

RESEARCH ARTICLE

# Rapid Prediction of Bacterial Heterotrophic Fluxomics Using Machine Learning and Constraint Programming

Stephen Gang Wu<sup>1</sup>, Yuxuan Wang<sup>2</sup>, Wu Jiang<sup>3</sup>, Tolutola Oyetunde<sup>1</sup>, Ruilian Yao<sup>4</sup>, Xuehong Zhang<sup>4</sup>, Kazuyuki Shimizu<sup>5</sup>, Yinjie J. Tang<sup>1\*</sup>, Forrest Sheng Bao<sup>6\*</sup>

**1** Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri, United States of America, **2** Department of Computer Science and Engineering, Ohio State University, Columbus, Ohio, United States of America, **3** Boxed Wholesale, Edison, New Jersey, United States of America, **4** State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, People's Republic of China, **5** Institute of Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan, **6** Department of Electrical and Computer Engineering, University of Akron, Akron, Ohio, United States of America

\* [yinjie.tang@seas.wustl.edu](mailto:yinjie.tang@seas.wustl.edu) (YJT); [forrest.bao@gmail.com](mailto:forrest.bao@gmail.com) (FSB)



**OPEN ACCESS**

**Citation:** Wu SG, Wang Y, Jiang W, Oyetunde T, Yao R, Zhang X, et al. (2016) Rapid Prediction of Bacterial Heterotrophic Fluxomics Using Machine Learning and Constraint Programming. *PLoS Comput Biol* 12(4): e1004838. doi:10.1371/journal.pcbi.1004838

**Editor:** Christos A. Ouzounis, Hellas, GREECE

**Received:** September 21, 2015

**Accepted:** March 1, 2016

**Published:** April 19, 2016

**Copyright:** © 2016 Wu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The dataset used to train MFlux is being updated all the time and it is downloadable via our project website <http://MFlux.org>. All other relevant data are in the manuscript and Supporting Information.

**Funding:** This work was funded by NSF DBI 1356669 [http://www.nsf.gov/awardsearch/showAward?AWD\\_ID=1356669](http://www.nsf.gov/awardsearch/showAward?AWD_ID=1356669). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Wu Jiang is an employee of Boxed Labs, whose business is not related to

## Abstract

<sup>13</sup>C metabolic flux analysis (<sup>13</sup>C-MFA) has been widely used to measure *in vivo* enzyme reaction rates (i.e., metabolic flux) in microorganisms. Mining the relationship between environmental and genetic factors and metabolic fluxes hidden in existing fluxomic data will lead to predictive models that can significantly accelerate flux quantification. In this paper, we present a web-based platform MFlux (<http://mflux.org>) that predicts the bacterial central metabolism via machine learning, leveraging data from approximately 100 <sup>13</sup>C-MFA papers on heterotrophic bacterial metabolisms. Three machine learning methods, namely Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), and Decision Tree, were employed to study the sophisticated relationship between influential factors and metabolic fluxes. We performed a grid search of the best parameter set for each algorithm and verified their performance through 10-fold cross validations. SVM yields the highest accuracy among all three algorithms. Further, we employed quadratic programming to adjust flux profiles to satisfy stoichiometric constraints. Multiple case studies have shown that MFlux can reasonably predict fluxomes as a function of bacterial species, substrate types, growth rate, oxygen conditions, and cultivation methods. Due to the interest of studying model organism under particular carbon sources, bias of fluxome in the dataset may limit the applicability of machine learning models. This problem can be resolved after more papers on <sup>13</sup>C-MFA are published for non-model species.

## Author Summary

Metabolic information is important for disease treatment, bioprocess optimization, environmental remediation, biogeochemical cycle regulation, and our understanding of life's

bacterial metabolism. All other authors declare no conflict of interest.

origin and evolution.  $^{13}\text{C}$ -MFA can quantify microbial physiology at the level of metabolic reaction rates. To speed up microbial characterizations and fluxomic studies, we hypothesize that genetic and environmental factors generate specific fluxome patterns that can be recognized by machine learning. Aided by constraint programming and quadratic optimization, our platform based on machine learning (ML) can predict meaningful metabolic information about bacterial species in their environments. Further, it can offer constraints to improve the accuracy of flux balance analysis. This study infers that the bacterial metabolic network has a certain degree of rigidity in allocating carbon fluxes, and different microbial species may share common regulatory strategies for balancing carbon and energy metabolisms. As a proof of concept, we demonstrate that the use of data-driven artificial intelligence (AI) approaches, e.g., ML, may assist mechanistic based models to elucidate the topology of microbial fluxomes.

## Introduction

With the advent of systems biology tools, such as genomics, transcriptomics, proteomics, and metabolomics during the last decade, the understanding of intracellular metabolisms from genotype to phenotype has been dramatically boosted. Notably,  $^{13}\text{C}$  metabolic flux analysis ( $^{13}\text{C}$ -MFA) enables the quantification of metabolic reaction rates *in vivo* [1]. It determines carbon metabolic fluxes using the mass isotopomer distribution (MID) of proteinogenic amino acids or free metabolites from  $^{13}\text{C}$  labeling experiments.  $^{13}\text{C}$ -MFA is considered as a reliable measurement of central metabolic reaction rates [2], which has demonstrated its power in discovering novel pathways [3, 4], validating gene functions [3], verifying engineered strains [5, 6], and revealing energy metabolism of host strains [7]. In the past decade, advanced parallel bioreactor systems, mass spectrometry, and computational tools resolving metabolic fluxes have been developed [8–11], which improved the precision of flux profiles [12] and extended  $^{13}\text{C}$ -MFA's application to the non-stationary metabolic phase [13, 14]. On the other hand, broad applications of  $^{13}\text{C}$ -MFA are still hindered because  $^{13}\text{C}$  experiments, biomass analysis, and flux calculations are expensive and time-consuming [15]. Moreover, some microbial systems may not be amenable to  $^{13}\text{C}$ -MFA if they require complex nutrients or their genome annotation is incomplete [16]. Before performing  $^{13}\text{C}$ -MFA on non-model species, laborious work is needed to examine extracellular metabolites, to characterize unknown pathways, and to analyze biomass compositions.

This study aims to employ an artificial intelligence (AI) approach called machine learning (ML) to investigate bacterial fluxomics patterns. ML is a powerful tool in systems biology [17] and has demonstrated successes in omics studies [18, 19]. For example, the precision of genome annotation on the model species *C. elegans* has been significantly enhanced by employing a simplified Support Vector Machine (SVM) method. Researchers have reached an accuracy of 75% on controversial genes [20]. At the transcriptomics level, ML approaches have been frequently used for disease identification. For instance, SVM has successfully recognized the gene expression patterns of hepatocellular carcinoma (HCC) [21], diffuse large B-cell lymphoma (DLBCL) [22] and ovarian cancer [23]. At the proteomics level, Supek *et al.* have employed a combined approach by integrating the Principal Component Analysis (PCA) method with SVM, to enhance analytic power in identifying “fingerprint” proteins (i.e., unique proteins in each tissue) from different horseradish tissues (leaf, teratoma, and tumor) grown *in vitro* [24]. In metabolomics, an SVM method can resolve the NMR data of metabolites in urine samples from different groups of people (healthy vs. pneumonia) [25]. In metabolic modeling,

Karp's group have adopted ML algorithms to predict the existence of various pathways for metabolic network reconstruction in different organisms [19].

The general idea of ML is to statistically build a numerical predictive *model* or an *estimator* which is a function  $f: X \mapsto y$  that maps a vector of numbers called the *feature vector* to a vector of numbers called the *target* or the *label*. In many cases, the target is a 1-D vector, or a scalar. One may consider the feature vector as the input and the target as the output of the model. If the target takes discrete values, we call the ML model a *classifier*. Otherwise, a *regressor*. A commonly used classifier is binary classifier, where the cardinality (size) of the target set  $y$  is 2, e.g.,  $y = \{+1, -1\}$ . In this paper, we build a regressor  $f: \mathbb{R}^n \mapsto \mathbb{R}$  for each flux, where  $\mathbb{R}$  stands for the set of all real numbers. In *supervised ML*, a pair of a feature vector and a target form a training *sample*. Given a finite set of  $N$  samples  $\{(X_1, y_1), \dots, (X_N, y_N)\}$ , an ML algorithm will find such a function, usually through solving a numerical optimization problem, to minimize the predictive error. Samples used to train a model form the *training set* while those for testing the performance form the *test set*. Given a new piece of data, numerically represented as a vector  $X_{new}$ , the model  $f$  will predict the target  $f(X_{new})$ , e.g., a flux value given reaction parameters where are represented by the vector  $X_{new}$  in this paper. The models learned through ML are usually not analytical models that can be represented using equations. Rather, they are numerical operators. For example, an artificial neural network (ANN) model can be represented by a series of weight and bias matrices, each of which is for one layer. A poor model can only predict well on the training set as if it only “remembers” the training samples, while a good model can learn the patterns among data and still be accurate on samples it has never “seen”. Hence, researchers make the training and test sets mutually exclusive. A mechanism called *cross validation* is used to ensure the mutual exclusiveness of training and test sets while making full use of the dataset.

A distinct advantage for ML applications is that they can reduce the need for costly experimental supplies and time-consuming benchwork. Despite the progress in utilizing ML methods in systems biology, there is no similar application in the fluxomics field to predict the flux profile. Therefore, we conceived the idea of integrating ML strategies with fluxomics research. To efficiently employ ML methods, a dataset with a sufficient number of samples is a prerequisite. Recently, a  $^{13}\text{C}$ -MFA dataset named CeCaFDB has been constructed, which includes more than 100 papers mostly on prokaryotic species [26]. Based on this dataset, five categorical and sixteen continuous features were initiated to describe the environmental and genetic factors involved in  $^{13}\text{C}$ -MFA of bacterial species. Unlike most omics projects employing ML approaches, this work built regressors rather than classifiers: 29 lumped central metabolic fluxes were adopted as the outputs to describe the central carbon metabolism of bacteria species. A 10-fold cross validation evaluated the performance of different algorithms. Furthermore, we included a knowledge-based system to check whether user inputs were biologically meaningful. Lastly, quadratic programming was employed to adjust the fluxes predicted by ML to satisfy stoichiometric constraints. Our web-based platform MFlux provides reasonable predictions for central metabolic flux profiles on 30 bacteria species, and it can be accessed online at <http://mflux.org> along with the training data. Although our platform is still in the early phase, our trial to integrate AI approaches with mechanistic models will have broad impacts on both systems biology and metabolic engineering fields.

## Methods

### Data collection

The dataset used to build MFlux are constructed from the literature. The total uptake rate of carbon sources is normalized as 100; all other fluxes are normalized based on the uptake rate of carbon sources. We obtained  $^{13}\text{C}$ -MFA information for bacterial species from the CeCaFDB

dataset and added a few recent papers (approximately 120 papers in total, as of January 2015). <sup>13</sup>C-MFA data related to photosynthetic bacteria was excluded in ML study because of their diverse CO<sub>2</sub> fixation pathways, light-sensitive fluxomes, and insufficient sampling sizes for ML. For photosynthetic species, MFlux currently only reports a general description of their fluxomic features based on corresponding references.

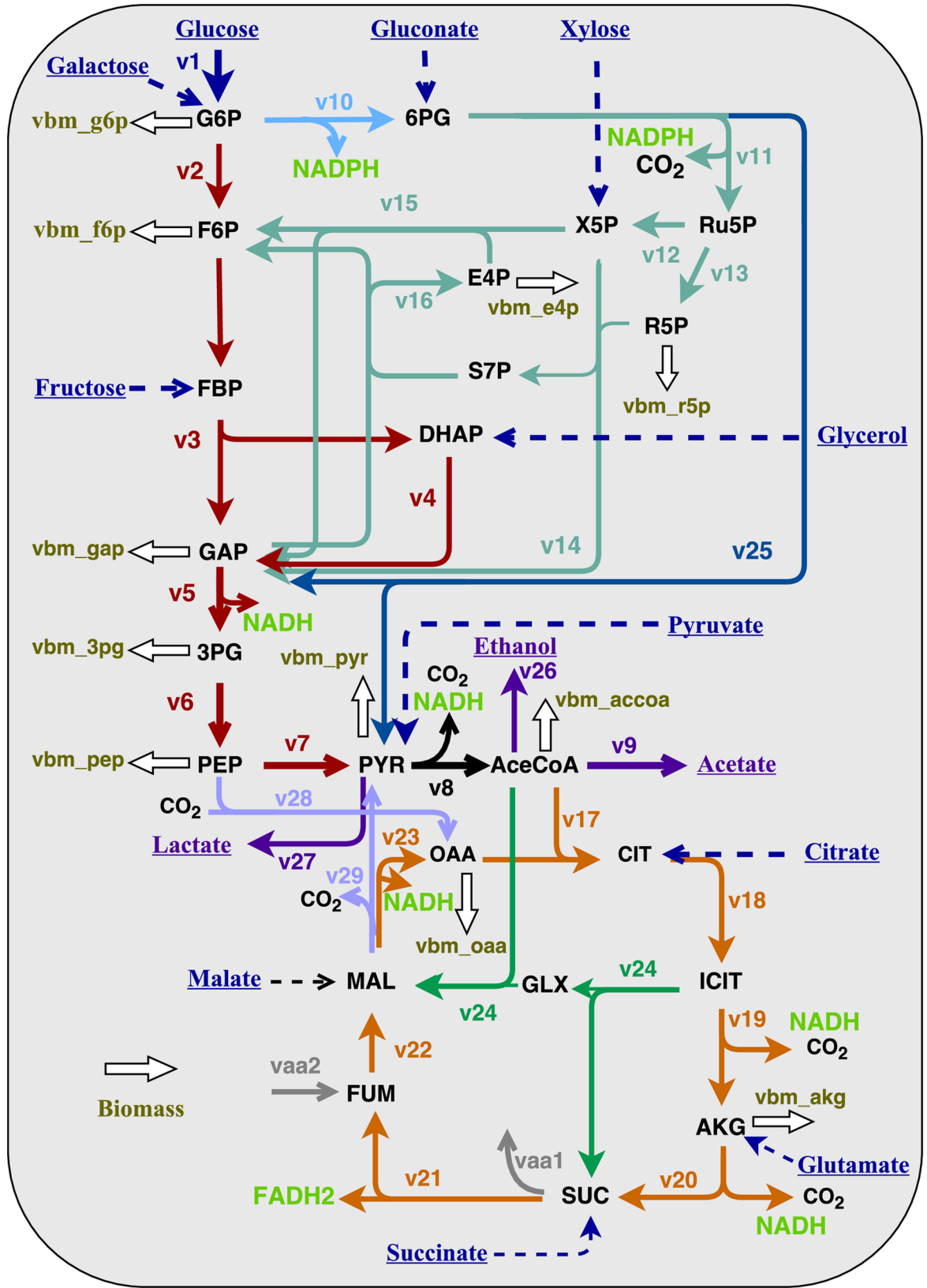
In heterotrophic microorganisms, interconversions between glycolysis metabolites (phosphoenolpyruvate and pyruvate) and TCA cycle metabolites (oxaloacetate and malate) involve a set of anaplerotic reactions (e.g., phosphoenolpyruvate carboxylase, phosphoenolpyruvate carboxykinase, pyruvate carboxylase, and malic enzyme) serving as a key switch point for carbon flux distribution in bacteria [27]. These reactions balancing both carbon and cofactors may be employed by different microbial species. For example, *E. coli* anaplerotic pathways involve phosphoenolpyruvate carboxylase and malic enzyme, while *Bacillus* species furnish pyruvate carboxylase (the pyruvate shunt). In the case of *Corynebacterium*, both phosphoenolpyruvate carboxylase and pyruvate carboxylase are functional [28, 29]. These anaplerotic pathways can re-route fluxes when central pathway such as pyruvate kinase is knocked out. To ease the ML efforts, the anaplerotic pathways were lumped into two routes that exchanges fluxes between the TCA cycle and the glycolysis nodes: ( $MAL \rightarrow PYR + CO_2$  and  $PEP + CO_2 \rightarrow OAA$ ). This simplification also considered the fact that <sup>13</sup>C-MFA has poor resolutions on anaplerotic fluxes because various combinations of these reactions could generate similar labeling patterns in amino acids [30].

## Feature vector for ML

As mentioned earlier, supervised ML builds models based on the samples, each of which is a pair of a feature vector and a target. Based on published <sup>13</sup>C-MFA methodologies and microbial physiologies, we proposed five categorical features: species, nutrient types, oxygen conditions, engineering method, genetic background, and cultivation methods. There were two considerations when choosing those features. First, genetic modifications can significantly reorganize fluxomes. To improve the predictability on mutant strains, our platform allows toggling on or off certain central pathways (by manually setting the flux boundaries) in engineered strains. Second, the factor of cultivation method aims to reveal fluxome differences between shake flask cultures (a pseudo-steady state approach) and bioreactor cultures (a well-controlled fermentation or chemostat cultivation). Meanwhile, we introduced sixteen continuous features: growth rate, substrate uptake rate, and the ratio of multiple substrate uptakes (glucose, fructose, galactose, gluconate, glutamate, citrate, xylose, succinate, malate, lactate, pyruvate, glycerol, acetate and NaHCO<sub>3</sub>, as shown in Fig 1). Since the features include both categorical and continuous ones, *one-hot encoders* were used to convert categorical feature values into real numbers. Each feature was then standardized into zero mean and unit variance as assumed by many ML approaches. For each predicted flux, or the target/label in ML terminology, we scaled it into the interval [0, 1] by the min-max method. In addition to the min-max method, we also tested unit-variance-zero-mean standardization for scaling flux values, and the result was quite similar.

## Machine learning algorithms

The problem of predicting fluxes was modeled as a regression problem in ML where a computer program learns from existing data to estimate continuous variables. Twenty-nine regressors were trained to predict the 29 fluxes. We tested three widely-applied ML algorithms, including k-nearest neighbors (k-NN), decision tree, and SVM. To ensure a fair comparison, we performed a grid search for the best parameter set of each algorithm. The detailed



**Fig 1. A universal central metabolic pathway for bacteria.** The central carbon metabolic pathway is simplified into 29 fluxes, used as the outputs of our model.

doi:10.1371/journal.pcbi.1004838.g001

parameter sets for 29 SVM-based regression models can be found from our web page. The programming language used for this project was Python 2.7 and the numpy and scikit-learn modules were utilized for machine learning [31]. Program files for training the models and testing them are wrapped in [S1 Program](#). Full version including web-end code is released under GNU GPL v3 at <https://bitbucket.org/forrestbao/influx>

## Model evaluation and cross validation

Considering the limited number of samples in the current dataset, we adopted a 10-fold cross validation. An  $N$ -fold cross validation works as follows. All samples in our dataset are spliced into  $N$  equal parts. In each iteration,  $N - 1$  parts are used as the training set, while the remaining as the test set. In the next iteration, the test set will be rotated to another part of the data, and the training set will consist of all other samples. This procedure will stop when all parts of the data have been incorporated into the test set exactly once, and training set exactly  $N - 1$  times. Finally, the accuracy of the model can be calculated by checking the prediction result for each sample. For each flux, the error in cross validation is computed using Mean Squared Error (MSE).

## Stoichiometric constraints and boundary

One unique feature of our method is incorporating the overall mass balance through central metabolic pathways. The stoichiometric equations in [Fig 1](#) under steady state are summarized as follows:

$$\text{G6P} : v_1 = v_2 + v_{10} + vbm_{g6p} \quad (1)$$

$$\text{F6P/FBP} : v_2 + v_{15} + v_{16} + 100 \cdot ratio_{fructose} = vbm_{f6p} + v_3 \quad (2)$$

$$\text{DHAP} : v_3 + 100 \cdot ratio_{glycerol} = v_4 \quad (3)$$

$$\text{GAP} : v_3 + v_4 + v_{14} + v_{15} + v_{25} = v_5 + v_{16} + vbm_{gap} \quad (4)$$

$$\text{3PG} : v_5 = v_6 + vbm_{3pg} \quad (5)$$

$$\text{PEP} : v_6 = v_7 + v_{28} + vbm_{pep} \quad (6)$$

$$\text{PYR} : v_7 + v_{25} + v_{29} + 100 \cdot ratio_{pyruvate} = v_8 + v_{27} + vbm_{pyr} \quad (7)$$

$$\text{AceCoA} : v_9 + v_{17} + v_{24} + v_{26} + vbm_{accoa} = v_8 \quad (8)$$

$$\text{Ru5P} : v_{11} = v_{12} + v_{13} \quad (9)$$

$$\text{R5P} : v_{13} = v_{14} + vbm_{r5p} \quad (10)$$

$$\text{E4P} : v_{15} + vbm_{e4p} = v_{16} \quad (11)$$

$$\text{S7P} : v_{14} = v_{16} \quad (12)$$

$$\text{X5P} : v_{12} + 100 \cdot ratio_{xylose} = v_{14} + v_{15} \quad (13)$$

$$6PG : v_{10} + 100 \cdot ratio_{gluconate} = v_{11} + v_{25} \quad (14)$$

$$CIT : v_{17} + 100 \cdot ratio_{citrate} = v_{18} \quad (15)$$

$$ICIT : v_{18} = v_{19} + v_{24} \quad (16)$$

$$AKG : v_{19} + 100 \cdot ratio_{glutamate} = v_{20} + vbm_{akg} \quad (17)$$

$$SUC : v_{20} + v_{24} + 100 \cdot ratio_{succinate} = v_{21} + v_{aa1} \quad (18)$$

$$FUM : v_{21} + v_{aa2} = v_{22} \quad (19)$$

$$MAL : v_{22} + v_{24} + 100 \cdot ratio_{malate} = v_{23} + v_{29} \quad (20)$$

$$OAA : v_{23} + v_{28} = v_{17} + vbm_{oaa} \quad (21)$$

Specifically,  $v_1$  represents the flux from carbon substrate (either glucose or galactose) to G6P since both glucose and galactose can be catabolized to G6P,  $v_{aa1}$  and  $v_{aa2}$  represent fluxes involved in biomass building block synthesis or extracellular products, while  $vbm$  represents carbon fluxes going to biomass from different precursors [32].

A series of linear constraints can be derived from the stoichiometric equations above and used to restrain fluxes predicted by the ML methods:

$$v_1 - 100 \cdot (ratio_{glucose} + ratio_{galactose}) = 0 \quad (22)$$

$$v_3 - v_4 + 100 \cdot ratio_{glycerol} = 0 \quad (23)$$

$$v_{11} - v_{12} - v_{13} = 0 \quad (24)$$

$$v_{14} - v_{16} = 0 \quad (25)$$

$$v_{10} - v_{11} - v_{25} + 100 \cdot ratio_{gluconate} = 0 \quad (26)$$

$$-v_{17} + v_{18} - 100 \cdot ratio_{citrate} = 0 \quad (27)$$

$$-v_{12} + v_{14} + v_{15} - 100 \cdot ratio_{xylose} = 0 \quad (28)$$

$$-v_{18} + v_{19} + v_{24} = 0 \quad (29)$$

$$-v_{22} + v_{23} - v_{24} + v_{29} - 100 \cdot ratio_{malate} = 0 \quad (30)$$

Among equations listed above, Eq 22 indicates the case for co-metabolism of both C6 sugars. Meanwhile, a list of inequality constraints can be drawn, given that all biomass fluxes are

non-negative:

$$v_1 - v_2 - v_{10} \geq 0 \quad (31)$$

$$v_2 - v_3 + v_{15} + v_{16} + 100 \cdot \text{ratio}_{\text{fructose}} \geq 0 \quad (32)$$

$$v_3 + v_4 - v_5 + v_{14} + v_{15} - v_{16} + v_{25} \geq 0 \quad (33)$$

$$v_5 - v_6 \geq 0 \quad (34)$$

$$v_6 - v_7 - v_{28} \geq 0 \quad (35)$$

$$v_7 - v_8 + v_{25} - v_{27} + v_{29} + 100 \cdot \text{ratio}_{\text{pyruvate}} \geq 0 \quad (36)$$

$$v_8 - v_9 - v_{17} - v_{24} - v_{26} \geq 0 \quad (37)$$

$$v_{13} - v_{14} \geq 0 \quad (38)$$

$$-v_{15} + v_{16} \geq 0 \quad (39)$$

$$v_{19} - v_{20} + 100 \cdot \text{ratio}_{\text{glutamate}} \geq 0 \quad (40)$$

$$-v_{17} + v_{23} + v_{28} \geq 0 \quad (41)$$

$$-v_{21} + v_{22} \geq 0 \quad (42)$$

Among all inequality constraints, [Eq 39](#) works well except for the case of *zwf* knockout, where the direction of [Eq 39](#) could be reversed [\[33\]](#).

## Flux adjustment using stoichiometric constraints

We adopted a quadratic programming method similar to minimization of metabolic adjustment (MOMA) [\[34\]](#), to tune fluxes to satisfy the stoichiometric constraints. The CVXOPT package for Python was employed here for quadratic programming [\[35\]](#). The optimization problem is modeled as

$$\begin{aligned} \text{Minimize } f(\mathbf{v}) &= \sum_{i=1}^{29} (\text{Scaled}(v_i) - \text{Scaled}(\hat{v}_i))^2 \\ \text{Subject to } \mathbf{S} \cdot \mathbf{v} &= 0, \\ \mathbf{A} \cdot \mathbf{v} &\geq 0, \end{aligned} \quad (43)$$

where the vector  $\hat{\mathbf{v}} = [\hat{v}_1, \dots, \hat{v}_{29}]$  is the flux values predicted by ML, the vector  $\mathbf{v} = [v_1, \dots, v_{29}]$  is the flux values to be solved in this optimization problem, the function  $\text{Scaled}(\cdot)$  using Min-Max scaling to scale all fluxes into the range  $[0, 1]$ , the matrix  $\mathbf{S}$  is obtained from all equality constraints from [Eq 22](#) to [Eq 30](#), and the matrix  $\mathbf{A}$  is obtained from all inequality constraints from [Eq 31](#) to [Eq 42](#). Notably, the biomass composition for a same species varies significantly under various conditions. Therefore, the quadratic programming loses mass balance constraints toward biomass synthesis. The purpose of scaling fluxes into the same range is to avoid the bias because fluxes have different dynamic ranges. The objective function  $f(\mathbf{v})$  can be



rewritten into a standard quadratic programming problem using the following steps:

$$\begin{aligned}
 f(\mathbf{v}) &= \sum_{i=1}^{29} (\text{Scaled}(v_i) - \text{Scaled}(\hat{v}_i))^2 = \sum_{i=1}^{29} \left( \frac{v_i - \text{Min}_i}{\text{Max}_i - \text{Min}_i} - \frac{\hat{v}_i - \text{Min}_i}{\text{Max}_i - \text{Min}_i} \right)^2 \\
 &= 2 \cdot \sum_{i=1}^{29} \left( \frac{1}{2} \frac{v_i^2}{(\text{Max}_i - \text{Min}_i)^2} + \frac{-1 \cdot v_i \cdot \hat{v}_i}{(\text{Max}_i - \text{Min}_i)^2} + \frac{1}{2} \frac{\hat{v}_i^2}{(\text{Max}_i - \text{Min}_i)^2} \right)
 \end{aligned} \tag{44}$$

where  $\text{Min}_i$  and  $\text{Max}_i$  are the range of the  $i$ -th flux. Since the last term  $\frac{1}{2} \left( \frac{\hat{v}_i}{\text{Max}_i - \text{Min}_i} \right)^2$  and the coefficient 2 are constants, they can be omitted from the objective function. Hence, Eq 43 can be rewritten in standard quadratic programming form as

$$\begin{aligned}
 \text{Minimize } f(\mathbf{v}) &= \frac{1}{2} \sum_{i=1}^{29} \frac{(v_i)^2}{(\text{Max}_i - \text{Min}_i)^2} + \sum_{i=1}^{29} \frac{-1 \cdot v_i \cdot \hat{v}_i}{(\text{Max}_i - \text{Min}_i)^2} \\
 \text{Subject to } \mathbf{S} \cdot \mathbf{v} &= 0, \\
 \mathbf{A} \cdot \mathbf{v} &\geq 0.
 \end{aligned} \tag{45}$$

For the upper and lower boundaries of each flux, i.e.,  $\text{Max}_i$  and  $\text{Min}_i$ , we used the maximal and minimal values observed in multiple datasets as the default values. Users can manually set desired values for the upper/lower bound of any specific flux in MFlux webpage, or they can opt to not use any boundaries. For instance, users can simply set the boundary of a certain flux as zero if this specific gene is knocked out.

## Constraint programming and input checking

To ensure user inputs, e.g., growth rates, oxygen usage, and substrate uptake rates, are biologically meaningful, MFlux first checks the satisfiability (e.g., whether cell growth rate is realistic) of input values [36]. The biological meaningfulness is represented using constraint programming [37], where each input is treated as a variable of a given domain. A set of inputs lacking of biological meaning will cause those constraints to be unsatisfied and MFlux will report an error message to warn the user. The Python module python-constraint [38] is used as the constraint solver.

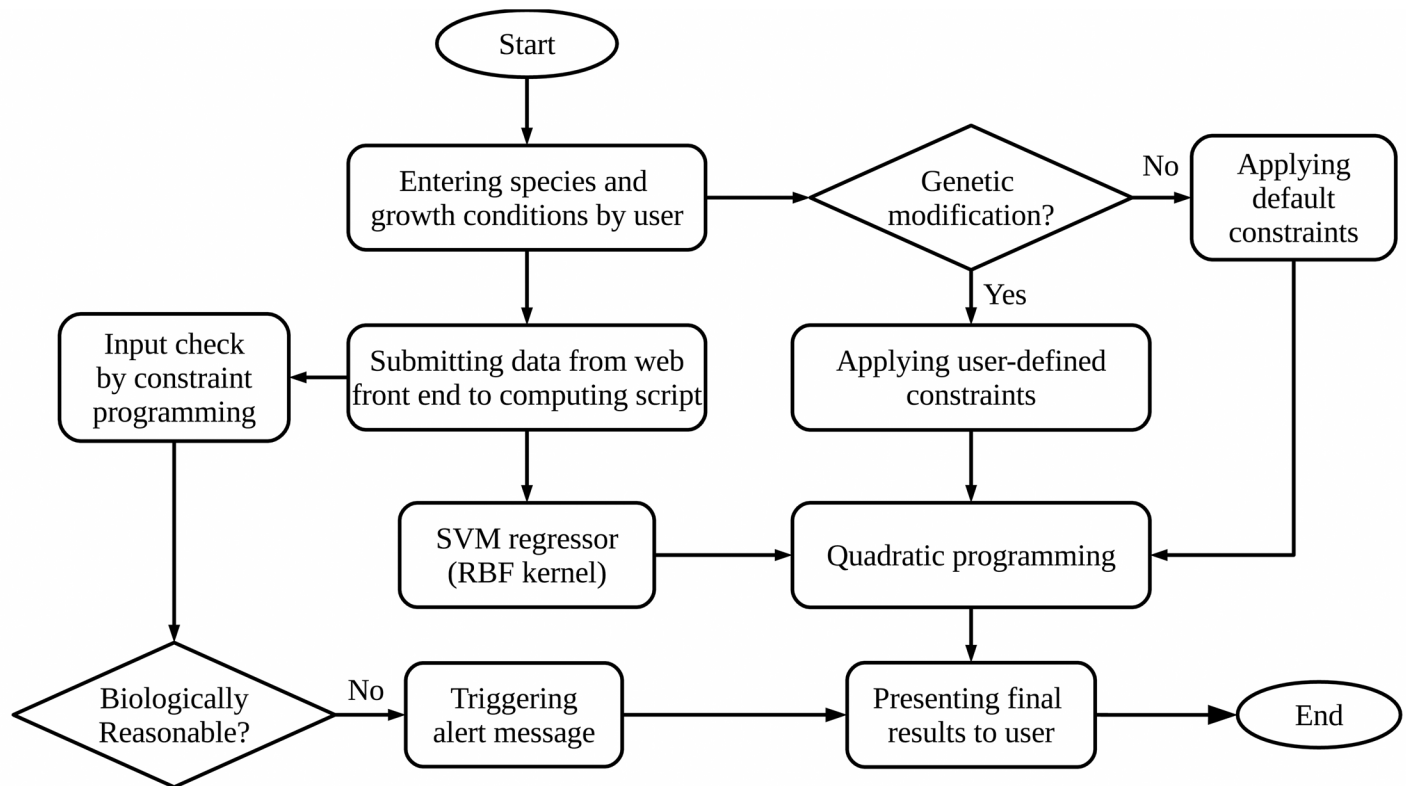
## Overall system design

Different parts of MFlux mentioned above are put together as illustrated in Fig 2. The prediction on 29 fluxes is done via an RBF-kernel SVM, whose outcome will be finalized by quadratic programming. Users can set boundary constraints to represent information about genes that are knocked out on the species, and such information will be used in quadratic programming. If parameters set by the user are not biologically meaningful, a warning message will be displayed. In the future, users will also have the option to enter flux constraints and settings of their own experiment to improve the prediction accuracy of MFlux.

## Results

### Pathway map and statistical analysis

The core metabolism of bacteria is summarized into a pathway map in Fig 1. Considering the availability of information, 29 major fluxes with 14 potential substrates were used to represent a universal heterotrophic carbon metabolism for non-photosynthetic bacteria species, which includes glycolysis, the tricarboxylic acid (TCA) cycle, the pentose phosphate (PP) pathway,

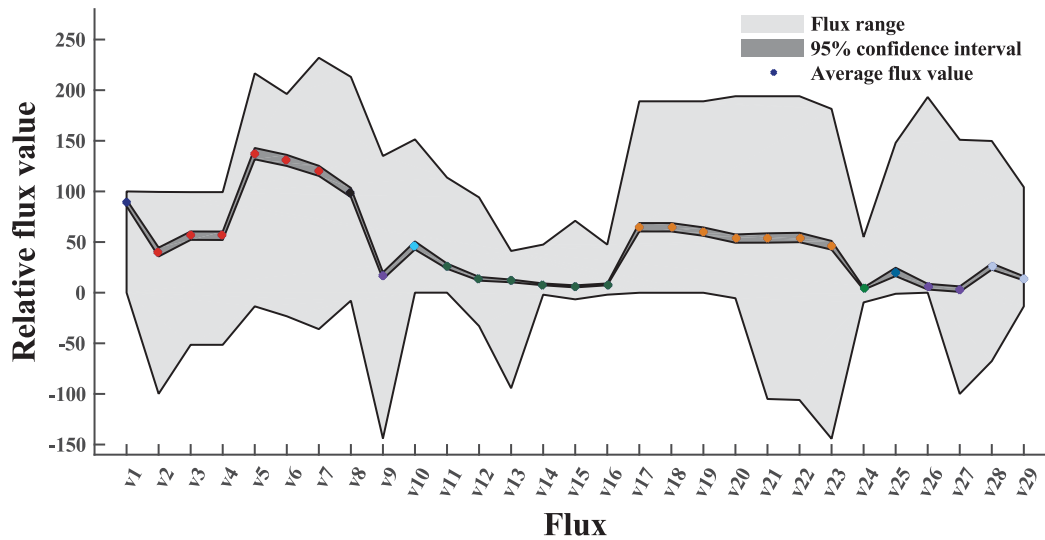


**Fig 2. The flowchart of MFlux algorithm.**

doi:10.1371/journal.pcbi.1004838.g002

the Entner–Doudoroff (ED) pathway, the glyoxylate shunt and the anaplerotic pathway. It is difficult for  $^{13}\text{C}$ -MFA to precisely resolve the anaplerotic pathway fluxes [39]. Information on the anaplerotic pathway is either incomplete or not precise in many publications in our dataset. Consequently, we simplified the anaplerotic pathway into two reversible fluxes. Similarly, we ignored several overflow fluxes which occasionally appear in  $^{13}\text{C}$ -MFA anaerobic metabolisms (e.g., the secretion of formate, butyrate, or pyruvate), because of lacking sufficient samples for machine learning. The omission of those fluxes can also partially explain the high prediction error in some fluxes (e.g.,  $v_8$ : Pyruvate  $\rightarrow$  Acetyl-CoA).

By statistical analysis, we determined the variation between each flux profile and the average flux profile from our  $^{13}\text{C}$ -MFA dataset. The average value, the range, and the 95% confidence interval for each flux are shown in Fig 3. The most conservative fluxes from our dataset include the non-oxidative pentose phosphate pathway and the glyoxylate shunt. The former pathway supplies precursors for bio-synthesizing amino acids (i.e., histidine, phenylalanine, and tyrosine) and nucleotides. The latter acts as an alternative carbon reserving path to the TCA cycle and is inhibited by the presence of glucose (most  $^{13}\text{C}$ -MFA is based on the glucose metabolism). All 29 fluxes are found to have a relatively narrow confidence interval compared to possible flux ranges, suggesting that fluxes of different bacteria species varies in a relatively small range. This is because most  $^{13}\text{C}$ -MFA studies are focusing on model species (e.g., *E. coli* and *B. subtilis*) and glucose based metabolism, while there are much less MFA efforts to study non-model species or metabolism of carbon substrates other than sugars (i.e., bias of fluxome research across).

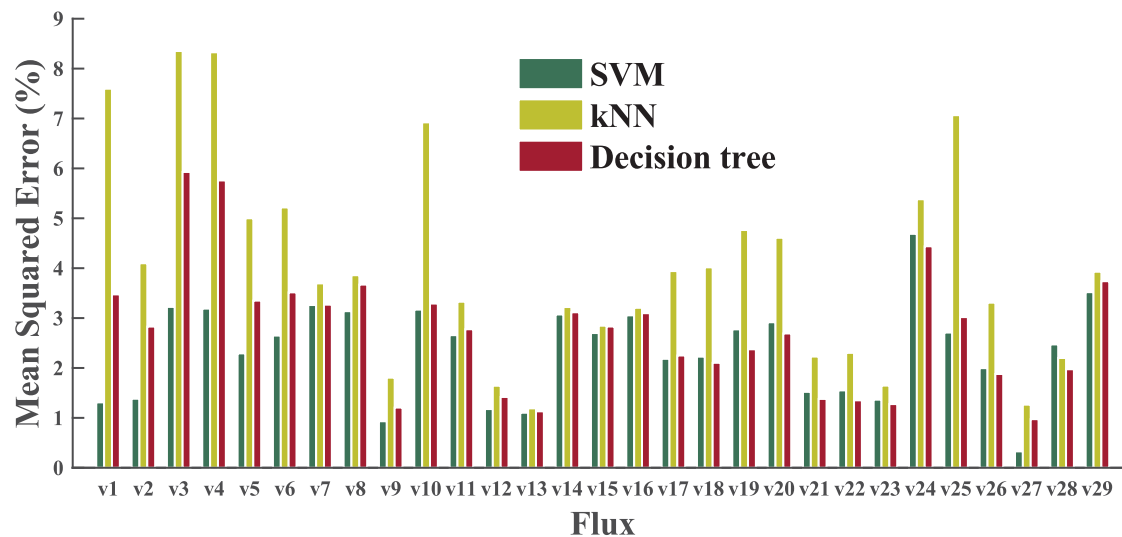


**Fig 3. Overview of central metabolic fluxes collected in our dataset.** “Flux range” represents the variation of each flux in the <sup>13</sup>C-MFA dataset. “95% confidence interval” indicates that 95% of flux data were within a small range. “Average flux value” is the average value in each flux based on all data in our <sup>13</sup>C-MFA dataset.

doi:10.1371/journal.pcbi.1004838.g003

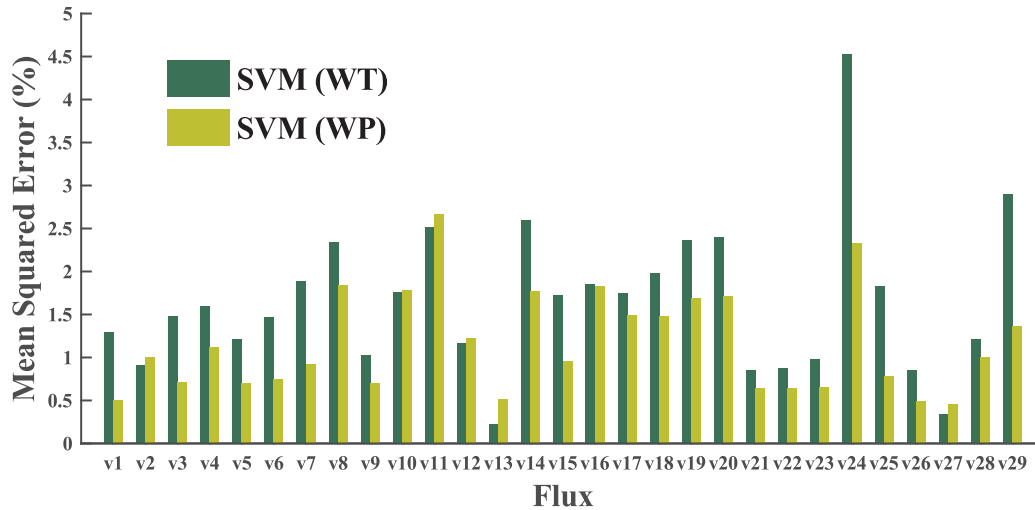
### Optimization of algorithms and parameters

To decide the most suitable ML algorithm, we first performed a grid search in the parameter space, using a dataset of wild type (WT) samples only. The best results of three different algorithms (for SVM, linear kernel only here) are presented in Fig 4. SVM makes better predictions than either the decision tree or k-NN on most fluxes. After this step, we carried out a second round of grid search to optimize parameters and improve the performance of SVM on the whole phenotype (WP) dataset (both WT and engineered). Both the linear kernel and radial bias function (RBF) kernel were included in this round of grid search.



**Fig 4. A comparison of three ML algorithms: SVM, k-NN, and decision tree.** The best cross-validation results on 29 fluxes are compared. All tests in this step were performed on the WT dataset only.

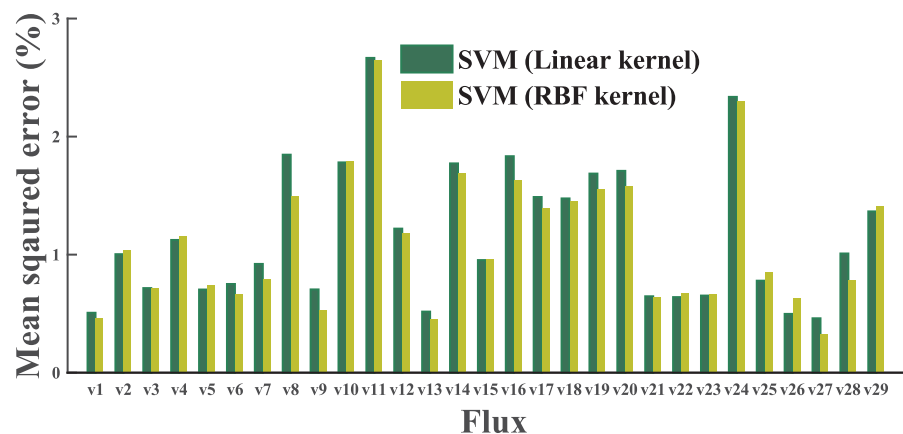
doi:10.1371/journal.pcbi.1004838.g004



**Fig 5. Best results by SVM on WT and WP datasets.** Grid searches are performed on both linear and RBF kernels. The results from WP dataset are much better than those from the WT dataset. The result indicated that the size of the dataset is an important factor affecting the predictive power of machine learning models.

doi:10.1371/journal.pcbi.1004838.g005

Better cross validation was expected from the SVM models trained on the WT dataset, rather than on the WP dataset, while sophisticated genetic variations are not included in the WT dataset. However, cross-validation results refuted our initial thought: the models from the WP dataset demonstrated significantly better performance than those trained on the WT dataset (data shown in Fig 5). This result can be interpreted as that the size of the training set is a major factor affecting the model quality, especially when the training set is relatively small (the sizes of WT and WP datasets are about 150 and 450 samples, respectively). We also compared the SVM results using the linear kernel with those using the RBF kernel, and the RBF kernel showed slightly better performance (Fig 6). The parameter set producing the most accurate cross-validation result was used to configure MFlux. Notably, prediction on  $v_{11}$  (the second step of the oxidative PP pathway) and  $v_{24}$  (the glyoxylate shunt) have relatively large variations. Two factors may contribute to this fact. Both  $v_{11}$  and  $v_{24}$  have relatively narrow ranges (see



**Fig 6. A comparison between linear-kernel SVM and RBF-kernel SVM.** The best cross-validation results of linear kernel and RBF kernel after grid searches on WP dataset are very similar. The RBF kernel is employed in the final model for flux prediction.

doi:10.1371/journal.pcbi.1004838.g006

Fig 1) and consequently even small numerical variations will generate larger relative errors for both fluxes. Meanwhile, genetic modifications may influence both  $v_{11}$  (e.g., *zwf* knockout [40]) and  $v_{24}$  (e.g., *ppc* knockout [41]) significantly. For instance, knocking out *zwf* in *E. coli* will cause a zero flux in  $v_{10}$  (the oxidative pentose phosphate pathway, OPP pathway) [42]. However, the lack of sufficient information on flux re-organization mechanisms in engineered microbes reduces ML predictability. This is because most engineered microbial fluxomics studies are focused on a few model species such as *E. coli*. To resolve this problem, the MFlux platform allows the users to manually set the boundaries of central fluxes to improve prediction quality (e.g., setting a zero flux through the OPP pathway for *E. coli zwf* mutant).

### Flux correction by quadratic programming

After parameter optimization, the SVM models of the best parameter sets can predict with relatively small error. However, the flux profile predicted by the ML method does not necessarily satisfy the inherent stoichiometric constraints of metabolic networks because the ML methods do not have big enough dataset at this stage to reflect this. The situation could get even worse where specific fluxes predicted by the ML algorithm may go beyond a biologically meaningful range (e.g., the predicted glyoxylate shunt flux  $v_{24}$  may have a negative value). To address those issues, we employed quadratic programming for flux correction as described in the Methods section. More rational results with improved accuracy are expected after flux correction. An essential assumption of this step is that ML predictions are relatively close to real values reported in the literature. This hypothesis is backed by our cross-validation results further validated in the following case studies.

### Case studies

To demonstrate the functionality of MFlux, we carried out tests on 20 cases, and the results are illustrated in Fig 7. Brief information for each case is listed in Table 1, and comprehensive results are included in S1 and S2 Tables. In general, MFlux can achieve decent flux predictions. Here we demonstrate two cases which are Cases 8 and 16.

In Case 8, *B. subtilis* strain uptakes the mixed substrates succinate and glutamate. To illustrate mixed substrates co-metabolisms, we tested MFlux with  $^{13}\text{C}$ -MFA data of *B. subtilis* reported by Chubukov *et al.* [44]. Microbial fermentation fed with multiple substrates of low price is promising for the biotechnology industry. However, there are very few quantitative analyses of this topic. In this test, we adopted the same set of parameters found in the literature

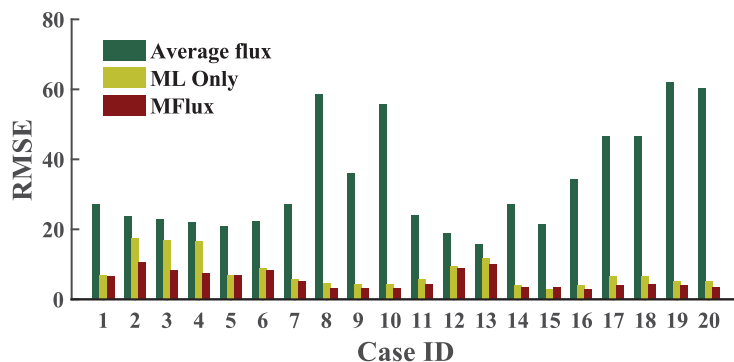


Fig 7. Summary of root mean squared error (RMSE) from 20 case studies: averaged flux from  $^{13}\text{C}$ -MFA dataset, ML-only, and MFlux (ML + quadratic programming). The average RMSE is 7.7 from ML-only, and 5.6 from MFlux. Detailed information is in S1 and S2 Tables.

doi:10.1371/journal.pcbi.1004838.g007

**Table 1. Summary of 20 cases of study.** Glc, glucose; Xyl, xylose; Lac, lactate; Ace, acetate; KO, knockout.

Species	Carbon source	Oxygen condition	Reactor	Genetic background	Case
<i>E. coli</i>	Glc	aerobic	tube	WT	1 [12]
<i>E. coli</i>	Glc	aerobic	baffled shake flask	<i>ppc</i> KO	2–4 [41]
<i>B. subtilis</i>	Glc	aerobic	shake flask, CSTR	WT, <i>spo0A</i> KO	5–7 [43]
<i>B. subtilis</i>	Multiple	aerobic	shake flask	mutant	8–11 [44]
<i>C. glutamicum</i>	Glc	aerobic	shake flask	WT	12 [45]
<i>C. glutamicum</i>	Glc	aerobic	shake flask	mutant	13 [46]
<i>P. denitrificans</i>	Glc	aerobic, microaerobic	fermentor	WT	14, 15 [47]
<i>G. thermoglucosidasius</i>	Glc	microaerobic	shake flask	WT	16 [28]
<i>Thermoanaerobacter sp.</i>	Xyl	anaerobic	batch (closed)	WT	17, 18 [48]
<i>D. vulgaris</i>	Lac	anaerobic	batch (closed)	WT	19 [49]
<i>G. metallireducens</i>	Ace	anaerobic	batch (closed)	WT	20 [3]

doi:10.1371/journal.pcbi.1004838.t001

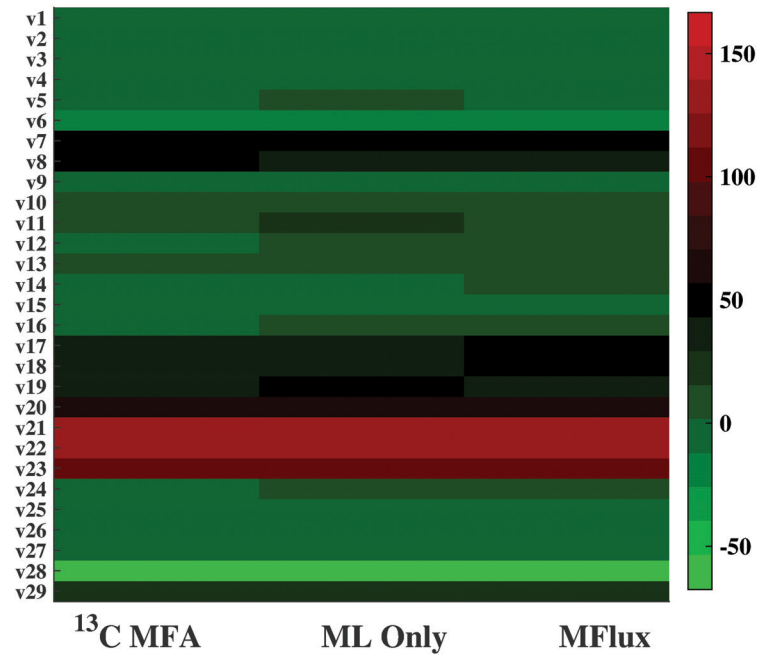
(S1 Table, Case 8) as the inputs of MFlux. For flux correction, we directly took the default boundary settings for quadratic programming. A comparison of flux profiles reported by <sup>13</sup>C-MFA, predicted by ML only, and predicted by MFlux (i.e., ML + quadratic programming) is illustrated in Fig 8. ML-only approach and MFlux accurately predict on most fluxes, closely matching the <sup>13</sup>C-MFA flux profiles with Root Mean Squared Error (RMSE) under 5. For ML, the predictions have large variation on specific fluxes (e.g.,  $v_{11}$ —oxidative PP pathway and  $v_{19}$ —TCA cycle). Quadratic programming can further adjust flux profiles and reduce deviations of flux predictions. The corrected flux profiles also meet the basic stoichiometric relationship of the metabolic network. The final prediction from MFlux shows improvement with RMSE reduces to 3.2.

In Case 16, *G. thermoglucosidasius* strain M10EXG grows under microaerobic conditions. *G. thermoglucosidasius* is a thermophilic and ethanol tolerant bacterium which can convert both hexose and pentose into ethanol [28]. To predict its central fluxomes, the parameter set used is listed in S1 Table, along with the default boundary settings for flux correction. A heat map (Fig 9) visualizes <sup>13</sup>C-MFA fluxes with ML-only fluxes and MFlux results. The results are encouraging: ML-only prediction gives an RMSE of 4.0, while MFlux uses both ML and quadratic programming to improve the prediction to an RMSE of only 3.0. Among the 20 case studies, the average flux set has very large variations (RMSE of 33.5) from actual <sup>13</sup>C-MFA fluxes (S2 Table). In this case, MFlux reduces the deviations of predicted fluxes from <sup>13</sup>C-MFA values.

For species with genetic modifications in major pathways (Cases 2, 3, 4, 12, and 13, *E. coli* and *C. glutamicum*), MFlux predictions have an RMSE between 5 and 10, higher than the RMSE for prediction of wild type strains. Since MFlux is currently unable to capture complex regulatory mechanisms of flux reorganization, human-computer interaction can be employed by manually tuning boundary values of certain fluxes to improve flux prediction quality. For example, knocking out *ppc* on *E. coli* may activate the glyoxylate shunt [41, 42]. The users can assign a non-zero lower boundary of the glyoxylate shunt when running MFlux.

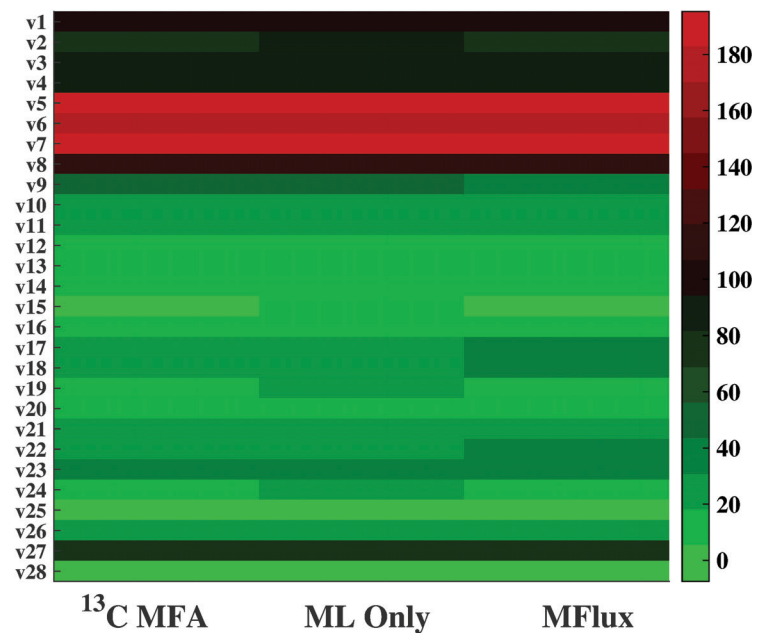
## Improving flux balance analysis of microbial metabolism via MFlux

Stoichiometry-based flux balance analysis (FBA) is an important mechanistic tool to predict unknown cell metabolism [50]. Accurate FBA prediction relies highly on setting the objective function and the flux constraints appropriately (based on thermodynamics or experimental analysis). Here, we compare FBA with MFlux for predicting *E. coli* metabolisms. The latest



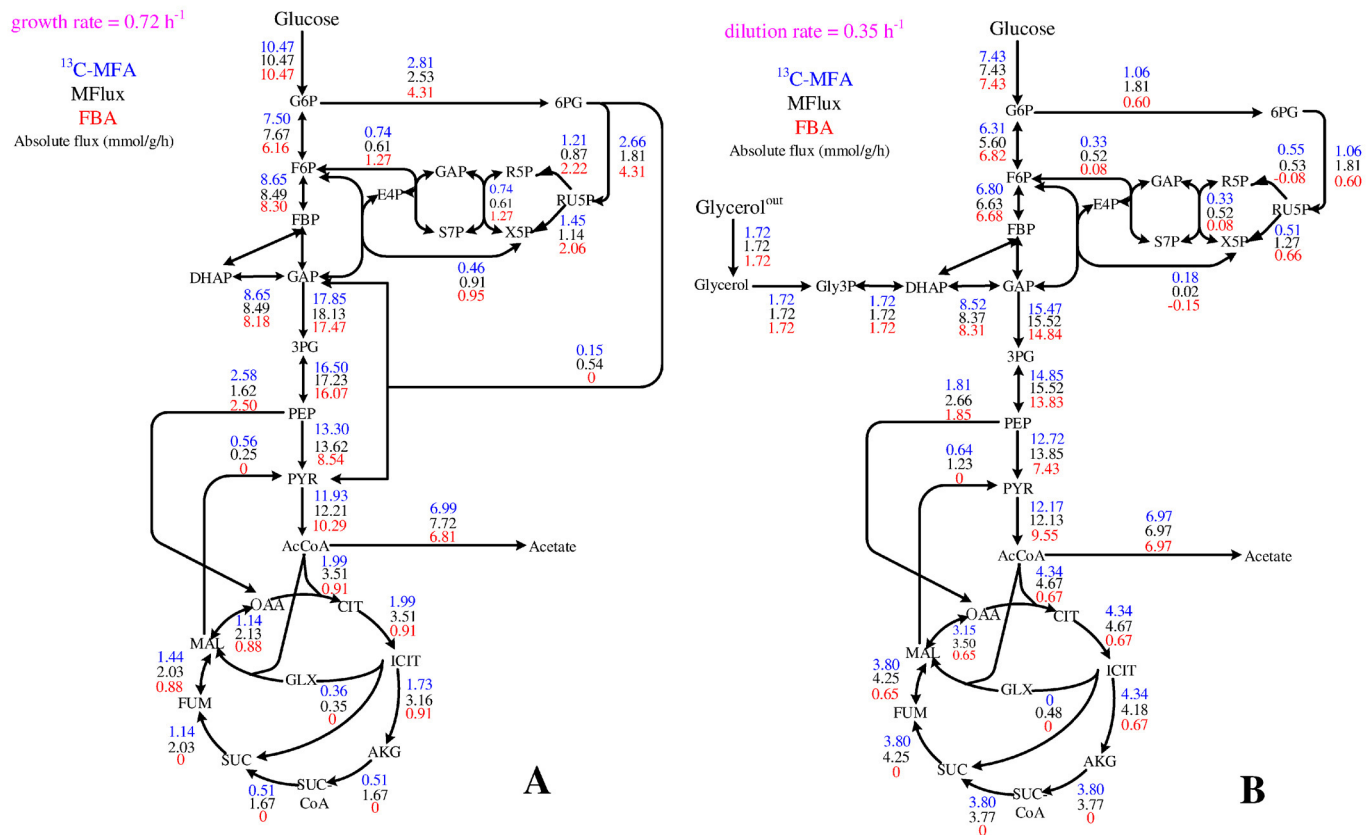
**Fig 8. A comparison of the  $^{13}\text{C}$ -MFA flux, the flux predicted by ML only, and the flux predicted by MFlux in Case 8.** *B. subtilis* was incubated in a shake flask (37 C, 300 rpm, aerobic condition), and supplied with labeled succinate and glutamate as carbon sources in M9 minimal medium. Detailed information is in [S1 Table](#).

doi:10.1371/journal.pcbi.1004838.g008



**Fig 9. A comparison of the  $^{13}\text{C}$ -MFA flux, the flux predicted by ML only, and the flux predicted by MFlux in Case 16.** *G. thermoglucosidasius* M10EXG was incubated in sealed bottles (micro-aerobic condition), supplied with glucose as a carbon source. Detailed information is in [S2 Table](#).

doi:10.1371/journal.pcbi.1004838.g009



**Fig 10. A comparison of the <sup>13</sup>C-MFA flux, the flux predicted by MFlux, and the flux predicted by FBA.** FBA analysis is simulated by an *E. coli* iJO1366 model (latest version) with default boundary settings from the reference [54]. The default values of growth associated maintenance energy (GAM) and non-growth associated maintenance energy (NGAM) were adopted. **A)** *E. coli* fluxome of glucose metabolism was precisely measured via parallel labeling experiments (a recent paper not in our dataset) [12]. **B)** *E. coli* fluxome of glycerol and glucose co-metabolism as measured by Drs. Yao and Shimizu (unpublished data). The *E. coli* strain was cultured in chemostat fermentor with a working volume of 1 L (37 C). The dilution rate in the continuous culture was 0.35 h<sup>-1</sup>. [<sup>1-13</sup>C] glucose and [<sup>1, 3-13</sup>C] glycerol were used for tracer experiments. The flux calculation is based on a previous method [42]. The RMSE from FBA is 22.5, while the RMSE from MFlux (this work) is 5.1. The COBRA toolbox running on MATLAB R2012b was employed for FBA/pFBA/geometricFBA simulation, and Gurobi 5.5 was used for linear programming. Detailed information is included in S2 Table.

doi:10.1371/journal.pcbi.1004838.g010

version of *E. coli* iJO1366 genome-scale model (2583 fluxes) was used [51]. Two comparative case studies were performed on *E. coli* fluxomes: one case for glucose based <sup>13</sup>C-MFA via parallel labeling experiments [12] and the other for glucose and glycerol co-utilization (unpublished data from the Shimizu Group). Neither of the test cases was included in the training set of MFlux. Given <sup>13</sup>C-MFA results as the control, MFlux results apparently have smaller RMSEs than FBA predictions. In the first case, the FBA has an RMSE of 11.3, while MFlux has an RMSE of 6.5 (Fig 10A). In the second case, the FBA has an RMSE of 22.5, while MFlux has an RMSE of 5.1 (Fig 10B). To circumvent variations caused by alternative solutions in FBA, we also employed pFBA and geometricFBA for both cases [52, 53] (S2 Table). In general, pFBA does not show better results compared with FBA for either case, while geometricFBA does not converge in our calculation.

FBA alone has given good predictions of growth rate as well as input and output fluxes, but not of intercellular fluxes. It is difficult to obtain actual P/O ratios, the ATP maintenance cost, the oxygen flux, and the transhydrogenase activities [55]. These energy/cofactor variables strongly affect the fluxes in the oxidative PP pathway (NADPH generation) and the TCA cycle (NADH, NADPH, and FADH<sub>2</sub> generation). Without proper flux constraints and



objective functions, it is more challenging for FBA to narrowly determine intracellular fluxomes in suboptimal metabolisms, especially for co-metabolism dual substrates because of the large solution space for the cell metabolism to optimize biomass growth using two substrates. As a complementary tool, MFlux may offer a quick metabolic overview and provide biologically meaningful flux boundaries to reduce FBA solution spaces when proper constraints for FBA are unavailable.

## Discussion

### Metabolic robustness of fluxome patterns among microbial species

“Robustness” was originally defined as the closed-loop process stability under perturbations in the control field. This definition is applicable to biochemical networks. To maintain the physiological output (i.e., the fluxome) within a desired range, microorganisms employ sophisticated control disciplines at different architecture levels, from the genome to the phenotype. In contrast to chaotic transcriptional profiles, the microbial fluxome shows robustness so that cells can survive in constantly-altering environments or in response to genetic mutations [56–58]. Metabolic rigidity at the flux level was first reported by Stephanopoulos in the early 1990s [59, 60]: NADPH is important for anabolism in the exponential growth phase, and the flux ratio around glucose-6-P node is rigid to form NADPH [60]. Moreover, 12 precursors from the central metabolism are required for biomass formation, which all have relatively small variations that are mainly dependent on biomass compositions. Due to both thermodynamic and mass balance constraints, cell metabolism aims to minimize variations in flux ratios under environmental perturbations. This rule also works for engineered microbes with moderately overexpressed pathways or strains from random mutations or deletions of non-essential genes. The feature of metabolic robustness facilitates ML applications.

Flux pattern recognition enables MFlux to predict metabolism of new species by learning from a small set of fluxome information from the same genus. For example, the metabolisms of *P. aeruginosa*, *P. fluorescens*, and *P. putida* have been studied by <sup>13</sup>C-MFA in the past decade [61–65]. The results show that different *Pseudomonas* species employ remarkably identical fluxomics types: they employ a highly active ED pathway for glycolytic metabolism and keep a low flux on the PP pathway for biomass synthesis, due to the lack of the *pfk* gene [66]. The ED pathway has less cost for protein formation than the Embden–Meyerhof–Parnas (EMP) pathway, yet only one ATP is formed per glucose [67, 68]. *Pseudomonas* species have slow cell growth rates and their aerobic metabolisms do not yield by-products. They also demonstrate a very active pyruvate shunt (*MAL* → *PYR*) and NADPH overproduction flux (a benefit for counteracting oxidative stress). On the other hand, the TCA cycle in *Pseudomonas* species show plasticity under genetic and environmental variations [69], and can respond to increased ATP and NADH demands under stress conditions [70].

For different bacterial species (e.g., *E. coli* and *Bacillus*), their fluxomes (e.g., glucose metabolisms) can be similar, because central fluxes in catabolism are regulated by energy and building block requirements that show much smaller variations than genome or transcriptional differences. On the other hand, change of carbon substrates may alternate flux distributions. For example, co-utilization of glucose and glycerol in *E. coli* cause significant reorganization of fluxomes. In a same microbial strain, different fluxome patterns can be employed for metabolizing different substrates (e.g., glucose-based fluxome vs acetate based fluxomes). Recognizing these metabolic patterns allows the use of a relatively small training set to perform a decent metabolic prediction of diverse metabolic types. Consequently, these common principles of certain classes of microorganisms can be captured by machine learning for fluxome predictions.

## Limitations of machine learning

There are several major challenges regarding MFlux. First, the  $^{13}\text{C}$ -MFA flux in literature may have errors and biases, which would be included in the learning/training process of MFlux and lead to further variations. For example, current  $^{13}\text{C}$ -MFA studies are not evenly distributed among a broad scope of microbial genera. Most reported MFAs are concentrated in a few model microbial species or metabolism of only a few substrates (mainly glucose), and thus our current ML cannot predict fluxomes well in certain cases. Such problem (model bias) can be resolved after more  $^{13}\text{C}$ -MFA papers for non-model species are included in the database and more constraints are implemented by our platform.

Second, the predictability of ML is limited to species and pathways that are already included in learning. More information and efforts are required to deal with cases of genetically modified strains with engineered pathways that hijack flux for synthesis of diverse commodity chemicals [13]. Currently,  $^{13}\text{C}$ -MFA has not widely used by synthetic biology community yet. In future versions of MFlux, new metabolic knowledge and rules should be applied for flux corrections.

Third, it is still difficult to incorporate regulation mechanisms into the current model. For instance, various catabolite repression mechanisms regulate the cell fluxome in the presence of multiple substrates (e.g., glucose shows catabolite repression for fast growing *E. coli* when both glucose and glycerol are available, Fig 10) [71]. These hierarchy regulations among substrate utilization can be dependent on growth rates or can differ among microbial species (*E. coli*, *Bacillus* and *Corynebacterium*).

Fourth, when oxygen is not available, fast bacterial sugar utilization will activate mixed acid fermentation (e.g., by utilizing lactate dehydrogenase and pyruvate formate lyase) to produce complicated overflow metabolites [13]. This mechanism is also furnished in yeast and mammalian cells. However,  $^{13}\text{C}$ -MFA studies on anaerobic metabolisms are much less frequent than on aerobic metabolisms. MFlux cannot predict the complicated patterns of overflow fluxes at this stage.

Fifth, our current dataset is still unable to support ML studies on phototrophic bacterial fluxomes. For phototrophic metabolism, its energy generation (ATP, NADH and NADPH) may not be controlled by substrate catabolism. Some phototrophic bacteria (e.g., cyanobacteria) have versatile autotrophic and photomixotrophic metabolism that is highly sensitive to light and substrate availability. Other phototrophs may even have  $\text{CO}_2$  fixation pathway (such as the reversed TCA cycle). Therefore, our MFlux platform could not make ML predictions but only reports a general description of metabolic features of these species.

Lastly, ML cannot directly estimate fluxes for carbon sources which are not part of the learning dataset. To predict fluxomes for new substrates, users need to assume that similar entry-points of carbon sources into the central metabolic network may cause similar flux distributions (e.g., sucrose has to be treated as a combination of glucose and fructose).

## Conclusion

This proof-of-concept study demonstrates that AI methods can facilitate fluxomics research with reasonable precision.  $^{13}\text{C}$ -MFA is a very small field of just hundreds of MFA research papers on microbial species published in the past two decades. In the long term, ML methods may solve this problem: with a large and reliable fluxomics dataset and more information from  $^{13}\text{C}$ -MFA and AI scientists, the future MFlux model can make broad-scope metabolism predictions. To sum up, MFlux presents the first platform introducing ML in the field of fluxomics and it will be continuously updated and improved. It will inspire the development of similar computational tools to advance omics and metabolic engineering fields [72].

## Supporting Information

**S1 Program. MFlux Computer Program (Source code).** Python scripts in a ZIP file. (ZIP)

**S1 Table. Results of 20 case studies.** Detailed information for 20 cases studies using MFlux, including literature references, input conditions,  $^{13}\text{C}$ -MFA flux, the flux profiles predicted by ML, and the flux profiles predicted by MFlux with additional constraints. (XLSX)

**S2 Table. Detailed information of the comparison with FBA/pFBA.** The information of constraints, objective function, and simulation results. (XLSX)

## Acknowledgments

The authors would like to appreciate helpful suggestions by Dr. Eric You Xu (Fitbit Inc.) and Dr. Lian He (Washington University).

## Author Contributions

Conceived and designed the experiments: SGW YJT. Performed the experiments: SGW YW WJ TO FSB. Analyzed the data: SGW RY XZ KS YJT FSB. Contributed reagents/materials/analysis tools: XZ YJT FSB. Wrote the paper: SGW KS YJT FSB.

## References

1. Winter G, Krömer JO. Fluxomics—connecting omics analysis and phenotypes. *Environmental Microbiology*. 2013; 15(7):1901–1916. doi: [10.1111/1462-2920.12064](https://doi.org/10.1111/1462-2920.12064) PMID: [23279205](https://pubmed.ncbi.nlm.nih.gov/23279205/)
2. Chen X, Alonso AP, Allen DK, Reed JL, Shachar-Hill Y. Synergy between  $^{13}\text{C}$ -metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in *E. coli*. *Metabolic Engineering*. 2011; 13(1):38–48. doi: [10.1016/j.ymben.2010.11.004](https://doi.org/10.1016/j.ymben.2010.11.004) PMID: [21129495](https://pubmed.ncbi.nlm.nih.gov/21129495/)
3. Tang YJ, Chakraborty R, Martín HG, Chu J, Hazen TC, Keasling JD. Flux analysis of central metabolic pathways in *Geobacter metallireducens* during reduction of soluble Fe(III)-nitrilotriacetic acid. *Applied and Environmental Microbiology*. 2007; 73(12):3859–3864. doi: [10.1128/AEM.02986-06](https://doi.org/10.1128/AEM.02986-06) PMID: [17468285](https://pubmed.ncbi.nlm.nih.gov/17468285/)
4. Tang JKH, You L, Blankenship RE, Tang YJ. Recent advances in mapping environmental microbial metabolisms through  $^{13}\text{C}$  isotopic fingerprints. *Journal of The Royal Society Interface*. 2012; 9(76):2767–2780. doi: [10.1098/rsif.2012.0396](https://doi.org/10.1098/rsif.2012.0396)
5. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, et al. Metabolic engineering of *Escherichia coli* for direct production of 1, 4-butanediol. *Nature Chemical Biology*. 2011; 7(7):445–452. doi: [10.1038/nchembio.580](https://doi.org/10.1038/nchembio.580) PMID: [21602812](https://pubmed.ncbi.nlm.nih.gov/21602812/)
6. Becker J, Zelder O, Häfner S, Schröder H, Wittmann C. From zero to hero—Design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production. *Metabolic Engineering*. 2011; 13(2):159–168. doi: [10.1016/j.ymben.2011.01.003](https://doi.org/10.1016/j.ymben.2011.01.003) PMID: [21241816](https://pubmed.ncbi.nlm.nih.gov/21241816/)
7. He L, Xiao Y, Gebreselassie N, Zhang F, Antoniewicz MR, Tang YJ, et al. Central metabolic responses to the overproduction of fatty acids in *Escherichia coli* based on  $^{13}\text{C}$ -metabolic flux analysis. *Biotechnology and Bioengineering*. 2014; 111(3):575–585. doi: [10.1002/bit.25124](https://doi.org/10.1002/bit.25124) PMID: [24122357](https://pubmed.ncbi.nlm.nih.gov/24122357/)
8. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metabolic Engineering*. 2007; 9(1):68–86. doi: [10.1016/j.ymben.2006.09.001](https://doi.org/10.1016/j.ymben.2006.09.001) PMID: [17088092](https://pubmed.ncbi.nlm.nih.gov/17088092/)
9. Weitzel M, Nöh K, Dalman T, Niedenführ S, Stute B, Wiechert W.  $^{13}\text{C}$ FLUX2—high-performance software suite for  $^{13}\text{C}$ -metabolic flux analysis. *Bioinformatics*. 2013; 29(1):143–145. doi: [10.1093/bioinformatics/bts646](https://doi.org/10.1093/bioinformatics/bts646) PMID: [23110970](https://pubmed.ncbi.nlm.nih.gov/23110970/)
10. Quek LE, Wittmann C, Nielsen LK, Krömer JO. OpenFLUX: efficient modelling software for  $^{13}\text{C}$ -based metabolic flux analysis. *Microbial Cell Factories*. 2009; 8:25. doi: [10.1186/1475-2859-8-25](https://doi.org/10.1186/1475-2859-8-25) PMID: [19409084](https://pubmed.ncbi.nlm.nih.gov/19409084/)

11. Zamboni N, Fischer E, Sauer U. FiatFlux—a software for metabolic flux analysis from <sup>13</sup>C-glucose experiments. *BMC Bioinformatics*. 2005; 6(1):209. doi: [10.1186/1471-2105-6-209](https://doi.org/10.1186/1471-2105-6-209) PMID: [16122385](https://pubmed.ncbi.nlm.nih.gov/16122385/)
12. Crown SB, Long CP, Antoniewicz MR. Integrated <sup>13</sup>C-metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli*. *Metabolic Engineering*. 2015; 28:151–158. doi: [10.1016/j.ymben.2015.01.001](https://doi.org/10.1016/j.ymben.2015.01.001) PMID: [25596508](https://pubmed.ncbi.nlm.nih.gov/25596508/)
13. Antoniewicz MR, Kraynie DF, Laffend LA, González-Lergier J, Kelleher JK, Stephanopoulos G. Metabolic flux analysis in a nonstationary system: fed-batch fermentation of a high yielding strain of *E. coli* producing 1, 3-propanediol. *Metabolic Engineering*. 2007; 9(3):277–292. doi: [10.1016/j.ymben.2007.01.003](https://doi.org/10.1016/j.ymben.2007.01.003) PMID: [17400499](https://pubmed.ncbi.nlm.nih.gov/17400499/)
14. Nöh K, Grönke K, Luo B, Takors R, Oldiges M, Wiechert W. Metabolic flux analysis at ultra short time scale: isotopically non-stationary <sup>13</sup>C labeling experiments. *Journal of Biotechnology*. 2007; 129(2):249–267. doi: [10.1016/j.jbiotec.2006.11.015](https://doi.org/10.1016/j.jbiotec.2006.11.015) PMID: [17207877](https://pubmed.ncbi.nlm.nih.gov/17207877/)
15. Tang YJ, Martin HG, Myers S, Rodriguez S, Baidoo EE, Keasling JD. Advances in analysis of microbial metabolic fluxes via <sup>13</sup>C isotopic labeling. *Mass Spectrometry Reviews*. 2009; 28(2):362–375. doi: [10.1002/mas.20191](https://doi.org/10.1002/mas.20191) PMID: [19025966](https://pubmed.ncbi.nlm.nih.gov/19025966/)
16. Zhuang WQ, Yi S, Bill M, Brisson VL, Feng X, Men Y, et al. Incomplete Wood–Ljungdahl pathway facilitates one-carbon metabolism in organohalide-respiring *Dehalococcoides mccartyi*. *Proceedings of the National Academy of Sciences*. 2015; 111(17):6419–6424. doi: [10.1073/pnas.1321542111](https://doi.org/10.1073/pnas.1321542111)
17. Tarca AL, Carey VJ, Chen X, Romero R, Draghici S. Machine learning and its applications to biology. *PLoS Computational Biology*. 2007; 3(6):e116. doi: [10.1371/journal.pcbi.0030116](https://doi.org/10.1371/journal.pcbi.0030116) PMID: [17604446](https://pubmed.ncbi.nlm.nih.gov/17604446/)
18. Kell DB. Metabolomics, modelling and machine learning in systems biology—towards an understanding of the languages of cells. *FEBS Journal*. 2006; 273(5):873–894.
19. Dale JM, Popescu L, Karp PD. Machine learning methods for metabolic pathway prediction. *BMC Bioinformatics*. 2010; 11(1):15. doi: [10.1186/1471-2105-11-15](https://doi.org/10.1186/1471-2105-11-15) PMID: [20064214](https://pubmed.ncbi.nlm.nih.gov/20064214/)
20. Rätsch G, Sonnenburg S, Srinivasan J, Witte H, Müller KR, Sommer RJ, et al. Improving the *Caenorhabditis elegans* genome annotation using machine learning. *PLoS Computational Biology*. 2007; 3(2):e20. doi: [10.1371/journal.pcbi.0030020](https://doi.org/10.1371/journal.pcbi.0030020) PMID: [17319737](https://pubmed.ncbi.nlm.nih.gov/17319737/)
21. Ye QH, Qin LX, Forgues M, He P, Kim JW, Peng AC, et al. Predicting hepatitis B virus–positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nature Medicine*. 2003; 9(4):416–423. doi: [10.1038/nm843](https://doi.org/10.1038/nm843) PMID: [12640447](https://pubmed.ncbi.nlm.nih.gov/12640447/)
22. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature Medicine*. 2002; 8(1):68–74. doi: [10.1038/nm0102-68](https://doi.org/10.1038/nm0102-68) PMID: [11786909](https://pubmed.ncbi.nlm.nih.gov/11786909/)
23. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*. 2000; 16(10):906–914. doi: [10.1093/bioinformatics/16.10.906](https://doi.org/10.1093/bioinformatics/16.10.906) PMID: [11120680](https://pubmed.ncbi.nlm.nih.gov/11120680/)
24. Supek F, Peharec P, Krsnik-Rasol M, Šmuc T. Enhanced analytical power of SDS-PAGE using machine learning algorithms. *Proteomics*. 2008; 8(1):28–31. doi: [10.1002/pmic.200700555](https://doi.org/10.1002/pmic.200700555) PMID: [18046695](https://pubmed.ncbi.nlm.nih.gov/18046695/)
25. Mahadevan S, Shah SL, Marrie TJ, Slupsky CM. Analysis of metabolomic data using support vector machines. *Analytical Chemistry*. 2008; 80(19):7562–7570. doi: [10.1021/ac800954c](https://doi.org/10.1021/ac800954c) PMID: [18767870](https://pubmed.ncbi.nlm.nih.gov/18767870/)
26. Zhang Z, Shen T, Rui B, Zhou W, Zhou X, Shang C, et al. CeCaFDB: a curated database for the documentation, visualization and comparative analysis of central carbon metabolic flux distributions explored by <sup>13</sup>C-fluxomics. *Nucleic Acids Research*. 2014; 43:D549–D557. doi: [10.1093/nar/gku1137](https://doi.org/10.1093/nar/gku1137)
27. Sauer U, Eikmanns BJ. The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiology Reviews*. 2005; 29(4):765–794. doi: [10.1016/j.femsre.2004.11.002](https://doi.org/10.1016/j.femsre.2004.11.002) PMID: [16102602](https://pubmed.ncbi.nlm.nih.gov/16102602/)
28. Tang YJ, Sapra R, Joyner D, Hazen TC, Myers S, Reichmuth D, et al. Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain. *Biotechnology and Bioengineering*. 2009; 102(5):1377–1386. doi: [10.1002/bit.22181](https://doi.org/10.1002/bit.22181) PMID: [19016470](https://pubmed.ncbi.nlm.nih.gov/19016470/)
29. Peters-Wendisch PG, Kreutzer C, Kalinowski J, Pátek M, Sahn H, Eikmanns BJ. Pyruvate carboxylase from *Corynebacterium glutamicum*: characterization, expression and inactivation of the *pyc* gene. *Microbiology*. 1998; 144(4):915–927. doi: [10.1099/00221287-144-4-915](https://doi.org/10.1099/00221287-144-4-915) PMID: [9579065](https://pubmed.ncbi.nlm.nih.gov/9579065/)
30. Toya Y, Ishii N, Nakahigashi K, Hirasawa T, Soga T, Tomita M, et al. <sup>13</sup>C-metabolic flux analysis for batch culture of *Escherichia coli* and its *pyk* and *pgi* gene knockout mutants based on mass isotopomer distribution of intracellular metabolites. *Biotechnology Progress*. 2010; 26(4):975–992. PMID: [20730757](https://pubmed.ncbi.nlm.nih.gov/20730757/)
31. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*. 2011; 12:2825–2830.

32. Leighty RW, Antoniewicz MR. COMPLETE-MFA: Complementary parallel labeling experiments technique for metabolic flux analysis. *Metabolic Engineering*. 2013; 20:49–55. doi: [10.1016/j.ymben.2013.08.006](https://doi.org/10.1016/j.ymben.2013.08.006) PMID: [24021936](https://pubmed.ncbi.nlm.nih.gov/24021936/)
33. Zhao J, Baba T, Mori H, Shimizu K. Effect of *zwf* gene knockout on the metabolism of *Escherichia coli* grown on glucose or acetate. *Metabolic Engineering*. 2004; 6(2):164–174. doi: [10.1016/j.ymben.2004.02.004](https://doi.org/10.1016/j.ymben.2004.02.004) PMID: [15113569](https://pubmed.ncbi.nlm.nih.gov/15113569/)
34. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*. 2002; 99(23):15112–15117. doi: [10.1073/pnas.232349399](https://doi.org/10.1073/pnas.232349399)
35. Andersen M, Dahl J, Liu Z, Vandenberghe L. Interior-point methods for large-scale cone programming. *Optimization for Machine Learning*. 2011;p. 55–83.
36. Towell GG, Shavlik JW. Knowledge-based artificial neural networks. *Artificial Intelligence*. 1994; 70(1):119–165. doi: [10.1016/0004-3702\(94\)90105-8](https://doi.org/10.1016/0004-3702(94)90105-8)
37. Marriott K, Stuckey P. *Programming with Constraints: An Introduction*. MIT Press; 1998.
38. Niemeyer G. python-constraint: Constraint Solving Problem solver for Python;. Available from <https://labix.org/python-constraint>.
39. Fischer E, Zamboni N, Sauer U. High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived <sup>13</sup>C constraints. *Analytical Biochemistry*. 2004; 325(2):308–316. doi: [10.1016/j.ab.2003.10.036](https://doi.org/10.1016/j.ab.2003.10.036) PMID: [14751266](https://pubmed.ncbi.nlm.nih.gov/14751266/)
40. Zhao J, Baba T, Mori H, Shimizu K. Global metabolic response of *Escherichia coli* to *gnd* or *zwf* gene-knockout, based on <sup>13</sup>C-labeling experiments and the measurement of enzyme activities. *Applied Microbiology and Biotechnology*. 2004; 64(1):91–98. doi: [10.1007/s00253-003-1458-5](https://doi.org/10.1007/s00253-003-1458-5) PMID: [14661115](https://pubmed.ncbi.nlm.nih.gov/14661115/)
41. Fong SS, Nanchen A, Palsson BO, Sauer U. Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *Journal of Biological Chemistry*. 2006; 281(12):8024–8033. doi: [10.1074/jbc.M510016200](https://doi.org/10.1074/jbc.M510016200) PMID: [16319065](https://pubmed.ncbi.nlm.nih.gov/16319065/)
42. Peng L, Arauzo-Bravo MJ, Shimizu K. Metabolic flux analysis for a *ppc* mutant *Escherichia coli* based on <sup>13</sup>C-labelling experiments together with enzyme activity assays and intracellular metabolite measurements. *FEMS Microbiology Letters*. 2004; 235(1):17–23. doi: [10.1111/j.1574-6968.2004.tb09562.x](https://doi.org/10.1111/j.1574-6968.2004.tb09562.x) PMID: [15158257](https://pubmed.ncbi.nlm.nih.gov/15158257/)
43. Tännler S, Decasper S, Sauer U. Maintenance metabolism and carbon fluxes in *Bacillus* species. *Microbial Cell Factories*. 2008; 7(1):19. doi: [10.1186/1475-2859-7-19](https://doi.org/10.1186/1475-2859-7-19) PMID: [18564406](https://pubmed.ncbi.nlm.nih.gov/18564406/)
44. Chubukov V, Uhr M, Le Chat L, Kleijn RJ, Jules M, Link H, et al. Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*. *Molecular Systems Biology*. 2013; 9(1):709. doi: [10.1038/msb.2013.66](https://doi.org/10.1038/msb.2013.66) PMID: [24281055](https://pubmed.ncbi.nlm.nih.gov/24281055/)
45. van Ooyen J, Noack S, Bott M, Reth A, Eggeling L. Improved L-lysine production with *Corynebacterium glutamicum* and systemic insight into citrate synthase flux and activity. *Biotechnology and Bioengineering*. 2012; 109(8):2070–2081. doi: [10.1002/bit.24486](https://doi.org/10.1002/bit.24486) PMID: [22392073](https://pubmed.ncbi.nlm.nih.gov/22392073/)
46. Bommareddy RR, Chen Z, Rappert S, Zeng AP. A de novo NADPH generation pathway for improving lysine production of *Corynebacterium glutamicum* by rational design of the coenzyme specificity of glyceraldehyde 3-phosphate dehydrogenase. *Metabolic Engineering*. 2014; 25:30–37. doi: [10.1016/j.ymben.2014.06.005](https://doi.org/10.1016/j.ymben.2014.06.005) PMID: [24953302](https://pubmed.ncbi.nlm.nih.gov/24953302/)
47. Wang ZJ, Wang P, Liu YW, Zhang YM, Chu J, Huang Mz, et al. Metabolic flux analysis of the central carbon metabolism of the industrial vitamin B12 producing strain *Pseudomonas denitrificans* using <sup>13</sup>C-labeled glucose. *Journal of the Taiwan Institute of Chemical Engineers*. 2012; 43(2):181–187. doi: [10.1016/j.jtice.2011.09.002](https://doi.org/10.1016/j.jtice.2011.09.002)
48. Hemme CL, Fields MW, He Q, Deng Y, Lin L, Tu Q, et al. Correlation of genomic and physiological traits of *Thermoanaerobacter* species with biofuel yields. *Applied and Environmental Microbiology*. 2011; 77(22):7998–8008. doi: [10.1128/AEM.05677-11](https://doi.org/10.1128/AEM.05677-11) PMID: [21948836](https://pubmed.ncbi.nlm.nih.gov/21948836/)
49. Tang Y, Pingitore F, Mukhopadhyay A, Phan R, Hazen TC, Keasling JD. Pathway confirmation and flux analysis of central metabolic pathways in *Desulfovibrio vulgaris* Hildenborough using gas chromatography-mass spectrometry and Fourier transform-ion cyclotron resonance mass spectrometry. *Journal of Bacteriology*. 2007; 189(3):940–949. doi: [10.1128/JB.00948-06](https://doi.org/10.1128/JB.00948-06) PMID: [17114264](https://pubmed.ncbi.nlm.nih.gov/17114264/)
50. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nature Biotechnology*. 2010; 28(3):245–248. doi: [10.1038/nbt.1614](https://doi.org/10.1038/nbt.1614) PMID: [20212490](https://pubmed.ncbi.nlm.nih.gov/20212490/)
51. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular Systems Biology*. 2011; 7(1):535. doi: [10.1038/msb.2011.65](https://doi.org/10.1038/msb.2011.65) PMID: [21988831](https://pubmed.ncbi.nlm.nih.gov/21988831/)

52. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*. 2010; 6(1):390. doi: [10.1038/msb.2010.47](https://doi.org/10.1038/msb.2010.47) PMID: [20664636](https://pubmed.ncbi.nlm.nih.gov/20664636/)
53. Smallbone K, Simeonidis E. Flux balance analysis: a geometric perspective. *Journal of Theoretical Biology*. 2009; 258(2):311–315. doi: [10.1016/j.jtbi.2009.01.027](https://doi.org/10.1016/j.jtbi.2009.01.027) PMID: [19490860](https://pubmed.ncbi.nlm.nih.gov/19490860/)
54. Orth JD, Palsson B. Gap-filling analysis of the iJO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions. *BMC Systems Biology*. 2012; 6(1):30. doi: [10.1186/1752-0509-6-30](https://doi.org/10.1186/1752-0509-6-30) PMID: [22548736](https://pubmed.ncbi.nlm.nih.gov/22548736/)
55. Wu SG, He L, Wang Q, Tang YJ. An ancient Chinese wisdom for metabolic engineering: Yin-Yang. *Microbial Cell Factories*. 2015; 14(1):39. doi: [10.1186/s12934-015-0219-3](https://doi.org/10.1186/s12934-015-0219-3) PMID: [25889067](https://pubmed.ncbi.nlm.nih.gov/25889067/)
56. Fischer E, Sauer U. Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nature Genetics*. 2005; 37(6):636–640. doi: [10.1038/ng1555](https://doi.org/10.1038/ng1555) PMID: [15880104](https://pubmed.ncbi.nlm.nih.gov/15880104/)
57. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*. 2007; 3(1):119. doi: [10.1038/msb4100162](https://doi.org/10.1038/msb4100162) PMID: [17625511](https://pubmed.ncbi.nlm.nih.gov/17625511/)
58. Tang YJ, Martin HG, Deutschbauer A, Feng X, Huang R, Llorca X, et al. Invariability of central metabolic flux distribution in *Shewanella oneidensis* MR-1 under environmental or genetic perturbations. *Biotechnology Progress*. 2009; 25(5):1254–1259. doi: [10.1002/btpr.227](https://doi.org/10.1002/btpr.227) PMID: [19610125](https://pubmed.ncbi.nlm.nih.gov/19610125/)
59. Stephanopoulos G. Metabolic fluxes and metabolic engineering. *Metabolic Engineering*. 1999; 1(1):1–11. doi: [10.1006/mben.1998.0101](https://doi.org/10.1006/mben.1998.0101) PMID: [10935750](https://pubmed.ncbi.nlm.nih.gov/10935750/)
60. Stephanopoulos G, Vallino JJ. Network rigidity and metabolic engineering in metabolite overproduction. *Science*. 1991; 252(5013):1675–1681. doi: [10.1126/science.1904627](https://doi.org/10.1126/science.1904627) PMID: [1904627](https://pubmed.ncbi.nlm.nih.gov/1904627/)
61. Lien SK, Niedenführ S, Sletta H, Nöh K, Bruheim P. Fluxome study of *Pseudomonas fluorescens* reveals major reorganisation of carbon flux through central metabolic pathways in response to inactivation of the anti-sigma factor MucA. *BMC Systems Biology*. 2015; 9(1):6. doi: [10.1186/s12918-015-0148-0](https://doi.org/10.1186/s12918-015-0148-0) PMID: [25889900](https://pubmed.ncbi.nlm.nih.gov/25889900/)
62. Fuhrer T, Fischer E, Sauer U. Experimental identification and quantification of glucose metabolism in seven bacterial species. *Journal of Bacteriology*. 2005; 187(5):1581–1590. doi: [10.1128/JB.187.5.1581-1590.2005](https://doi.org/10.1128/JB.187.5.1581-1590.2005) PMID: [15716428](https://pubmed.ncbi.nlm.nih.gov/15716428/)
63. Wierckx N, Ruijsenaars HJ, de Winde JH, Schmid A, Blank LM. Metabolic flux analysis of a phenol producing mutant of *Pseudomonas putida* S12: verification and complementation of hypotheses derived from transcriptomics. *Journal of Biotechnology*. 2009; 143(2):124–129. doi: [10.1016/j.jbiotec.2009.06.023](https://doi.org/10.1016/j.jbiotec.2009.06.023) PMID: [19560494](https://pubmed.ncbi.nlm.nih.gov/19560494/)
64. del Castillo T, Ramos JL, Rodríguez-Herva JJ, Fuhrer T, Sauer U, Duque E. Convergent peripheral pathways catalyze initial glucose catabolism in *Pseudomonas putida*: genomic and flux analysis. *Journal of Bacteriology*. 2007; 189(14):5142–5152. doi: [10.1128/JB.00203-07](https://doi.org/10.1128/JB.00203-07) PMID: [17483213](https://pubmed.ncbi.nlm.nih.gov/17483213/)
65. Blank LM, Ionidis G, Ebert BE, Bühler B, Schmid A. Metabolic response of *Pseudomonas putida* during redox biocatalysis in the presence of a second octanol phase. *FEBS Journal*. 2008; 275(20):5173–5190. doi: [10.1111/j.1742-4658.2008.06648.x](https://doi.org/10.1111/j.1742-4658.2008.06648.x) PMID: [18803670](https://pubmed.ncbi.nlm.nih.gov/18803670/)
66. Conway T. The Entner-Doudoroff pathway: history, physiology and molecular biology. *FEMS Microbiology Reviews*. 1992; 103(1):1–28. doi: [10.1111/j.1574-6968.1992.tb05822.x](https://doi.org/10.1111/j.1574-6968.1992.tb05822.x)
67. Bar-Even A, Flamholz A, Noor E, Milo R. Rethinking glycolysis: on the biochemical logic of metabolic pathways. *Nature Chemical Biology*. 2012; 8(6):509–517. doi: [10.1038/nchembio.971](https://doi.org/10.1038/nchembio.971) PMID: [22596202](https://pubmed.ncbi.nlm.nih.gov/22596202/)
68. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences*. 2013; 110(24):10039–10044. doi: [10.1073/pnas.1215283110](https://doi.org/10.1073/pnas.1215283110)
69. Berger A, Dohnt K, Tielen P, Jahn D, Becker J, Wittmann C. Robustness and plasticity of metabolic pathway flux among uropathogenic isolates of *Pseudomonas aeruginosa*. *PloS One*. 2014; 9(4). doi: [10.1371/journal.pone.0088368](https://doi.org/10.1371/journal.pone.0088368)
70. Ebert BE, Kurth F, Grund M, Blank LM, Schmid A. Response of *Pseudomonas putida* KT2440 to increased NADH and ATP demand. *Applied and Environmental Microbiology*. 2011; 77(18):6597–6605. doi: [10.1128/AEM.05588-11](https://doi.org/10.1128/AEM.05588-11) PMID: [21803911](https://pubmed.ncbi.nlm.nih.gov/21803911/)
71. Yao R, Hirose Y, Sarkar D, Nakahigashi K, Ye Q, Shimizu K. Catabolic regulation analysis of *Escherichia coli* and its *crp*, *mlc*, *mgsA*, *pgi* and *ptsG* mutants. *Microbial Cell Factories*. 2011; 10(67):1475–2859. doi: [10.1186/1475-2859-10-67](https://doi.org/10.1186/1475-2859-10-67) PMID: [21831320](https://pubmed.ncbi.nlm.nih.gov/21831320/)
72. Wu SG, Shimizu K, Tang JKH, Tang YJ. Facilitate Collaborations among Synthetic Biology, Metabolic Engineering and Machine Learning. *ChemBioEng Reviews*. 2016; 3(2):1–11. doi: [10.1002/cben.201500024](https://doi.org/10.1002/cben.201500024)