUt2r | 1.24 | 1.29 | 0.715 | 0.427 | 0.354 | 0 | 0.0413 | 0.938 |
| GLYCt | 6.22 | 2.43 | 1.28 | 3.54 | 1.04 | 1.81 | -1.49 | -1.81 |
| GLYK | 6.22 | 0 | 0.0313 | 0 | 1.39 | 1.72 | 0.442 | 0.928 |
| GLYO1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GNKr | 5.13 | 0.0319 | 0.0388 | 0.531 | 0.747 | 3 | 3.47 | 5.08 |
| HISTD | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| HISt2r | 0.0556 | 0.0605 | 0.0613 | 0.0621 | 0.0638 | 0.0654 | 0.0654 | 0.0605 |
| HSTPTr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ICDHyr | 15.6 | 6.69 | 3.69 | 3.54 | 3.27 | 4.82 | 4.24 | 6.45 |
| ICITt10 | 0.0692 | 0 | 0.0327 | 0.0773 | 0.0209 | 0.0814 | 0.0814 | 0.0753 |
| ICITt2 | -4.24 | -2.78 | -1.35 | -0.0773 | -2.09 | -1.57 | -0.0814 | -0.0753 |
| LCADi | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LDH_L | 0 | -2.42 | -2.91 | -1.76 | -4.13 | -8.07 | -7.08 | -7.54 |
| L_LACt2r | 1.66 | -1.14 | -1.74 | -0.574 | -2.88 | -6.69 | -5.7 | -6.66 |
| MALt10 | 1.66 | 1.28 | 1.17 | 1.19 | 1.25 | 1.38 | 1.38 | 0.879 |
| MALt2r | 24.8 | 20.6 | 18.8 | 19 | 20 | 22.1 | 22.1 | 14.1 |
| MALt4 | 0.0259 | 0.0201 | 0.0184 | 0.0185 | 0.0195 | 0.0216 | 0.0216 | 0.0137 |
| MDH | 29.9 | 23.6 | 15.9 | 20 | 17 | 13.8 | 16.3 | 16.9 |
| ME2 | 0.000367 | 0.000377 | 1.65 | 3.78E-04 | 0.000351 | 0 | 0.000448 | 0.000171 |
| MCITL2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| OXGDC | 0 | 0 | 0.816 | 0.783 | 2.15 | 0 | 0.0992 | 0 |
| PC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PDH | 24.9 | 19.7 | 17.4 | 17.7 | 16.8 | 17.9 | 20 | 18.8 |
| PFK | 7.38 | 4.25 | 4.84 | 6 | 7.98 | 9.09 | 9.23 | 8.64 |

| Reaction | 0 min | 5 min | 10 min | 15 min | 25 min | 45 min | 60 min | 90 min |
|---|---|---|---|---|---|---|---|---|
| PGCD | 0.301 | 0.328 | 0.332 | 0.345 | 0.345 | 0.354 | 0.354 | 0.328 |
| GND | 2.84 | -3.5 | -2.02 | -2.23 | 0.763 | 2.83 | 3.5 | 1.86 |
| PGI | 5.99 | 6.83 | 6.43 | 7.84 | 7.81 | 7.72 | 7.61 | 8.06 |
| PGK_1 | 7.6 | 6.96 | 9.31 | 10.3 | 15.5 | 21.6 | 21.2 | 20.9 |
| PGM_1 | 9.59 | 6.63 | 8.97 | 9.91 | 15.1 | 22.1 | 20.9 | 20.5 |
| PHETA1 | 0 | 0 | 0 | -0.134 | -0.137 | -0.141 | -0.141 | -0.13 |
| PPCK | 8.98 | 12.8 | 8.76 | 7.48 | 4.16 | 0.669 | 2.38 | 0.274 |
| PPS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00E+00 |
| PRAGSr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PTAr | 3.68 | 8.66 | 9.96 | 4.91 | 3.63 | 4.44 | 5.8 | 1.98 |
| PYK | 16.2 | 10.9 | 10.2 | 8.83 | 10.4 | 13.3 | 13.1 | 9.3 |
| PYRt2 | 8.26 | 4.94 | 3.28 | 4.9 | 4.45 | 6.07 | 6.6 | 8.26 |
| RPE | 1.61 | -2.35 | -1.35 | -1.58 | 0.425 | 1.62 | 1.93 | 0.931 |
| RPI | -1.23 | 1.16 | 0.673 | 0.644 | -0.338 | -1.21 | -1.57 | -0.932 |
| SERAT | 0.0335 | 0.0374 | 0 | 0 | 0.0203 | 0.0367 | 0.0357 | 0.0356 |
| SERD_L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SHSL1_1 | 0.00235 | 0.00554 | 0.00614 | 0.00359 | 0.00182 | 0 | 0 | 0 |
| SHSL2 | 0.0823 | 0.0887 | 0.128 | 0.129 | 0.113 | 0.0995 | 0.101 | 0.0905 |
| SHSL4r | 0 | 0 | 0.236 | 0.223 | 0.183 | 0 | 0 | 0 |
| SUCCt2r | -13.9 | -6.47 | -5.32 | -5.19 | -4.87 | -6.4 | -5.82 | -6.07 |
| SUCD1 | 3.41 | 1.85 | 0 | 0 | 0 | 0 | 0 | 1.96 |
| SUCOAS | 17.2 | 8.23 | 4.13 | 4.05 | 2.42 | 6.3 | 5.62 | 7.94 |
| TALA | 0.805 | -1.17 | -0.673 | -0.683 | 0.324 | 0.91 | 1.04 | 0.572 |
| THRD | 0.2 | 0.218 | 0.221 | 0.216 | 0.23 | 0.236 | 0.236 | 0.218 |
| THRD_L | 0.251 | 0.271 | 0 | 0.019 | 0.0871 | 0.295 | 0.294 | 0.274 |
| THRS | 0.578 | 0.626 | 0.361 | 0.376 | 0.462 | 0.68 | 0.679 | 0.63 |
| TKT1 | 0.805 | -1.17 | -0.673 | -0.683 | 0.324 | 0.91 | 1.04 | 0.572 |
| TKT2 | 0.805 | -1.17 | -0.675 | -0.902 | 0.1 | 0.713 | 0.888 | 0.359 |
| TPI | -0.578 | 0 | 3.13 | 1.61 | 7.42 | 11.5 | 11.1 | 11.9 |
| TRPAS1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TRPS1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TYRTA | 0 | 0 | -0.00211 | -0.0842 | -0.0864 | -0.0552 | -0.0104 | -0.082 |
| UNK5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

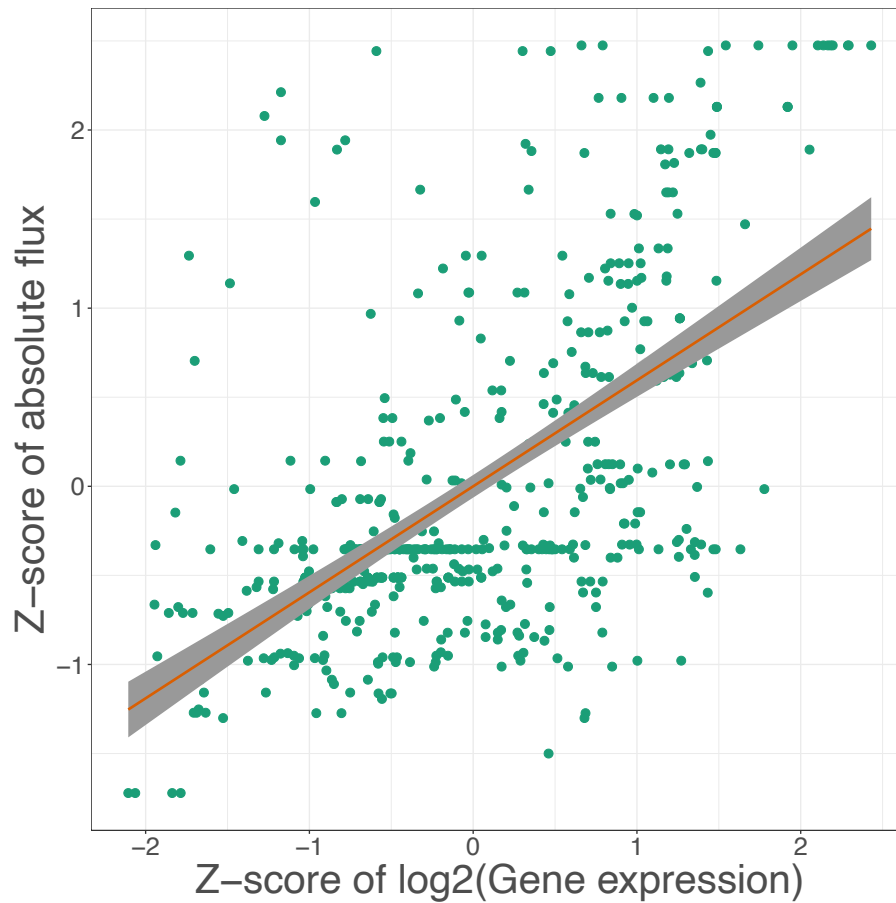*For full reaction name, see the publication of *B. subtilis* 168 (iYO844) model (King et al. 2016).

**Figure A1: The Scatterplot showing a relationship between $^{13}$C metabolic flux data and corresponding gene expression values.** Z-scores of the absolute values of the $^{13}$C tracer derived metabolic flux data from 40 reactions of the gold standard and corresponding gene expression values from the same carbon source conditions, and a linear regression line are plotted. The scatterplot shows a linear tendency, Pearson's correlation coefficient was significant (r = 0.59, p-value = 1.06E-53).

# Acknowledgements

First of all, I would like to express my gratitude to Prof. Dr. Rainer König for accepting me to his Network Modelling group and letting me pursue my doctoral degree here. His guidance, support, and attention were overwhelming. Also, I would like to thank Prof. Dr. Stefan Schuster for his help and advice when I needed it.

I would like to thank the rest of the thesis committee: Prof. Dr. Christoph Kaleta and Dr. Zita Soons, for kindly agreeing to review my thesis.

I would like to express my special thanks to Dr. Marcus Oswald and Franziska Hörhold for their support and helpful discussion during the research. It was a fantastic experience to work with both of you.

I would like to thank current and former members of the Network Modelling group for lively academic discussions during seminars, breaks, or retreats and all the fun leisure time.

I would like to thank my funding provider, Deutscher Akademischer Austauschdienst (DAAD), for giving me an invaluable opportunity to do the doctoral degree and provide support to ensure my comfortable stay in Germany. Additionally, I would like to thank Dr. Kitiporn Plaimas and Dr. Sittiporn Pattaradilokrat for their supports during the funding selection process.

Lastly, I would like to thank my family and friends for their supports during this chapter of my life. Thank you for always encouraging me to pursue my dream.

# Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass mir die Promotionsordnung der Medizinischen Fakultät der Friedrich-Schiller-Universität bekannt ist, ich die Dissertation selbst angefertigt habe und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben sind, mich folgende Personen bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts unterstützt haben: Prof. Dr. Rainer König und Prof. Dr. Stefan Schuster, die Hilfe eines Promotionsberaters nicht in Anspruch genommen wurde und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe und dass ich die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung nicht bei einer anderen Hochschule als Dissertation eingereicht habe.

München, 28.07.2021                                   Kulwadee Thanamit