

SOFTWARE

Open Access

# PFA toolbox: a MATLAB tool for Metabolic Flux Analysis



Yeimy Morales<sup>1\*</sup> , Gabriel Bosque<sup>2</sup>, Josep Vehí<sup>1</sup>, Jesús Picó<sup>2</sup> and Francisco Llaneras<sup>1</sup>

## Abstract

**Background:** Metabolic Flux Analysis (MFA) is a methodology that has been successfully applied to estimate metabolic fluxes in living cells. However, traditional frameworks based on this approach have some limitations, particularly when measurements are scarce and imprecise. This is very common in industrial environments. The PFA Toolbox can be used to face those scenarios.

**Results:** Here we present the PFA (Possibilistic Flux Analysis) Toolbox for MATLAB, which simplifies the use of Interval and Possibilistic Metabolic Flux Analysis. The main features of the PFA Toolbox are the following: (a) It provides reliable MFA estimations in scenarios where only a few fluxes can be measured or those available are imprecise. (b) It provides tools to easily plot the results as interval estimates or flux distributions. (c) It is composed of simple functions that MATLAB users can apply in flexible ways. (d) It includes a Graphical User Interface (GUI), which provides a visual representation of the measurements and their uncertainty. (e) It can use stoichiometric models in COBRA format. In addition, the PFA Toolbox includes a User's Guide with a thorough description of its functions and several examples.

**Conclusions:** The PFA Toolbox for MATLAB is a freely available Toolbox that is able to perform Interval and Possibilistic MFA estimations.

**Keywords:** Metabolic Flux Analysis, Interval MFA, Possibilistic MFA, Constraint-based modelling

## Background

The problem of estimating unknown metabolic fluxes in living cells has been tackled using several methodologies. MFA is one of the most extensively and successfully applied approaches to estimating fluxes [1]. Usually MFA refers to <sup>13</sup>C-MFA which uses stable isotopically labeled substrates (e.g., <sup>13</sup>C-labeled glucose) combined with stoichiometric balancing to estimate the metabolic fluxes in steady state systems [2, 3]. However, in this study we refer to non-<sup>13</sup>C-MFA methods. These methods mainly rely on measurements of external fluxes (uptake and production rates) to estimate the flux state of cells. Traditional MFA methods present some limitations when accounting for irreversible reactions [4], underdetermined problems [5], and lack of measurements [6]. To reduce these limitations we have developed Interval [7] and Possibilistic [8] MFA methods, which are well-

sued methodologies for scenarios with limited available data. Their main benefits are the following [6–10]: (a) They can consider the irreversibility of the reactions and other inequality constraints. (b) They are able to represent the measured fluxes as intervals and even distributions to describe the uncertainty of the system. (c) They provide interval estimates, which are more reliable and more informative than pointwise solutions, particularly when multiple flux values are possible. (d) They are able to perform estimations in scenarios of high uncertainty or lack of measurements, being those estimates as reliable as possible. In addition, (e) Possibilistic MFA allows the detection and handling of inconsistencies between a model and a set of measurements. The PFA Toolbox provides all these features while preserving computational efficiency.

In the last years, several published works have used these methodologies to perform interval estimations of metabolic fluxes [9, 11–18] and consistency analysis with Possibilistic MFA [9, 17, 18]. Interval MFA was also implemented in FASIMU [16]. However, any

\* Correspondence: yeimy.morales@udg.edu

<sup>1</sup>MICElab, IIIA, Universitat de Girona, Campus Montilivi, P4, Girona 17071, Spain

Full list of author information is available at the end of the article



intermediate user of MATLAB, Mathematica, R, etc. can easily implement Interval MFA. The easily implementation of Interval MFA has led to be used more often than Possibilistic MFA, which requires more mathematical development and additional linear optimizations. The PFA Toolbox presented here simplifies the use of both methods.

The PFA Toolbox provides a comprehensive set of MATLAB functions to easily and quickly apply Interval and Possibilistic MFA. The PFA Toolbox is completely free and open source; users are welcome to modify and adapt the toolbox code to build their own particular functions to fulfill specific requirements under the mild conditions described in the accompanying license. In the following subsections, we briefly describe the methods implemented in the toolbox: Interval MFA and Possibilistic MFA. A detailed description of both methods can be found in [6].

### Interval MFA

Interval MFA is a simple yet powerful extension of traditional MFA methods. It starts with a stoichiometric model or providing model-based constraints, denoted in the sequel as MOC, defined by a stoichiometric matrix  $\mathbf{N}$  and a set of irreversibility constraints. These together define a space of feasible steady-state flux distributions [19, 20] (matrices and vector are denoted in bold):

$$MOC = \begin{cases} \mathbf{N} \cdot \mathbf{v} = 0 \\ \mathbf{D} \cdot \mathbf{v} \geq 0 \end{cases} \quad (1)$$

where, considering a system with  $n$  metabolites and  $r$  reactions,  $\mathbf{N} \in \mathbf{R}^{(n \times r)}$  and  $\mathbf{D} \in \mathbf{R}^{(r \times r)}$  is a diagonal matrix with  $\mathbf{D}_{ii} = 1$  if the flux is reversible (0 otherwise), and  $\mathbf{v} \in \mathbf{R}^{(r)}$  is the vector of metabolic fluxes. The values of  $\mathbf{v}$  that are solution of (1) define a flux distribution.

Consider now a subset  $\mathbf{v}_m \in \mathbf{R}^m$  of measured fluxes in  $\mathbf{v}$  with  $m$  typically much smaller than  $r$ . Following the interval approach, we represent each measured flux as an interval with inequalities:

$$\mathbf{v}_m^m \leq \mathbf{v}_m \leq \mathbf{v}_m^M \quad (2)$$

where  $\mathbf{v}_m^m$  and  $\mathbf{v}_m^M$  are vectors with the minimum and maximum possible values that the measured fluxes  $\mathbf{v}_m$  can take due to measurement's uncertainty.

Equations (1–2) describe a constraint-based model (CB) that defines the space of feasible fluxes. From this CB, the interval of feasible (possible) values for any flux  $v_i$  in the flux distribution  $\mathbf{v}$  can be obtained solving two Linear Programming (LP) problems, as follows:

$$\begin{aligned} v_i^m &= \min v_i \text{ s.t. } v \in \begin{cases} MOC \\ \mathbf{v}_m^m \leq \mathbf{v}_m \leq \mathbf{v}_m^M \end{cases} \\ v_i^M &= \max v_i \text{ s.t. } v \in \begin{cases} MOC \\ \mathbf{v}_m^m \leq \mathbf{v}_m \leq \mathbf{v}_m^M \end{cases} \end{aligned} \quad (3)$$

This procedure provides an interval estimate for any flux of interest. These interval estimates are particularly useful in the two situations of having imprecise measurements and/or when few measures are available. Extra details about Interval MFA can be found in [6, 7, 10].

### Possibilistic MFA

Possibilistic MFA may be seen as a more flexible and powerful extension of Interval MFA. The methodology is based on two ideas: (a) Representing knowledge with constraints satisfied to a certain degree, thus transforming the feasibility of a potential solution into a gradual notion of “possibility” that accounts for uncertainty, and (b) using computationally efficient optimization-based methods, such as Linear Programming, to query for the “most possible” solutions. This methodology is able to face two different problems: (a) To evaluate the consistency between a model and a set of measurements, and (b) to obtain rich estimates of metabolic fluxes. Instead of pointwise estimates, it computes interval estimations for a desired degree of possibility and for entire possibility distributions.

Possibilistic MFA starts with a set of model-based constraints (MOC) defined in (1).

In this case, however, instead of using the simple inequalities (2), the measurements are incorporated in possibilistic terms by means of a set of constraints and two non-negative slack variables that represent the measurement's uncertainty. These constraints, which we call measurement constraints (MEC), can be expressed as:

$$MEC = \begin{cases} \mathbf{w}_m = \mathbf{v}_m + \boldsymbol{\varepsilon}_1 - \boldsymbol{\mu}_1 + \boldsymbol{\varepsilon}_2 - \boldsymbol{\mu}_2 \\ \boldsymbol{\varepsilon}_1, \boldsymbol{\mu}_1 \geq 0 \\ 0 \leq \boldsymbol{\varepsilon}_2 \leq \boldsymbol{\varepsilon}_2^{max} \\ 0 \leq \boldsymbol{\mu}_2 \leq \boldsymbol{\mu}_2^{max} \end{cases} \quad (4)$$

where  $\mathbf{v}_m$  is the vector of the actual values of the measured fluxes, and  $\mathbf{w}_m$  is the vector of the measured values for them. Both differ due to errors and imprecisions. This uncertainty is represented by the slack variables  $\boldsymbol{\varepsilon}_1, \boldsymbol{\mu}_1, \boldsymbol{\varepsilon}_2$  and  $\boldsymbol{\mu}_2$ . The bounds  $\boldsymbol{\varepsilon}_2$  and  $\boldsymbol{\mu}_2$  define a band of fully possible values for  $\mathbf{v}_m$  around the measured values  $\mathbf{w}_m$ . The components  $\boldsymbol{\varepsilon}_1$  and  $\boldsymbol{\mu}_1$  are penalized in a cost index (5) to assign a decreasing possibility to larger errors. Each candidate solution of (1) and (4) can be denoted as  $\delta = \{\mathbf{v}, \mathbf{w}_m, \boldsymbol{\varepsilon}_1, \boldsymbol{\mu}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\mu}_2\}$ .

Now, we define a function,  $\pi(\delta): \Delta \rightarrow [0,1]$  that assigns possibility  $\pi$  in  $[0, 1]$  to each solution, ranging from impossible to fully possible. A simple way to build this

function is using a linear cost index  $J$  to penalize large deviations between the actual values of the fluxes and their measured ones:

$$J = \alpha \cdot \epsilon_1 + \beta \cdot \mu_1 \tag{5}$$

The possibility of each solution is defined as:

$$\pi(\delta) = \exp(-J(\delta)) \delta \in MEC \cap MOC \tag{6}$$

Where  $\alpha$  and  $\beta$  are row vectors of accuracy coefficients or weights that define each measurement's a priori accuracy. These weights need to be defined by the user, e.g., if sensor error is «symmetric»,  $\alpha$  and  $\beta$  should be defined to be equal.

From this point, Possibilistic MFA calculates different estimates by solving LP problems. You can compute the set of flux values with maximum possibility (a pointwise estimation) or a more informative estimation with intervals or flux distributions.

**Pointwise estimations**

The simplest outcome of a Possibilistic MFA problem is a pointwise estimate. It corresponds to the flux values with the maximum possibility (minimum cost), which are obtained by minimizing  $J$  and solving the LP problem:

$$J_{\min} = \min_{\epsilon, \mu, v} J = \alpha \cdot \epsilon_1 + \beta \cdot \mu_1 \text{ s.t. } \{MOC \cap MEC\} \tag{7}$$

The solution flux vector  $v$ , that we call  $v_{mp}$ , contains the most possible values that are consistent with both the model and the measurements.

This pointwise estimation may be unreliable when multiple solutions are reasonably possible. In these instances, distributions and interval estimates can be computed instead.

**Interval estimates**

The interval estimate  $[v_Y^m, v_Y^M]$  for a flux  $v$ , with a conditional possibility higher than  $\gamma$ , can be computed solving two extra LP's:

$$v_Y^m = \min_{\epsilon, \mu, v} \begin{cases} MOC \cap MEC \\ J - J_{\min} < -\ln \gamma \end{cases} \tag{8}$$

The upper bound is defined by replacing minimum for maximum.

**Distributions as estimates**

The complete possibility distribution of a flux can also be obtained for marginal and conditional possibilities. Marginal possibilities provide the degree of possibility of each value for a given flux. Conditional distributions are equivalent to normalizing the marginal possibility distribution to a maximum equal to one.

Possibilistic MFA was casted as a linear optimization problem, for which widely known and efficient tools exist. This great computational performance makes the methodology suitable—in principle—for large-scale metabolic networks.

More information about the methods and a deeper discussion about the strengths and limitations of each approach can be found in our previous works [6–8, 10] and in the toolbox User's Guide (<http://kikollan.github.io/PFA-Toolbox/>).

**Implementation**

The PFA Toolbox has been developed to run in MATLAB. Its core is a set of MATLAB functions that solve each step in a typical MFA problem. The code for all functions is provided with the toolbox. The PFA Toolbox also includes a Graphical User Interface (GUI) to represent the measurements in possibilistic terms. The GUI runs within MATLAB.

The toolbox requires solving LP problems, and those are solved with a flexible and efficient external optimizer, YALMIP [21]. We provide a copy of YALMIP within the PFA Toolbox, but further information about it can be found at the YALMIP website [22]. YALMIP can use different LP solvers, and so does the PFA Toolbox. Three LP solvers were tested: IBM ILOG CPLEX by IBM [23], GLPK [24], and Linprog, the LP solver included in MATLAB. However, we do not recommend the use of Linprog, which proved unreliable, especially for larger MFA problems. Instead, CPLEX or GLPK showed excellent performance. CPLEX has a 90-day free evaluation version, and can be used free for research and academic purposes. GLPK is freely available.

**Results and discussion**

In this section, we show how to use the PFA Toolbox for MATLAB. A list of the functions provided by the toolbox is shown in Table 1. These functions simplify the process of (1) defining the MFA problem, (2) computing different types of estimates (pointwise, interval or distributions) and (3) plotting the results. There is also a function to plot the measurements defined in possibilistic terms, and a GUI to define those measurements. Advanced users can modify and extend each function.

The main features of the PFA Toolbox are the following:

- » It gives reliable MFA estimations even in uncertain or underdetermined scenarios (those where only a few fluxes can be measured).
- » It provides MFA estimations accounting for measurement's imprecision.
- » It provides functions to plot interval estimates and distributions.
- » It is composed of simple, free and open functions.

**Table 1** List of functions in the PFA Toolbox

Initialization	
<code>initPFAtoolbox</code>	It starts the PFA Toolbox
1: MFA problem formulation	
<code>define_MOC</code>	It defines the model-based constraints
<code>define_PossMeasurements</code>	It represents the measured fluxes
<code>define_MEC</code>	It defines the measured-based constraints
2: Computing estimations	
<code>solve_maxPoss</code>	It calculates the most possible set of flux values
<code>solve_maxPossIntervals</code>	It calculates the interval of most possible flux values
<code>solve_PossInterval</code>	It calculates the interval of flux values with the desired possibility
3: Plotting the estimations	
<code>plot_PossMeasurements</code>	It plots measurements in possibilistic terms
<code>plot_distribution</code>	It plots the distribution of a given flux
<code>plot_intervals</code>	It plots interval estimates of a given flux
4: Other	
<code>Solve_possintervalYMP</code>	Advanced function; read its help.
<code>solve_Interval</code>	It solves an Interval MFA problem

A step-by-step protocol to apply Interval or Possibilistic MFA is presented in Fig. 1.

In addition, the toolbox is developed to use stoichiometric models with the format of the COBRA Toolbox (Constraint-Based Reconstruction and Analysis). This format is widely used due to the popularity of COBRA. As an alternative, the user can simply define a model by providing a stoichiometric matrix.

The main features of the toolbox are shown in the next three examples. Additional examples and a thorough description of all functionalities of the toolbox are provided in the User's Guide. The details about the mathematical methods implemented in the toolbox can be found in [7, 8, 10], and in [6].

#### Example of flux estimation under data scarcity

We use a toy metabolic network to illustrate how to use the PFA Toolbox in scenarios of data scarcity. The first step is to formulate the problem. Consider the metabolic network shown in Fig. 2a. The network has six fluxes and three balanced metabolites. One of the fluxes is reversible. Additionally, the fluxes  $v_4$  and  $v_6$  have been measured, with values  $w_4 = 9.5$  mmol/h, and  $w_6 = 10.5$  mmol/h.

The MFA problem consists in the estimation of all six fluxes. Notice, however, that traditional MFA cannot be performed because the problem is undetermined: any pointwise estimate will be only a particular solution of a group of possible ones [5]. The methods in the PFA Toolbox tackle this situation and provide reliable and informative estimates.

In this case, we choose to apply Possibilistic MFA to estimate the fluxes. The first step to solve the problem is to define the model-based constraints (MOC). Stoichiometric model can be directly defined in the code or be provided in COBRA format.

The next step is the addition of measurements and their uncertainties (in this example, we assume that the measurement  $w_4$  is very accurate, but  $w_6$  is not. In agreement with the problem formulation, we assign values to the slack variables  $\mu_2$  and  $\varepsilon_2$ , and the weights  $\alpha$  and  $\beta$  (details about this process can be found in the User's Guide).

Once the MOC and MEC constraints have been defined, the third step is to obtain the estimates. Possibilistic MFA methodology calculates three types of estimations. In this case, we compute three interval estimates for each flux, for conditional possibilities of 0.5, 0.8 and 1.

Finally, we plot the interval estimates using the function `plot_intervals`. The metabolic network and the main features of the algorithm to solve the problem with the PFA Toolbox are shown in Fig. 2. Figure 3a shows the interval estimations for each dataset. Notice that even if only two measurements are available, the estimation is reliable.

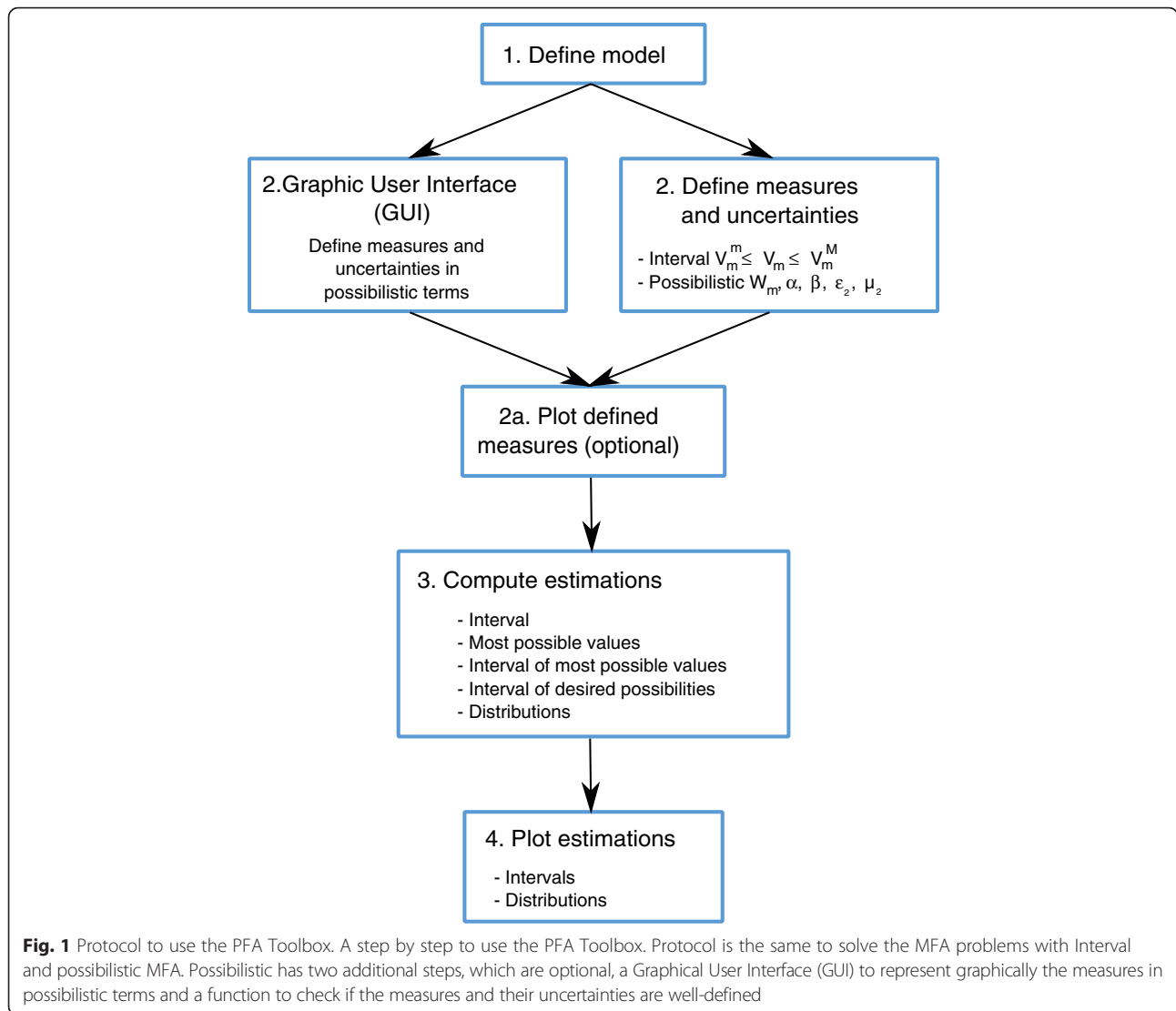
This same procedure can be applied to obtain other types of estimates, such as the complete possibility distribution for a flux. Those computations can be performed using the function `solve_PossInterval`. The obtained distributions are for conditional possibilities (see [8] for a detailed explanation of the notion of conditional possibility). These possibilistic distributions can be plotted with the function `plot_distribution`. As an example, Fig. 3b shows the distribution estimation for all the six fluxes. The results show, for instance, that the most possible value for  $v_1$  is 2.75 mmol/h ( $\pi = 1$ ), that  $v_1$  being equal to 6.1 mmol/h is a less possible situation ( $\pi = 0.6$ ), and that a  $v_1$  being larger than 18 mmol/h is very unlikely ( $\pi < 0.1$ ).

The model and the code for all the computations are provided as (Additional file 1a).

Note: to apply Interval MFA a similar protocol can be followed. The main difference is that the measures will be represented as intervals instead of being represented in possibilistic terms.

#### Example of flux estimation: biomass growth of *Pichia pastoris*

In this example, we estimate the growth of several chemostat cultures of *P. pastoris*. For each chemostat only a few extracellular fluxes are measured (mainly



substrates uptakes and secretion rates) and the aim is to estimate the cellular growth.

The constraint-based model for *P. pastoris* used is presented in [18] (see Additional file 2). It is a relatively small representation including only the main catabolic pathways considering the uptake of the usual carbon sources: methanol, glucose and glycerol. The stoichiometric model contains 37 metabolites and 48 reactions, with reversibility accounted for. The stoichiometric matrix and all the measurements can be found in the (Additional file 3) [31-35].

We select to apply Possibilistic MFA to perform the estimation. As before, we start by defining the MOC and MEC constraints. In this example, we assign the same uncertainty to all the measurements: a deviation of 5 % around the measured value is assumed to be fully possible, while a deviation larger than 20 % is assumed to be an event of low possibility ( $\pi = 0.1$ ). The next step is

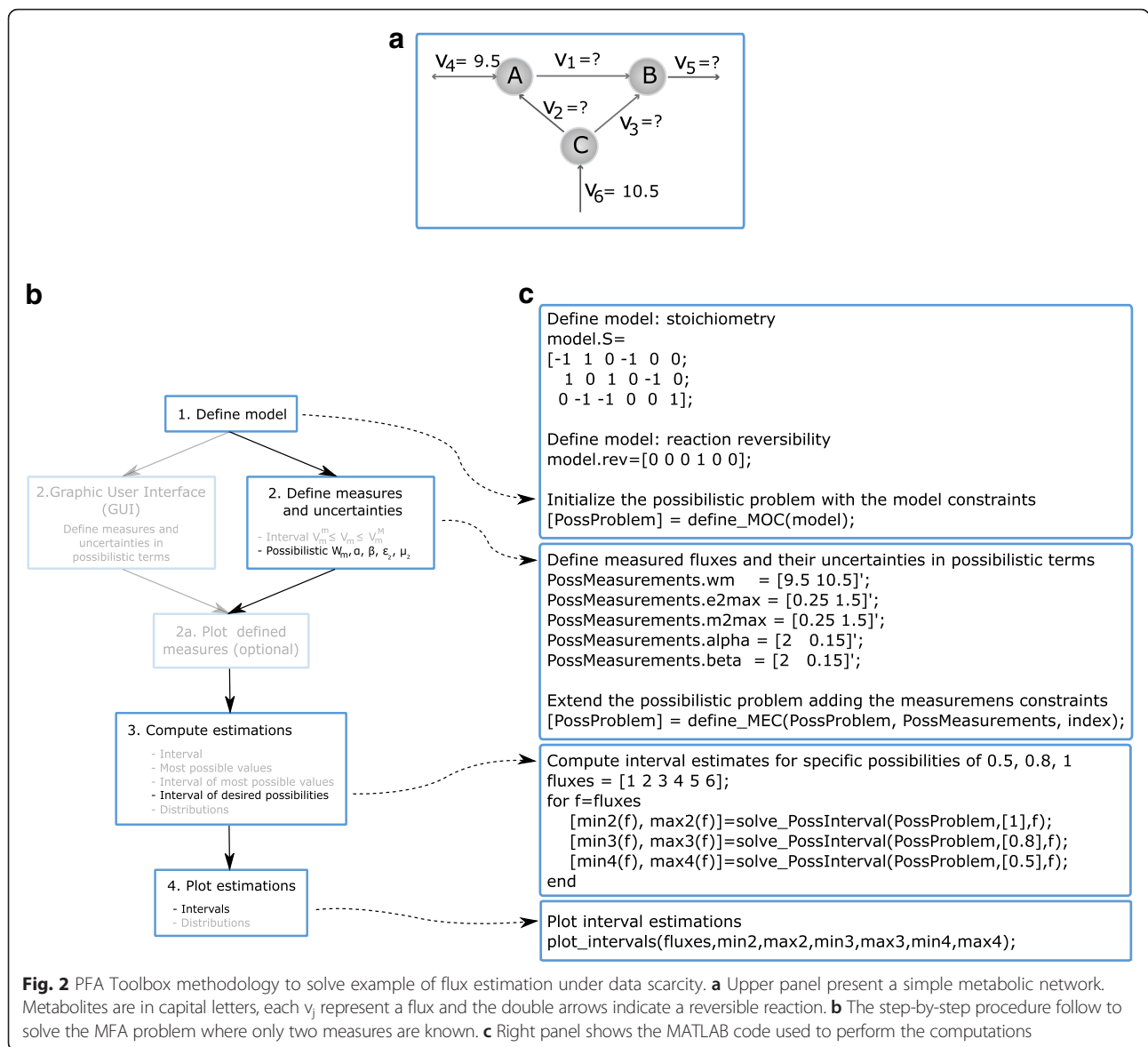
to estimate the growth for each experiment. We compute three interval estimates for conditional possibilities of 0.99, 0.5 and 0.1. Finally, we plot the interval estimates, results are shown in Fig. 4a.

The estimations show good agreement with the experimental growth rates (as expected, since this model and the data have been tested previously). Notice that the interval estimates not only predict the growth rates but also provide an indication of the estimation reliability. The complete code for all computations can be found in the (Additional file 1b).

**Example of flux estimation: growth of *Escherichia coli***

Here we use a well-known model of *E. coli*, taken from [25] and illustrated in the (Additional file 4). It is a relatively compact model containing 72 metabolites and 95 reactions. We consider six chemostat experiments of *E. coli* growing in glucose [26]. The





datasets contain information only for a handful of extracellular measurements (growth rate, substrate uptake, oxygen uptake, CO<sub>2</sub> production and acetate and pyruvate secretion). The model and the measurements can be found in the (Additional file 5).

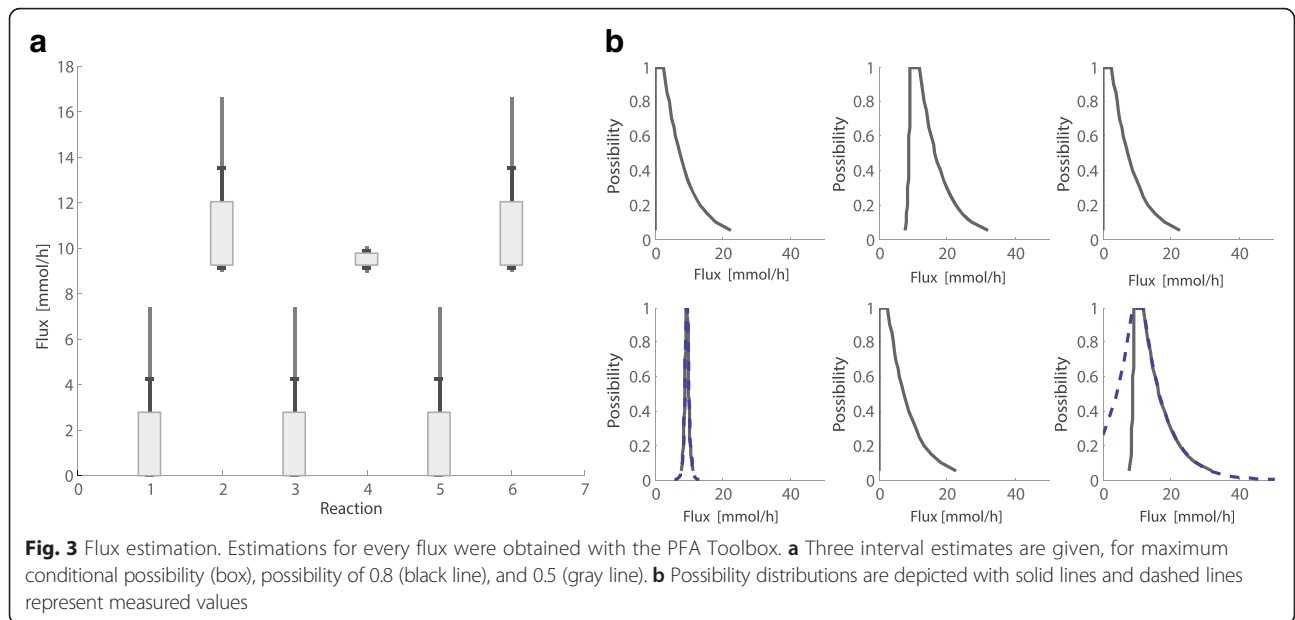
Possibilistic MFA is applied again to estimate the growth rate for all six scenarios. The problem is similar to the previous one, and we assume the same uncertainty for each measurement. However, we now consider a larger model for a different and widely used organism. The computation procedure is analogous to the one previously described. The complete code for all computations can be found in the (Additional file 1c).

The flux estimates computed with the toolbox are compatible with the actual growth rate in all scenarios (Fig. 4b). Notice, however, that the estimates are wider

than in the first example (no-growth is possible in all of them, but the maximum possible growth is near the actual one). The model is larger and the available measurements are not enough to determine completely the flux state of cells. This illustrates one limitation of Interval and Possibilistic MFA: the estimates are only as precise as the uncertainty and the available measurements allow.

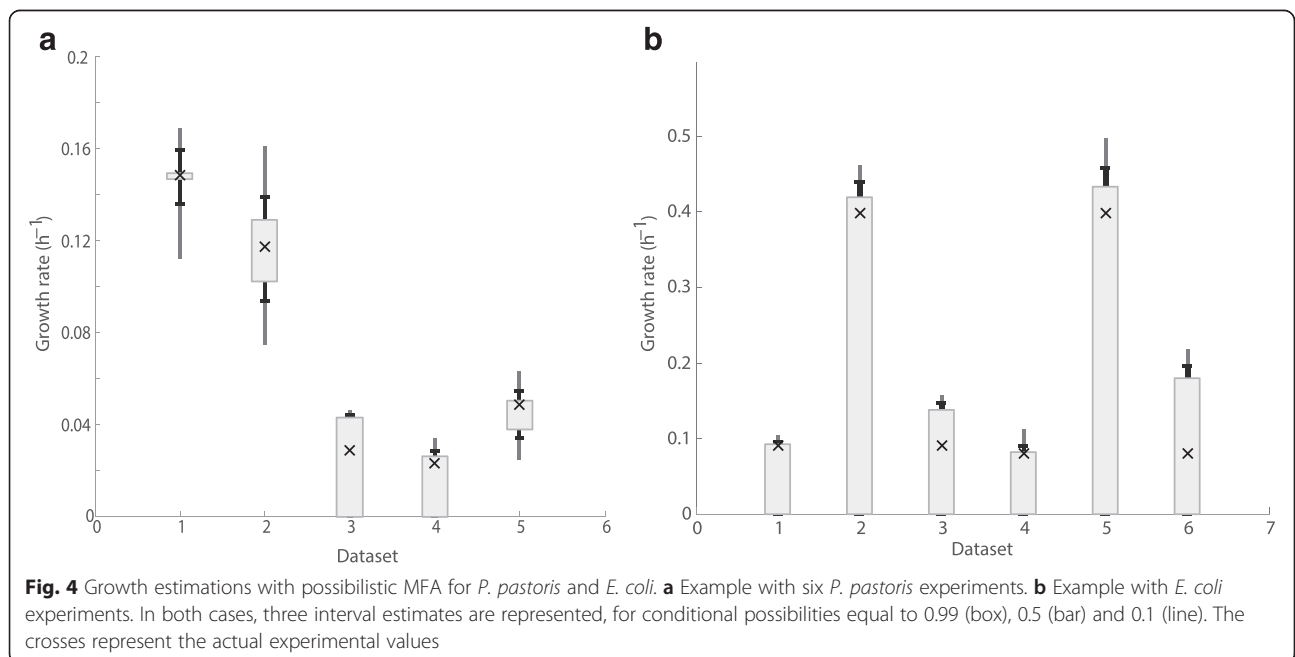
#### Example of consistency analysis with *P. pastoris*

The last example illustrates how the PFA Toolbox can be used for another purpose: to evaluate the degree of consistency between a given model and a set of experimental measurements. Consider the data of six chemostat experiments with *P. pastoris* taken from the literature (Table 2). We test how consistent the data for each experiment are against the model of *P. pastoris*



described previously. We assume that the model is reliable and therefore it can be used to evaluate the validity of each dataset. Notice that this is a strong assumption, valid here for the purpose of this example. It is indeed possible to perform the exact opposite analysis: to obtain several experimental datasets and use them to assess the quality of a metabolic model. We use Possibilistic MFA to validate the model of *P. pastoris* [9, 18]. The objective of the analysis performed here is to detect if there are (larger than expected) errors in the measurements.

We start as in previous examples by defining MOC and MEC constraints. The next step is to compute the estimation. In this example, we compute the most possible solution for each experiment with the *solve\_max-Poss* function. This provides the maximum possibility flux vector and the associated degree of possibility ( $\pi_{mp}$ ) between  $[0, 1]$  of the most possible solution. This value provides an indication of the agreement between the model-based constraints (MOC) and the measurements constraints (MEC).



**Table 2** Experimental data for six chemostat experiments with *Pichia pastoris* and an analysis of its consistency against a model

Reference	$\mu$ Cmmol <sup>a</sup> /g/h	$Q_{Glu}$ mmol/g/h	$Q_{Glyc}$ mmol/g/h	$Q_{Met}$ mmol/g/h	$Q_{Pyr}$ mmol/g/h	$Q_{Cit}$ mmol/g/h	$Q_{EtOH}$ mmol/g/h	OUR mmol/g/h	CPR mmol/g/h	$\pi_{mp}^b$
[31]	6.17	0.00	2.75	0.00	0.00	0.00	0.00	3.62	2.35	0.16
[32]	3.27	0.81	0.00	1.09	0.00	0.00	0.00	4.02	2.68	1.00
[33]	2.38	0.00	1.21	0.00	N.A.	N.A.	N.A.	1.65	1.22	1.00
[33]	4.89	0.00	2.40	0.00	N.A.	N.A.	N.A.	3.12	2.29	1.00
[34]	1.40	0.00	0.00	2.55	N.A.	N.A.	N.A.	2.16	1.15	1.00
[34]	0.94	0.00	0.00	1.87	N.A.	N.A.	N.A.	1.67	0.93	1.00

<sup>a</sup>Cmmol = Carbon mmol <sup>b</sup>Dimensionless value of the possibility of the most possible flux distribution

A possibility equal to one is interpreted as a complete consistency; a lower value implies that there are errors in one (or more) of the measurements or in the model. The complete MATLAB code for this computation can be found in (Additional file 1b).

The results presented in Table 2 show that all datasets except one are highly consistent with the model. The dataset 1 has a low degree of possibility (lower 0.2). This suggests that one or more of the measured fluxes in that experiment is unreliable and may contain errors.

All the computations of these four examples were performed with the PFA Toolbox. The computations take approximately 13 s in a 64-bit Windows PC (Intel Core™ i5 2.5 GHz processor), using MATLAB R2012a with IBM ILOG CPLEX Optimizer as the solver for Linear Programming problems.

#### Notes on computational efficiency and large networks

The methods used by the PFA Toolbox, Possibilistic MFA and interval MFA, have been cast as linear optimization problems, and thus they can be solved with computational efficiency. This makes these methodologies suitable for large-scale metabolic networks. For instance, when tested on a genome-scale *E. coli* model (iJO1366) that contains 2583 reactions [27], the PFA Toolbox is able to get estimates for all 2507 fluxes with three degrees of possibility (i.e., solving 3x2507 LP problems). Computing those estimates required 120 min in an AMD A10–5800 K with Radeon HD graphic (3.80 GHz) PC and 8 GB of RAM with GLPK optimizer. This suggests that the PFA Toolbox may be able to solve MFA flux estimations of large models with good results and reasonable computational cost.

There is, however, a limitation regarding MFA-wise methods when estimating fluxes in large networks: there may be too many flux vectors compatible with the (few) available measurements [28]. Unlike traditional methods, those proposed here may still be of use in this situation. Possibilistic MFA and Interval MFA capture all the equally possible flux states (or “similarly” possible) by means of possibilistic distributions or intervals. If there is a wide range of candidates, however, the estimation

may be only slightly informative. If this is the case, one could decide to incorporate a rational assumption, as done in FBA methods [29, 30].

## Conclusions

We have presented the PFA Toolbox for MATLAB. This toolbox provides a set of MATLAB functions to apply Interval MFA and Possibilistic MFA in a simple and flexible way. The PFA Toolbox is completely free and open source, and can be modified by its users. The toolbox implements MFA-wise methods to perform metabolic flux estimations that are particularly well suited to deal with scenarios of high uncertainty and scarce measurements, which are common in industry.

## Availability and requirements

**Project name:** PFA Toolbox version 1.0.0.

**Project home page:** <http://kikollan.github.io/PFA-Toolbox/>

**Operating systems:** platform independent.

**Programming language:** MATLAB

**Other requirements:** –

**License:** Own license.

**Any restriction to use by non-academics:** none.

## Additional files

**Additional file 1:** Code for the examples A.rar file with the MATLAB files code to perform the examples described below with Example of flux estimation under data scarcity (a), *P. pastoris* (b) and *E. coli* (c). (RAR 5 kb)

**Additional file 2:** Metabolic network of *P. pastoris*. Metabolic network for the *Pichia pastoris* model. For the sake of clarity, the reactions representing biomass growth and ATP balance have not been included in the scheme. (PDF 1082 kb)

**Additional file 3:** Stoichiometric matrix and experimental data for *Pichia pastoris*. A Microsoft Excel spreadsheet file with i) the list of reactions and metabolites, ii) the stoichiometric matrix of *P. pastoris* and iii) the experimental datasets taken from the literature. This includes measurements of biomass, substrates uptakes (glycerol, glucose, and methanol), Oxygen Uptake Rate (OUR), CO<sub>2</sub> production (CPR), and formation of byproducts (ethanol, citrate, and pyruvate). (XLSX 48 kb)

**Additional file 4:** Metabolic network of *Escherichia Coli*. Metabolic network for the *Escherichia coli* model. (PDF 86 kb)



**Additional file 5:** Stoichiometric matrix and experimental data for *Escherichia coli*. A Microsoft Excel spreadsheet file with i) the stoichiometric matrix of *E. coli* and ii) the experimental datasets taken from the literature. This includes measurements of biomass, glycerol, OUR, CPR and pyruvate. (XLSX 62 kb)

### Abbreviations

CB, constraint-based model; COBRA, Constraint-Based Reconstruction and Analysis; FASIMU, Flux-balance Analysis based Simulations; GLPK, GNU Linear Programming kit; GUI, Graphical User Interface; IBM ILOG CPLEX, High-performance mathematical programming solver for linear programming; LP, Linear Programming; MEC, Measurement constraints; MFA, Metabolic Flux Analysis; MOC, model-based constraints; PFA, Possibilistic Flux Analysis; YALMIP, Modelling language for advanced modeling and solution of optimization problems

### Acknowledgements

We acknowledge Ignacio Ribelles for contributing to programming the MATLAB functions and for writing the GUI.

### Funding

This research has been partially supported by the Spanish Government (FEDER-CICYT: DPI 2014–55276–C5–1–R). Yeimy Morales is grateful for the BR Grants of the University of Girona (BR2012/26). Gabriel Bosque Chacón is recipient of a doctoral fellowship from the Spanish Government (BES-2012–053772).

### Availability of data and materials

All data are included in the manuscript, and the associated supplementary material and links provided.

### Authors' contributions

FLL and JP developed the idea for the toolbox, and with JV, they designed the research and coordinated the project. FLL designed the toolbox implementation and wrote the first version of the code. YM contributed to the code, documented it and wrote the user's documentation. YM and GB developed the examples and debugged the toolbox. YM drafted the first manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Consent for publication

Not applicable.

### Ethics approval and consent to participate

Not applicable.

### Author details

<sup>1</sup>MICElab, IIIA, Universitat de Girona, Campus Montilivi, P4, Girona 17071, Spain. <sup>2</sup>Institut Universitari d'Automàtica i Informàtica Industrial, Universitat Politècnica de València, Camino de Vera s/n, Edificio 5C, 46022 Valencia, Spain.

Received: 24 October 2015 Accepted: 1 June 2016

Published online: 11 July 2016

### References

- Sauer U, Hatzimanikatis V, Bailey J, Hochuli M, Szyperski T, Wuethrich K. Metabolic fluxes in riboflavin-producing *Bacillus subtilis*. *Nature biotechnology*. 1997;15(5):448–52.
- Wittmann C. Metabolic flux analysis using mass spectrometry. In: *Tools and Applications of Biochemical Engineering Science*. Berlin: Springer; 2002. p. 39–64.
- Antoniewicz M. Methods and advances in metabolic flux analysis: a mini-review. *J Ind Microbiol Biot*. 2015;42(3):317–25.
- Araújo-Bravo MR, Shimizu JK. An improved method for statistical analysis of metabolic flux analysis using isotopomer-mapping matrices with analytical expressions. *J Biotech*. 2003;05:117–33.
- Klamt S, Schuster S, Gilles D. Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotechnol Bioeng*. 2002;77(7):734–51.
- Llaneras F. Interval and possibilistic methods for constraint-based metabolic models, PhD Thesis. Universidad Politècnica de Valencia: Departamento de Ingeniería de Sistemas y Automática; 2011.
- Llaneras F, Picó J. An interval approach for dealing with flux distributions and elementary modes activity patterns. *J Theor Biol*. 2007;246(2):290–308.
- Llaneras F, Sala A, Picó J. A possibilistic framework for constraint-based metabolic flux analysis. *BMC Syst Biol*. 2009;3(1):79.
- Tortajada M, Llaneras F, Picó J. Validation of a constraint-based model of *Pichia pastoris* metabolism under data scarcity. *BMC Syst Biol*. 2010;4(1):115.
- Llaneras F, Picó J. A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient. *BMC Bioinformatics*. 2007;8(1):421.
- Iyer VV, Ovacik MA, Androulakis IP, Roth CM, Ierapetritou MG. Transcriptional and metabolic flux profiling of triadimefon effects on cultured hepatocytes. *Toxicology and applied pharmacology*. 2010;248(3):165–77.
- Zamorano F, Wouwer A, Bastin G. Detailed metabolic flux analysis of an underdetermined network of CHO cells. *J Biotechnol*. 2010;150(4):497–508.
- Iyer V, Yang H, Ierapetritou M, Roth C. Effects of glucose and insulin on HepG2-C3A cell metabolism. *Biotechnol Bioeng*. 2010;107(2):347–56.
- Iyer V, Androulakis I, Roth C, Ierapetritou M. Effects of Triadimefon on the Metabolism of Cultured Hepatocytes. In: *Bioinformatics and BioEngineering (BIBE), IEEE International Conference on*. 2010. p. 118–23.
- Orman MA, Arai K, Yarmush ML, Androulakis IP, Berthiaume F, Ierapetritou MG. Metabolic flux determination in perfused livers by mass balance analysis: effect of fasting. *Biotechnology and bioengineering*. 2010;107(5):825–35.
- Hoppe A, Hoffmann S, Gerasch A, Gille C, Holzhütter H. FASIMU: flexible software for flux-balance computation series in large metabolic networks. *BMC bioinformatics*. 2011;12(1):28.
- González J, Folch-Fortuny A, Llaneras F, Tortajada M, Picó J, Ferrer A. Metabolic flux understanding of *Pichia pastoris* grown on heterogenous culture media. *Chemometr Intell Lab*. 2014;134:89–99.
- Morales Y, Tortajada M, Picó J, Vehí J, Llaneras F. Validation of an FBA model for *Pichia pastoris* in chemostat cultures. *BMC System Biol*. 2014;8(1):142.
- Stephanopoulos GN, Aristidou AA, Nielsen J. *Metabolic Engineering: Principles and Methodologies*. San Diego, USA: Academic; 1998.
- Heijden R, Romein B, Heijnen J, Hellinga C, Luyben K. Linear constraint relations in biochemical reaction systems: I & II. *Biotech Bioeng*. 1994; 43(1):3–10.
- Lofberg J. YALMIP: A toolbox for modeling and optimization in MATLAB. In: *IEEE International Symposium on Computer Aided Control Systems Design*. 2004. p. 284–9.
- YALMIP Home Page [<http://users.isy.liu.se/johanl/yalmip/>]. Accessed 11 May 2016.
- IBM ILOG CPLEX- High-performance mathematical programming engine. [<http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/>]. Accessed 11 May 2016.
- GLPK (GNU Linear programming kit) [<http://www.gnu.org/software/glpk/>]. Accessed 11 May 2016.
- Orth D, Fleming M, Palsson B. Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. *EcoSal Plus*. 2010;4:1.
- Emmerling M, Dauner M, Ponti A, Fiaux J, Hochuli M, Szyperski T, Wüthrich K, Bailey J, Sauer U. Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. *Journal of bacteriology*. 2002;184(1):152–64.
- Orth J, Conrad T, Na J, Lerman J, Nam H, Feist A, Palsson B. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular systems biology*. 2011;7(1):535.
- Bonarius H, Schmid G, Tramper J. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends in Biotechnology*. 1997;15(8):308–14.
- Palsson BØ. *Systems biology: properties of reconstructed networks*. New York: Cambridge University Press; 2006.
- Schilling C, Covert M, Famili I, Church G, Edwards J, Palsson B. Genome-scale metabolic model of *Helicobacter pylori* 26695. *Journal of Bacteriology*. 2002;184(16):4582–93.
- Solà A, Jouhnen P, Maaheimo H, Sánchez-Ferrando F, Szyperski T, Ferrer P. Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol

- mixtures in chemostat cultures at low and high dilution rates. *Microbiol.* 2007;153:281–90.
32. Solà A. Estudi del metabolisme central del carboni de *Pichia pastoris*, PhD Thesis. Universitat Autònoma de Barcelona: Escola Tècnica Superior d'Enginyeria; 2004.
  33. Jungo C, Rerat C, Marison IW, von Stockar U. Quantitative characterization of the regulation of the synthesis of alcohol oxidase and of the expression of recombinant avidin in a *Pichia pastoris* Mut + strain. *Enzyme Microb Technol.* 2006;39:936–44.
  34. Tortajada M. Process development for the obtention and use of recombinant glycosidases: expression, modelling and immobilization, PhD Thesis. Universidad Politécnica de Valencia: Departamento de Ingeniería de Sistemas y Automática; 2012.
  35. Jordà J, de Jesus SS, Peltier S, Ferrer P, Albiol J. Metabolic flux analysis of recombinant *Pichia pastoris* growing on different glycerol/methanol mixtures by iterative fitting of NMR-derived <sup>13</sup>C-labelling data from proteinogenic amino acids. *New Biotechnol.* 2014;31(1):120–32.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

