frontiers in Genetics

ORIGINAL RESEARCH
published: 21 February 2022
doi: 10.3389/fgene.2022.815476

# Totoro: Identifying Active Reactions During the Transient State for Metabolic Perturbations

*Mariana Galvão Ferrarini*[1,2], *Irene Ziska*[1,3], *Ricardo Andrade*[1,4], *Alice Julien-Laferrière*[5], *Louis Duchemin*[1], *Roberto Marcondes César Jr.*[4], *Arnaud Mary*[1,3], *Susana Vinga*[6] *and Marie-France Sagot*[1,3]*

[1]*Laboratoire de Biométrie et Biologie Évolutive, UMR 5558, CNRS, Université de Lyon, Université Lyon 1, Villeurbanne, France,* [2]*Univ Lyon, INRAE, INSA-Lyon, BF2I, UMR 203, Villeurbanne, France,* [3]*INRIA Grenoble Rhône-Alpes, Villeurbanne, France,* [4]*Institute of Mathematics and Statistics (IME), University of São Paulo, São Paulo, Brazil,* [5]*Soladis GmbH, Basel, Switzerland,* [6]*INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal*

**Motivation:** The increasing availability of metabolomic data and their analysis are improving the understanding of cellular mechanisms and how biological systems respond to different perturbations. Currently, there is a need for novel computational methods that facilitate the analysis and integration of increasing volume of available data.

**Results:** In this paper, we present TOTORO a new constraint-based approach that integrates quantitative non-targeted metabolomic data of two different metabolic states into genome-wide metabolic models and predicts reactions that were most likely active during the transient state. We applied TOTORO to real data of three different growth experiments (pulses of glucose, pyruvate, succinate) from *Escherichia coli* and we were able to predict known active pathways and gather new insights on the different metabolisms related to each substrate. We used both the *E. coli* core and the iJO1366 models to demonstrate that our approach is applicable to both smaller and larger networks.

**Availability:** TOTORO is an open source method (available at https://gitlab.inria.fr/erable/totoro) suitable for any organism with an available metabolic model. It is implemented in C++ and depends on IBM CPLEX which is freely available for academic purposes.

**Keywords: metabolomics, metabolic networks, transient state, metabolic perturbation, omics integration**

OPEN ACCESS

***Edited by:***
*Marco Pellegrini,*
*Italian National Research Council, Italy*

***Reviewed by:***
*Philipp Wendering,*
*University of Potsdam, Germany*
*Samuel Seaver,*
*Argonne National Laboratory (DOE), United States*

***Correspondence:**
*Marie-France Sagot*
*marie-france.sagot@inria.fr*

***Specialty section:***
*This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics*

***Received:*** *15 November 2021*
***Accepted:*** *24 January 2022*
***Published:*** *21 February 2022*

***Citation:***
*Galvão Ferrarini M, Ziska I, Andrade R, Julien-Laferrière A, Duchemin L, César RM, Mary A, Vinga S and Sagot M-F (2022) Totoro: Identifying Active Reactions During the Transient State for Metabolic Perturbations. Front. Genet. 13:815476. doi: 10.3389/fgene.2022.815476*

## 1 INTRODUCTION

The increasing availability of metabolomic data and their analysis are currently enhancing our knowledge on diverse biological mechanisms and elucidating how cells and organisms respond to different perturbations (Sevin et al., 2015). Metabolomics can be used to obtain a metabolic profile that characterizes the physiological response of a cell, tissue or organism to a stress or to a general perturbation (Roessner and Bowne, 2009), and experiments ranging from shorter-term responses (such as stress response programs) to longer-term responses (such as acclimation) are broadly available for diverse species. Different network-based strategies for metabolomic data analysis have been recently reviewed in (Perez de Souza et al., 2020) and amongst others, such strategies can be used to establish associations between metabolites or to integrate them into metabolic pathways.

Metabolic profiles are often analyzed and interpreted with the help of bioinformatic software such as METEXPLORE (Cottret et al., 2018; Frainay et al., 2019), METABOANALYST (Xia et al., 2015; Chong et al., 2018) or 3OMICS (Kuo et al., 2013) that can identify the set of metabolites with a significant change in their concentration. The metabolomic data are projected on the annotated metabolic pathways in order to highlight the processes that may be linked to the observed changes. The

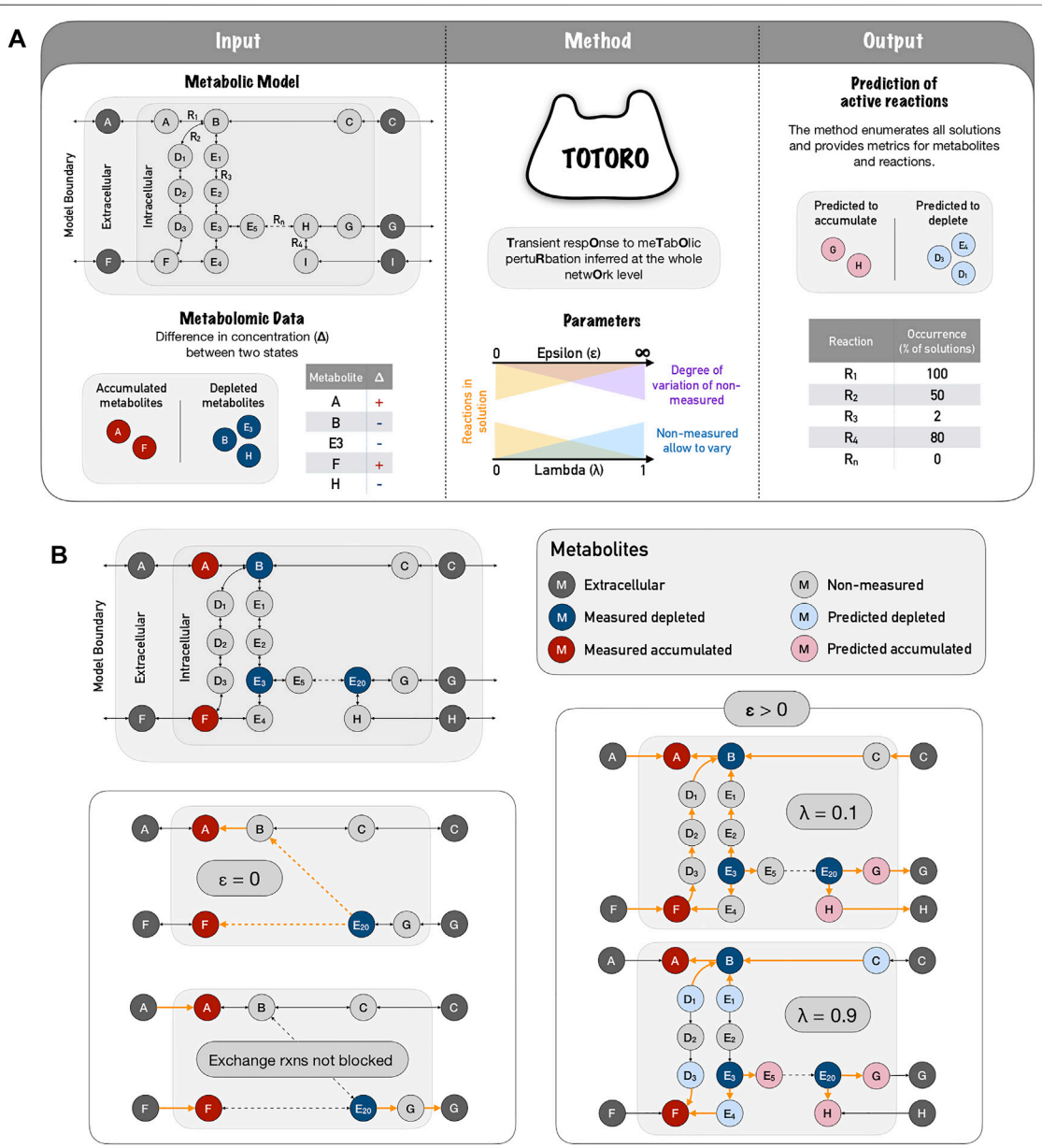


**FIGURE 1 |** TOTORO method explained. **(A)** TOTORO is able to integrate a metabolic model with metabolomic data in order to predict active reactions during the transient state between two conditions (or simply after a perturbation). The inputs of TOTORO are an SBML metabolic model, and a list of intervals for the difference in concentration (Δ) for each measured metabolite. In the metabolic model panel, grey circles depict metabolites and arrows depict reactions. In the metabolic data panel, accumulated metabolites are depicted in red circles, depleted metabolites are depicted in blue circles. The method TOTORO then requires two additional user-defined parameters to fine tune the results, namely $\lambda$ and $\epsilon$. TOTORO provides as output the predicted variation of metabolites and reactions that were most likely active between the two states in each enumerated solution as well as metric files grouping all enumerated solutions. In the figure, reaction occurrence is depicted as a percentage in all enumerated solutions. **(B)** The fine-tuning of parameters $\lambda$ and $\epsilon$ are provided within a toy network, in which active reactions are showed in orange and dashed arrows indicate several reactions in a row. When we don't allow an accumulation of non-measured metabolites ($\epsilon = 0$), the method will try to connect the input deltas of distant and possibly unrelated metabolites; and in the case exchange reactions are not blocked, the method will most likely propagate the accumulation or depletion towards outside of the boundaries of the model. When accumulation is allowed ($\epsilon > 0$) a low lambda ($\lambda = 0.1$) will favor solutions in which fewer non-measured metabolites accumulate or deplete, and will include a larger number of reactions within the solutions. As we raise the parameter lambda ($\lambda = 0.9$), we favor local and smaller solutions.

aforementioned software also try to integrate different kinds of omic data (such as transcriptomic, metabolomic or proteomic data) in order to give a deeper understanding of the studied mechanisms (Cambiaghi et al., 2017). Different approaches were reviewed in (Rosato et al., 2018; Ivanisevic and Want, 2019; Stanstrup et al., 2019) and software for the enrichment analysis of metabolomic data were evaluated and their results compared in (Marco-Ramell et al., 2018). However, metabolic pathways have subjective definitions and can differ between databases (Ginsburg, 2009). Additionally, this kind of analysis can make it hard to identify the connections between metabolites since they can be part of many pathways and it is thus possible to miss paths which traverse several biological pathways.

Another approach is to use graph-based methods that allow to consider the whole metabolism as an integrated system focusing on the parts that are connecting the metabolites of interest. Usually, these methods rely mainly on the network structure, chemical information and on an input list of metabolites (Frainay and Jourdan, 2017). Another example can be seen in (Acuña et al., 2012; Milreu et al., 2014), with the enumeration of metabolic stories. A metabolic story is defined by the authors as the set of reactions that summarize the flow of matter from a set of source metabolites to a set of target metabolites and is characterized as a maximum directed acyclic subgraph connecting the metabolites of interest. One of the drawbacks of this approach is that a metabolic story is acyclic and thus, it is not possible to obtain sets of reactions that contain cycles. Nevertheless, cycles are common in metabolic networks and this assumption does not reflect reality. Additionally, the method does not take into account the stoichiometry of the reactions, which can lead to a set of unfeasible reactions in practice.

Metabolite concentrations have also been used to assess the responses to small perturbations in the context of constraint-based models (Palsson, 2000; Covert and Palsson, 2003; Klamt et al., 2014), and has been reviewed in detail by (Topfer et al., 2015). While standard flux balance analysis (FBA) tries to predict the flux distribution for one specific steady-state condition, dynamic FBA, as described in (Mahadevan et al., 2002), has been extensively used in smaller models to predict the evolution of the fluxes and of the metabolite concentrations over time. In (Reznik et al., 2013), the authors provide a method derived from the classical FBA framework, and showed that the variables of the dual problem (the so-called shadow prices, which correspond to the sensitivity of FBA to imbalances in the flux) can indicate if a metabolite is a growth-limiting metabolite in FBA. In (Bordbar et al., 2017) the authors describe the unsteady-FBA method (uFBA), created to integrate dynamic time-course metabolomics with a constraint-based metabolic model, allowing a bypass into the steady-state assumption for intracellular metabolites that are measured. In (Rohwer and Hofmeyr, 2008; Christensen et al., 2015), methods are presented to identify regulatory metabolites and paths by varying *in silico* their known concentrations in a measured steady-state using supply-demand analysis. Therefore, these methods are based on the response of an organism to a relatively small perturbation and on the influence of the metabolite concentrations on the reaction rates of the system to return to the original equilibrium.

In this paper, we focus not on the metabolite pools in one condition but on the difference of the obtained measurements between two conditions, which could be measured either within shorter or longer timeframes, depending on the biological question to be addressed. We also do not need neither comprehensive time-course datasets nor coupled data from the relative expression of genes or proteins, which are much harder to obtain. Our main hypothesis is that the difference of metabolite pools between two metabolic states can provide information on the transient state, that is, on the transition between the two measured conditions.

Similar problems have been studied in the literature. In (Sajitz-Hermstein et al., 2016), the authors provide a method (IREMET-FLUX) to integrate relative metabolomic measurements in order to make predictions about differential fluxes. They use a constraint-based approach which minimizes the distance between the two flux vectors of the two different states based on the ratio between the measured metabolite concentrations in both conditions. For both states, steady-state is assumed for the flux vectors. However, the authors identify differential fluxes between the two conditions whereas we aim at finding reactions that are likely active during the transient state. In (Case et al., 2016), the authors investigated reachability problems in chemical reaction networks. Given two different states of the network, the goal is to identify a path that leads the network from the first state to the second one. They prove that this problem can be solved in polynomial time. However, they also discuss that a variant of this problem in which the maximum size of the path is fixed is more difficult to solve. Our approach overcomes this limitation at the same time that it minimizes the number of active reactions in the solutions, since we are interested in identifying only the parts of the network that are potentially active during the transient state. Even though other methods could be adapted to answer this problem, our objective is much simpler, requiring less computational complexity. By reformulating our problem in a simpler way we can also address larger genome-scale metabolic models, instead of focusing on smaller portions of the metabolism (e.g., core models).

We use constraint-based modeling to enumerate sets of reactions that explain the changes in concentrations for some measured metabolites, i.e., how the system moved from a state to another. We implemented our approach in a software we called TOTORO (for "Transient respOnse to meTabOlic pertuRbation inferred at the whole netwOrk level"), that is publicly available at https://gitlab.inria.fr/erable/totoro, along with the test datasets presented in this study. It is implemented in C++ and depends on IBM CPLEX which is freely available for academic purposes. We also tested our method with data from pulse experiments with different carbon sources (glucose, pyruvate and succinate) in *Escherichia coli*.

## 2 METHODS

A metabolic network can be represented as a weighted directed hypergraph $H(\mathcal{V}, \mathcal{R}, \mathcal{S})$ where $\mathcal{V}$ is the set of vertices, $\mathcal{R}$ the set of hyperarcs and $\mathcal{S}$ the stoichiometric matrix representing weights

on the hyperarcs. Each $c \in \mathcal{V}$ represents a metabolite of the network and each hyperarc $r \in \mathcal{R}$ a reaction that connects two sets of disjoint metabolites $Subs_r$, $Prod_r$ with $Subs_r, Prod_r \subseteq \mathcal{V}$. To each hyperarc, a set of weights is associated representing the stoichiometric coefficients of the metabolites participating to the corresponding reaction. These weights are given by the stoichiometric matrix $\mathcal{S}$ which is a $m \times n$ matrix where each column represents a reaction and each row a different metabolite. It contains the stoichiometric coefficients which are positive if a metabolite is produced by a reaction and negative if it is consumed.

The set $X \subseteq \mathcal{V}$ contains all measured metabolites. The metabolomic data is given as a list which, for each measured metabolite in $X$, contains an interval. This interval describes by how much the internal metabolite concentration changed between two different states. Usually, small deviations for the measurements are available which can be used to calculate the minimum and the maximum possible difference between the internal metabolite concentrations. Furthermore, all reversible reactions of the network are split into forward and backward reactions.

We are interested in solving the following problem: Given a network $H$ and a list containing the changes for some metabolite concentrations before and after a perturbation, we want to identify sets of reactions that were involved in diverting the system from the initial state before the perturbation to the state after the perturbation (**Figure 1A**). Here, we present a constraint-based approach to solve this problem where the change of concentrations ($\Delta$) between two states is represented as an interval.

## 2.1 Core Method

The variation of the concentrations in time of the metabolites in $X$ can be written as:

$$\frac{dX}{dt} = (\mathcal{S} \cdot v)_X. \tag{1}$$

In this equation, $v$ is a flux vector and the $(\cdot)_X$ operator means that only the entries of the vector corresponding to the metabolites in $X$ are taken into account. We use $[X]_t$ to denote the concentration for the metabolites in $X$ at time point $t$. Considering two points $t_0$ and $t_1$ in time and $\Delta_X = [X]_{t_1} - [X]_{t_0}$, one can write:

$$\Delta_X = \mathcal{S} \cdot \varphi. \tag{2}$$

In this case, each entry of the vector $\varphi$ can be interpreted as the overall number of moles that passed through the reaction $j$ during the time interval $[t_0, t_f]$ which corresponds to the area under the reaction rate curve in this time interval:

$$\varphi_j = \int_{t_0}^{t_1} v_j(t) \cdot dt. \tag{3}$$

Due to biological and technical variability that can arise from different replicates of the same experiment, we assume that the measured variations in concentrations of the metabolites in $X$ are represented by an interval rather than using a fixed number:

$$\Delta_X = [\Delta_X^{\min},\ \Delta_X^{\max}]. \tag{4}$$

Furthermore, for the non-measured metabolites, we do not know if their concentration changed or not. Therefore, similarly to the approach of UFBA (Bordbar et al., 2017) and their 'node relaxation' to allow for changes in non-measured metabolites, we assume that a variation ($\epsilon$) is possible for all non-measured metabolites $\bar{X} = \mathcal{V} \backslash X$:

$$\Delta_{\bar{X}} = [\epsilon^{\min},\ \epsilon^{\max}]. \tag{5}$$

Based on these assumptions, we can model the production or consumption of metabolites between two states by the following constraints:

$$\begin{aligned} &\Delta^{\min} \le \mathcal{S} \cdot \varphi \le \Delta^{\max} \\ &0 \le \varphi_j \le u_j \qquad\qquad \forall j \in \mathcal{R}. \end{aligned} \tag{6}$$

All $\varphi_j$ are positive and have an upper bound $u_j$. We have that $\Delta^{\min}$ is a vector composed of $\Delta_X^{\min}$ and $\epsilon^{\min}$ while $\Delta^{\max}$ is composed of $\Delta_X^{\max}$ and $\epsilon^{\max}$.

As showed above, in our formulation, the variable $\varphi$ can only be zero or have a positive value. For this, we use an additional constraint as explained in **Section 2.2** in order to prevent both forward and reverse senses of reversible reactions from being picked in any given solution. However, this means that we do not know if the activity of the corresponding reaction was increased or decreased during the shift compared to the initial steady state. We only know that if $\varphi_j$ is zero in the solution, reaction $j$ is proposed as inactive during the shift while if $\varphi_j$ has a non-zero value, reaction $j$ is proposed as active during the shift. Hence, we are only interested in the reactions that have a non-zero $\varphi$ because we want to identify the part of the metabolic network that was active during the metabolic shift. These reactions are represented by the support of the vector $\varphi$.

## 2.2 Minimizing the Number of Reactions and the Variation of the Concentrations for the Non-Measured Metabolites

Since the number of possible paths that can explain the measured metabolic shifts can be very large, we will focus on finding the smallest solutions with regard to the number of active reactions that still explain the metabolic shift. This corresponds to the parsimonious assumption that the fewest possible resources are used or the smallest changes are made. Thus, we are interested in identifying minimum sets of reactions that play a major role in the metabolic shift. For each reaction $j$, a binary variable $y_j$ is then introduced that is set to zero if and only if the corresponding $\varphi_j$ is zero and therefore, the reaction is not part of the solution. In this way, these variables will correspond to the support vector of $\varphi$ and it will be sufficient to minimize their sum:

$$\begin{aligned} &y_j = 0 \leftrightarrow \varphi_j = 0 \quad \forall j \in \mathcal{R} \\ &y_j \in \{0, 1\}. \end{aligned} \tag{7}$$

Additionally, to prevent that both a reaction $j$ and its reversible $\bar{j}$ can be picked at the same time for one solution, the following constraint is used:

$$y_j + y_{\bar{j}} \le 1 \quad \forall (j, \bar{j}) \in \mathcal{R}. \tag{8}$$

To minimize the number of reactions that are part of the solution, the objective function is written as:

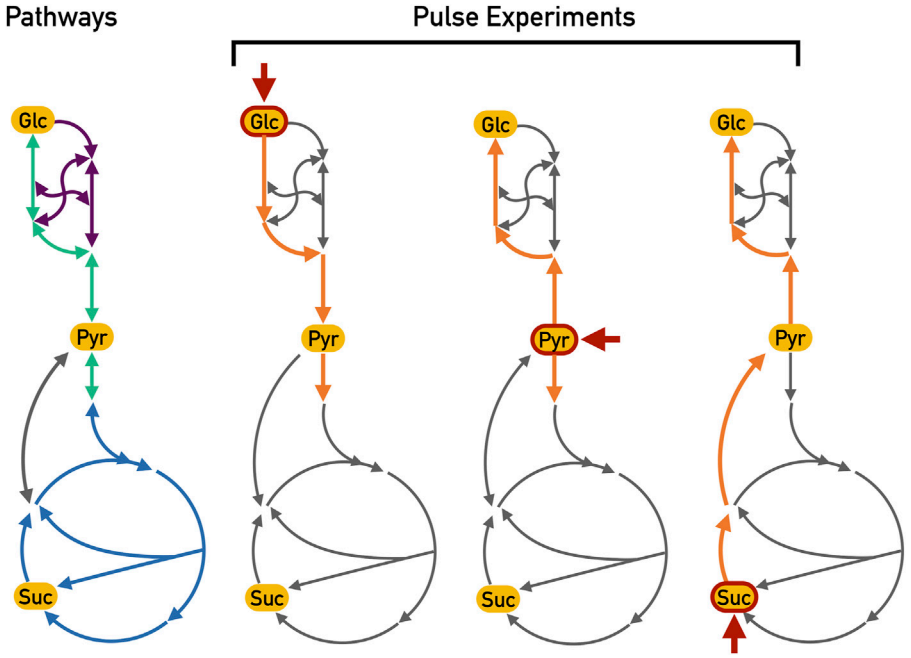


**FIGURE 2 |** Expected active reactions for different pulse experiments. These essential reactions along with their expected directions are highlighted in orange whereas other non-essential reactions (but which nonetheless could be chosen) are depicted in grey. Each pulse is indicated by the short red arrow (Glc: glucose; Pyr: pyruvate and Suc: Succinate). During the glucose pulse, the glycolysis reactions (depicted in green) should be active in order to generate ATP from the hydrolysis of glucose. On the other hand, the pyruvate and succinate pulse experiments should show gluconeogenesis activation (also depicted in green but in the opposite sense), generating glucose-6-phosphate from these two carbon sources. Furthermore, the TCA cycle (depicted in blue) can be fed from pyruvate during the pyruvate and glucose pulses. During the succinate pulse, the overflow in the TCA cycle should lead to the production of pyruvate with a subsequent activation of gluconeogenesis to produce biomass precursors. The pentose phosphate pathway (depicted in purple) is most likely active in all pulses in order to generate biomass precursors; however, since this pathway is a mere interconversion of carbohydrates, there is no particular expectation as to the actual direction of these reactions.

$$\min \sum_{j=1}^{m} y_j. \tag{9}$$

However, we are not only interested in minimizing the number of reactions in the solution but also in minimizing the variation in concentration for the non-measured metabolites $\bar{X}$. Since the measured compounds are usually the more important ones for analyzing the biological experiment, it is reasonable to aim for solutions where other compounds do not accumulate or deplete a lot. This leads to the following minimization:

$$\min \sum_{i \in \bar{X}} |(\mathcal{S} \cdot \varphi)_i|. \tag{10}$$

On the other hand, we are trying to explain as much change in the concentration as possible for the measured metabolites:

$$\max \sum_{i \in X} |(\mathcal{S} \cdot \varphi)_i|. \tag{11}$$

To combine both ideas in one objective function, a weight $\lambda$ is used for both objectives:

$$\min \lambda \sum_{j=1}^{m} y_j + (1-\lambda) \sum_{i \in \bar{X}} |(\mathcal{S} \cdot \varphi)_i| - (1-\lambda) \sum_{i \in X} |(\mathcal{S} \cdot \varphi)_i|. \tag{12}$$

The value for $\lambda$ should lie between 0 and 1. Finding a good balance between these two objectives can be challenging but necessary to identify meaningful biological solutions (for a schematic representation of TOTORO, see **Figure 1A**). A toy network example is provided in **Figure 1B** to show the influence of parameters $\lambda$ and $\epsilon$ on the solutions. This will be further discussed in the following sections.

Summing up, the mixed-integer linear program (MILP) that is implemented in our software TOTORO is the following:

$$\begin{aligned} \min_{\varphi, y} \quad & \lambda \sum_{j=1}^{m} y_j + (1-\lambda) \sum_{i \in \bar{X}} |(\mathcal{S} \cdot \varphi)_i| - (1-\lambda) \sum_{i \in X} |(\mathcal{S} \cdot \varphi)_i| \\ s.t \quad & \Delta^{\min} \le \mathcal{S} \cdot \varphi \le \Delta^{\max} \\ & 0 \le \varphi_j \le u_j && \forall j \in \mathcal{R} \\ & y_j = 0 \leftrightarrow \varphi_j = 0 && \forall j \in \mathcal{R} \\ & y_j + y_{\bar{j}} \le 1 && \forall (j, \bar{j}) \in \mathcal{R} \\ & y_j \in \{0,1\}; \lambda \in (0,1); u_j, \varphi_j \in \mathbb{R}. \end{aligned} \tag{13}$$

## 2.3 Enumerating Different Solutions

To enumerate different solutions, once a solution is found, it must be excluded for the next iteration. Two solutions are different if they do not contain the same reactions. We are using the following constraint where $y^*$ is a previously found solution vector:

$$\sum_{j \in \mathcal{R}:\ y_j^* = 1} y_j \le \sum_{j=1}^{m} y_j^* - 1. \tag{14}$$

This prevents that the exact same combination of reactions gets chosen again. Afterwards, we can solve the updated MILP again to compute a different solution. We repeat this process until no more new solutions can be found or until a desired number of solutions has been computed.

## 2.4 Dealing With Source/Sink Reactions and Non-Measured Metabolites

Source and sink reaction (i.e., reactions that have only products or only substrates) of the network should be blocked to avoid that changes in the concentration are just transferred outside of the network where they cannot be taken into account by the objective function. However, no information is lost if source and sink reactions are blocked. If the substrates of a sink reaction are accumulated or the products of a source reaction are depleted in a solution, this indicates that the corresponding source/sink reaction is active. Their use is limited by the chosen $\epsilon$ but it can be set to a very low or large value to imitate an infinite source or sink. Hence, specific sources or sinks can be added to the problem by specifying a large negative $\Delta^{\min}$ or a large positive $\Delta^{\max}$ for certain metabolites, but the method will remain robust to small variations, as long as the range of this parameter remains within a similar order of magnitude of the values of the measured metabolites.

However, if the minimization of the number of active reactions is prioritized ($\lambda \approx 1$) and the value of $\epsilon$ for the non-measured metabolites is higher than the one for the measured metabolites, the changes in concentration of the measured metabolites can simply be distributed to (accumulated on or taken from) the nearby non-measured metabolites (**Figure 1B**, $\epsilon > 0$, $\lambda = 0.9$) and prevents that larger sub-hypergraphs are chosen (which would instead connect several measured metabolites and explain how the depletion of one measured metabolite leads to the accumulation of another measured metabolite, or vice-versa). However, this can be addressed by decreasing the value of $\lambda$ in the objective function and thereby giving more weight to the portion of the function that minimizes the accumulation in non-measured metabolites **Figure 1B**, $\epsilon > 0$, $\lambda = 0.1$). This should result in solutions that are larger but that connect the measured metabolites better than when only the number of reactions is minimized. Furthermore, based on other experimental data, the user might choose smaller values of $\epsilon$, or constrain it to the highest measured metabolite to further restrict the accumulation/depletion of the non-measured metabolites.

# 3 RESULTS

To evaluate our approach, we used data from different pulse experiments with different carbon sources in *E. coli* as presented in (Taymaz-Nikerel et al., 2013). The authors measured the internal concentrations for several metabolites for a glucose baseline and for glucose, pyruvate and succinate pulse experiments. These data were used to apply the method on the *E. coli* core model (Orth et al., 2010) and the *E. coli* iJO1366 model (Orth et al., 2011) available from the BiGG database (King et al., 2015b). The *E. coli* core model consists of 72 metabolites and 95 reactions, the *E. coli* iJO1366 model of 1,805 metabolites and 2,583 reactions.

We were interested in the difference between the glucose baseline and the pseudo-steady state which was achieved in about 30–40s after each pulse. In (Taymaz-Nikerel et al., 2013), the authors provided the internal concentrations for the baseline, including the deviations for their measurements and the fold changes for the three different pseudo-steady states which we used to calculate the internal concentrations for each pseudo-steady state. In (Taymaz-Nikerel et al., 2013), deviations for the measured concentration of the glucose baseline are given that were derived from several replicates of the same experiment. We used them to be able to calculate the minimum difference $\Delta_X^{\min}$ and maximum difference $\Delta_X^{\max}$ in the concentrations between the glucose baseline and each pseudo-steady state. A detailed explanation can be found in the **Supplementary Material Section S1.1**. The calculated $\Delta_X^{\min}$ and $\Delta_X^{\max}$ for all three pulse experiments can be found in the **Supplementary Tables S1–S3**.

We used all measured metabolites that are present in the network and that had a significant change in their concentration as input. It should be noted that a change for each given metabolite must be either positive or negative. For further details, see the **Supplementary Material Section S1.1**.

Furthermore, source and sink reactions cannot be chosen as part of the solution and therefore glucose, pyruvate and succinate were added as sources for the corresponding pulse experiments. Oxygen was added as another source because in (Taymaz-Nikerel et al., 2013), the authors identified increased oxygen uptake rates during the pulse experiment. To allow unlimited growth, the biomass was added as sink.

The expected active reactions in the core metabolism of *E. coli* are displayed in **Figure 2** for each pulse experiment.

## 3.1 *E. coli* Core Model

At first, the method was applied using the *E. coli* core model. To better understand how the different parts of our model impacted the solutions, we did several runs with different values for $\lambda$ (0.1, 0.5, and 0.9) and $\epsilon$ (5 and 10) for each pulse experiment. Although a single solution should be enough to identify some pathways responsible for the shift, we wanted to see if we could also obtain alternative pathways. Furthermore, we wanted to investigate how the solutions evolve when they are slightly suboptimal. For each different parameter setting, 100 different solutions were therefore enumerated. The results are displayed using *Escher* (King et al., 2015a) in the Supplementary Figures S1to S18.

In general, we could observe that solutions with $\lambda = 0.1$ were preferable since usually the goal is to have a final solution which is overall more connected. In this way, we were able to extract connected sub-hypergraphs that resemble complete biological pathways which played a role during the metabolic shifts. This was the case for all three pulse experiments. A higher $\lambda$ led to solutions that were less connected since the optimizer prioritizes solutions with fewer active reactions, and depending on the case, it might be harder to interpret these solutions biologically. Nevertheless, the user is able to fine-tune the number of reactions in the final solution and the degree of connectivity (for instance, if the goal is to highlight only parts of the complete metabolic network instead of finding a connected sub-hypergraphs).

By adjusting the parameters $\lambda$ and $\epsilon$, TOTORO could propose connected sub-hypergraphs for all three pulse experiments. The

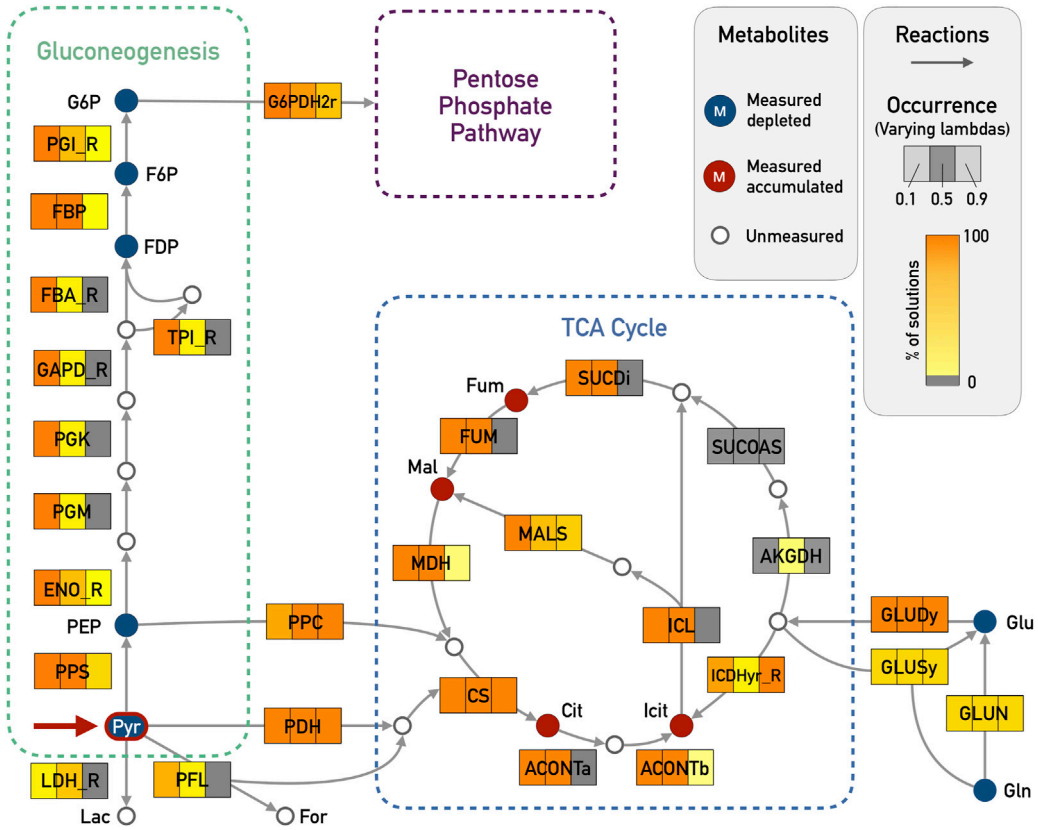


**FIGURE 3 |** *E. coli* core model - Results for Gluconeogenesis and TCA Cycle in the pyruvate pulse (red arrow in Pyr) with $\epsilon = 5$ and varying $\lambda$ (0.1, 0.5, and 0.9). The metabolites that were given as input are highlighted in blue if the corresponding input deltas were below zero and red if they were above zero. Reactions that are highlighted in orange were chosen in almost all of the enumerated solutions, while light yellow corresponds to very few occurrences (less than 5%). Reactions in gray were not chosen in any solution. The expected reactions of the gluconeogenesis and part of the TCA cycle are active in all 100 solutions for $\lambda = 0.1$. The reversible reactions of the gluconeogenesis were chosen in the correct direction. For simplicity reasons, side compounds and cofactors were excluded from the figure. Abbreviations for metabolites: G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; FDP, fructose-biphosphate; PEP, phosphoenolpyruvate; Pyr, pyruvate; Lac, lactate; For, formate; Mal, malate; Fum, fumarate; Cit, citrate; Icit, isocitrate; Glu, glutamate; Gln, glutamine; Abbreviations for reaction names (_R indicates the reverse direction of a reversible reaction within the original model): G6PDH2r, glucose 6-phosphate dehydrogenase; PGI, glucose-6-phosphate isomerase; FBP, fructose-bisphosphatase; FBA_R, fructose-bisphosphate aldolase; TPI, triose-phosphate isomerase; GAPD, glyceraldehyde-3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; ENO, enolase; PPS, phosphoenolpyruvate synthase; LDH, D-lactate dehydrogenase; PFL, pyruvate formate lyase; PPC, phosphoenolpyruvate carboxylase; PDH, pyruvate dehydrogenase; CS, citrate synthase; ACONTa, Aconitase (half reaction A); ACONTb, Aconitase (half reaction B); ICDHyr, Isocitrate dehydrogenase; AKGDH, 2-Oxoglutarate dehydrogenase; SUCOAS, Succinyl-CoA synthetase; SUCDi, Succinate dehydrogenase; FUM, fumarase; MDH, malate dehydrogenase; ICL, isocitrate lyase; MALS, malate synthase; GLUDy, glutamate dehydrogenase; GLUSy, glutamate synthase; GLUN, glutaminase.

predicted solutions did not use co-factors as shortcuts through the network. We therefore did not modify our method further to treat co-factors separately.

### 3.1.1 Pyruvate Pulse

For the pyruvate pulse, we expected that the activity of the TCA cycle would increase and that reactions for gluconeogenesis would be active (see **Figure 2**). Both observations could be reproduced with a $\lambda = 0.1$ (see **Figure 3** for a comparison of the values of $\lambda$), while higher values of lambda constrained the solutions locally around the measured metabolites. For $\lambda = 0.9$, neither the TCA cycle nor the gluconeogenesis pathway were proposed to be active. Setting $\lambda$ to 0.5 already improved the results: the TCA cycle was proposed as active but the complete gluconeogenesis pathway was only recovered in less than 50% of the solutions.

The four measured metabolites citrate, isocitrate, L-malate and fumarate had positive input deltas and could thus be used as sinks. The results showed how the TCA cycle can be fed from pyruvate either by the phosphoenolpyruvate carboxylase (PPC) or by the combination of pyruvate dehydrogenase (PDH) and citrate synthase (CS). Furthermore, the pathway from pyruvate to glucose 6-phosphate (G6P) was active in 100% of solutions for $\lambda = 0.1$. The pathway from pyruvate to G6P contains nine reactions including seven reversible ones: glucose-6-phosphate isomerase (PGI), fructose-bisphosphate aldolase (FBA_R), triose isomerase (TPI), glyceraldehyde-3-phosphate dehydrogenase (GAPD), phosphoglycerate kinase (PGK), phosphoglycerate mutase (PGM) and enolase (ENO). Especially here, it is important to state that all these reversible reactions were predicted in the correct direction going from pyruvate towards G6P. The core

**TABLE 1 |** Comparison of different objective values for the best runs for each experiment. Since we are not fixing the objective value of the first solution in our optimization problem, the objective values for the subsequent solutions can be worse. In this table, we are comparing the difference in the objective values between the first solution and the 100th solution. In addition to the absolute differences, also the percentage of how much the objective value worsened compared to the first solution is displayed. The underlying optimization problem is a minimization problem. Therefore, smaller objective values are better.

| Pulse experiment | 1st sol. | 100th sol. | Abs. diff. | % |
|---|---|---|---|---|
| Pyruvate ($\lambda = 0.1$, $\epsilon = 5$) | −32.139 4 | −30.663 5 | 1.48 | 5.5 |
| Glucose ($\lambda = 0.1$, $\epsilon = 1.2$) | 5.383 0 | 6.558 2 | 1.18 | 21.8 |
| Succinate ($\lambda = 0.1$, $\epsilon = 5$) | −158.177 0 | −157.576 0 | 0.60 | 0.4 |

network results can be seen in **Supplementary Figures S1–S6**, with varying $\lambda$ and $\epsilon$. These figures were created using *Escher* (King et al., 2015a).

We do not fix the objective value in our optimization problem after obtaining the first solution but in every iteration, the minimization problem is solved again after excluding the newly found solution. This means that the next solution can have the same objective value but it is also possible that the objective value is worse than in the previous iteration. In this particular case, the 100th solution had an objective value that was only 5.5% worse than the objective value of the first solution (see **Table 1**) which shows that, as concerns optimality, all 100 solutions were very similar. They also had very similar active reactions. Comparing the 100 enumerated solutions for $\lambda = 0.1$ and $\epsilon = 5$, a total of 43 reactions with a specific direction were chosen in all solutions. Out of these 43 reactions, 24 were chosen in every solution (including reactions in the TCA cycle and the gluconeogenesis pathway). This means that certain core pathways were consistently picked also in slightly suboptimal solutions. Looking at only the ten best solutions, already 38 out of the 43 reactions were identified. The missing reactions were mostly part of the pentose phosphate pathway which also contains reactions that were part of the solution only in a few cases. Even with only ten solutions, we were able to obtain the alternative pathways feeding the TCA cycle (PPC/PDH). This indicates that it is not necessary to enumerate a large amount of solutions to get significant results and to identify alternative biological pathways.

To check the robustness of the method against small perturbations, we tested within the pyruvate pulse the results of TOTORO for the values of $\lambda = 0.1$ and 0.9, excluding one metabolite at a time, recomputing the results, and computing the distances to the results on the complete metabolite set for reaction occurrence (in terms of absolute difference of occurrences). Overall, the results for both $\lambda = 0.1$ and 0.9 differed from less than 5% to around 20%. In general, the results were robust ($< 10\%$ in average distance) for 70–80% of the metabolites tested (with $\lambda = 0.1$, and 0.9, respectively), but we noticed that excluding metabolites with a higher neighborhood connectivity (such as glutamate and glutamine) had a greater impact on the final results. These results show that even though the distances were small, the amount of information provided by different metabolites varied widely.

Moreover, we tested 10 random sets of measured metabolites (**Supplementary Tables S4, S5**), with a varying number of excluded metabolites to detect at which point the method would not behave as with the complete dataset. In accordance with the previous results for single exclusions, and within the tests with less than 50% of the measured metabolites excluded, the smallest distance ($\approx 10\%$) to the complete results came from a random dataset which included glutamate, glutamine and pyruvate (to ensure the carbon source uptake). As expected, when more than 50% of the measured metabolites were excluded, we detected a much higher difference ($\approx 40\%$) between the results from the complete dataset and those from the random datasets.

### 3.1.2 Glucose Pulse

For the glucose pulse, we expected that reactions that are part of the glycolysis pathway would be active as they convert glucose into pyruvate generating energy. Consequently, the TCA cycle should also be fed (see **Figure 2**). For $\lambda = 0.9$ and 0.5, the active reactions proposed by TOTORO were disconnected and it was not possible to identify active pathways. We believe that the results coming from this pulse were less insightful since the bacteria were already grown in glucose prior to the pulse, which in turn might be a reason why the changes in metabolites were not as informative as the other pulses. This was for the most part corrected if more metabolites were added as input to TOTORO when using the complete network as presented in **Section 3.2**. This also shows the importance of careful experimental design and how subtle perturbations may generate results that are not always homogeneous.

Even for $\lambda = 0.1$ and $\epsilon = 5$, only disconnected parts of the network were active (see **Supplementary Figure S9**). Since we were interested in testing the method to obtain more connected sub-hypergraphs, we decided to fine-tune the solutions by lowering the value of $\epsilon$ as much as possible. The result for $\epsilon = 2$ and 1.2 can be found in **Supplementary Figures S10, S11**, respectively. Lowering the value of $\epsilon$ to 1.1 rendered the underlying optimization problem infeasible. For $\epsilon = 1.2$, we got solutions that linked intermediate metabolites of the glycolysis pathway to the TCA cycle through the PPC reaction. In some solutions, the TCA cycle was also fed by PDH and CS to account for the accumulation of citrate. As previously mentioned, when the solutions are disconnected and this is unwanted, decreasing the value of $\epsilon$ can sometimes help to obtain more connected solutions. However, this should be used carefully in order to avoid linking unrelated and distant metabolites, which might not be meaningful biologically.

The 100 solutions were very similar ($\lambda = 0.1$, $\epsilon = 1.2$). They accounted for a total of 47 reactions (with distinct directions) and 30 of these appeared in all solutions. Similarly to the pyruvate pulse, the difference in these solutions were mostly based on a few reactions that are not part of the main pathways (glycolysis/TCA cycle). One critical observation is that the D-glucose transport reaction (GLCpts) was not part of every solution although glucose should be used as important source. As previously mentioned, the bacteria were already grown in glucose prior to the glucose pulse, which is possibly a reason why glucose was already internalized prior to the initial pulse. When comparing the objective values for these 100 solutions, the absolute difference between the first solution and the 100th one was similar to the one observed

for the pyruvate pulse (see **Table 1**). However, proportionally this value was 21.8% worse than for the first solution. When we repeated the run for $\lambda = 0.1$ and $\epsilon = 1.2$ with 50 iterations, the D-glucose transport reaction was part of 42 solutions. For ten iterations, this reaction was picked in all ten solutions. Hence, the glucose transport reaction was active in solutions with the best objective values. This showed that although the solutions remained very similar, there was a decline in their quality. And similarly to the pyruvate pulse, we saw that it is not necessary to enumerate a large amount of solutions.

### 3.1.3 Succinate Pulse

After the succinate pulse, part of the TCA cycle should always be active. Furthermore, the gluconeogenesis pathway should be active to produce G3P and glucose-6-phosphate from succinate. Again, the results for $\lambda = 0.5$ and 0.9 led to smaller solutions that were more disconnected (see **Supplementary Figures S13–S16**). Therefore, we focused on the analysis of the results for $\lambda = 0.1$ (see **Supplementary Figures S17, S18**). For both $\epsilon = 5$ and 10, succinate entered the TCA cycle and turned into oxaloacetate. TOTORO proposed two possibilities to output the excess of the TCA cycle: Either phosphoenolpyruvate (PEP) was produced by PEP carboxykinase (PPCK) or by PEP synthase (PPS) using pyruvate as intermediate substrate. Subsequently, PEP was, as expected, transformed to G3P. The lower right part of the TCA cycle predicted as active can be explained by the fact that the concentration of L-glutamate decreased and the concentration of citrate increased. The active reaction in this part connected these two metabolites. Furthermore, reactions of the pentose phosphate pathway were proposed as active and the biomass precursors R5P, E4P, and G3P were produced.

The results for $\epsilon = 5$ and 10 were very similar. For example, one difference was that for $\epsilon = 10$, the reverse D-lactate dehydrogenase (LDH) was predicted to be active in 56 solutions which led to a small accumulation of D-lactate. It does make sense biologically because in general, D-lactate is one of the main products of the fermentation but we do not have the measurements for the concentration of D-lactate for this pulse experiment to actually verify this observation. However, in total, the differences were negligible and in contrast to the glucose pulse, the parameter $\epsilon$ had a lower impact on the outcome.

Again, the core reactions of all 100 solutions were very similar. In total, 41 reactions (with distinct directions) appeared in all 100 solutions (for $\lambda = 0.1$, $\epsilon = 5$). We observed that 22 of these were always active (mostly in the gluconeogenesis pathway and part of the TCA cycle). The objective values for all 100 solutions were extremely close (see **Table 1**).

## 3.2 *E. coli* iJO1366 Model

Based on the results for the *E. coli* core model, we only did runs with $\lambda = 0.1$ for the *E. coli* iJO1366 model. The inputs were updated because this network contains more metabolites and therefore, more measured metabolites could be added. The amount of iterations was decreased to ten because the runtime in the larger network is significantly higher and we had already established in the core model that it was not necessary to enumerate a larger amount of solutions. To decrease the runtime for each solution, CPLEX was configured differently. The relative MIP gap tolerance was set to 0.05 which means that the solver will stop an iteration if a solution is found that is within 5% of the optimal. This allows for a faster result and we could see in the core model that the first 100 solutions tended to be very similar. This means that even if we are enumerating slightly suboptimal solutions, we should be able to compute solutions that are very similar to the actual optimal solution. If the 5% limit is not reached after 48 h, the iteration is stopped. The memory usage of CPLEX was limited to 10 GB.

The runtime for the different pulse experiments differed a lot. The results for the pyruvate and glucose pulses were computed on a cluster. For the pyruvate pulse, the 5% limit was reached only in three iterations (see **Supplementary Table S6**). All other iterations were stopped after 48 h. However, all solutions obtained were within 7% of the optimum. Thus, we still took them into account when analyzing the predicted active reactions. In none of the iterations for the glucose pulse, the 5% limit was reached. The obtained solutions were within 8.5% of the optimal value (see **Supplementary Table S7**).

In contrast to the pyruvate and the glucose pulses, the 5% limit was reached in all iterations for the succinate pulse and computing all ten solutions took less than 5 min on a personal machine (2.90 GHz Intel i7-7820HQ CPU, 16 GB RAM). This shows that the constraints describing the input deltas in the MILP have a large influence on the difficulty of the optimization problem, and thus also on the runtime.

However, although the obtained solutions were suboptimal, the active reactions predicted by TOTORO for the core metabolism were similar to the best results of the *E. coli* core model for all three pulse experiments. For instance, in the pyruvate pulse results, out of 12 reactions in the TCA cycle within the large network, 8 were also present in the core model. In total, 5 were chosen in 100% of the solutions in the same direction in both core and large networks. The complete network was also able to correct the only inconsistency within the TCA cycle for the core network: the direction of the reaction ICDHyr, which shows the advantage of relying on complete networks whenever available. For the glycolysis/gluconeogenesis pathways, out of 12 reactions, 9 were also included in the core model. In total, 6 reactions were chosen in 100% and 1 in more than 80% of the solutions in the same direction in both networks. TOTORO predicted as active for pyruvate, glucose and succinate (in at least 1 solution) a total of 221, 284, and 189 reactions respectively. Moreover, 52% of the reactions were chosen across all iterations in the pyruvate pulse dataset, 81% in the succinate pulse dataset and 62% in the glucose pulse dataset.

The additional measurements that were added as input deltas for the large network were mostly amino acids (see **Supplementary Tables S1–S3**). In (Waschina et al., 2016), the authors show for the example of amino acid production in *E. coli* how the production cost for individual amino acids can depend on the available carbon source, and reactions close to the entry point of the carbon source might have considerably higher fluxes. A schematic representation of this is provided in **Figure 4A**. Indeed, from the experimental data, alanine and valine only accumulated during the pyruvate pulse, and were depleted

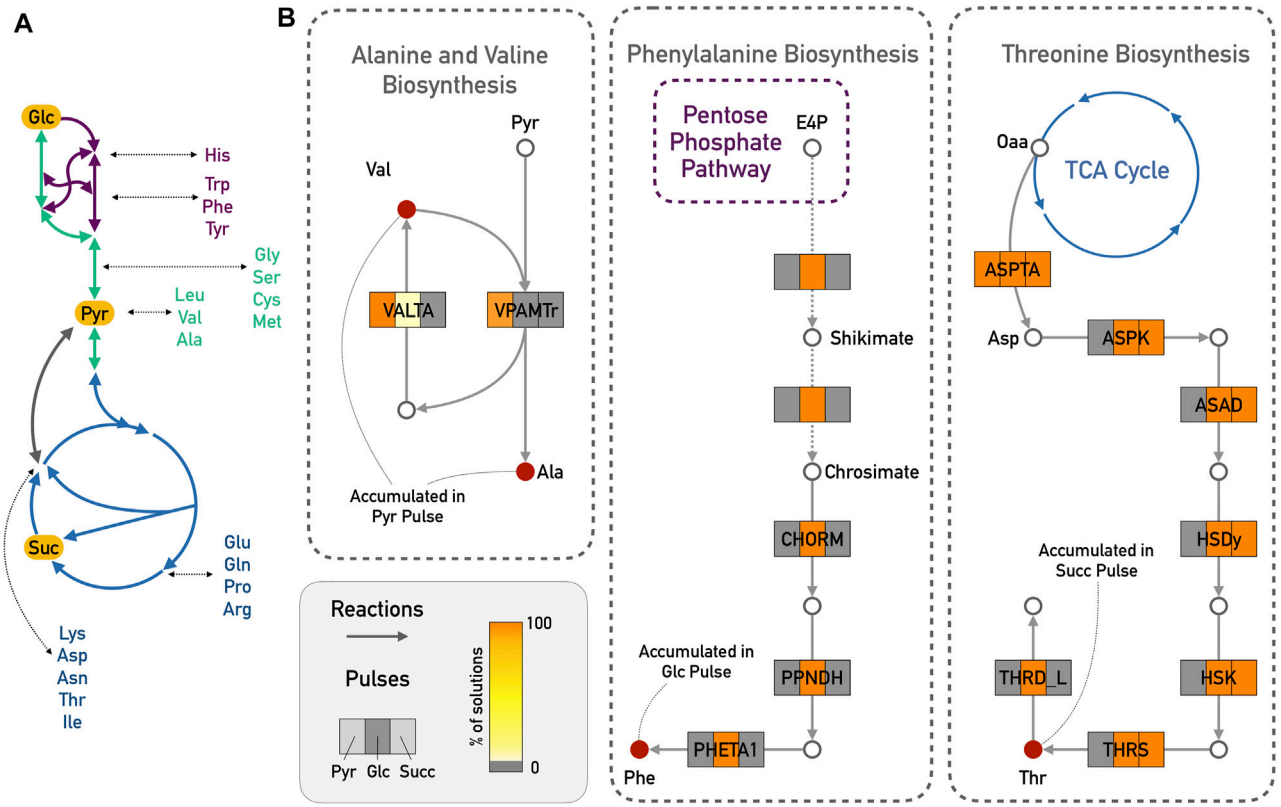


**FIGURE 4 |** Amino acid biosynthesis in the *E. coli* iJO1366 model. **(A)** Schematic representation of carbon sources with closely related amino acids. Glycolysis/Gluconeogenesis in green; TCA cycle in blue and Pentose Phosphate Pathway in purple. **(B)** TOTORO results explaining the accumulation of valine (Val) and alanine (Ala) in the pyruvate (Pyr) pulse; accumulation of phenylalanine (Phe) from the glucose (Glc) pulse and accumulation of threonine (Thr) from the succinate (Succ) pulse. For simplicity reasons, side compounds and cofactors were excluded from the figure. Dashed arrows indicate several reactions from the shikimate and chorismate pathways. Abbreviations for reaction names are as follows: VALTA, valine transaminase; VPAMTr, valine-pyruvate aminotransferase; CHORM, chorismate mutase; PPNDH, prephenate dehydratase; PHETA1, phenylalanine transaminase; ASPTA, aspartate transaminase; ASPK, aspartate kinase; ASAD, aspartate-semialdehyde dehydrogenase; HSDy, homoserine dehydrogenase; HSK, homoserine kinase; THRS, threonine synthase; THRD_L, L-threonine deaminase.

with the other two carbon sources. Pyruvate is a direct precursor for valine production. We therefore expected that reactions of the alanine and valine biosynthesis should play a greater role in the predicted results for pyruvate compared to the other two pulses. TOTORO predicted an activation of the pathway from pyruvate to alanine and valine, which resulted in the accumulation of these amino acids (**Figure 4B**). In accordance with the predictions in (Waschina et al., 2016), another example is the accumulation of threonine during the succinate pulse. Threonine and succinate are closely connected, and TOTORO predicted active reactions leading to its biosynthesis and accumulation in the succinate pulse (**Figure 4B**). Compared to the results for succinate, TOTORO predicted more active reactions consuming threonine during the glucose pulse, and no reactions producing it in the pyruvate pulse, resulting in the depletion of this amino acid with those carbon sources. Moreover, only during the glucose pulse, phenylalanine was accumulated, and TOTORO proposed the complete pathway for the phenylalanine biosynthesis as active when compared to the pyruvate and succinate pulses (**Figure 4B**), in accordance with the predictions in (Waschina et al., 2016) of lower cost to produce this amino acid with glucose as carbon source.

## 4 DISCUSSION

TOTORO was able to predict expected pathways as active based on the differences in the measured concentrations for some internal metabolites for both the *E. coli* core and complete models. We show that in general, it is preferable to use smaller values of $\lambda$ (e.g., $\lambda = 0.1$) though the method is not critically sensible to this setup, being robust to small perturbations. However, it is worth noting that a higher $\lambda$ can lead to smaller solutions which might be biologically irrelevant. Here, we focused in extracting connected sub-hypergraphs that explained the changes in concentration between two different conditions. We also show that a reduction of $\epsilon$ can also be used to obtain more connected solutions. However, there might be situations where the user might be interested in only local changes around the measurements. In this context, it might be advantageous to choose higher values for $\lambda$ and $\epsilon$. We did not encounter problems specific to co-factors which is a known problem when looking for shortest paths in metabolic networks. This is probably due to the fact that we are not only minimizing the number of active reactions in the solutions but also focusing on the changes in the metabolite

concentrations. By splitting reversible reactions, TOTORO was able to predict distinct directions for them.

Both in the core network and in the larger network, we were able to recover biologically meaningful pathways. Additionally, although the larger network contains more reactions and we added more input deltas, the predictions for the core metabolism of *E. coli* were fairly similar to the results for the core network. We also showed a particular case in which the perturbation was subtle, and the results from the complete model were more insightful than the ones from the core model. It must be however noted that the predictions do depend on the measured metabolites. If for large parts of the network, no metabolite concentrations are measured, TOTORO will likely not be able to find active pathways for these parts of the network.

Moreover, we could also see that it is not necessary to enumerate a high number of solutions which is especially important when larger networks are used and the runtime of TORORO increases. We enumerated 100 different solutions for the core network. However, in our case, the enumerated solutions were very similar and a large amount of reactions appeared in all 100 solutions. Therefore, already one (or few) solution(s) would have been sufficient to infer the most important reactions that were proposed to be active.

## 5 CONCLUSION

In this paper, we presented TOTORO, a method that identifies active reactions during the transient state based on the differences in the concentrations for some measured metabolites from two different conditions and we showed its prediction power on the example of different pulse experiments in *E. coli*. It is important to note that even though we provided several biologically trivial results, TOTORO only used metabolomic data as basis for these predictions, without any other source of bias such as defined metabolic pathways. Our method was also able to handle full networks which take into account model stoichiometry, and we did not perform any type of filtering for cycles, reversible reactions or co-factors.

With the current technologies, it gets more common to have different kinds of data available which creates a need for methods that combine, for instance, metabolomic, transcriptomic and proteomic data. We have recently developed a method for integration of metabolic networks and transcriptomic data (Pusa et al., 2019) and we intend in the future to adapt our approaches to be able to integrate multiple kinds of omic data, similarly to what was proposed in (Pandey et al., 2019) for thermodynamic, transcriptomic and metabolomic data, and in (Kleessen et al., 2015) for transcriptomic and metabolomic data. On a larger scale, it might be interesting also to consider whether some measures used in (hyper)graph theory such as connectivity or (hyper)path length might be related to the parameters used and thus provide an automatic and perhaps more reliable way of setting them. Notice that achieving this would be even more challenging in the case of hypergraphs for which such measures might have to be adapted.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

M-FS, AJ-L, RA, MF, SV: Conception of the work; IZ, RA, AJ-L, AM, LD, RC: Constructed the code; MF, IZ, RA: Analysis of datasets; IZ, MF, AJ-L, M-FS: Wrote the manuscript with input of other authors; All authors approved the last version of the manuscript.

## FUNDING

This work was supported by a postdoctoral fellowship from the Agence Nationale de la Recherche (ANR-GREEN 17_CE20_0031_01) granted to MF; Fundação para a Ciẽncia e a Tecnologia (UIDB/50021/2020); Fundação de Amparo à Pesquisa do Estado de São Paulo (#15/13430-9, #15/22308-2) and was part of a PhD funded by Inria. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 951970 (OLISSIPO project).

## ACKNOWLEDGMENTS

This work was performed using the computing facilities of the CC LBBE/PRABI.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.815476/full#supplementary-material

## REFERENCES

Acuña, V., Birmelé, E., Cottret, L., Crescenzi, P., Jourdan, F., Lacroix, V., et al. (2012). Telling Stories: Enumerating Maximal Directed Acyclic Graphs with a Constrained Set of Sources and Targets. *Theor. Computer Sci.* 457, 1–9. doi:10.1016/j.tcs.2012.07.023

Bordbar, A., Yurkovich, J. T., Paglia, G., Rolfsson, O., Sigurjónsson, Ó. E., and Palsson, B. O. (2017). Elucidating Dynamic Metabolic Physiology through Network Integration of Quantitative Time-Course Metabolomics. *Sci. Rep.* 7, 46249. doi:10.1038/srep46249

Cambiaghi, A., Ferrario, M., and Masseroli, M. (2017). Analysis of Metabolomic Data: Tools, Current Strategies and Future Challenges for Omics Data Integration. *Brief Bioinform.* 18, bbw031–510. doi:10.1093/bib/bbw031

Case, A., Lutz, J. H., and Stull, D. M. (2016). "Reachability Problems for Continuous Chemical Reaction Networks," in International Conference on Unconventional Computation and Natural Computation (Springer), 1–10. doi:10.1007/978-3-319-41312-9_1

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., et al. (2018). MetaboAnalyst 4.0: towards More Transparent and Integrative Metabolomics Analysis. *Nucleic Acids Res.* 46, W486–W494. doi:10.1093/nar/gky310

Christensen, C. D., Hofmeyr, J.-H. S., and Rohwer, J. M. (2015). Tracing Regulatory Routes in Metabolism Using Generalised Supply-Demand Analysis. *BMC Syst. Biol.* 9, 89. doi:10.1186/s12918-015-0236-1

Cottret, L., Frainay, C., Chazalviel, M., Cabanettes, F., Gloaguen, Y., Camenen, E., et al. (2018). Metexplore: Collaborative Edition and Exploration of Metabolic Networks. *Nucleic Acids Res.* 46, W495–W502. doi:10.1093/nar/gky301

Covert, M. W., and Palsson, B. O. (2003). Constraints-based Models: Regulation of Gene Expression Reduces the Steady-State Solution Space. *J. Theor. Biol.* 221, 309–325. doi:10.1006/jtbi.2003.3071

Frainay, C., Aros, S., Chazalviel, M., Garcia, T., Vinson, F., Weiss, N., et al. (2019). Metaborank: Network-Based Recommendation System to Interpret and Enrich Metabolomics Results. *Bioinformatics.* 35, 274–283. doi:10.1093/bioinformatics/bty577

Frainay, C., and Jourdan, F. (2017). Computational Methods to Identify Metabolic Sub-networks Based on Metabolomic Profiles. *Brief Bioinform.* 18, 43–56. doi:10.1093/bib/bbv115

Ginsburg, H. (2009). Caveat Emptor: Limitations of the Automated Reconstruction of Metabolic Pathways in Plasmodium. *Trends Parasitology.* 25, 37–43. doi:10.1016/j.pt.2008.08.012

Ivanisevic, J., and Want, E. J. (2019). From Samples to Insights into Metabolism: Uncovering Biologically Relevant Information in Lc-Hrms Metabolomics Data. *Metabolites.* 9, 308. doi:10.3390/metabo9120308

King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., and Palsson, B. O. (2015a). Escher: a Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *Plos Comput. Biol.* 11, e1004321. doi:10.1371/journal.pcbi.1004321

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., et al. (2015b). BiGG Models: A Platform for Integrating, Standardizing and Sharing Genome-Scale Models. *Nucleic Acids Res.* 44, D515–D522. doi:10.1093/nar/gkv1049

Klamt, S., Hädicke, O., and von Kamp, A. (2014). "Stoichiometric and Constraint-Based Analysis of Biochemical Reaction Networks," in Large-scale networks in engineering and life sciences (Springer), 263–316. doi:10.1007/978-3-319-08437-4_5

Kleessen, S., Irgang, S., Klie, S., Giavalisco, P., and Nikoloski, Z. (2015). Integration of Transcriptomics and Metabolomics Data Specifies the Metabolic Response of Chlamydomonas to Rapamycin Treatment. *Plant J.* 81, 822–835. doi:10.1111/tpj.12763

Kuo, T.-C., Tian, T.-F., and Tseng, Y. J. (2013). 3omics: a Web-Based Systems Biology Tool for Analysis, Integration and Visualization of Human Transcriptomic, Proteomic and Metabolomic Data. *BMC Syst. Biol.* 7, 64. doi:10.1186/1752-0509-7-64

Mahadevan, R., Edwards, J. S., and Doyle, F. J. (2002). Dynamic Flux Balance Analysis of Diauxic Growth in *Escherichia coli*. *Biophysical J.* 83, 1331–1340. doi:10.1016/s0006-3495(02)73903-9

Marco-Ramell, A., Palau-Rodriguez, M., Alay, A., Tulipani, S., Urpi-Sarda, M., Sanchez-Pla, A., et al. (2018). Evaluation and Comparison of Bioinformatic Tools for the Enrichment Analysis of Metabolomics Data. *BMC bioinformatics.* 19, 1. doi:10.1186/s12859-017-2006-0

Milreu, P. V., Klein, C. C., Cottret, L., Acuña, V., Birmelé, E., Borassi, M., et al. (2014). Telling Metabolic Stories to Explore Metabolomics Data: a Case Study on the Yeast Response to Cadmium Exposure. *Bioinformatics.* 30, 61–70. doi:10.1093/bioinformatics/btt597

Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., et al. (2011). A Comprehensive Genome-scale Reconstruction of *Escherichia coli* Metabolism-2011. *Mol. Syst. Biol.* 7, 535. doi:10.1038/msb.2011.65

Orth, J. D., Fleming, R. M., and Palsson, B. O. (2010). Reconstruction and Use of Microbial Metabolic Networks: the Core escherichia Coli Metabolic Model as an Educational Guide. *EcoSal plus.* 4 (1). doi:10.1128/ecosalplus.10.2.1

Palsson, B. (2000). The Challenges of In Silico Biology. *Nat. Biotechnol.* 18, 1147–1150. doi:10.1038/81125

Pandey, V., Hadadi, N., and Hatzimanikatis, V. (2019). Enhanced Flux Prediction by Integrating Relative Expression and Relative Metabolite Abundance into Thermodynamically Consistent Metabolic Models. *Plos Comput. Biol.* 15, e1007036. doi:10.1371/journal.pcbi.1007036

Perez de Souza, L., Alseekh, S., Brotman, Y., and Fernie, A. R. (2020). Network Based Strategies in Metabolomics Data Analysis and Interpretation: from Molecular Networking to Biological Interpretation. *Expert Rev. Proteomics* 17 (4), 243–255. doi:10.1080/14789450.2020.1766975

Pusa, T., Ferrarini, M. G., Andrade, R., Mary, A., Marchetti-Spaccamela, A., Stougie, L., et al. (2019). MOOMIN - Mathematical explOration of 'Omics Data on a MetabolIc Network. *Bioinformatics.* 36, 514–523. doi:10.1093/bioinformatics/btz584

Reznik, E., Mehta, P., and Segrè, D. (2013). Flux Imbalance Analysis and the Sensitivity of Cellular Growth to Changes in Metabolite Pools. *Plos Comput. Biol.* 9, e1003195–13. doi:10.1371/journal.pcbi.1003195

Roessner, U., and Bowne, J. (2009). What Is Metabolomics All about? *Biotechniques.* 46, 363–365. doi:10.2144/000113133

Rohwer, J. M., and Hofmeyr, J.-H. S. (2008). Identifying and Characterising Regulatory Metabolites with Generalised Supply-Demand Analysis. *J. Theor. Biol.* 252, 546–554. doi:10.1016/j.jtbi.2007.10.032

Rosato, A., Tenori, L., Cascante, M., De Atauri Carulla, P. R., Martins dos Santos, V. A. P., and Saccenti, E. (2018). From Correlation to Causation: Analysis of Metabolomics Data Using Systems Biology Approaches. *Metabolomics.* 14, 37. doi:10.1007/s11306-018-1335-y

Sajitz-Hermstein, M., Töpfer, N., Kleessen, S., Fernie, A. R., and Nikoloski, Z. (2016). Iremet-Flux: Constraint-Based Approach for Integrating Relative Metabolite Levels into a Stoichiometric Metabolic Models. *Bioinformatics.* 32, i755–i762. doi:10.1093/bioinformatics/btw465

Sévin, D. C., Kuehne, A., Zamboni, N., and Sauer, U. (2015). Biological Insights through Nontargeted Metabolomics. *Curr. Opin. Biotechnol.* 34, 1–8. doi:10.1016/j.copbio.2014.10.001

Stanstrup, J., Broeckling, C., Helmus, R., Hoffmann, N., Mathé, E., Naake, T., et al. (2019). The Metarbolomics Toolbox in Bioconductor and beyond. *Metabolites.* 9, 200. doi:10.3390/metabo9100200

Taymaz-Nikerel, H., De Mey, M., Baart, G., Maertens, J., Heijnen, J. J., and van Gulik, W. (2013). Changes in Substrate Availability in escherichia Coli lead to Rapid Metabolite, Flux and Growth Rate Responses. *Metab. Eng.* 16, 115–129. doi:10.1016/j.ymben.2013.01.004

Töpfer, N., Kleessen, S., and Nikoloski, Z. (2015). Integration of Metabolomics Data into Metabolic Networks. *Front. Plant Sci.* 6, 49. doi:10.3389/fpls.2015.00049

Waschina, S., D'Souza, G., Kost, C., and Kaleta, C. (2016). Metabolic Network Architecture and Carbon Source Determine Metabolite Production Costs. *Febs J.* 283, 2149–2163. doi:10.1111/febs.13727

Xia, J., Sinelnikov, I. V., Han, B., and Wishart, D. S. (2015). MetaboAnalyst 3.0-making Metabolomics More Meaningful. *Nucleic Acids Res.* 43, W251–W257. doi:10.1093/nar/gkv380

**Conflict of Interest:** Author AJ-L was employed by Soladis GmBH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Galvão Ferrarini, Ziska, Andrade, Julien-Laferrière, Duchemin, César, Mary, Vinga and Sagot. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*