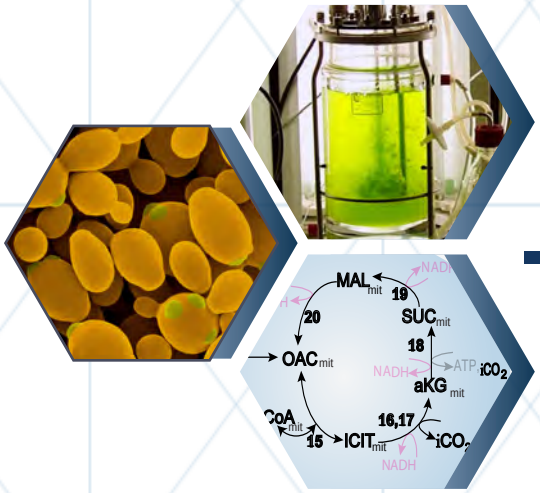
CONSTRAINT-BASED METABOLIC MODELS AND THEIR APPLICATION IN INDUSTRIAL BIOTECHNOLOGY

Yeimy Liceth Morales Pérez

Universitat de Girona



Doctoral Thesis

CONSTRAINT-BASED MODELS AND THEIR APPLICATION IN INDUSTRIAL BIOTECHNOLOGY

Yeimy Liceth Morales Pérez

2016

Doctoral thesis

Constraint-based metabolic models and their application in industrial biotechnology.

Yeimy Liceth Morales Pérez

2016

Universitat de Girona

Doctoral Thesis

Constraint-based metabolic models and their application in industrial biotechnology.

Yeimy Liceth Morales Pérez

2016

Doctoral Programme in Technology

Supervised By:

Dr. Josep Vehí Casellas and Dr. Francisco Llaneras Estrada

Work submitted to the University of Girona in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Universitat de Girona

Dr. Josep Vehí Casellas from Universitat de Girona

Dr. Francisco Llaneras Estrada from Universitat de Girona

DECLARE

The work entitled *Constraint-based metabolic models and their application in industrial biotechnology p*resented by Yeimy Liceth Morales Pérez to obtain the degree in Doctor of Phylosophy has been developed under our supervision.

For all intents and purposes, we hereby sign this document.

Josep Vehi Casellas

Francisco Llaneras Estrada

Girona, May 2016

‘’Somewhere, something incredible is waiting to be known.”

Carl Sagan

AGRADECIMIENTOS

En este momento, al estar frente al ordenador escribiendo estas líneas, llegan a mí un torrente de recuerdos que han enmarcado esta aventura de la tesis doctoral. Debo decir que a pesar de que en algunos momentos me sentí desalentada y muchas veces confundida ahora mismo todos los recuerdos que vienen a mí son muy positivos. Principalmente por la fortuna que he tenido al estar rodeada en este tiempo de gente maravillosa.

Quiero iniciar agradeciéndoles a mis directores. A Kiko Llaneras por presentarme este maravilloso mundo de la biología de sistemas. Por su ayuda incondicional, su rigurosidad y por toda sus enseñanzas. A Josep Vehi, por todo su apoyo y orientación en todo este proceso. Definitivamente sus aportes han hecho que esta tesis sea una realidad.

Me gustaría agradecer a mis co-autores, Jesús Picó, Marta Tortada y a Gabriel del Bosque. Gracias por su rigurosidad, dedicación y por todos los valiosos aportes que contribuyeron a mejorar sustancialmente cada uno de los trabajos.

A los chicos del Lunch time con los que compartí disertaciones de todo tipo de temas durante las horas de comida. Que hubiese sido del día a día sin ese momento. Especialmente agradezco a Roberto Petite, Imanol, Raissel, Robert y Xavi.

También quiero agradecer a las personas que se han convertido en grandes amigos en este tiempo a Yenny, Mauricio, Maira, Fabian, Lien, gracias por acogerme como una integrante más de la familia, por las cenas, los cafés y por todos los buenos momentos compartidos; a Anna Carreras gracias especiales por las charlas y por las escaladas y a Leandro por su valiosa amistad y por ser mi compañero de conciertos.

A mis amigos en Colombia Camilo, Liliana Viviana, por tener siempre una palabra de aliento cuando la melancolía llamó a mi puerta.

A los compañeros de Micelab, por darme una mano siempre que lo necesite.

A Pere, por su apoyo, su paciencia e incondicionalidad.

Finalmente pero nunca menos importante a mi familia, por entenderme y apoyarme siempre en mis decisiones, a mi madre por su valentía y especialmente a mis hermanas y mi sobrina por creer firmemente en mí.

Adicionalmente un agradecimiento especial a Dennis Kunkel Ph.D. at Dennis Kunkel Microscopy, Inc por la imagen para la portada.

¡Gracias a todos!

FULL LIST OF PUBLICATIONS

This thesis is based on a compendium of the followed publications:

1. Morales, Y., Tortajada, M., Picó, J., Vehí, J., Llaneras, F. (2014). Validation of an FBA model for *Pichia pastoris* in chemostat cultures. Published in *BMC Systems Biology*, *8*(1), 142. (Impact factor: 2.44 Position: 13/57 in mathematical & computational biology. Q1).

2. Morales, Y., Tortajada, M., Picó, J., Vehí, J., Llaneras, F. (2016). An FBA model of *Pichia pastoris* to predict protein production and growth over methanol and glycerol, based on energetic/ATP allocation. Submitted to *Biotechnology and Bioengineering* (Impact factor: 4.126 Position: 24/163 in Biotechnology & applied microbiology. Q1).

3. Morales, Y., Bosque, G., Vehí, J., Picó, J., Llaneras, F. (2016). PFA Toolbox: a MATLAB tool for Metabolic Flux Analysis. Accepted in *BMC Systems Biology*. (Impact factor: 2.44 Position: 13/57 in mathematical & computational biology. Q1).

LIST OF ABBREVIATIONS

ACCOA	Acetyl-coenzyme-A
ACD	Acetaldehyde
ACE	Acetate
AKG	Alpha-ketoglutarate
CO2	Carbon dioxide
CIT	Citric acid
DHA	Dihydroxyacetone
DHAP	Dihydroxyacetone phosphate
E4P	Erytrose-4-phosphate
EtOH	Ethanol
F6P	Fructose-6-phosphate
FBA	Flux balance analysis
FBP	Fructose-1,6-biphosphate
GAP	Glyceraldehyde-3-phosphate
GLC	Glucose
GOL	Glycerol
G6P	Glucose-6-phosphate
HCHO	Formaldehyde
ICIT	Isocitrate
O2	Oxygen
MAL	Malate
MeOH	Methanol
MFA	Metabolic Flux Analysis
MOC	Model of Constraints
MEC	Modes of error Constraints
NADH	Nicotinamide adenine dinucleotide, reduced
NADPH	Nicotinamide adenine dinucleotide phosphate
O2	Oxygen
OAC	oxaloacetate
PEPcyt	phosphoenolpyruvate
PG3cyt	3-phosphoglycerate
OAC	Oxaloacetate
OUR	Oxygen Uptake Rate
OTR	Oxygen Transfer Rate
PEP	Phosphoenolpyruvate
PG3	3-phosphoglycerate
PYR	Pyruvate
S7P	Septulose-7-phosphate
SUC	Succinate
R5P	Ribose-5-phosphate
RU5P	Ribulose-5-phosphate
XU5P	Xylulose-5-phosphate
XU5P	Xylulose-5-phosphate

LIST OF FIGURES

TABLE OF CONTENTS

ABSTRACT

This thesis is focused on the application of small constraint-based models to analyze and predict the behavior of wild type and modified strains of *Pichia pastoris*. The presented work deals with the common limitations that industrial environment imposes: measurements are scarce, models are not detailed, the modelled organisms are not always well-known and, in most cases, they are genetically modified.

The results have been divided in three articles. The first presents the validation of a small FBA (*flux balance analysis*) model of unmodified *P. pastoris* cells, based on the assumption of "maximizing growth" as evolved biological objective for the cells. The model has been validated in heterogeneous experimental situations.

In the second article, I exploit a feature of constraint-based models: they are easily extendable. In particular, the FBA model has been extended to represent and predict the behavior of genetically modified cells of *P. pastoris* producing a recombinant protein. The new model represents the energetic requirements of the protein production process, and also the impact that protein production has over the cells growth. The model predictions for growth and even for protein production have been validated against multiple experimental datasets.

Finally, a software toolbox is presented. It implements two MFA-wise methods to get estimations from small, constraint-based models in uncertain scenarios. These implementations simplify and extend the application of MFA (Metabolic flux analysis) when measurements are scarce and imprecise.

The thesis is an application of small, constraint-based models to *P. pastoris*. It illustrates how these models can be a valuable tool to analyze, estimate or predict the behavior of unmodified and modified *P. pastoris* cells. The approaches followed in this work account for some of the limitations of industrial environments, and thus, they may be of use when modelling other microorganisms of industrial interest.

RESUM

Aquesta tesi està enfocada en l'aplicació de petits models basats en restriccions amb la finalitat d'analitzar i predir el comportament de socas salvatges i genèticament modificades de *Pichia pastoris*. El treball presentat afronta les limitacions comunes que els ambients industrials imposen: les mesures són escasses, els models no són detallats, els organismes modelats no són sempre ben coneguts i en molts casos han sigut modificats genèticament.

Els resultats han sigut dividits en tres articles. El primer presenta la validació d'un petit model FBA (flux balance analysis) per organismes no modificats de *P. Pastoris* basat en la suposició de ''maximitzar el creixement'' com a objectiu biològic de l'evolució de les cèl·lules. El model ha sigut validat en situacions experimentals heterogènies.

En el segon article, he explotat una característica dels models basats en restriccions: aquests models es poden ampliar per a representar i predir el comportament de cèl·lules genèticament modificades de *P. Pastoris* produint una proteïna recombinant. El nou model representa els requeriments energètics del procés de producció de proteïna, a més de l'impacte que té aquest procés sobre el creixement cel·lular. Les prediccions del model per a creixement i inclús per a la producció de proteïna han sigut validades utilitzant múltiples conjunts de mesures experimentals.

Finalment, es presenta una eina de software. Aquesta implementa dos mètodes MFA-wise per a obtenir estimacions de petits models basats en restriccions en escenaris amb incertesa. Aquesta implementació facilita i estén l'aplicació de MFA (metabolic flux analysis) quan les mesures són escasses i imprecises.

La tesi és una aplicació de petits models basts en restriccions a *P. pastoris*. Aquesta il·lustra com aquests models poden ser una eina útil per analitzar, estimar o predir el comportament de cèl·lules de *P. pastoris* modificades o salvatges. Els enfocaments seguits en aquest treball consideren algunes de les limitacions d'ambients industrials i en conseqüència, aquests podrien utilitzar-se quan es modelin altres organismes d'interès industrial.

RESUMEN

Esta tesis está enfocada en la aplicación de pequeños modelos basados en restricciones con el fin de analizar y predecir el comportamiento de cepas salvajes y genéticamente modificadas de *Pichia pastoris*. El trabajo presentado afronta las limitaciones comunes que los ambientes industriales imponen: las mediciones son escasas, los modelos no son detallados, los organismos modelados no son siempre bien conocidos y en muchos casos han sido modificados genéticamente.

Los resultados han sido divididos en tres artículos. El primero presenta la validación de un pequeño modelo FBA (flux balance analysis) para organismos no modificados de *P. pastoris* basado en la suposición de ''maximizar el crecimiento'' como objetivo biológico de la evolución de las células. El modelo ha sido validado en situaciones experimentales heterogéneas.

En el segundo artículo, he explotado una característica de los modelos basado en restricciones: estos modelos se pueden ampliar para representar y predecir el comportamiento de células genéticamente modificadas de *P. pastoris* produciendo una proteína recombinante. El nuevo modelo representa los requerimientos energéticos del proceso de producción de proteína, además del impacto que tiene este proceso sobre el crecimiento celular. Las predicciones del modelo para crecimiento e incluso para la producción de proteína han sido validadas usando múltiples conjuntos de medidas experimentales.

Finalmente, se presenta una herramienta software. Esta implementa dos métodos MFA-wise para obtener estimaciones de pequeños modelos basados en restricciones en escenarios con incertidumbre. Esta implementación facilita y extiende la aplicación de MFA (Metabolic flux analysis) cuando las mediciones son escasas e imprecisas.

La tesis es una aplicación de pequeños modelos basados en restricciones a *P. pastoris.* Esta ilustra cómo estos modelos pueden ser una herramienta útil para analizar, estimar o predecir el comportamiento de células de *P. pastoris* modificadas o salvajes. Los enfoques seguidos en este trabajo consideran algunas de las limitaciones de ambientes industriales y en consecuencia, estos tal vez pueden ser de uso cuando se modelen otros organismos de interés industrial.

INTRODUCTION

MOTIVATION

Biotechnology industry involves the use of living cells to make or improve products, plants, animals, and develop or modify microorganisms for specific purposes. It is an increasing sector, and mathematical modeling is playing a role of growing importance on it. Mathematical model limit the need of experimentation –which is slow and expensive– allows the exploration of different scenarios in short time and makes it possible to address a number of difficult questions (Viceconti et al., 2016).

Many modelling approaches have been developed. One of those is the study through models of the cellular metabolism. Those models look at the cell as chemical factories where raw materials (substrates) are transformed to produce other (sometimes useful) substances and energy through several biochemical reactions. The set of all the biochemical reactions that take place within the cell define what is typically called a metabolic network.

Constraint-based models have emerged as a valuable approach that exploits those metabolic networks. These models consider a series of known constraints –e.g. stoichiometry, thermodynamics, cell capacities, regulatory restraints, etc.– to elucidate which functional states can and cannot be achieved by a cell. Among the different approaches that exist for the analysis of metabolic networks, constraint-based modeling has attracted attention because it does not require detailed knowledge or data to be useful. For example, they do not require kinetic information for every metabolic reaction, so they limit the needs of experimental information to identify appropriate kinetic parameters. This feature has allowed to apply models that range from small metabolic networks to large, genome-scale metabolic models – being the latest ones the most commonly used–. Additionally, constraint-based models can be easily extended just by incorporating new or better knowledge as new constraints.

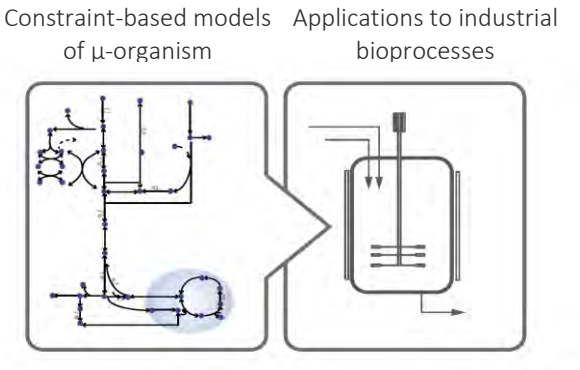


Figure 1 Models of biological systems and applications

Notice that constraint-based models only describe possible flux states, i.e., combinations of metabolic reactions activity that cell could show. However, to predict which particular state among those possible will be exhibit at given conditions, the constraint-based model needs to be coupled with other constraints or a metabolic objective. Flux Balance Analysis (FBA) is the most successful methodology, or family of methodologies, that predicts a particular cell state adding an assumption of optimal cell behavior. Basically, FBA predictions are based on assuming that cells, due to evolutionary pressure, have evolved to be optimal in a particular (and known) way.

Biotechnological companies have interest in FBA models because they may allow predicting the behavior of cellular cultures, and particularly, the synthesis of end-products that are not typically (or easily) measured on-line. This may allow, for instance, to early detect a batch process that is not producing the product of interest, or to optimize media conditions to maximize productivity.

However, some difficulties arise when using constraint-based and FBA models in many industrial environments imposes: (a) lack of available measures, (b) models are not detailed, and (c) organisms of interest are not always well-known and in most cases, genetically modified. This work is devoted to apply small, constraint-based models and FBA models in that context, and for a particular organism of industrial interest: the methylotrophic yeast *Pichia pastoris*.

P. pastoris is a recognized platform to produce several high-value recombinant proteins. It is used both in laboratory research and industrial manufacture. The advantages of this system include the following: a) *P. pastoris* cultures achieves high cell densities on a defined minimal basal salts medium (Cregg et al., 1995), b) it enables the utilization of different metabolic pathways (Baumann et al., 2008), c) it allows efficient post-translational modifications (Digan et al., 1988), d) it produces less secretion of endogenous proteins while expressing secreted recombinant protein (Laroche et al., 1994) and

e) it includes a strong, well-regulated methanol induced promoter (Cregg et al.,1993; Zhang et al., 2000). So far, more than 500 proteins have been expressed using this system (Cos et al., 2006). However, it is still a less studied organism when is compared with other recombinant protein expression systems, such as *Escherichia coli.*

OBJECTIVES

This thesis is mainly devoted to develop and apply FBA models of an important cell factory, P*ichia pastoris*, under the limitations that industrial environments impose: data scarcity, reduced models, and genetically modified organisms.

More specifically, the objectives pursued in this work are the following:

a) To develop a small and simple FBA model of *P. pastoris*, under the hypothesis of maximum growth, and test its ability to predict the behavior of wild type strains.

The first objective is to develop and validate a small FBA model using only the hypothesis of maximum growth. Small or medium metabolic models are not always properly validated. However, validation of small models is particularly important because they are just simplifications of the whole metabolism. In addition the underlying hypothesis of maximum growth is a strong assumption regarding how cells regulate their metabolic fluxes. This assumption may not be accurate to predict behaviors of wild type cells of *P. pastoris* with small metabolic network. For this reason, is worthy to test the growth maximization hypothesis and its ability to predict behavior of wild type strains of *P. pastoris* in different experimental conditions. In this way is possible to identify how accurate is to assume in those scenarios that *P. pastoris* objective is maximizing its growth.

b) To extend the previous FBA model to genetically modified strains of *P. pastoris*, making it possible to predict the production of recombinant protein over methanol and mixes methanol-glycerol.

Genetically modified organisms of *P. pastoris* are those used in industrial applications. Cells are modified to express different proteins of industrial interest, such as, α-L-arabinofuranosidasa. The interest herein, is to develop an FBA model able to predict the production of those proteins, which are the key variable. The model should be able to predict also the biomass growth rate.

c) To simplify the use of two constraint-based methodologies to estimate metabolic fluxes of cells through interval and possibilistic metabolic flux analysis.

The third objective is to develop a tool to easily apply two MFA-wise methodologies especially suited for data scarcity scenarios, that will be useful in my work with *P. pastoris*.

RESEARCH CONTEXT

Mathematical models in bioprocesses

Bioprocess engineering involves all the industrial processes that employ living organisms –mainly microbial cells– to convert raw materials into products. Bioprocessing is an essential part in many chemical, food and pharmaceutical industries. Although the use of microorganisms to make fermented foods and alcoholic beverages have a long history, bioprocess engineering has had a bigger growth in recent years thanks to the developments in genetic and molecular biology. The use of genetically modified organisms gave birth to an increasingly interest because the wide range of products, since industrial alcohol and organic solvents, to high valued chemicals such as antibiotics, therapeutic proteins and vaccines. Due to the complexity of microbial metabolism, the large number of interacting reactions and the complex regulation, there has been a growing focus on the use of mathematical models of its study (Almquist et al., 2014).

Usually mathematical models are defined for a specific system; what it is meant by system is a portion of the universe that we limited for its study. In bioprocess the models have been mostly focus in two main systems. On the one hand, models that are interested in the global behavior of the cell culture where the system is the bioreactor and on the other hand, models focused in cell and cell population.

Most models focused on the global behavior of the process are unstructured. In these models, cell-growing —biomass– is considered as a structure-less entity, ignoring cell physiology and regarding it as black-box (Owens, 1981). The dynamic in these models is represented by ordinary differential equations, such as mass balance for substrates, cells and products (Schügerl & Bellgardt, 2012), taking into account input, output and inner variables of the bioreactor (Caramihai & Severin, 2013) and kinetic expressions, which correlate the rate of formation or consumptions of substrates, cells and products with biomass growth rate.

Although microbial growth is a complex phenomenon, and is function of several variables (e.g. temperature, concentration of the species, pH, etc.) simple rate empirical ex-

pressions as Monod equation have been commonly used to modeled (Han & Levenspiel, 1988). Those empirical expressions include specific parameters that should be fitted with experimental data that are time consuming and complex in nature (Baltes et al., 1994). However, even unstructured models represent an oversimplification of microbial systems, these have been applying profusely mainly due to its simplicity (Garcia et al., 1995).

In bioprocess engineering, models have been used for different purposes: as off-line measured estimations (Barrigón et al., 2012; Veloso et al.,2009), on-line monitoring and control (Bastin.,1990; Ban Impe, & Bastin, 1995; Craven et al., 2013; Skupin et al.,2015), process optimization (Banga et al.,2003; 2005; Hjersted & Henson, 2006; Li et al., 2015; Camacho et al.,2007), fault detection (Nomikos & MacGregor, 1995; Gins et al.,2012), or experimental design (Galvanauskas et al.,1998).

The other perspective of mathematical models in bioprocess has been focused on the cell and on the process inside. This thesis is focused on these types of models where a cell is viewed from a holistic point of view and the main effort is on understanding the system structure and its dynamics. Since the availability of the full genome sequence, considerable amounts of data have been offered, allowing the study of the cell as a system that interacts with its components. This frame of reference is called system biology (Palsson, 2006; Kitano, 2002). The aim of system biology is to use all those information and compiling in networks of biochemical reactions to understand the genotype-phenotype relations in living systems as cells, tissues, organism, among others, through experimental observations and theoretical knowledge, relying in mathematical models and computational techniques (Stelling, 2004).

Models in system biology have been improving the knowledge about cellular behavior and have been used as a tools to: a) generate experimentally testable hypotheses on underlying mechanisms as well as predictions of cellular behavior, thereby iteratively producing refined models and insight into the system (Stelling, 2004; Kitano, 2002), b) understanding complex cellular systems for the creation of new phenotypes through metabolic engineering or synthetic biology (de Jong et al.,2012), c) quantification of the intracellular fluxes in the cell metabolism (Vallino & Stephanopoulos, 1989), d) translate the large amount of data obtained by high-throughput technologies into a better understanding of the underlying biological phenomena (Draghici, et al.,2007), and e) providing advanced monitoring in bioprocess and mechanisms that systematically control the state of the cell to minimize malfunctions and provide potential therapeutic targets for treatment of disease (Kitano, 2002).

Constraint-based models

Mathematical modeling in system biology has different approaches according to the features of the network included in the model. This generates different types of mathematical models, from global views of cellular systems to detailed descriptions involving different levels of cellular organizations, including genes, proteins, metabolism or signaling pathways.

Constraint-based models or stoichiometric models are one of the most popular approaches (Bordbar et al., 2014). These models use the metabolic network, which contain the biochemical reactions involved in the cell metabolism. Metabolism can be compared to a chemical engine that drives the living process. Through the utilization of a vast repertoire of enzymatic reactions and transport processes, unicellular and multicellular organisms can process and convert thousands of organic compounds into the several biomolecules necessary to support their existence (Schilling et al., 1999).

The set of metabolic reactions of a cell can be represented by a graph-oriented network where the nodes represent the metabolites and its edges represent the reaction rate or metabolic fluxes. The stoichiometric coefficients of the chemical reactions embedded in the metabolic network can be translated into mathematical terms through the stoichiometric matrix **N**, where the columns correspond to the reactions and the rows, to metabolites.

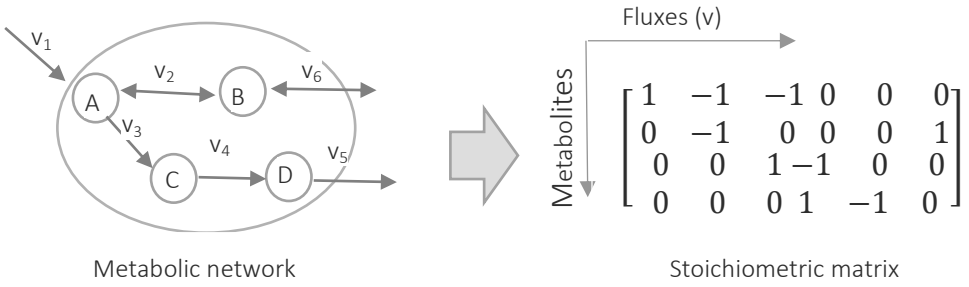


Figure 2 Representation of a metabolic network, the nodes represent internal metabolites, and edges metabolic fluxes v. Two types of metabolic fluxes can be defined, internal metabolic fluxes and exchange fluxes

In order to represent the change of all the metabolites, it is possible to establish a mass balance around an intracellular metabolite i:

$$\frac{d(c_{i,}, x)}{dt} = \sum a_{i,j} \cdot j - \mu \cdot c_i \qquad (1)$$

Where, a_{ij} is the stoichiometric coefficient for the metabolite i in the reaction j, v_j is the specific flux through each reaction and μ is the specific growth rate. The expression (1) is represented in compact form for all the metabolites by (2)

$$\frac{d(\boldsymbol{c}, x)}{dt} = \boldsymbol{N} \cdot \boldsymbol{v} \cdot x - \mu \cdot \boldsymbol{c} \quad (2)$$

Expanding the derivative term

$$\frac{d(\boldsymbol{c} \cdot x)}{dt} = x \cdot \frac{d\boldsymbol{c}}{dt} + \boldsymbol{c} \cdot \frac{dx}{dt} \quad (3)$$

Being $\frac{dx}{dt}$ the growth rate of the biomass, which can be expressed as (4) (Provost, 2004).

$$\frac{dx}{dt} = \mu \cdot x \quad (4)$$

By substituting (3) and (4) in (2), the mass balance equations around intracellular metabolites can be rewritten as (5).

$$\frac{d\boldsymbol{c}}{dt} = \boldsymbol{N} \cdot \boldsymbol{v} - \mu \cdot \boldsymbol{c}. \quad (5)$$

This equation represents the change of the concentration of each internal metabolite over time. However, the kinetic information used to explore the dynamic of a system is particularly hard to come by and therefore; stoichiometric models assume that most of intracellular metabolites are in steady state. This assumption can be used considering the fact that internal metabolites have faster dynamics and will not be accumulated inside the cell, assuming a pseudo steady state. Additionally, the second term of the equation represents the effect of dilution as a result of growth. This term can be disregarded as the concentration of subcellular compound is much smaller that fluxes affecting the same metabolites (Stephanopoulos et al., 1998). Under these assumptions the matricidal set can be described by a system of linear equations.

$$\boldsymbol{N} \cdot \boldsymbol{v} = 0 \quad (6)$$

This system of equations describes a null space, where each stoichometrically feasible steady state is represented by a flux vector $\boldsymbol{v}$. In most of the metabolic models there are more reactions or fluxes –*n*– than metabolites –*m*–, (*n*>*m*). In other words, there are more unknown variables than equations. So there is no unique solution to this system of equations (Orth et al., 2010). The existence of several possible solutions to the system is consistent with the fact that cell can behave differently depending on the environmental conditions (Llaneras, 2011). However not all those possible states can be achieved for an organism in a specific genetic or environment conditions, because other type of con-

strains limit its behavior (Palsson, 2006; Fong & Palsson, 2004). Constraint-based models are an extension of stoichiometric models, where imposing known constraints limits the range of attainable flux distributions or metabolic phenotypes that can be achieved by an organism (Mahadevan & Schilling, 2003).

Different types of constraints can be imposed. Among the most used ones, it can be mentioned: a) the reaction directionality, b) capacity of enzymes/transporters, c) measured fluxes, d) thermodynamics laws, and e) kinetic constraints (Jankowski, et al.,2008; Flamholz et al.,2012; Fleming et al.,2009; Vazquez et al., 2008; Hoppe et al.,2007; Ederer & Gilles, 2007; Visser et al.,2000). Other more specific constraints have been introduced to better explain specific situations like the available space on the cytoplasmic membrane to represent the respire-fermentation physiology (Zhuang et al., 2011).

Constraint-based modeling has been explored with different methodologies. These can be classified into two main categories: a) methodologies to analyze the entire flux space, as pathway analysis with linear algebra or those that use convex analysis as elementary modes (Papin et al.,2004; Machado et al.,2012); and b) those that are interested in a particular flux state as metabolic flux analysis (MFA), which allow estimate the current flux state by adding some measurements (Klamt, 2002; Lohr et al., 2014; Llaneras et al.,2009), and flux balance analysis (FBA) which predict the state exhibit by cells at given conditions (Orth et al.,2010; Brochado et al.,2012). This thesis is focuses on the last type of methodologies.

CONTRIBUTIONS

As a result of this thesis the following contributions have been made:

- To develop a small and simple FBA model of *P. pastoris*, under the hypothesis of maximum growth, and test its ability to predict the behavior of wild type strains. This contribution is presented in a journal article: “Validation of an FBA model for *Pichia pastoris* in chemostat cultures”, (BMC Systems Biology published in BMC Systems Biology. (2014) 8:142).

- A FBA model for genetically modified strains of *P. pastoris* able to predict simultaneously protein production and biomass growth rate on methanol and mixes methanol-glycerol. This contribution is presented in a journal article: “An FBA model of *Pichia pastoris* to predict protein production and growth over methanol and glycerol, based on energetic/ATP allocation” (Submitted to Biotechnology and Bioenginering).

- A MATLAB software toolbox to simplify the use of Interval and Possibilistic MFA. The software is presented in a journal article: "PFA Toolbox: a MATLAB tool for Metabolic Flux Analysis" (/Accepted in BMC Systems Biology). The software is freely available online at: http://kikollan.github.io/PFA-Toolbox/

Following, a summary of the publications developed to accomplish the main goal of this thesis is presented. A copy of each of these publications is providing in the overview section.

«Validation of an FBA model for *Pichia pastoris* in chemostat cultures». A research article published in BMC Systems Biology, (JCR quartile: Q1; Impact factor: 2.435 in 2014; 13 among 57 journals in Mathematical & computational Biology).

Constraint-based metabolic models and FBA have been extensively used in the last years to investigate the behavior of cells and also as basis for different industrial applications. This work provides a validation of a small-sized FBA model of the yeast P. pastoris. The main objective is to test how accurate is the hypothesis of maximum growth to predict the behavior of P. pastoris in a range of experimental environments. In this paper has been verified the model ability to predict the cells behavior in different conditions without introducing measurements, experimental parameters, or any additional constraint, just by assuming that cells will make the best use of the available resources to maximize its growth.

In particular, the FBA model ability has been tested to: (a) predict growth yields over single substrates (glucose, glycerol, and methanol), (b) predict growth rate, substrate uptakes, respiration rates, and by-product formation in scenarios where different substrates are available (glucose, glycerol, methanol, or mixes of methanol and glycerol), (c) predict the different behaviors of P. pastoris cultures in aerobic and hypoxic conditions for each single substrate. In every case, experimental data from literature was used as validation. We conclude that predictions based on growth maximization are reasonably accurate, but still far from perfect.

The deviations are significant in scenarios where P. pastoris grows on methanol, suggesting that the hypothesis of maximum growth could be not dominating in these situations. However, predictions are much better when glycerol or glucose is used as substrate. In these scenarios, even if the presented FBA model is small and imposes a strong assumption regarding how cells will regulate their metabolic fluxes, it provides reasonably good predictions in terms of growth, substrate preference, product formation, and respiration rates.

«An FBA model of Pichia pastoris to predict protein production and growth over methanol and glycerol, based on energetic/ATP allocation» A research article submitted in Biotechnology and Bioengineering. (JCR quartile: Q1; impact factor: 4.126 in 2014; 24 among 163 in Biotechnology & applied microbiology).

One of the main observations in the previous model for wild type cells was that predictions in protein producer scenarios presented significant deviations. However in the previous model was not taking into account the recombinant protein production. Maximizing recombinant protein production is a key feature of Pichia pastoris cultures. In this work, we present a Flux Balance Analysis (FBA) model with growth maximization assumption, able to predict recombinant protein production simultaneously with growth rate in chemostat cultures of Genetically Modified Organisms (GMOs) of P. pastoris.

The model has been developed for pure methanol and mixed feeds (methanol-glycerol) as substrates. In a previous work, we predicted the behavior of wild type P. pastoris strains with a small constraint-based metabolic model and FBA. Herein, we add a mass balance to this small model to represent ATP consumption, and a hypothesis about how modified cells distribute their ATP resources –which are used only for growth in wild type cells- for growth and recombinant protein synthesis. The predictions were validated with several experimental scenarios from literature.

The results show the model is able to predict with reasonable accuracy protein production and growth rate. Some features of the model are the following: a) is remarkably accurate despite cells were expressing different heterologous proteins in a variety of operating conditions, b) requires little information to perform its predictions —just the availability of substrates in the environment— and c) is a small and simple model with only a few parameters that have been tuned from experimental data. The approach followed in this work is of interest in industrial environments and pilot laboratories where experimental data are not abundant, as it provides valuable predictions using few data and a small and simple metabolic model.

«PFA Toolbox: A MATLAB tool for Metabolic Flux Analysis»

A software article accepted in in BMC Systems Biology, (JCR quartile: Q1; Impact factor: 2.435 in 2014; 13 among 57 journals in Mathematical & computational Biology).

Metabolic Flux Analysis (MFA) is a methodology 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, as in industrial environments. The methodologies implemented in this Toolbox —Interval MFA and Possibilistic MFA— are well suited to face those scenarios.

The presented PFA Toolbox for MATLAB simplifies the use of Interval MFA and Possibilistic MFA. 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) provides tools to easily plot the results as interval estimates or flux distributions, (c) is composed of simple functions that MATLAB users can apply in flexible ways, (d) includes a Graphical User Interface (GUI), which provides a visual representation of the measurement and its uncertainty and (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 of its use.

The PFA Toolbox for MATLAB is a freely available toolbox that is able to easily perform Interval and Possibilistic MFA estimations. In addition as an annex is presented the Users' Manual of the PFA toolbox.

RESULTS AND DISCUSSION

The developed thesis has proposed and applied FBA models to the important cell factory *P. pastoris*. These models took into account the common limitations that industrial environment imposes: measurements are scarce, models are not detailed, the modelled organisms are not always well-known and, in most cases, are genetically modified

A FBA model for wild type microorganisms of P. pastoris.

The presented model for wild type *P. pastoris* is a relatively small representation including only the main catabolic pathways, considering the uptake of the usual carbon sources: methanol, glucose and glycerol. Additionally, the model included the assumption that cells will make the best use of the available resources to maximize their growth.

It is noteworthy that predictions achieved by the model were reasonably accurate, even not perfect. Larger deviations were presented with respect to the experimental data in scenarios of *P .pastoris* growing on methanol or mixed glycerol-methanol. Among the possible explanations are: (a) the model is not detailed enough, and some reactions regarding methanol pathways could be missing, (b) the model represents wild-type strains and it does not account for the alterations that occur due to the production of recombinant protein in genetically modified organisms and (c) also, it could be possible that the hypothesis of maximizing growth is not as suitable in the case of methanol growth.

However, even if: (a) The model is a small representation of the whole metabolism of *P. pastoris*, (b) No parameter fitting was included, and in addition (c) a strong assumption about how cells regulate their fluxes was imposed, the FBA model was able to provide reasonable predictions regarding; growth, substrate preference, product formation, and respiration rates in many heterogeneous experimental scenarios. Those results show that the proposed small FBA model can be a valuable tool to get reasonably predictions of the cells behavior, especially in environments where little information is available.

An FBA model for Genetically Modified microorganisms of P. pastoris.

Here, the constraint-based model capability of being easily extended was exploded by adding new constraints to the wild type FBA model. The model presented a straightforward approach to represent and to predict the behavior of genetically modified cells of *P. pastoris* producing a recombinant protein. The proposed model includes a mass bal-

ance of ATP and a hypothesis of how cells allocate their ATP resources to growth and produce recombinant protein. Therefore, the presented hypothesis states that there is a constant portion of energetic resources that cells devote to growth and protein production.

In wild type strains, ATP is mainly devoted to biomass growth and maintenance task. But modified cells of *P. pastoris* cannot avoid the recombinant protein production as they grow; this new process that cells have to face drains energy towards product formation and so, growth is penalized.

The proposed model was able to predict recombinant protein production and respirations rates over methanol and methanol-glycerol mixtures for GMO of *P. pastoris*. However some scenarios presented divergences with respect to the experimental data. A possible explanation is that the presented proportion may not be linear or it can be mediated by other factors that were not accounted in the model. Nevertheless, the presented model was somehow accurate enough to get predictions of protein production and growth rate without adding many experimental measurements, which make this approach useful in industrial environments.

A new software tool to facilitate the use of interval and possibilistic Metabolic Flux Analysis (MFA). Several methodologies based on constraint-based models have been proposed in recent years. However, the success of their further application by different researches could be limited by deficiency of computational tools. In consequence, a flexible and easy-to-use computational tool that is able to lead users to these methods is required. The presented PFA Toolbox for MATLAB®, simplifies the use of two MFA-wise methodologies that are particularly useful in scenarios of data scarcity: Interval MFA and Possibilistic MFA. The PFA Toolbox is composed of simple functions that MATLAB users can easily apply to solve their MFA problems. Furthermore, users can modify and adapt the toolbox code to build their own particular functions and fulfill their specific requirements. Finally, it is worthy to mention that the PFA Toolbox is completely free and open source.

In summary, the work presented in this thesis shows that relatively small, constraint-based and FBA models provide a way to get reasonably good predictions for wild type and modified organisms of *P. pastoris*, even in scenarios of uncertainty and lack of data. The predictions provided by those models were not perfectly accurate, and sometimes only qualitatively valid. But still, the results may be useful in many situations, because the models require little information to be built. They have only a few parameters to be

fit, and rely on small, well-known metabolic models and simple hypothesis, such as growth maximization and constant ATP allocation of resources.

Lack of data is a common scenario. The presented tools and models can be easily developed to provide predictions of the cells behavior, and be of use to early detect abnormal batches, optimize the cultivation conditions, or monitor a process on-line, all interesting features pursued in industry. In this work the focus has been in *Pichia pastoris*, but similar approaches may be of use with other microbial systems.

CONCLUSIONS AND FUTURE

This thesis has been mainly devoted to develop and apply FBA models of an important cell factory, *Pichia pastoris*, under some of the difficulties that arise in industrial environments. Conclusions and lines for future research are presented below.

A small constraint-based model of *P. pastoris* for wild type microorganisms has been presented. The model simply considers the central metabolism of *P. pastoris*, and the assumption that cells will make the best use of the available resources to maximize their growth. The results show that a small FBA model for wild type strains can be a valuable tool to get reasonably predictions of the cells behavior, especially in environments with little information available.

An FBA model for genetically modified cells of *P. pastoris* has been presented, which is able to predict growth and recombinant protein production over methanol and glycerol. The model was able to predict recombinant protein production and biomass growth for GMO of *P. pastoris*. The model includes a mass balance of ATP, and a hypothesis of how the cells allocate the ATP to growth and protein production. The model is simple yet accurate enough, what I believe that make it valuable for industrial environments.

A hypothesis of how genetically modified *P. pastoris* cells will allocate the energy resources to growth and protein production was proposed. The hypothesis states that there is a relatively constant portion of energetic resources that cells devote to growth and to produce recombinant protein. Based on this hypothesis, the behavior of modified P pastoris cells can be predicted, at least approximately. Modified cells of *P. pastoris* cannot avoid the recombinant protein production as they grow, because that process is coupled to cellular growth. This new metabolic processes that occur within GMO cells impact their growth and behavior. The presented hypothesis allows to represent this metabolic burden in *P. pastoris* cells. The predictions of protein production based on the hypothesis are far from perfect, but still useful in many scenarios.

A new software tool to facilitate the use of interval and possibilistic Metabolic Flux Analysis (MFA) has been developed. The presented PFA Toolbox for MATLAB®, simplifies the use of two MFA-wise methodologies that are particularly useful in scenarios of data scarcity. The PFA Toolbox is composed of simple functions that MATLAB users can easily apply to solve their MFA problems. Furthermore, users can also modify those functions to suit specific requirements. The PFA Toolbox will be free.

FUTURE RESEARCH

This thesis contributes to the development of FBA models and their use for the prediction of uncertain biological systems. Two small FBA models have been developed for wild type and genetically modified cells of *P. pastoris*. These models were based on the assumption that cells have evolved to maximize its growth. This assumption has been proved useful in many works. However, cells face many different and complex environments, which leads to different, more complex, subtler, and eventually competing objectives. For future work it could be interesting to investigate other FBA assumptions, including non-linear o multi-objective functions, which may represent better the strategies that cells acquire through evolution.

Another limitation of the developed models is that they have been only validated with chemostat cultures. However, batch cultures are common in industry. It may be useful to investigate the application of FBA along time to simulate a batch culture.

Finally, the model could be extended to consider other carbon sources. In this work, a small metabolic network of *P. pastoris* has been used. The model considers the uptake of three main carbon sources: glucose, glycerol, and methanol. Nevertheless other carbon sources have been used as carbon sources for this yeast as alanine, sorbitol, mannitol and trehalose. An extension to our constraint-based model including some of these carbon sources could be made to investigate how accurate the approaches implemented in this thesis are with respect to other carbon sources.

REFERENCES

1. Agren, R., Liu L., Shoaie, S., Vongsangnak, W., Nookaew, I., Nielsen, J. (2013). The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. PLoS Computational Biol. 9(3), e1002980.
2. Ahmad, M., Hirz, M., Pichler, H., Schwab, H. (2014). Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. Applied microbiology and biotechnology. 98(12), 5301-5317.
3. Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., Jirstrand, M. (2014). Kinetic models in industrial biotechnology–Improving cell factory performance. Metabolic engineering. 24, 38-60.
4. Antoniewicz, M. (2015) Methods and advances in metabolic flux analysis: a mini-review. J Ind Microbiol Biot. 42(3), 317-325.
5. Araúzo-Bravo, M. R., Shimizu, J. K. (2013). An improved method for statistical analysis of metabolic flux analysis using isotopomer-mapping matrices with analytical expressions. J Biotech. 105,117-133.
6. Baltes, M., Schneider, R., Sturm, C., Reuss, M. (1994). Optimal experimental design for parameter estimation in unstructured growth models. Biotechnology Progress. 10(5), 480-488.
7. Ban Impe, J. F., Bastin, G. (1995). Optimal adaptive control of fed-batch fermentation processes. Control Engineering Practice. 3(7), 939-954.
8. Banga, J. R., Balsa-Canto, E., Moles, C. G., Alonso, A. A. (2005). Dynamic optimization of bioprocesses: Efficient and robust numerical strategies. Journal of Biotechnology. 117(4), 407-419.
9. Banga, J. R., Balsa-Canto, E., Moles, C. G., Alonso, A. A. (2003). Dynamic optimization of bioreactors: a review. Proceedings-Indian national science academy part A. 69(3/4), 257-266.
10. Barrigón, J. M., Ramon, R., Rocha, I., Valero, F., Ferreira, E. C., Montesinos, J. L. (2012). State and specific growth estimation in heterologous protein production by *Pichia pastoris*. AIChE Journal. 58(10), 2966-2979.
11. Bastin, G., Dochain, D. (1990). On-line estimation and adaptive control of bioreactors. Process measurement and control (1)
12. Baumann, K., Maurer, M., Dragosits, M., Cos, O., Ferrer, P., Mattanovich, D. (2008). Hypoxic fed-batch cultivation of Pichia pastoris increases specific and volumetric

productivity of recombinant proteins. Biotechnology and bioengineering. 100(1), 177-183.

13. Baumann, K., Carnicer, M., Dragosits, M., Graf, A., Stadlmann, J., Jouhten, P., Maaheimo, H., Gasser, B., Albiol, J., Mattanovich, D., Ferrer, P. (2010) A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. BMC Syst Biol. 4,141.
14. Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. Ø., Herrgard, M. J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nature protocols. 2(3), 727-738.
15. Baumann, K., Maurer, M., Dragosits, M., Cos, O., Ferrer, P., Mattanovich, D. (2008). Hypoxic fed ‐ batch cultivation of Pichia pastoris increases specific and volumetric productivity of recombinant proteins. Biotechnology and bioengineering, 100(1), 177-183.
16. Bonarius, H., Schmid, G., Tramper, J. (1997) Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. Trends in Biotechnology. 15(8), 308–314.
17. Bordbar, A., Monk, J. M., King, Z. A., Palsson, B. Ø. (2014). Constraint-based models predict metabolic and associated cellular functions. Nature Reviews Genetics, 15(2), 107-120.
18. Bordel, S. (2013). Experimental evidence suggests the existence of evolutionary conserved global operation principles governing microbial metabolism. Sci Rep 2. 3, 1–7.
19. Brochado, A. R., Andrejev, S., Maranas, C. D., Patil, K. R. (2012). Impact of stoichiometry representation on simulation of genotype-phenotype relationships in metabolic networks. PLoS computational biology. 8(11), e1002758.
20. Camacho, J., Picó, J., Ferrer, A. (2007). Self-tuning run to run optimization of fed-batch processes using unfold-PLS. AIChE journal. 53(7), 1789-1804.
21. Caramihai, M., Severin, I. (2013). Bioprocess modeling and control. Biomass Now: Sustainable Growth and Use (Matovic, DB, ed.), InTech, Rijeka.
22. Carnicer M., Baumann K., Töplitz I., Sanchez F., Mattanovich D., Ferrer P., Albiol J. (2009) Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. Microb Cell Fact. 8(65).
23. Caspeta, L., Shoaie, S., Agren, R., Nookaew, I., Nielsen, J. (2012). Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and in silico evaluation of their potentials. BMC systems biology. 6(1), 24

24. Çelik, E., Çalık, P., Oliver, S. G. (2009). A structured kinetic model for recombinant protein production by Mut+ strain of *Pichia pastoris* Chemical Engineering Science. 64(23), 5028-5035.
25. Cereghino, G. P., Cereghino, J. L., Ilgen, C., Cregg, J. M. (2002). Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr Opin Biotechnol. 13(4), 329–332.
26. Charoenrat, T., Ketudat-Cairns, M., Stendahl-Andersen, H., Jahic, M., Enfors, S. O. (2005). Oxygen-limited fed-batch process: an alternative control for *Pichia pastoris* recombinant protein processes. Bioprocess Biosyst Eng. 27(6), 399–406.
27. Chiruvolu, V., Eskridge, K. M., Cregg, J. M., Meagher, M. M. (1998). Effects of glycerol concentration and pH on growth of recombinant Pichia pastoris. Appl Biochem Biotechnol. 75,163–173.
28. Cos, O., Ramón, R., Montesinos, J., Valero, F. (2006). Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: a review. Microbial cell factories. 5(1), 1.
29. Chung, B., Selvarasu, S., Camattari, A., Ryu, J., Lee, H., Ahn, J., Lee, D. Y. (2010). Research Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement. Microb Cell Fact. 9(50).
30. Craven, S., Shirsat, N., Whelan, J., Glennon, B. (2013). Process model comparison and transferability across bioreactor scales and modes of operation for a mammalian cell bioprocess. Biotechnology progress. 29(1), 186-196.
31. Cregg, J. M., Higgins, D. R. (1995). Production of foreign proteins in the yeast *Pichia pastoris.* Canadian journal of botany. 73(S1), 891-897.
32. Cregg, J. M., Vedvick, T. S., Raschke, W. C. (1993). Recent advances in the expression of foreign genes in *Pichia pastoris*. Nature Biotechnology. 11(8), 905-910.
33. De Battista, H., Picó, J., Picó-Marco, E. (2006). Globally stabilizing control of fed-batch processes with Haldane kinetics using growth rate estimation feedback. Journal of Process Control, 16(8), 865-875.
34. D'Anjou, M. C., Daugulis, A. (2001). A rational approach to improving productivity in recombinant *Pichia pastoris* fermentation. Biotechnol Bioeng. 72, 1-11.
35. D'Huys, P. J., Lule, I., Van Hove, S., Vercammen, D., Anné, J., Bernaerts, K., Van Impe, J. (2010). Flux balance analysis and flux spectrum analysis for Streptomyces lividans batch fermentations. Book of Abstracts-Study Day Interuniversity Attraction Pole IAP VI/4.
36. De Jong, B., Siewers, V., Nielsen, J. (2012). Systems biology of yeast: enabling technology for development of cell factories for production of advanced biofuels. Current opinion in biotechnology. 23(4), 624-630.

37. De Oliveira Dal'Molin, C. G., Nielsen, L. K. (2013). Plant genome-scale metabolic reconstruction and modelling. Current opinion in biotechnology. 24(2), 271-277.
38. Digan, M. E., Tschopp, J., Grinna, L., Lair, S. V., Craig, W. S., Velicelebi, G. Thill, G. P. (1988). Secretion of heterologous proteins from the methylotrophic yeast, *Pichia pastoris*. Developments in industrial microbiology (USA).
39. Dikicioglu, D., Wood, V., Rutherford, K. M., McDowall, M. D., Oliver, S. G. (2014). Improving functional annotation for industrial microbes: a case study with *Pichia pastoris*. Trends in biotechnology. 32(8), 396-399.
40. Draghici, S., Khatri, P., Tarca, A. L., Amin, K., Done, A., Voichita, C., Romero, R. (2007). A systems biology approach for pathway level analysis. Genome research. 17(10), 1537-1545.
41. Dragosits, M., Stadlmann, J., Albiol, J., Baumann, K., Maurer, M., Gasser, B., Sauer, M., Altmann, F., Ferrer, P., Mattanovich, D. (2009). The effect of temperature on the proteome of recombinant *Pichia pastoris*. J Proteome Res. 8(3), 1380–1392.
42. Dubois, D., Fargier, H., Prade, H. (1996). Possibility theory in constraint satisfaction problems: Handling priority, preference and uncertainty. Applied Intelligence. 6(4), 287-309.
43. Ebenhöh, O., Heinrich, R. (2001). Evolutionary optimization of metabolic pathways. Theoretical reconstruction of the stoichiometry of ATP and NADH producing systems. Bulletin of mathematical biology. 63(1), 21-55.
44. Ederer, M., Gilles, E. D. (2007). Thermodynamically feasible kinetic models of reaction networks. Biophysical journal. 92(6), 1846-1857.
45. Edwards, J. S., Covert M., Palsson B. (2002).Metabolic modelling of microbes: the flux-balance approach. Environmental Microbiology. 4, 133-140.
46. Emmerling, M., Dauner, M., Ponti, A., Fiaux, J., Hochuli, M., Szyperski, T., Sauer, U. (2002). Metabolic flux responses to pyruvate kinase knockout in Escherichia coli. Journal of bacteriology. 184(1), 152-164.
47. Flamholz, A., Noor, E., Bar-Even, A., Milo, R. (2012). eQuilibrator– the biochemical thermodynamics calculator. Nucleic acids research. 40(D1), D770-D775.
48. Fleming, R. M., Thiele, I., Nasheuer, H. P. (2009). Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to Escherichia coli. Biophysical chemistry. 145(2), 47-56.
49. Fliedl, L., Grillari, J., Grillari-Voglauer, R. (2015). Human cell lines for the production of recombinant proteins: on the horizon. New biotechnology. 32(6), 673-679.
50. Fong, S. S., Palsson, B. Ø. (2004). Metabolic gene–deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. Nature genetics. 36(10), 1056-1058.

51. Galvanauskas, V., Simutis R., Volk, N., Lübbert, A. (1998). Model based design of a biochemical cultivation process. Bioprocess and Biosystems Engineering. 18,227–234.
52. Garcia-Ochoa, F., Santos, V. E., Alcon, A. (1995). Xanthan gum production: an unstructured kinetic model. Enzyme and microbial technology. 17(3), 206-217.
53. Garcia, X., Ferrer, P., Montesinos, J., Valero, F. (2013). Fed-batch operational strategies for recombinant Fab production with *Pichia pastoris* using the constitutive GAP promoter. Biochem Eng J. 79, 172–181.
54. Gins, G., Vanlaer, J., Van Impe, J. F. (2012). Discriminating between Critical and Non-critical Disturbances in (Bio) Chemical Batch Processes Using Multimodel Fault Detection and End-Quality Prediction. Industrial & Engineering Chemistry Research. 51(38), 12375-12385.
55. Glick, B. R. 1995. Metabolic load and heterologous gene expression. Biotechnology advances. 13(2), 247-261.
56. GLPK (GNU Linear programing kit) [http://www.gnu.org/software/glpk/]
57. González, J., Folch-Fortuny, A., Llaneras, F., Tortajada, M., Picó, J., Ferrer, A. (2014) Metabolic flux understanding of *Pichia pastoris* grown on heterogenous culture media. Chemometr Intell Lab. 134:89-99.
58. Han, K., Levenspiel, O. (1988). Extended Monod kinetics for substrate, product, and cell inhibition. Biotechnology and Bioengineering. 32(4), 430-447.
59. Heijden R.T., Romein B., Heijnen J.J., Hellinga C., Luyben K.C. (1994). Linear Constraint Relations in Biochemical Reaction Systems: I & II. Biotechnology & Bioengineering. 43(1), 3–10.
60. Heyland, J., Fu, J., Blank, L. M., Schmid, A. (2011). Carbon metabolism limits recombinant protein production in Pichia pastoris. Biotechnology and bioengineering. 108(8), 1942-1953.
61. Hjersted, J. L., Henson, M. A. (2006). Optimization of Fed-Batch Saccharomyces cerevisiae Fermentation Using Dynamic Flux Balance Models. Biotechnology progress, 22(5), 1239-1248.
62. Holmes, W. J., Darby, R. A., Wilks, M. D., Smith, R., Bill, R. M. (2009). Developing a scalable model of recombinant protein yield from *Pichia pastoris*: the influence of culture conditions, biomass and induction regime. Microbial cell factories. 8(1), 35.
63. Holzhütter, H. G. (2004). The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. European Journal of Biochemistry. 271(14), 2905-2922.

64. Hoppe, A., Hoffmann, S., Holzhütter, H. G. (2007). Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. BMC systems biology. 1(1), 23.
65. Hoppe, A., Hoffmann, S., Gerasch, A., Gille, C., Holzhütter, H. (2011). FASIMU: flexible software for flux-balance computation series in large metabolic networks. BMC bioinformatics. 12(1), 28.
66. Ibarra, R. U., Edwards, J.S., Palsson, B. Ø. (2002). Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature. 420, 186–189.
67. IBM ILOG CPLEX- High-performance mathematical programming engine. [http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/]
68. Inan, M., Meagher, M.M. (2001). Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of Pichia pastoris. J Biosci Bioeng. 92(6), 585–589.
69. Inan, M., Meagher, M. M. (2001). The effect of ethanol and acetate on protein expression in *Pichia pastoris*. J Biosci Bioeng. 92(4), 337–341.
70. Iyer, V. V., Ovacik, M. A., Androulakis, I. P., Roth, C. M., Ierapetritou, M. G. (2010). Transcriptional and metabolic flux profiling of triadimefon effects on cultured hepatocytes. Toxicology and applied pharmacology. 248(3), 165-177.
71. Iyer, V., Yang, H., Ierapetritou, M., Roth, C. (2010). Effects of glucose and insulin on HepG2 - C3A cell metabolism. Biotechnol Bioeng. 107(2), 347-356.
72. Iyer, V., Androulakis, I., Roth, C., Ierapetritou, M. (2010). Effects of Triadimefon on the Metabolism of Cultured Hepatocytes. In BioInformatics and BioEngineering (BIBE), IEEE International Conference on. 118-123.
73. Jankowski, M. D., Henry, C. S., Broadbelt, L. J., Hatzimanikatis, V. (2008). Group contribution method for thermodynamic analysis of complex metabolic networks. Biophysical journal. 95(3), 1487-1499.
74. Jordà, J., de Jesus, S. S., Peltier, S., Ferrer, P., Albiol, J. (2014). Metabolic flux analysis of recombinant *Pichia pastoris* growing on different glycerol/methanol mixtures by iterative fitting of NMR-derived 13C-labelling data from proteinogenic amino acids. New biotecnol. 31(1), 120-132.
75. Jordà, J., Suarez, C., Carnicer, M., Pierick, A., Heijnen, J., Van Gulik, W., Ferrer, P., Albiol, J., Aljoscha, W. (2013). Glucose-methanol co-utilization in *Pichia pastoris* studied by metabolomics and instationary 13C flux analysis. BMC Sys Biol. 7(17), 2–16.
76. Jungo, C., Marison, I., Stockar, U. (2007). Mixed feeds of glycerol and methanol can improve the performance of Pichia pastoris cultures: a quantitative study based on

concentration gradients in transient continuous cultures. J. Biotechnol. 128, 824–837.
77. Jungo, C., Rerat, C., Marison, I. W., von Stockar, U. (2006). 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. 39, 936-944.
78. Kim, S., Warburton, S., Boldogh, I., Svensson, C., Pon, L., d'Anjou, M., Choi, B. K. (2013). Regulation of alcohol oxidase 1 (AOX1) promoter and peroxisome biogenesis in different fermentation processes in Pichia pastoris. J Biotechnol. 166(4), 174–181.
79. Kitano, H. (2002). Computational systems biology. Nature. 420(6912), 206-210.
80. Klamt, S., Schuster, S., Gilles, E. D. (2002). Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple non-sulfur bacteria. Biotechnology and Bioengineering. 77(7), 734-751.
81. Klapa, M. I., Stephanopoulos, G. (2000). Metabolic flux analysis. InBioreaction Engineering. Springer Berlin Heidelberg. 106-124.
82. Knorr, A. L., Jain, R., Srivastava, R. (2007). Bayesian-based selection of metabolic objective functions. Bioinformatics. 23(3), 351-357.
83. Li, S., Tang, B., Xu, Z., Chen, T., Liu, L. (2015). Fermentation Optimization and Unstructured Kinetic Model for Cellulase Production by Rhizopus stolonifer var. reflexus TP-02 on Agriculture By-Products. Applied Biochemistry and Biotechnology. 177(8), 1589-1606.
84. Llaneras Estrada, F. (2011) Interval and Possibilistic Methods for Constraint-Based Metabolic Models, In PhD Thesis. Universidad Politécnica de Valencia: Departamento de Ingeniería de Sistemas y Automática.
85. Llaneras, F., Picó, J. (2007). An interval approach for dealing with flux distributions and elementary modes activity patterns. Journal of theoretical biology, 246(2), 290-308.
86. Llaneras, F., Picó, J. (2007). A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient. BMC bioinformatics. 8(1), 421.
87. Llaneras, F., Picó, J. (2008). Stoichiometric modelling of cell metabolism. Journal of Bioscience and Bioengineering. 105(1), 1-11.
88. Llaneras, F., Sala, A., Picó, J. (2009). Applications of possibilistic reasoning to intelligent system monitoring: a case study. In Control Applications, (CCA) & Intelligent Control, (ISIC), IEEE. 171-176.

89. Lofberg, J. (2004). YALMIP. A Toolbox for Modeling and Optimization in MATLAB, In Flux-Balance Approach. In IEEE International Symposium on Computer Aided Control Systems Design. Edited by IEEE. Taipei: IEEE. 284–289.
90. Lohr, V., Hädicke, O., Genzel, Y., Jordan, I., Büntemeyer, H., Klamt, S., Reichl, U. (2014). The avian cell line AGE1. CR. pIX characterized by metabolic flux analysis. BMC biotechnology. 14(1), 72.
91. Machado, D., Soons, Z., Patil, K. R., Ferreira, E. C., Rocha, I. (2012). Random sampling of elementary flux modes in large-scale metabolic networks. Bioinformatics, 28(18), i515-i521.
92. Macauley-Patrick, S., Fazenda, M. L., McNeil, B., Harvey, L. M. (2005). Heterologous protein production using the Pichia pastoris expression system. Yeast. 22, 249–270.
93. Mahadevan, R., Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metabolic engineering. 5(4), 264-276.
94. Morales, Y., Tortajada, M., Picó, J., Vehí, J., Llaneras, F. (2014). Validation of an FBA model for *Pichia pasto*ris in chemostat cultures. BMC systems biology. 8(1), 142.
95. Muñoz, D. F. M., Enciso, N. A. A., Ruiz, H. C., Avellaneda, L. A. B. 2008. A simple structured model for recombinant IDShr protein production in Pichia pastoris. Biotechnology letters. 30(10), 1727-1734.
96. Nocon, J., Steiger, M. G., Pfeffer, M., Sohn, S. B., Kim, T. Y., Maurer, M., Mattanovich, D. 2014. Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. Metabolic engineering. 24, 129-138.
97. Nomikos, P., & MacGregor, J. F. (1995). Multivariate SPC charts for monitoring batch processes. Technometrics, 37(1), 41-59.
98. Olivier B, Rohwer J, Hofmeyr, J. (2005). Modelling cellular systems with PySCeS. Bioinformatics. 21(4), 560-561.
99. Orman, M. A., Arai, K., Yarmush, M. L., Androulakis, I. P., Berthiaume, F., Ierapetritou, M. G. (2010). Metabolic flux determination in perfused livers by mass balance analysis: effect of fasting. Biotechnology and bioengineering. 107(5), 825-835.
100. Orth, J. D., Thiele, I., Palsson, B. Ø. (2010). What is flux balance analysis? Nature biotechnology, 28(3), 245-248.
101. Owens, J. D. (1981). Annual reports on fermentation processes: Volume 3. Edited by D. Perlman. Academic Press, New York. 1979. 346.
102. Palsson, B. Ø. (2000). The challenges of in silico biology. *Nature biotechnology*, 18(11):1147–50.
103. Palsson, B. Ø. (2006). Systems biology: properties of reconstructed networks. New York, USA: Cambridge University Press New York.

104. Papin, J. A., Stelling, J., Price, N. D., Klamt, S., Schuster, S., Palsson, B. Ø. (2004). Comparison of network-based pathway analysis methods. Trends in biotechnology. 22(8), 400-405.
105. Provost, A., Bastin, G. (2004). Dynamic metabolic modelling under the balanced growth condition. Journal of Process Control. 14(7), 717–728.
106. Ren, H. T., Yuan, J. Q., Bellgardt, K. H. (2003). Macrokinetic model for methylotrophic *Pichia pastoris* based on stoichiometric balance. Journal of biotechnology. 106(1), 53-68.
107. Sala, A., Albertos, P. (2001). Inference error minimization: fuzzy modelling of ambiguous functions. Fuzzy Sets and Systems, 121(1), 95-111.
108. Sala, A. (2008). Encoding fuzzy possibilistic diagnostics as a constrained optimization problem. Information Sciences, 178(22), 4246-4263.
109. Santos, S. (2008). Análisis cuantitativo y modelización del metabolismo de la levadura *Pichia pastoris*. Doctoral dissertation. Universitat Autónoma de Barcelona. Departamento de Ingeniería Química.
110. Sauer, U., Hatzimanikatis, V., Bailey, J., Hochuli, M., Szyperski, T., Wuethrich, K. (1997) Metabolic fluxes in riboflavin-producing Bacillus subtilis. Nature biotechnology. 15(5), 448-452.
111. Schellenberger, J., Que, R., Fleming, R., Thiele, I., Orth, J., Feist, A., Palsson, B. (2001). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. Nature protocols. 6(9), 1290-1307.
112. Schilling, C. H., Schuster, S., Palsson, B. O., Heinrich, R. (1999). Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnology progress. 15(3), 296-303.
113. Schügerl K, Bellgardt KH (2000). Bioreaction Engineering: Modelling and Control. Heidelberg, Germany: Springer-Verlag.
114. Schilling, C., Covert, M., Famili, I., Church, G., Edwards, J., Palsson, B. (2002). Genome-scale metabolic model of Helicobacter pylori 26695. Journal of Bacteriology. 184(16), 4582–93.
115. Schuetz, R., Kuepfer, L., Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol Syst Biol. 3, 1–15.
116. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U. (2012). Multidimensional optimality of microbial metabolism. Science. 336, 601–604
117. Schuster, S., Pfeiffer, T., Fell, D. A. (2007). Is maximization of molar yield in metabolic networks favoured by evolution? J Theor Biol. 252, 497–504.

118. Segrè, D., Vitkup, D., Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. Proceedings of the National Academy of Sciences. 99(23), 15112-15117.
119. Sendín, J. O., Alonso, A. A., Banga, J. R. (2008). Multi-Objective Optimization of Biological Networks for Prediction of Intracellular Fluxes. In 2nd International Workshop on Practical Applications of Computational Biology and Bioinformatics (IWPACBB). Berlin Heidelberg: Springer. 197–205.
120. Shlomi, T., Berkman, O., Ruppin, E. (2005) .Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proceedings of the National Academy of Sciences of the United States of America. 102(21), 7695-7700.
121. Sinha, J., Plantz, B. A., Zhang, W., Gouthro, M., Schlegel, V., Liu, C. P., Meagher, M. M. (2003). Improved Production of Recombinant Ovine Interferon-τ by Mut+ Strain of *Pichia pastoris* using an Optimized Methanol Feed Profile. Biotechnology progress. 19(3), 794-802.
122. Skupin, P., Metzger, M. (2015). Oscillatory Behavior Control in Continuous Fermentation Processes. IFAC-PapersOnLine, 48(8), 1114-1119.
123. Solà A, Jouhten P, Maaheimo H, Sánchez-Ferrando F, Szyperski T, Ferrer P. (2007). Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. Microbiol. 153:281-290.
124. Solà A. (2004). Estudi del metabolisme central del carboni de Pichia pastoris. PhD Thesis. Universitat Autònoma de Barcelona, Escola Tècnica Superior d'Enginyeria.
125. Stelling, J. (2004). Mathematical models in microbial systems biology. Current opinion in microbiology, 7(5), 513-518.
126. Stephanopoulos, G., Aristidou, A. A., Nielsen, J. (1998). Metabolic engineering: principles and methodologies. Academic press.
127. Tortajada, M., Llaneras, F., Picó, J. (2010). Validation of a constraint-based model of *Pichia pastoris* growth under data scarcity. BMC Syst Biol. 4(115), 1–11.
128. Tortajada, M., Llaneras, F., Ramón, D., Picó, J. (2012) Estimation of recombinant protein production in *Pichia pastoris* based on a constraint-based model. J Process Control. 22(6), 1139–1151.
129. Tortajada, M. (2012) Process Development for the obtention and use of Recombinant Glycosidases: Expression, Modelling and Immobilization. In PhD Thesis. Universidad Politécnica de Valencia: Departamento de Ingeniería de Sistemas y Automática.
130. Vallino, J. J., Stephanopoulos, G. (1989). Flux determination in cellular bioreaction networks: applications to lysine fermentations. Frontiers in bioprocessing. 1, 205-219.

131. VanGulik, W. M., Antoniewicz, M. R., Delaat, W. T. A. M., Vinke, J. L., Heijnen, J. J. (2001). Energetics of growth and penicillin production in a high-producing strain of Penicillium chrysogenum. Biotechnology and bioengineering. 72(2), 185-193.
132. Varma, A., Boesch, B. W., Palsson, B. Ø. (1993). Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates. Appl Environ Microbiol. 59(8):2465–2473.
133. Varma, A., Palsson, B. Ø. (1993). Metabolic capabilities of Escherichia coli: I. Synthesis of biosynthetic precursors and cofactors. J Theor Biol. 165(4):477–502.
134. Varma, A., Palsson B. Ø. (1994). Metabolic flux balancing: basic concepts, scientific and practical use. Nature Biotechnology. 12(10), 994–998.
135. Vazquez, A., Beg, Q. K., Ernst, J., Bar-Joseph, Z., Barabási, A. L., Boros, L. G., Oltvai, Z. N. (2008). Impact of the solvent capacity constraint on E. coli metabolism. BMC systems biology, 2(1), 7.
136. Veloso, A. C. A., Rocha, I., Ferreira, E. C. (2009). Monitoring of fed-batch *E. coli* fermentations with software sensors. Bioprocess and biosystems engineering. 32(3), 381–388.
137. Viceconti, M., Henney, A., Morley-Fletcher, E. (2016). In silico Clinical Trials: how computer will transform biomedical industry. Technical report.
138. Visser, D., van der Heijden, R., Mauch, K., Reuss, M., Heijnen, S. (2000). Tendency modeling: a new approach to obtain simplified kinetic models of metabolism applied to Saccharomyces cerevisiae. Metabolic Engineering. 2(3), 252-275.
139. Wintermute, E. H., Lieberman, T. D., Silver, P. A. (2013). An objective function exploiting suboptimal solutions in metabolic networks. BMC Syst Biol. 7, 98.
140. Wittmann, C. Metabolic flux analysis using mass spectrometry. (2002) In: Tools and Applications of Biochemical Engineering Science. Berlin: Springer. 39-64.
141. Yves Laroche, V. S., De Meutter, J., Messens, J., Lauwereys, M. (1994). High– Level Secretion and Very Efficient Isotopic Labeling of Tick Anticoagulant Peptide (TAP) Expressed in the Methylotrophic Yeast, Pichia pastoris. Bio/Technology. *12*, 1119.
142. Zamorano F, Wouwer A, Bastin, G. (2010). Detailed metabolic flux analysis of an underdetermined network of CHO cells. J Biotechnol. 150(4), 497-508.
143. Zhang, W., Bevins, M. A., Plantz, B. A., Smith, L. A., Meagher, M. M. (2000). Modeling Pichia pastoris growth on methanol and optimizing the production of a recombinant protein, the heavy-chain fragment C of botulinum neurotoxin, serotype A. Papers in Biotechnology, 18. Methylotrophic Yeast, Pichia pastoris. Bio/Technology. 12, 1119.
144. Zhang, W., Liu, C. P., Inan, M., Meagher, M. M. (2004). Optimization of cell density and dilution rate in Pichia pastoris continuous fermentations for production of re-

combinant proteins. Journal of Industrial Microbiology and Biotechnology. 317, 330-334.

145. Zhuang, K., Vemuri, G. N., Mahadevan, R. (2011). Economics of membrane occupancy and respiro-fermentation. Molecular systems biology. 7(1), 500.

ANNEX

Following, a copy of the full text of the publications developed in this thesis is presented.

Article I

Morales, Y., Tortajada, M., Picó, J., Vehí, J., Llaneras, F. (2014). Validation of an FBA model for *Pichia pastoris* in chemostat cultures.

Article II

Morales, Y., Tortajada, M., Picó, J., Vehí, J., Llaneras, F. (2016). An FBA model of *Pichia pastoris* to predict protein production and growth over methanol and glycerol, based on energetic/ATP allocation.

Article III

Morales, Y., Bosque, G., Vehí, J., Picó, J., Llaneras, G. (2016). PFA Toolbox: a MATLAB tool for Metabolic Flux Analysis

ARTICLE I

VALIDATION OF AN FBA MODEL FOR *PICHIA PASTORIS* IN CHEMOSTAT CULTURES

Published in *BMC Systems Biology*. (JCR quartile: Q1; Impact factor: 2.435 in 2014; 13 among 57 journals in Mathematical & computational Biology). DOI: 10.1186/s12918-014-0142-y

Morales et al. BMC Systems Biology (2014) 8:142
DOI 10.1186/s12918-014-0142-y

BMC Systems Biology

RESEARCH ARTICLE Open Access

Validation of an FBA model for *Pichia pastoris* in chemostat cultures

Yeimy Morales[1*], Marta Tortajada[2], Jesús Picó[3], Josep Vehí[1] and Francisco Llaneras[1]

Abstract

Background: Constraint-based metabolic models and *flux balance analysis (FBA)* have been extensively used in the last years to investigate the behavior of cells and also as basis for different industrial applications. In this context, this work provides a validation of a small-sized FBA model of the yeast *Pichia pastoris*. Our main objective is testing how accurate is the hypothesis of maximum growth to predict the behavior of *P. pastoris* in a range of experimental environments.

Results: A constraint-based model of P. pastoris was previously validated using metabolic flux analysis (MFA). In this paper we have verified the model ability to predict the cells behavior in different conditions without introducing measurements, experimental parameters, or any additional constraint, just by assuming that cells will make the best use of the available resources to maximize its growth. In particular, we have tested FBA model ability to: (a) predict growth yields over single substrates (glucose, glycerol, and methanol); (b) predict growth rate, substrate uptakes, respiration rates, and by-product formation in scenarios where different substrates are available (glucose, glycerol, methanol, or mixes of methanol and glycerol); (c) predict the different behaviors of P. pastoris cultures in aerobic and hypoxic conditions for each single substrate. In every case, experimental data from literature are used as validation.

Conclusions: We conclude that our predictions based on growth maximisation are reasonably accurate, but still far from perfect. The deviations are significant in scenarios where *P. pastoris* grows on methanol, suggesting that the hypothesis of maximum growth could be not dominating in these situations. However, predictions are much better when glycerol or glucose are used as substrates. In these scenarios, even if our FBA model is small and imposes a strong assumption regarding how cells will regulate their metabolic fluxes, it provides reasonably good predictions in terms of growth, substrate preference, product formation, and respiration rates.

Keywords: Constraint- based model, Flux balance analysis, Possibilistic metabolic flux analysis, *Pichia pastoris*

Background

Pichia pastoris is a methylotrophic yeast widely recognized as a suitable expression system for basic research and industrial application [1]. More than 500 proteins have been expressed using this system due to (a) the possibility to grow cultures to very high cell densities. (b) The existence of methanol-inducible alcohol oxidase promoters (AOX). (c) its ability to produce post-translational modifications, and (d) the good protein yield/cost ratio.

As any other living cell, *P. pastoris* cells are complex systems, but they can be represented as an array of reactions that convert raw materials into energy and building blocks. These collections of chemical reactions form a metabolic network; and these metabolic networks can be encoded in an ***mxn*** matrix, with ***m*** metabolites and ***n*** reactions, called stoichiometric matrix [2-4]. From these networks, a constraint-based model can be derived by imposing a mass balance around the metabolites assumed to be balanced —mostly internal ones—, and by constraining those reactions that are assumed to be irreversible. This way, a constraint-based model defines a space of feasible flux distributions, *i.e.*, a space of all the metabolic behaviors that the cells can show in different conditions [5,6]. These models have the advantage of not requiring knowledge about kinetic parameters, which are rarely known for most intracellular reactions.

The space of feasible flux distribution can be still reduced by adding more constraints, such as context-

* Correspondence: yeimy.morales@udg.edu
[1]MICElab, IIIA, Universitat de Girona, Campus Montilivi, P4, Girona 17071, Spain
Full list of author information is available at the end of the article

dependent assumptions. As a result, there are several methodologies employed with different purposes and making use of different mathematical frameworks, but they all have in common the use of a constraint-based modeling approach [5].Two popular approaches are metabolic flux analysis (MFA) and flux balance analysis (FBA). MFA combines the constraint-based model with a set of experimental measurements, usually of extracellular fluxes, to perform estimations [7]. FBA also uses a constraint-based model, but it incorporates an assumption of optimal cell behavior [2,8-10].

In particular, FBA is a framework to get predictions from a constraint-based model using optimization [2,6,8,11]. FBA predictions are based on assuming that cells, due to evolutionary pressure, have evolved to be optimal in a particular (and known) way. This approach reduces the space of feasible flux distributions generated by the constraint-based model by incorporating «input» constraints —typically bounds for the uptake fluxes, based on known capacities or the availability of substrates—, and defining an objective function based on an assumption of optimal cell behavior. Often, the objective function chosen is the maximization of the biomass growth rate [12,13]. However, many other objective functions have been proposed, such as the maximization of ATP production rate [14] or the minimization of total flux [15].

Even if FBA predictions based on the hypothesis of maximal growth rate have been shown to be reasonably accurate in several studies, their limitations have been also investigated [16]. It has been argued that the assumption is well justified in many cases, but not in all situations [10]. Similar conclusions were drawn by Shuetz et al., when the authors performed a systematic evaluation of different objective functions in order to predict intracellular fluxes of *E. coli* cultures by invoking optimality principles [13]. They found that no single objective function was able to accurately predict the behavior that cells shown in all the conditions. These limitations are the basis to investigate more sophisticated objective functions and also for dealing with multiple criteria simultaneously, by means of Pareto surface and other analytical tools [17,18].

In this paper, we present the validation of a FBA (constraint-based) model of *P. pastoris* based on a small-sized metabolic network. In line with previous works done with small models of other organisms, such as *E. coli* [19,20], *S. cerevisiae* [21,22] or *Aspergillus niger* [23], with a less studied organism as *P. pastoris*. Our main objective is testing how accurate is the hypothesis of maximum growth rate to predict the cells behavior in a range of experimental environments. The underlying constraint-based model of *P. pastoris* was previously validated against experimental data using MFA [24]. Now we will test the FBA model ability to give reasonable predictions without incorporating measurements, just by assuming that cells will make the best use of the available resources.

Methods

Constraint based metabolic model

Along this paper, a constraint-based model of *P. pastoris* has been used. The model is a modified version of the one previously described and validated in [24,25]. It is a standard constraint-based model, as those described in [5] or [2]. The model was derived from a set of central metabolic reactions. These reactions are then translated into constraints by assuming that intracellular metabolites are at steady-state (and disregarding the dilution effect). Then, another set of inequality constraints is incorporated by imposing irreversibility to some reactions. This procedure results in a set of model constraints (MOC) that defines a space of feasible steady state flux distributions, as follows:

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

Where **N** is a stoichiometric matrix, with *m* metabolites and *n* reactions, the vector **v** is the vector of reaction fluxes, which represent the mass flow through each of the *n* reactions in the network. The matrix **D**, is a diagonal matrix with $\mathbf{D}_{ij} = 1$ if the flux is irreversible and null otherwise.

Consistency analysis of experimental data

To validate our model predictions, several experimental datasets corresponding to different *P. pastoris* chemostat experiments have been collected from literature. Each dataset contains experimental measurements of several extracellular fluxes (*e.g.,* biomass growth, glucose uptake rate, oxygen uptake rate, *etc.*). However, these experiments came from different sources, correspond to cultures of different strains, and have been obtained following different experimental protocols. For this reason the consistency of each dataset has been evaluated beforehand using two different methods: (a) a simple carbon balance, and (b) a possibilistic consistency analysis against our stoichiometric model.

Carbon balance

The consistency of each experimental dataset has been evaluated checking that the measurements fulfilled a C-mol balance. This test could only be performed when measurements for the main uptake and production fluxes of carbon sources were available, which generally means that all substrates (glucose, glycerol and methanol), biomass and CO_2 rates were measured, as well as the main possible byproducts (ethanol, pyruvate, and citrate). The actual

elemental composition of biomass and ash content were taken into account whenever available; otherwise a mean composition was used. A general elemental composition for recombinant protein was taken from [3]. In those cases where heterologous protein was measured, it was included in the carbon balance; however, as the carbon content was small, it was neglected in those datasets where protein production was unknown.

In summary, for 52 datasets the carbon balance was checked based on measurements of glucose, glycerol, methanol, CO_2, biomass, protein, pyruvate, ethanol, and citrate (note: in some cases the byproducts were not measured, but reported negligible). For datasets 17, 18 and 50–52 protein production was unknown, but its carbon content was assumed to be negligible. Finally, datasets 29 to 45 and 53 to 55 could not be checked because the CO_2 production rate was unknown.

Possibilistic MFA

As a complementary test, and also to deal with those experimental datasets lacking a carbon-balance, we perform a different consistency analysis based on Possibilistic MFA. The method was described in [5,26] and applied in [24,25]. Details can be found in those works, but a short description follows. First, we describe the Possibilistic MFA method, and then we explain how it can be used to perform a consistency analysis.

Possibilistic MFA takes into account that experimental measurements are imprecise and do not exactly satisfy the constraints in (1). All measurements are thus considered relatively uncertain, as follows: $\mathbf{w_m} = \mathbf{v_m} + \mathbf{e_m}$, where $\mathbf{e_m}$ is a vector containing the errors (or deviations) between the actual fluxes **and** their measured values. Similarly, these measurement errors can be represented with two sets of non-negative variables, ε and μ:

$$M\varepsilon C = \begin{cases} \mathbf{w_m} = \mathbf{v_m} + \varepsilon_1 - \mu_1 + \varepsilon_2 - \mu_2 \\ \varepsilon_1, \mu_1 \geq 0 \\ 0 \leq \varepsilon_2 \leq \varepsilon_2^{\max} \\ 0 \leq \mu_2 \leq \mu_2^{\max} \end{cases} \quad (2)$$

Each candidate solution of (1) and (2) can be denoted as δ. Then, we (as users) define a function that assigns possibility in [0, 1] to each solution, ranging between impossible and fully possible. A simple way is using a linear cost index as:

$$\mathbf{J}(\delta) = \alpha \cdot \varepsilon_1 + \beta \cdot \mu_1 \quad (3)$$

Then, the possibility of each solution can be defined as:

$$\pi(\delta) = \exp(-\mathbf{J}(\delta)) \qquad \delta \subset \Delta \quad (4)$$

Where α y β are row vectors of user defined, sensor accuracy coefficients. The results can be interpreted as "$\mathbf{v_m} = \mathbf{w}$ is fully possible; the more $\mathbf{v_m}$ and $\mathbf{w}$ differ, the less possible such situation is". In particular, and for all our computations, the bounds $\varepsilon_2^{\max}$ and $\mu_2^{\max}$ have been chosen to define an interval of fully possible values around the measured ones (±5% deviation); while the weights α and β have been chosen to a decreasing possibility to larger deviations (e.g., deviations larger than ±20% have a possibility of lower than $\pi = 0.1$). More details can be found in [25].

At this point, Possibilistic MFA provides flux estimates accounting for uncertainty. For instance, the simplest flux estimate $\mathbf{v}_{\mathrm{mp}}$ in δ_{mp} is given by a maximum possibility (minimum cost) solution of the constraint satisfaction problem (1)-(2), which can be obtained solving a linear programming (LP) problem.

$$\mathbf{J^{min}} = \min_{\varepsilon,\mu,v} \mathbf{J} \quad s.t \{M\mathcal{O}C \cap M\varepsilon C \quad (5)$$

This most possible solution given by (5) has an associated degree of possibility:

$$\pi^{mp} = \exp(-J^{min}) \quad (6)$$

This value in [0, 1] provides our consistency check. This value π^{mp} is the possibility of the most possible flux distribution. It is grading the degree of consistency between different measurements, and between the measurements (2) and the model constraints in (1). A possibility equal to one must be interpreted as a complete consistency, while lower values imply that there is some error in measurements or in the model.

Finally, there is a similar way of express the degree of consistency provided by the possibilistic method. In this case, we calculate the percentage of measurements error (in $\varepsilon_2^{\max}$, $\mu_2^{\max}$) that must be allowed to find a solution with possibility equal to 1. We denote this degree of "assumed error" as AE index. Clearly, the larger this index is, the more inconsistent measurements are. For example, an AE index of 10% implies that a 10% of flexibility is required around all the measurements to find a solution that fulfills simultaneously the measurements and model constraints.

Note: This consistency analysis assumes that model constraints are accurate; but let us remark that the FBA hypothesis, which will be evaluated along this paper, has not been included so far. The model used in the consistency analysis was validated before and has been proved to be relatively reliable [24,25].

Flux balance analysis

Several flux balance analysis (FBA) simulations have been performed. As stated in the backgrounds section, FBA is a methodology to get predictions from a constraint-based model by assuming that the cells behave optimally. In this way, predictions are obtained by

solving an optimization problem: maximize the (hypothetical) cells objective function subject to the constraints that are imposed by the model.

If the objective function is linear and the constraints are linear equalities and inequalities —which is the case for all our computations—, the FBA problem can be formulated as a linear programming problem. In this case, predictions can be obtained following a simple and efficient four-step procedure.

First: define a set of model constraints (MOC), such as in (1). These constraints are always the same for a given organism, independently of its environment and particular circumstances.

Second: incorporate context-dependent constraints, which represent the scenario that the modeled organism is facing in a particular case. For example, these constraints define which substrates are available or if there is oxygen in the media. In general, these constraints will be inequalities:

$$\mathbf{v}_u^{min} \geq \mathbf{v}_u \geq \mathbf{v}_u^{max} \tag{7}$$

Third: define a biologically relevant objective function Z that is assumed to represent the cells objective, as result of evolutionary pressure. In all our computations this objective will be to maximize growth. The objective function is defined as follows (where d is column vector of size n with zeros in every position but the one corresponding to the biomass growth):

$$Z = \mathbf{d} \cdot \mathbf{v} \tag{8}$$

Fourth: finally, predictions are obtained by solving a linear programing problem to compute the flux distribution that makes the optimal use of the available resources, (*i.e.*, that maximizes the objective function Z).

$$\mathbf{v}^{opt} = \max_{\mathbf{v}} Z \; s.t \left\{ \begin{array}{c} \boldsymbol{N} \cdot \mathbf{v} = 0 \\ \boldsymbol{D} \cdot \mathbf{v} \geq 0 \\ \mathbf{v}_u^{min} \geq \mathbf{v}_u \geq \mathbf{v}_u^{max} \end{array} \right\} \tag{9}$$

All FBA computations have been performed with MATLAB (MathWorks Inc., 2009) and YALMIP Toolbox [27].

Results and discussion

P. pastoris constraint-based model building

Along this paper, a small-sized, constraint-based model of *P. pastoris* shown in Figure 1 will be used. The model is a modified version of the one previously described and validated in [24], which was based in a previous model by Dragosits *et al.* [28] it is a standard constraint-based model, whose generalities are described in [5] or [2].

As a constraint based model, it was derived from the knowledge about *P. pastoris* metabolic network. The model is not a comprehensive representation of *P. pastoris* metabolism, but it includes the main catabolic pathways (Embden-Meyerhoff-Parnas pathway, citric acid cycle, pentose phosphate and fermentative pathways), considers the uptake of several carbon sources (glucose, glycerol, and methanol) and accounts for biomass growth and ATP balance. Metabolites such as NAD, AcCoA, oxaloacetate, or pyruvate are considered in both cytosolic and mitochondrial pools.

Two new reactions have been incorporated to the model described in [24] in the pyruvate metabolism and in the mitochondrial transport. The new reactions are:

Reaction 36: $ATP + Oxaloacetate \rightarrow ADP + Phosphoenolpyruvate + CO2$.

Reaction 37: $Acetyl - CoAmit \leftrightarrow Acetyl - CoA$.

The model contains 47 metabolites and 48 metabolic reactions. There are 37 internal metabolites that are assumed balanced, which define a 37x48 stoichiometric matrix **N** with 11 degrees of freedom. All internal reactions are considered irreversible, except for reactions; 2–8, 15, 22–27, 29, 34, 37 and 44. The matrix and the list of reactions are given in the Additional file 1.

P. pastoris FBA models

Along this paper the word "model" is used to denote two different representations of *P. pastoris.* The first one is the constraint-based model of *P. pastoris* that we have already defined which contains only information regarding its central metabolism and reactions irreversibilities. The second type of model emerges when we combine this constraint-based model with a biological objective for the cells (maximizing growth), so that we obtain a complete FBA model as defined in the methods sections. Please recall that the main goal of this paper is to evaluate the validity of the second model, i.e., the validity of assuming that *P. pastoris* cells objective is maximizing its growth rate. Hereinafter, we will denote this second model as *FBA model.*

Recompilation and analysis of experimental data

Thus, the main goal of this paper is to validate the predictions of an FBA model. To do that, experimental datasets from different chemostat experiments have been collected from literature. We collected data from 72 chemostat experiments that correspond to *P. pastoris* cultures growing on methanol, glycerol, glucose or mixtures of these substrates. Each dataset is defined by a set of experimental measurements of several extracellular fluxes (*e.g.,* biomass growth, glucose uptake rate, oxygen uptake rate, *etc.*). The number of available measurements in each dataset is not always the same, mostly because gas measurements are sometimes unavailable. Most datasets correspond to recombinant strains, resulting in the production of a heterologous protein. All datasets can be found in Additional file 2.

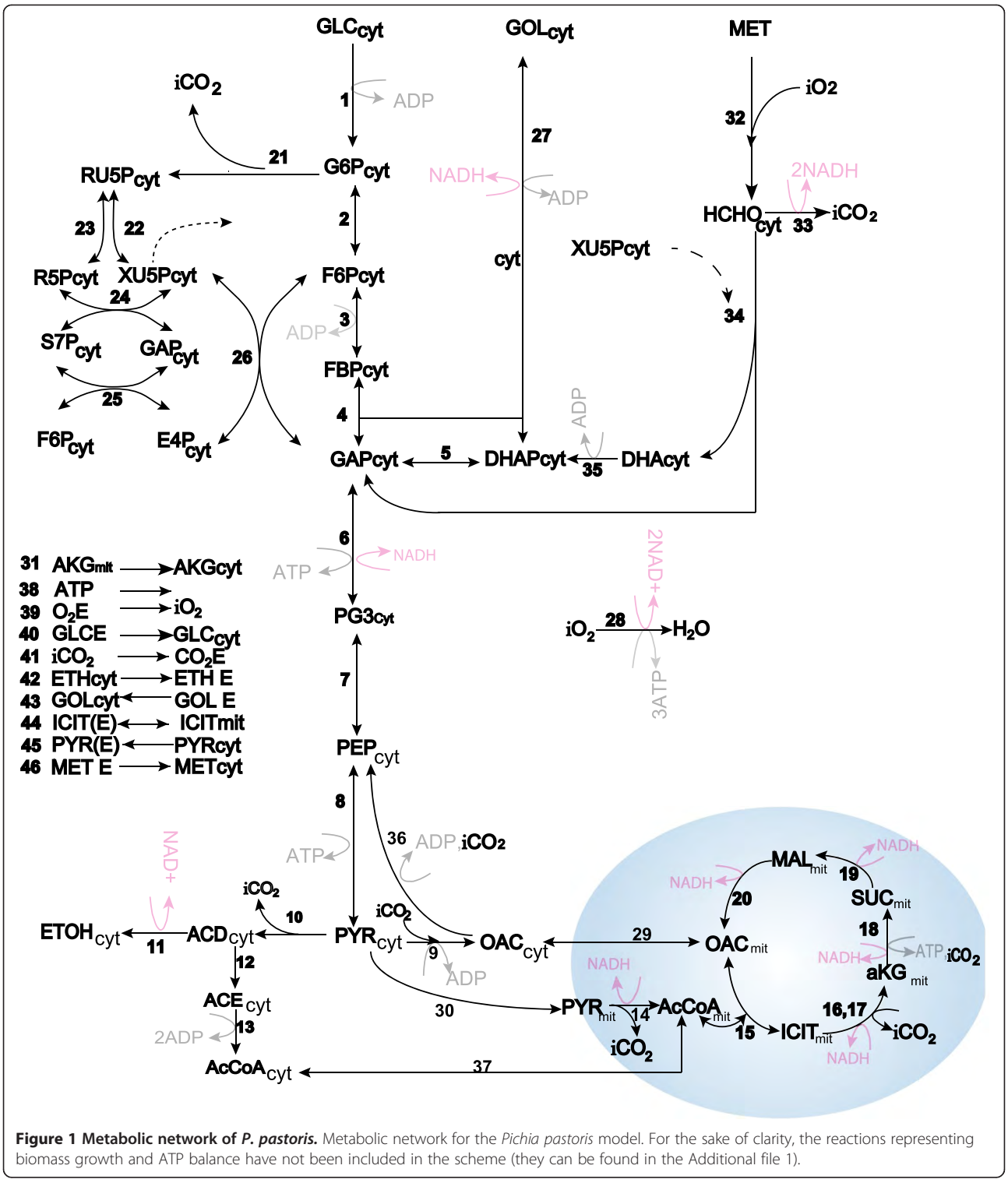


Figure 1 Metabolic network of *P. pastoris*. Metabolic network for the *Pichia pastoris* model. For the sake of clarity, the reactions representing biomass growth and ATP balance have not been included in the scheme (they can be found in the Additional file 1).

Please notice that the experimental datasets come from different sources and correspond to experiments with different strains and different experimental protocols. For this reason, before using them, the consistency of each dataset has been evaluated using two different methods: (a) a simple carbon-balance, and (b) a possibilistic consistency analysis against our stoichiometric model. Both methods are described in detail in the methods section. The complete results of these analyses can be found in the Additional file 2.

The carbon-balance test of consistency could only be performed with 52 datasets for which CO_2 measurements were available. The consistency is reasonably good for the majority of the tested datasets, with a deviation minor than 10% in carbon content for datasets; 1–4, 7–14, 46–48, 50, 51, 56–72. Only a few datasets (5, 6, 15, 24–28, 49) have a deviation higher than 10%.

To provide further validation of the data, and deal with those datasets which consistency cannot be evaluated with a carbon balance, a possibilistic MFA consistency test was also applied. Again, most of the datasets are highly consistent with the model: 72% are fully possible and only 4 in 72 datasets have an AE index larger than 15% —this includes the intrinsic uncertainty of any measure (*e.g.* calibration errors, offsets, etc.).

As a result of the analysis, datasets 5, 6, and 15 have been classified as inconsistent with both methods. This result suggests that measurement errors are likely in those datasets. We have decided to keep all datasets in our further analysis, but these ones will be labeled as less trustworthy data.

Validation 1: prediction of growth and yields on single substrates

Several validation tests will be performed in subsequent sections in order to validate our *P. pastoris* FBA model. First, we will check if the model is able to predict growth on several substrates (glucose, glycerol and methanol). Then, we will check if the theoretical biomass yields on these substrates are in agreement with the actual yields that *P. pastoris* shows in experimental conditions.

Simulation procedure

To predict the biomass yield we compute a set of FBA simulations, one per each substrate (glucose, glycerol, and methanol). In each simulation all substrate uptakes were fixed to be zero (thus representing the substrate unavailability) except one, which was fixed to be 1 mmol/g/h (the exact value is not important, since we will be calculating yields). Oxygen uptake was assumed to be unlimited. This way we represent a scenario where one single substrate is being consumed, no other substrates are available, and oxygen is not limited. The assumed cells objective is maximizing growth.

In summary, we are predicting how *P. pastoris* cells will be using each substrate in the selected scenarios, according to our model constraints and the assumption of growth maximization as evolutionary objective.

We performed our simulations to get the optimal flux distribution that is the model prediction (see methods). Then we compute biomass growth yields (Yx/s) based on the flux values of the optimal solution. These values are finally compared with experimental yields taken from literature. We also included the yields reported in a genome-scale model of *P. pastoris* [29]. The comparison is presented in Table 1.

Results

We first checked that, as expected, our FBA model is able to sustain growth on all three single substrates. Glucose, glycerol and methanol are sufficient in their own to produce all precursors and energy requirement for growth. According to the model, the best carbon source was glucose (with a yield of 3.97 Cmol dcw/mmol) followed by glycerol (2.26 Cmol dcw/mmol), and finally methanol (0.66 Cmol dcw/mmol). This ranking is in agreement with data previously reported [30], supporting the idea that the set of reactions considered in our model is capturing relatively well the main metabolic pathways *P. pastoris*.

Furthermore, the predicted biomass yields for all three substrates are found to be in reasonably good agreement with the average experimental yields of our 72 datasets, and also with the values reported for Caspeta's genome-scale model. This provides a first validation for the model constraints and also for the hypothesis of maximal growth as cells objective, as it seems able to capture (partially, at least) the metabolic regulation that *P. pastoris* has evolved and which determines its behavior in the presence of these substrates. Notice, however, that the predicted yields tend to be larger than the experimental ones. The best agreement is shown with glycerol and glucose (around 13% overestimation), but deviation is significant with methanol (around 50% overestimation).

We suggest three tentative hypotheses to explain these last results.

Firstly, the simplicity of our model makes us disregard other operating constraints (*e.g.,* thermodynamics, availability

Table 1 *P. pastoris* yields in single substrates

	Methanol			Glucose			Glycerol		
	Yx/s Cmmol mmol-1	$Y_{S/O2}$ mmol mmol-1	$Y_{S/CO2}$ mmol mmol-1	Yx/s Cmmol mmol-1	$Y_{S/O2}$ mmol mmol-1	$Y_{S/CO2}$ mmol mmol-1	Yx/s Cmmol mmol-1	$Y_{S/O2}$ Mmol mmol-1	$Y_{S/CO2}$ mmol mmol-1
FBA (this work)	0.66	0.83	0.34	3.97	1.97	2.03	2.26	1.21	0.74
FBA (Caspeta)	0.49	1.43	0.49	3.91	1.53	1.96	2.23	0.95	0.68
Exp. (average)	0.42 ± 0.09	1.06 ± 0.06	0.55 ± 0.02	3.41 ± 0.66	1.44 ± 0.58	1.84 ± 0.4	1.99 ± 0.17	1.33 ± 0.27	1.01 ± 0.18

of other nutrients, *etc.*) additional to stoichiometric and irreversibility constraints that could also influence the actual capabilities of the microorganism, resulting in actual yields lower that predicted.

Secondly, our model is not accounting for recombinant protein production, which occurs in the majority of the experiments used for validation, and which is known to affect *P. pastoris's* use of available resources (and generally, but not always, to result in lower growth).

Finally, the assumption of growth maximization may not perfectly capture the actual cells evolutionary objectives (which may be more subtle and complex). This seems particularly likely when methanol is the substrate, since the deviation is larger in these scenarios.

All these three issues will be discussed in more depth in subsequent sections, where more data will be available.

Validation 2: FBA predictions in real scenarios

For the next validation of our FBA model, we will define scenarios where some substrates are available (glucose, methanol, or mixes of ethanol and glycerol). Then, we will use the FBA model to predict if and how these substrates will be consumed. These scenarios correspond to our 72 datasets, so we will have data to validate the model predictions. Predictions of growth, substrate uptake, respiration rates and byproduct formation rates will be validated against experimental data in each case.

Simulation procedure

Each scenario is defined by the availability of each substrate (glucose, glycerol and methanol), which is represented by binding their uptake to a maximum value equal to the experimental one, as reported in the corresponding dataset ($v_i \leq v_{i,measured}$). Notice that the uptake flux values are not fixed, but just bounded. To represent the unavailability of substrates their uptake flux is fixed to be zero. The oxygen uptake rate was not restricted, thus assuming that it was not the limiting factor (notice that this makes the prediction more difficult: if oxygen was indeed a limitation in some scenarios, our model will not have this information about the environment that cells are facing). As before, the objective function used in the FBA model is growth maximization.

Results

Prediction of growth, substrate uptake, respiration rates, and byproduct formation rates are given in Figure 2 and Table 2 for each scenario. As shown in Figure 2 and Table 2, predictions of growth and substrate uptake are remarkably accurate in scenarios growing on glycerol and glucose. It seems clear that growth maximization is a quite reasonable assumption in these scenarios. It seems that substrates tend to be used through pathways that result in almost optimal growth. Notice also that byproduct formation is not predicted in any scenario, which is also in agreement with the experimental evidence.

Predictions of oxygen uptake rate and carbon production rate are less accurate. This may pinpoint modeling errors (in the model constraints or in the assumption of maximizing growth), but also errors in gas measurements: these measurements are generally less reliable, since they are based on determinations of the exhaust gases flow and concentration, which are prone to substantial experimental deviations.

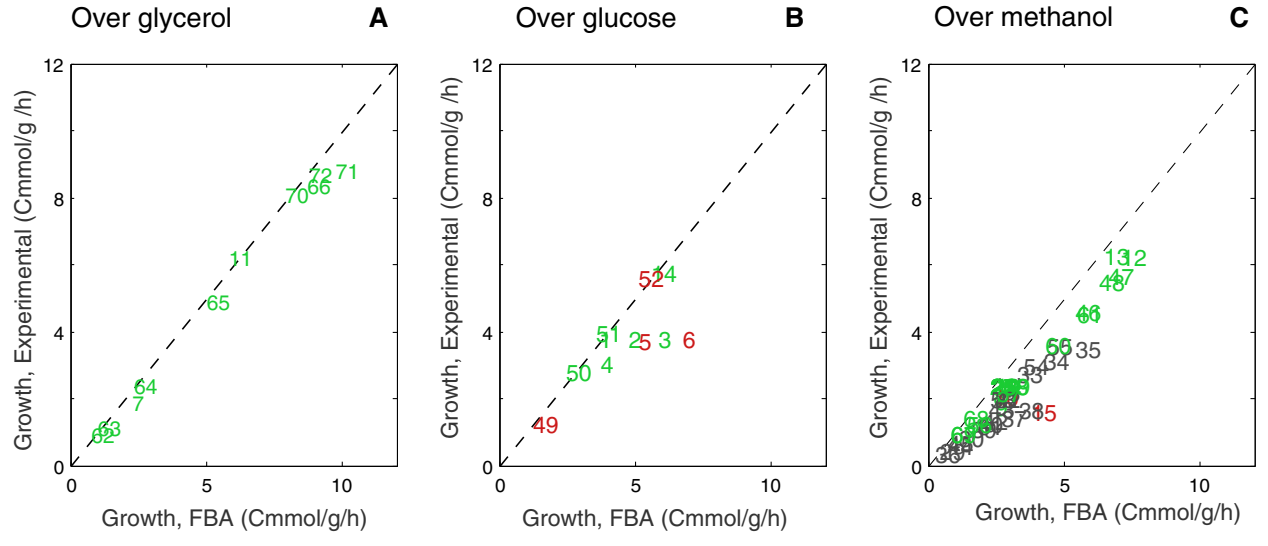


Figure 2 FBA growth predictions vs. experimental growth. Comparison of predicted growth and experimental growth for each dataset with different substrates as carbon sources: **A)** glycerol, **B)** glucose, **C)** methanol and methanol-glycerol mixtures. Green labels represent consistent datasets, whereas red ones are those classified as inaccurate. Gray represents those in which the carbon balance could not be checked.

Table 2 FBA predicted fluxes vs. experimental fluxes

Glycerol										
Code	**μ**		**Glycerol**		**Oxygen**		**CO_2**		**By-products**	
	msd[1]	**ptd[2]**	**msd**	**ptd**	**msd**	**ptd**	**msd**	**ptd**	**msd**	**ptd**
7	1.88	2.46	1.09	1.09	2.16	1.32	1.56	0.81	0.00	0.00
11	6.17	6.21	2.75	2.75	3.62	3.33	2.35	2.04	0.00	0.00
62	0.90	1.16	0.52	0.52	0.82	0.62	0.62	0.38	NR[3]	0.00
63	1.11	1.39	0.62	0.62	0.87	0.75	0.65	0.46	NR	0.00
64	2.38	2.74	1.21	1.21	1.65	1.47	1.22	0.90	NR	0.00
65	4.89	5.42	2.40	2.40	3.12	2.91	2.29	1.78	NR	0.00
66	8.37	9.13	4.04	4.04	4.77	4.89	3.40	3.00	NR	0.00
70	8.07	8.31	3.68	3.68	3.99	4.45	2.96	2.73	NR	0.00
71	8.79	10.16	4.50	4.50	4.71	5.44	4.70	3.34	NR	0.00
72	8.66	9.17	4.06	4.06	4.19	4.91	3.53	3.01	NR	0.00
NRMSE[4]	12%		0%		14%		23%			
Median error[5]	13%		0%		13%		24%			
Glucose										
Code	**μ**		**Glucose**		**Oxygen**		**CO_2**		**By-products**	
	msd	**ptd**	**msd**	**ptd**	**msd**	**ptd**	**msd**	**ptd**	**msd**	**ptd**
1	3.72	3.83	0.96	0.96	2.00	1.90	2.09	1.95	≥0.02[a]	0.00
2	3.72	3.74	0.94	0.94	1.78	1.85	1.87	1.90	≥0.04[a]	0.00
3	3.74	3.64	0.92	0.91	1.69	1.80	1.75	1.85	0.00	0.00
4	3.01	3.93	0.99	0.99	2.12	1.95	2.37	2.01	≥0.03[b]	0.00
5	3.71	5.29	1.33	1.33	1.57	2.62	2.03	2.69	≥0.3[c]	0.00
6	3.73	6.92	1.74	1.74	0.54	3.43	1.65	3.52	≥1.0[c]	0.00
14	5.74	6.00	1.51	1.51	2.71	2.97	3.18	3.06	0.00	0.00
49	1.03	1.71	0.43	0.43	0.33	0.85	0.74	0.87	NR	0.00
50	2.57	2.78	0.70	0.70	0.78	1.38	1.15	1.42	NR	0.00
51	3.99	3.93	0.99	0.99	1.75	1.95	2.27	2.01	NR	0.00
53	5.26	5.56	1.40	1.40	1.34	2.76	2.12	2.84	NR	0.00
NRMSE		29%		0%		64%		32%		
% Median error		6%		0%		11%		15%		
Methanol										
Code	**μ**		**Methanol**		**Oxygen**		**CO_2**		**By-products**	
	msd	**ptd**	**msd**	**ptd**	**msd**	**ptd**	**msd**	**ptd**	**msd**	**ptd**
15	1.60	4.18	6.31	6.31	7.56	5.23	3.44	2.13	0.00	0.00
28	2.32	2.66	4.02	4.02	4.22	3.33	2.33	1.36	0.00	0.00
36	0.31	0.66	0.99	0.99		0.82		0.33	0.00	0.00
37	1.39	3.09	4.66	4.66		3.86		1.57	0.00	0.00
38	1.62	3.73	5.64	5.64		4.67		1.90	0.00	0.00
39	1.04	2.00	3.02	3.02		2.50		1.02	0.00	0.00
40	1.20	2.25	3.39	3.39		2.81		1.14	0.00	0.00
41	1.93	2.88	4.34	4.34		3.60		1.47	0.00	0.00
42	1.31	2.40	3.62	3.62		3.00		1.22	≥0.01[b]	0.00
43	1.66	2.67	4.02	4.02		3.33		1.36	≥0.01[b]	0.00
44	0.54	1.15	1.73	1.73		1.44		0.59	0.00	0.00

Table 2 FBA predicted fluxes vs. experimental fluxes *(Continued)*

45	0.66	1.11	1.67	1.67		1.38		0.56	0.00	0.00
53	1.97	2.69	4.06	4.06		3.36		1.37	NR	0.00
54	2.96	3.93	5.93	5.93		4.91		2.00	NR	0.00
55	3.54	4.76	7.18	7.18		5.95		2.42	NR	0.00
56	1.22	1.80	2.72	2.72	2.93	2.26	1.57	0.92	NR	0.00
57	2.12	2.94	4.44	4.44	4.70	3.68	2.48	1.50	NR	0.00
58	2.31	3.17	4.79	4.79	5.05	3.97	2.72	1.62	NR	0.00
59	2.34	3.21	4.85	4.85	5.08	4.02	2.68	1.64	NR	0.00
60	3.53	4.71	7.12	7.12	7.22	5.89	3.76	2.40	NR	0.00
61	4.47	5.90	8.90	8.90	8.67	7.37	4.46	3.01	NR	0.00
NRMSE		61%		0%		51%		45%		
% Median error		45%		0%		21%		39%		

Glycerol methanol mixtures

Code	μ		Glycerol		Methanol		Oxygen		CO_2		By-products	
	msd	ptd	msd	ptd	msd	ptd	msd	ptd	msd	ptd	msd	ptd
8	2.07	2.56	0.95	0.95	0.63	0.63	2.70	1.67	1.70	0.92	0.00	0.00
9	1.72	2.65	0.74	0.74	1.48	1.48	3.90	2.12	2.10	1.05	0.00	0.00
10	2.02	2.83	0.57	0.57	2.33	2.33	4.85	2.62	2.21	1.21	0.00	0.00
12	6.18	7.49	2.77	2.77	1.87	1.87	7.19	4.90	4.18	2.69	0.00	0.00
13	6.24	6.84	2.23	2.23	2.73	2.73	7.20	4.96	3.60	2.58	0.00	0.00
19	2.32	2.84	0.67	0.67	2.01	2.01	3.21	2.47	1.77	1.18	0.00	0.00
20	2.32	2.80	0.51	0.51	2.49	2.49	3.46	2.68	1.89	1.22	0.00	0.00
21	2.32	2.78	0.43	0.43	2.73	2.73	3.58	2.78	1.97	1.24	0.00	0.00
22	2.32	2.75	0.31	0.31	3.09	3.09	3.76	2.93	2.09	1.27	0.00	0.00
23	2.32	2.74	0.28	0.28	3.18	3.18	3.79	2.97	2.09	1.28	0.00	0.00
24	2.32	2.72	0.18	0.18	3.49	3.49	3.96	3.11	2.17	1.31	0.00	0.00
25	2.32	2.69	0.13	0.13	3.62	3.62	3.96	3.16	2.21	1.32	0.00	0.00
26	2.32	2.69	0.11	0.11	3.69	3.69	4.02	3.19	2.25	1.33	0.00	0.00
27	2.32	2.68	0.09	0.09	3.74	3.74	4.06	3.21	2.25	1.33	0.00	0.00
29	0.39	0.86	0.27	0.27	0.38	0.37		0.64		0.33	0.00	0.00
30	0.77	1.56	0.54	0.54	0.50	0.50		1.07		0.57	0.00	0.00
31	1.16	2.25	0.82	0.81	0.63	0.63		1.50		0.82	0.00	0.00
32	1.93	2.89	1.09	1.09	0.66	0.66		1.86		1.03	0.00	0.00
33	2.71	3.69	1.36	1.36	0.94	0.94		2.42		1.32	0.00	0.00
34	3.09	4.66	1.90	1.90	0.55	0.55		2.75		1.60	0.00	0.00
35	3.48	5.81	2.45	2.45	0.44	0.44		3.32		1.96	0.00	0.00
46	4.54	5.83	2.53	2.53	0.18	0.18	4.78	3.21	3.25	1.94	0.00	0.00
47	5.63	7.06	2.61	2.61	1.76	1.76	5.35	4.61	3.09	2.53	0.00	0.00
48	5.44	6.72	2.22	2.22	2.58	2.58	5.73	4.82	3.33	2.52	0.00	0.00
NRMSE		32%		0%		0%		34%		39%		
% Md error		19%		0%		0%		23%		39%		

[1]Measured values from dataset. [2]Predicted values. [3]Non reported values. [4]Root mean square deviation normalized. [5]Median of percentage errors.
Note: The datasets 1, 2, 4, 5, 6, 42 and 43 reported small quantities of byproducts. [a]Ethanol and citrate, [b]citrate only, [c]ethanol, citrate and pyruvate.

It is also noticeable that discrepancies in methanol scenarios are larger than those in other substrates, with a median error of 45% for biomass growth (for 19% in mixes of glycerol-methanol, 12% in glycerol, and 6% in glucose). Again, this indicates that the FBA model is less precise in scenarios in which methanol is consumed. As we have already mentioned in the former section, there are several possible reasons for this behavior: (i) our underlying constraint-based model may have errors or limitation in the methanol pathways, *e.g.*, reactions and other constraints may be missing, (b) our model is not considering the resources devoted to produce recombinant protein, and (c) the hypothesis of maximizing growth could be less suitable in the case of methanol, since it is a less frequent substrate in the environment for which *P. pastoris* is selectively adapted.

Let us discuss in more depth what could explain these deviations between predicted and actual cells behavior.

The first reason to explain why predicted values are larger than the measured ones is that our model is only accounting for stoichiometric and irreversibility constraints, but there could be other operating constraints such as thermodynamic constraints or biochemical restrictions resulting from regulation (*e.g.* feedback inhibition of enzymes limiting the optimal use of substrates). This applies for all three substrates; however the overestimation in methanol is larger than in glycerol and glucose, suggesting that our stoichiometric model could be not accounting for relevant skills in the methanol metabolism. For example, phenomena such as accumulation of formaldehyde and hydroxide peroxide at high methanol concentrations may result in cell growth impairment as both oxidized products of methanol are toxic for the cell [31]. Biogenesis of peroxisomes, the central metabolism organelle for assimilation and dissimilation of methanol greatly disturbs cellular content, as it can occupy 90% of the cell volume during growth in methanol [32,33]. It should also be mentioned that the biomass equation in the model was adapted from other yeast (*S. cerevisiae*) and growth conditions (glucose as the only carbon source) [28]. Exclusive growth on methanol might also represent a highly specific cellular condition that would require the development of a biomass equation of its own for an improved predictive accuracy.

However, it is still remarkable that even if our model is a raw representation of the whole metabolism and even if metabolism is only part of all phenomena occurring within cells, imposing these constraints seems to be enough to allow reasonably accurate predictions.

A second reason to explain the deviation is that the assumption of growth maximization does not perfectly represent the evolutionary objectives of these cells. This is particularly plausible in the case of methanol, because it is a less common (or frequent) substrate in nature for *P. pastoris.* If this is the case, it would be an efficient evolutionary strategy to not completely regulate every metabolic reaction if methanol is the only available substrate in a given moment, because these conditions will not remain long time, and therefore the metabolic cost of regulate and deregulate every reaction could be an inefficient effort. This reasoning is in agreement with the hypothesis that a specific flux distribution at a certain condition might be chosen to minimize adjustment efforts to other conditions, as proposed in [17]. In addition, as methanol assimilation is a highly specific capability for this yeast, not seen in most species, it could be the case that optimal growth is not required to overtake competitors in an already favorable environment.

Finally, it must be taken into account that our model is not considering recombinant protein production. This can also explain why the predicted growth tends to be larger than the observed one. Metabolic precursors and energetic resources required to produce recombinant protein, as the stress that this production provokes in cells, are not taken into account in our predictions —instead, we are implicitly assuming that recombinant strains behave as a wild type strains, and thus no heterologous protein is produced—. These phenomena penalize substrate uptake, and thus growth, and will possibly impact also growth in terms of yield (although there is evidence suggesting the opposite in scenarios where glucose is the substrate [34]). If these phenomena related with protein production were taken into account in our model, the predicted growth might be lower and show a better agreement with experimental data.

In summary, our FBA model, which couples a constraint based model with the hypotheses of maximization of growth, shows an acceptable agreement with the experimental data of dozens of chemostat cultures of *P. pastoris,* especially when glycerol and glucose are the carbon sources. Several issues must be highlighted in this regard: (1) heterogeneity within the evaluated experimental conditions (different sources, microbial strains, recombinant proteins, culture conditions), where, in addition, measurement accuracy will not always be perfect; (2) our model does not consider all constraints operating in the system, but only (partial) stoichiometry and irreversibility; (3) we are assuming that cells behavior is optimal in one particular sense —growth—, what is an extreme and rough assumption; and (4) we are not considering the effects that protein production may have on cells behavior. These factors are clearly important. Anyhow, it is remarkable that even thought this model is a crude representation of whole metabolism, and metabolism is also a limited part of all cellular phenomena, those constraints seem to be relevant enough to result in reasonably accurate predictions.

Validation 3: predicting behavior under oxygen limitation

To continue the validation of our *P. pastoris* FBA model, we will investigate its behavior in aerobic and hypoxic conditions. First, we will check if the model is able to predict the qualitative behavior of cells for each single substrate.

Simulation procedure

We will predict the behavior of *P. pastoris* in microaerobic and aerobic conditions for each single substrate. To study growth over glucose, the glucose uptake was limited to be less than 1 mmol/g/h, while methanol and glycerol uptakes were fixed to be zero. Then we performed a set of FBA simulations with increasing levels of available oxygen (*i.e.,* the oxygen uptake rate will be successively limited to be less or equal than 0.01, 0.02 … *etc.* up to 10 mmol/g/h). This way, a range of scenarios is represented, where glucose can be consumed, no other substrate is available, and oxygen changes from scarce, to abundant. In all these simulations the cells objective was maximizing growth. This exercise was repeated in three scenarios where only one substrate was available at a time. This way, we predict the aerobic and hypoxic behavior of *P. pastoris* over each single substrate to check if it correctly fits with actual cells behavior.

Results

The model predictions for each single substrate and different oxygen conditions are shown in Figure 3. Each graph shows the substrate uptake rate, the biomass growth rate, and byproduct production. Comparing the results, it can be observed that that glucose is predicted to be the most efficient substrate both in aerobic and microaerobic conditions (it achieves a better yield, as we already knew). Methanol will be the least efficient substrate, both in aerobic and microaerobic conditions.

Figure 3A also shows that our FBA model predicts that growth on glucose will be qualitatively different depending on oxygen availability. In microaerobic conditions, glucose is consumed via fermentative pathways (although some respiration is occurring as can be seen in Figure 3B), and thus ethanol is produced as a byproduct. These predictions are in accordance with the experimental evidence previously reported [35,36]. In those studios *P. pastoris* growth on glucose shows a facultative anaerobic behavior with oxygen limitation; however this leads to byproduct formation, especially ethanol, and also arabinitol [37]. Little information is known about the impact of oxygen availability on the physiology of recombinant yeasts, but it is well described that *P. pastoris* growth is higher in respiratory rather than fermentative mode [38]. Oxygen limitation strongly affects the core metabolism by causing energy deprivation, affecting growth, and cells have to readjust their metabolic fluxes from cellular respiration to fermentation [39].

According to our predictions, the maximum ethanol production rate will be achieved with an oxygen uptake around 0.2 mmol/g/h per 1 mmol/g/h of glucose ($Y_{EtOH/Glu}$ = 1.53 mmol/mmol, $Y_{x/glu}$ =1.17 Cmmol/mmol). If more oxygen is available, there is a switch from fermentative to respirative pathways —which are more efficient in terms of biomass yield, but require more oxygen—, and therefore ethanol production tends to be lower. This also makes sense from a biological standpoint. If oxygen uptake is larger than 1.96 mmol/g/h per 1 mmol/g/h of glucose, ethanol will no longer be produced, because oxygen is now in excess, and glucose can be completely consumed via respirative pathways ($Y_{EtOH/Glu}$ = 0.00 mmol/mmol, Yx/s = 3.97 Cmmol/mmol). In this situation, the optimal growth is achieved by directing fluxes through pathways that do not involve ethanol production.

Figure 3B shows that our predictions for growth on (only) glycerol depend also on oxygen availability. The results are analogous to those obtained with glucose: ethanol is produced when oxygen is scarce, because fermentative pathways are active, but at lower rates that those predicted with glucose [40]. This agrees with the experimental evidence: even if glycerol is typically considered a non-fermentable carbon source in *P. pastoris,* residual ethanol production has been reported both in batch and fed-batch cultures [41,42]. It could be hypothesized that this lower tendency of *P. pastoris* to fermentation over glycerol with respect to glucose may be due to the extra NAD^+ that glycerol uptake requires (in reaction 27).

Conversely, as it is shown in Figure 2C, the behavior of *P. pastoris* is different when growth is sustained on methanol: ethanol is never produced as byproduct even if oxygen is limited. Despite oxygen scarcity, our model always predicts that methanol will be consumed via respirative pathways, and never by fermentative metabolism. One obvious reason is that oxygen is required to metabolize methanol (by reaction 32), and therefore fermenting methanol is an inefficient way of getting NADH or ATP, because respiration (reaction 28) provides a better alternative—more economical in terms of oxygen—to get these resources. According to our model methanol fermentation is possible, but inefficient, and thus it is not predicted to occur.

Validation 4: predicting substrate preferences and a behavior in hypoxic conditions

To continue the analysis of the previous section, we will now check if the model correctly predicts the preferences among multiple substrates that *P. pastoris* cells exhibit when facing an environment where oxygen is limited.

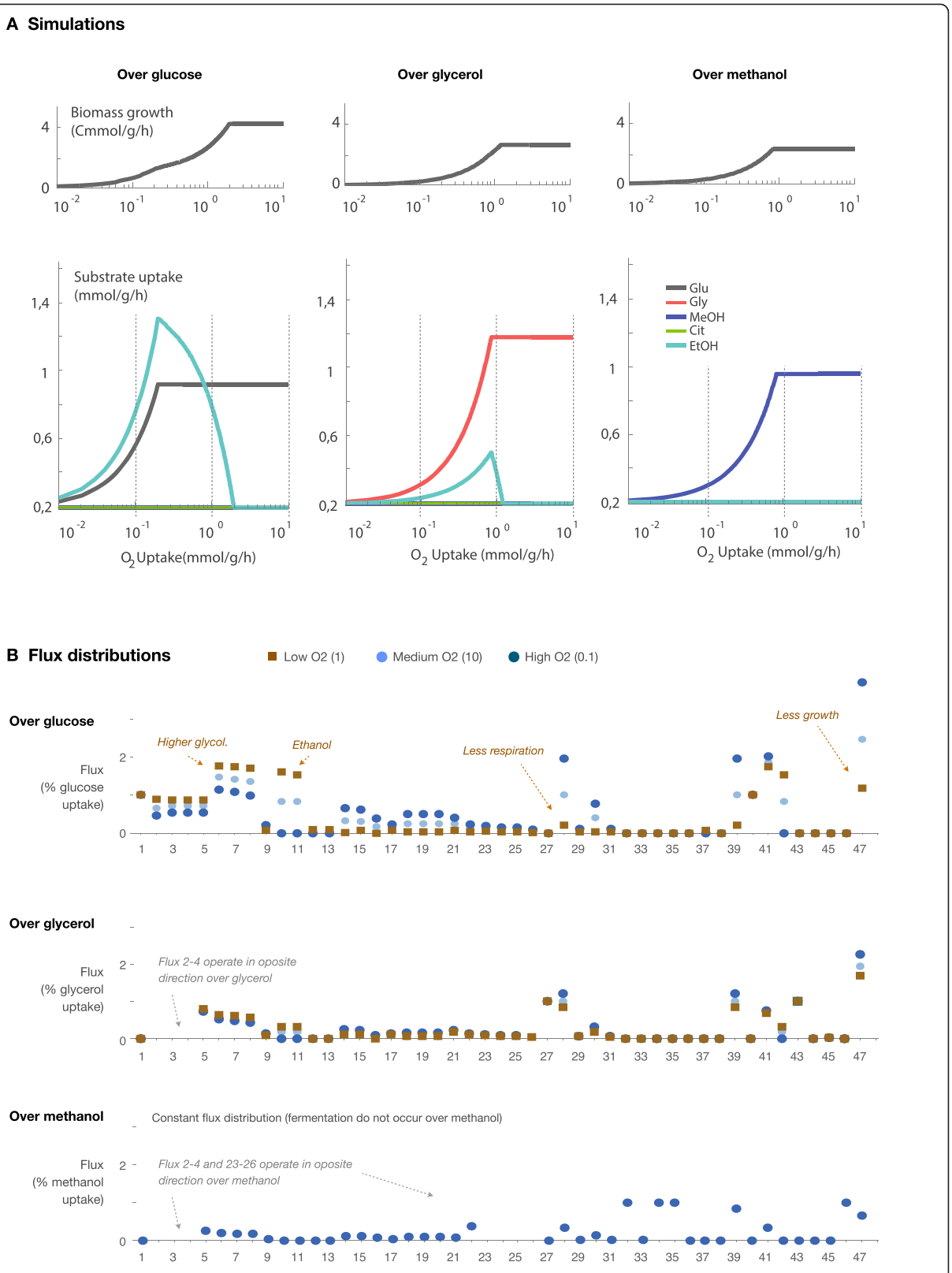


Figure 3 FBA predicted behavior under oxygen limitation. A) Biomass growth (upper panel) and substrate uptake and byproduct production (lower panel) predicted for *P. pastoris* cultures growing over a) glucose, b) glycerol, and c) methanol. **B)** Flux distributions predicted for *P. pastoris* cultures growing over glucose, glycerol, and methanol in different oxygen conditions.

Simulation procedure

In this simulation all three substrates were assumed to be available simultaneously. Glucose, glycerol and methanol were all limited to be less than 1 mmol/g/h. Then we performed a set of FBA simulations with increasing levels of available oxygen (*i.e.*, oxygen uptake rate was successively limited to be less or equal than 0.01, 0.02 … *etc*. up to 10 mmol/g/h). This way, we represent a range of scenarios where all substrates are available and oxygen ranges from scarce to abundant. In all these simulations the cells objective was maximizing growth. In these scenarios *P. pastoris* cells could consume the three substrates, but a preference could be shown because oxygen was limited. This way, the substrate preference of *P. pastoris* will be predicted.

Results

The results for the battery of simulations are shown in Figure 4A. According to our FBA model, if methanol, glycerol and glucose are simultaneously fed, but oxygen is limited (less than 0.28 mmol/g/h per 1 mmol/g/h of glucose), *P. pastoris* shows a preference for glucose as carbon source. Glucose is consumed, while the others substrates are not. Simply, if oxygen availability limits the substrate uptakes, the most efficient source (in terms of yield) will be preferred. If more oxygen is available, the model predicts that glycerol will be the next substrate to be consumed, and methanol the last one. These results are in concordance with the preferences reported by Inan & Meagner —they observed that if glycerol, acetate, ethanol and methanol were present, the order of utilization was glycerol, ethanol, acetate, and finally methanol [30].

Now, let us elaborate about the four situations that our model predicts depending on how much oxygen is available. See Figure 4B and C for details about each phase.

Phase I. Cells use the first available oxygen to grow on glucose, showing a fermentative behavior that result in ethanol as by-product (pathway 1 in Figure 4B and C). This prediction is in good agreement with experimental results [35]. This behavior is shown until the oxygen is sufficient to metabolize all the available flux of glucose.

Phase II. If some more oxygen is available, glucose is still the only substrate being consumed, but now partially through respirative pathways. This implies that there is a partial metabolic switch in order to start using pathways that allow for an optimal use of glucose (in terms of growth), but that require more oxygen than those exhibited in hypoxic conditions (Phase I). As a result, the production of ethanol slightly decays. This behavior is only shown for a small range of oxygen levels: if they increase above 0.29 mmol/g/h per 1 mmol/g/h of glucose, then glycerol starts to be consumed.

Phase III. When the oxygen uptake is larger than 1.13 mmol/g/h per 1 mmol/g/h of glucose and glycerol, the FBA prediction is that glucose and glycerol will be consumed simultaneously. There is now enough oxygen to consume all the available glucose, so the "excess" is devoted to consume glycerol, while ethanol will appear as a byproduct in larger quantities —indicating that both substrates are mainly consumed through fermentative pathways (pathways 1 and 3 in Figure 4B)—.The production of ethanol and other byproducts in cultures with glycerol and glucose as carbon sources has also been reported in experimental observations [40]. The switch between phases II and III, which cannot be consequence of substrates (which do not change), could be related with NADH and ATP acting as limitants via oxygen restriction.

Phase IV. If oxygen is even more abundant, the next transition is that glycerol and glucose will be still consumed, but using the more efficient respirative pathways (the change occurs from pathways 1 and 3 to 2 and 4 in Figure 4). As a result, ethanol production tends to zero as oxygen availability increases.

Phase V. Finally, if there is more than enough oxygen to consume all the glucose and glycerol via respirative pathways, methanol is predicted to be consumed. Since methanol is the least productive substrate, the model prediction is that it will only be consumed if there are no other substrates available, or if oxygen is in high excess.

These results show that if methanol, glycerol and glucose are simultaneously fed in a limited scenario (in this case by the available oxygen), our FBA model predicts that *P. pastoris* will show a preference for glucose, followed by glycerol, and finally methanol, what is in agreement with experimental observations [41]. Notice that our FBA model is based solely on metabolic constraints and the hypothesis of maximal growth, and includes no knowledge about regulation, signaling or any other processes occurring within the cells. Remarkably, the optimality assumption is sufficient to predict (i) the substrate preference, and (ii) the use of fermentative or respiratory pathways, without representing the complex regulative machinery that cells have evolved in order to govern these processes.

Nevertheless, our FBA predictions fail in predicting co-consumptions of substrates in phases III to V. When the preferred substrate is limited (glucose) but oxygen is still available, our model predicts that the second best substrate will be consumed (glycerol). Yet, this behavior is not shown in actual batch cultures. As it is well known, when glucose, glycerol, and methanol are accumulated in culture media, they will be consumed sequentially due to enzyme regulation through catabolite repression (if the cells sense the presence of glucose, a regulation process will occur to inhibit the catabolic

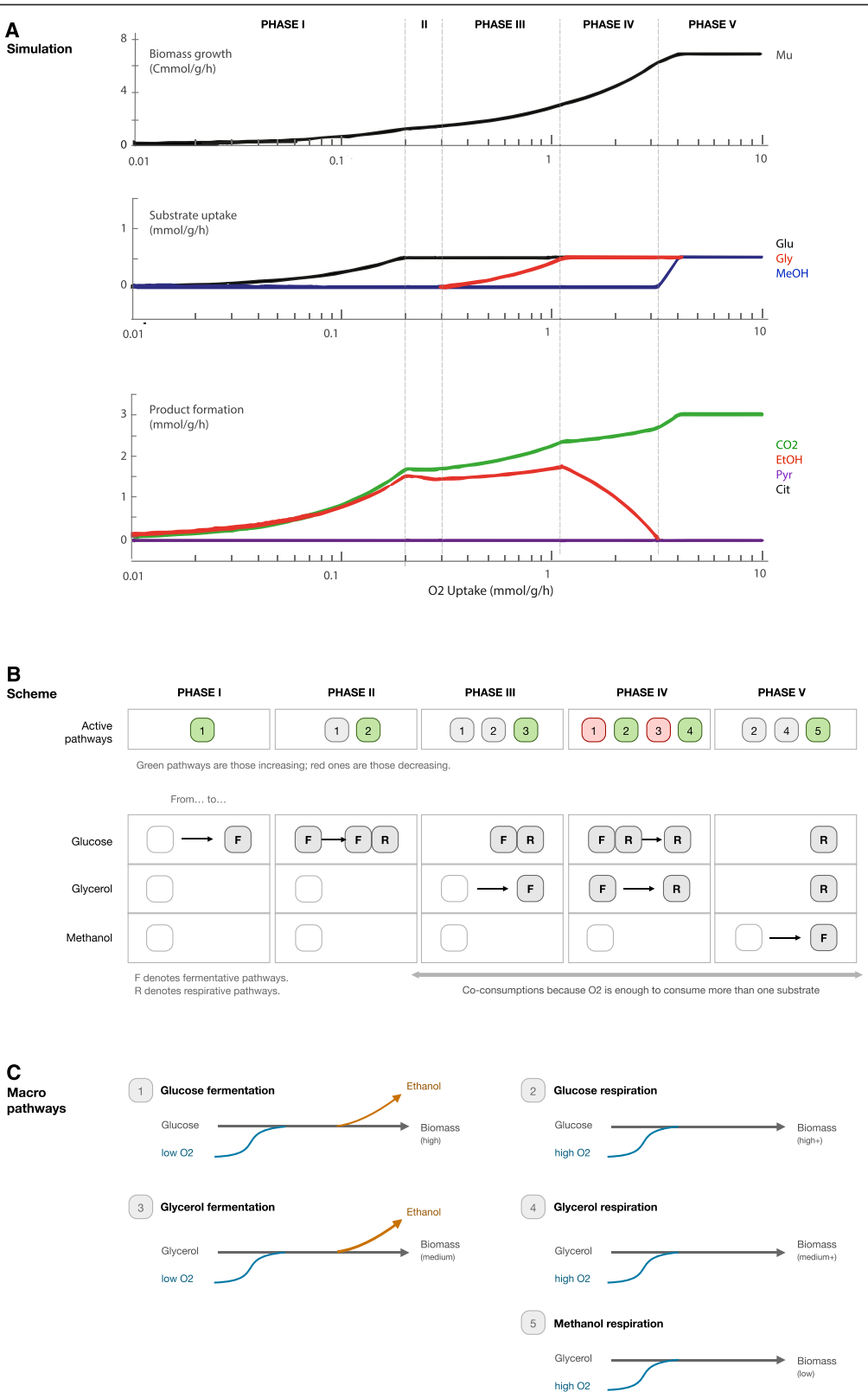


Figure 4 Behavior under O2 limitation with multiple substrates. A) Predictions of *P. pastoris* growth (up), uptakes (middle) and byproducts (down) in scenarios where the three substrates are available but oxygen ranges from scarce, to limiting condition, to abundant. **B)** Schematic active pathways in each different phase. **C)** Schematic macro pathways showed with each substrate.

pathways of glycerol and methanol). The same phenomena occur when glycerol (but not glucose) is available; methanol uptake pathways will be inhibited. This catabolic regulation —which occurs at transcriptional level— is the mechanism that cells have evolved in order to implement the substrate preference that we have predicted to result in optimal growth.

But why our FBA model predicts co-consumptions when oxygen is available in excess? Or better, why cells have not evolved a machinery to show this behavior if it is predicted to be more efficient? The explanation, in our opinion, could be in our model setting, which is not accounting for other constraints limiting the "biological activity" in a broad sense, such as transport processes, enzyme production, scarcity of cellular anabolic machineries (*e.g.*, ribosomes), *etc*. If oxygen or a single substrate acts as limitant, our predictions are reasonable; however, if those limits are not active at certain conditions, our model lacks the remaining constraints and tends to predict more growth (or, in general, "biological activity") that the one actually possible. In other words, if we include in our model any kind of limiting factors, the predictions tend to be in agreement with actual cells behavior, but when these limiting factors are missing, our predictions will predict more activity than the actual one, as it happens with co-consumptions.

Finally, notice that in fed-batch cultures —where the catabolic regulation will not occur because the substrate is not accumulated and therefore cells are unable to sense its presence— *P. pastoris* cultures indeed show co-consumptions as those predicted by our FBA model. The glucose–glycerol co-consumption has been previously observed in fed-batch cultures [40], and also glycerol-methanol [41,43] and glucose-methanol [44].

Note that our objective with this last validation procedure was to get predictions from the original, raw model at different substrate environments before fine-tuning the model without considering regulation or kinetics. At this point, the limits of our simple FBA model are known, we may consider adding a minimum layer of regulation to incorporate knowledge that the model is lacking. The advantage is that now this can be done with a minimal complexity approach —that is, adding as little complexity as possible in order to further increase the model accuracy—, while keeping the optimal growth hypothesis as the main driving force of our FBA model.

Conclusions

We have validated a small-sized FBA model of *P. pastoris* metabolism using experimental data from the literature. Our purpose was to test the model ability to give reasonable predictions in a wide range of experimental conditions without tuning the model, just applying an FBA hypothesis of maximal growth over a constraint-based model that accounts only for simple stoichiometric and reversibilities. We have intentionally avoided fine-tuning any parameter related to biomass composition, ATP assimilation, substrate preference, reaction kinetics, regulation phenomena, etc.

The computations along the paper show that our *P. pastoris* FBA model is able to (i) predict growth yields over single substrates; (b) predict growth, substrate uptake, respiration rates, and byproduct formation in scenarios with different substrates; (c) predict the behavior of *P. pastoris* in aerobic and hypoxic conditions over single substrates; and (d) predict the substrate preference under oxygen limitation.

In general, the results show that FBA model predictions based on growth maximization are reasonably accurate in many situations, particularly when glucose and glycerol are the carbon sources. The divergences with respect to the experimental data become larger in scenarios growing on methanol. We have already discussed how different causes could explain this. One possible explanation is that our model is not detailed enough. Another explanation is that our model, which represents wild-type strains, disregards the alterations that occur in modified organisms due to the production of recombinant protein. Finally, it could be that the hypothesis of maximizing growth is not as suitable growing on methanol growth as it is when cells uptake glucose or glycerol. Another limitation of our model occurs in scenarios of multiple substrates and no oxygen limitation, when it predicts co-consumptions that are not seen in actual cultures. Probably, the reason is that our model is lacking other constraints that operate in those situations. At this point, the model can be extended to improve its predictive capacity. First, methanol pathways can be detailed and the biomass equation could be revised in those conditions. Second, the expression of recombinant protein could be addressed to better represent modified organisms. Finally, we want to consider adding a layer of regulation into the model in order to better predict the cells behavior in scenarios where multiple carbon sources are available.

Nevertheless, even if (i) our FBA model is a small one, (ii) it has no parameter tuned, and (iii) it imposes a strong assumption regarding how cells regulate their metabolic fluxes (maximizing growth), it is able to provide reasonably good predictions regarding growth, substrate preference, product formation, and respiration rates in many heterogeneous experimental scenarios. In our opinion, these results suggest that small FBA models can be a valuable tool in scenarios of data scarcity —where measurable fluxes are scarce, models are small and general, and experimental data is not abundant—, which are common circumstances in industrial environments and pilot laboratories.

Additional files

Additional file 1: *P. pastoris* Metabolic Network, Excel file with the list of reactions, metabolites and stoichiometric matrix.

Additional file 2: Experimental datasets. Excel file with all the 72 experimental datasets taken from the literature. This file includes measurement of biomass, substrates uptakes (glycerol, glucose, and methanol), Oxygen Uptake Rate (OUR), CO_2 production (CPR), and formation of byproducts (ethanol, citrate, and pyruvate) and Consistency analysis results [45-52].

Competing interest
The authors declare that they have no competing interests.

Authors' contributions
FLL and YM designed the research and conceptualized the manuscript. YM and MT collected the experimental data; YM and FLL performed the computations; all the authors analyzed the results. JV and JP coordinated the project. YM, FLL and MT drafted the first manuscript. All authors contribute to the final manuscript. All authors read and approved the final manuscript.

Acknowledgements
This research has been partially supported by the Spanish Government (cicyt: DPI 2011-28112-C04-01, DPI 2013-46982-C2-2-R) and Biopolis S.L. (R.C.055/12). Yeimy Morales is grateful for the BR Grant of the University of Girona (BR2012/26). The authors are grateful to the company Biopolis S.L. for his support to this research.

Author details
[1]MICElab, IIIA, Universitat de Girona, Campus Montilivi, P4, Girona 17071, Spain. [2]BIÓPOLIS S.L., C/Catedrático Agustín Escardino Benlloch, 9, 46980 Paterna, Valencia, Spain. [3]Institut Universitari d Automàtica i Informàtica Industrial, Universitat Politècnica de València, Camino de Vera s/n, Edificio 5C, 46022 Valencia, Spain.

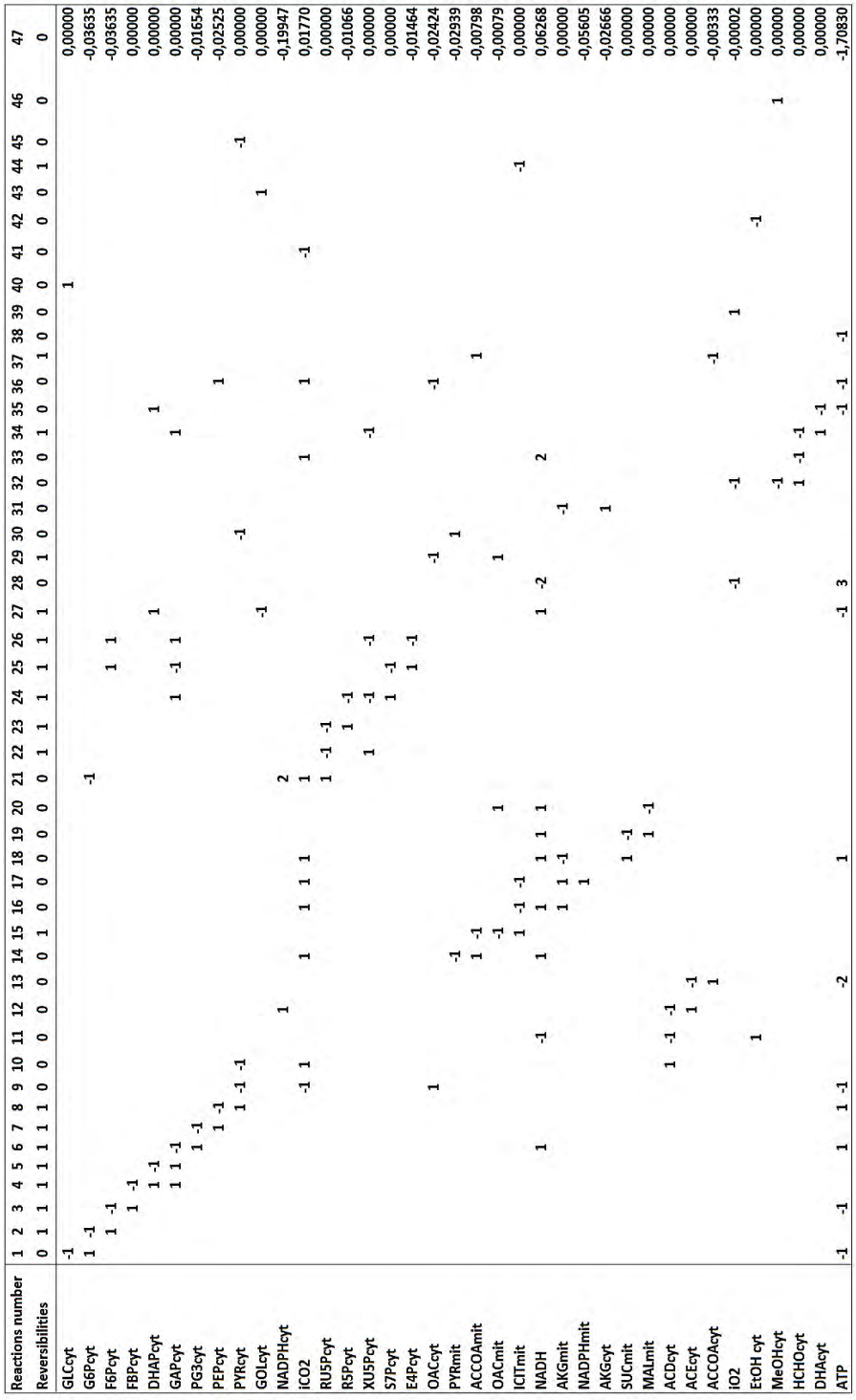
Received: 5 September 2014 Accepted: 17 December 2014
Published online: 24 December 2014

References
1. Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM: **Heterologous protein production using the *Pichia pastoris* expression system.** *Yeast* 2005, **22:**249–270.
2. Palsson BO: *System Biology, Properties of Reconstructed Networks.* New York USA: Cambridge University Press; 2006.
3. Stephanopoulos GN, Aristidou AA, Nielsen J: *Metabolic Engineering: Principles and Methodologies.* San Diego, USA: Academic Press; 1998.
4. Klamt S, Schuster S, Gilles ED: **Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria.** *Biotechnol Bioeng* 2002, **77:**734–775.
5. Llaneras F, Picó J: **Stoichiometric modelling of cell metabolism.** *J Biosci Bioeng* 2008, **105**(1):1–11.
6. Orth JD, Thiele I, Palsson BO: **What is flux balance analysis?** *Nat Biotech* 2010, **28:**245–248.
7. Van der Heijden RT, Heijnen JJ, Hellinga C, Romein B, Luyben KC: **Linear constraint relations in biochemical reaction systems.** *I Biotechnol Bioeng* 1994, **43**(1):3–10.
8. Varma A, Palsson BO: **Metabolic flux balancing: basic concepts, scientific and practical use.** *Nat Biotechnol* 1994, **12**(10):994–998.
9. Edwards JS, Covert M, Palsson B: **Metabolic modelling of microbes: the flux-balance approach.** *Environ Microbiol* 2002, **4:**133–140.
10. Schuster S, Pfeiffer T, Fell DA: **Is maximization of molar yield in metabolic networks favoured by evolution?** *J Theor Biol* 2007, **252:**497–504.
11. Ibarra RU, Edwards JS, Palsson BO: ***Escherichia coli K-12* undergoes adaptive evolution to achieve in silico predicted optimal growth.** *Nature* 2002, **420:**186–189.
12. Wintermute EH, Lieberman TD, Silver PA: **An objective function exploiting suboptimal solutions in metabolic networks.** *BMC Syst Biol* 2013, **7:**98.
13. Schuetz R, Kuepfer L, Sauer U: **Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli.** *Mol Syst Biol* 2007, **3:**1–15.
14. Ebenhöh O, Heinrich R: **Evolutionary optimization of metabolic pathways. Theoretical reconstruction of the stoichiometry of ATP and NADH producing systems.** *Bull Math Biol* 2001, **63**(1):21–55.
15. Holzhütter HG: **The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks.** *Eur J Biochem* 2004, **271:**2905–2922.
16. Bordel S: **Experimental evidence suggests the existence of evolutionary conserved global operation principles governing microbial metabolism.** *Sci Rep* 2013, **3:**1–7.
17. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U: **Multidimensional optimality of microbial metabolism.** *Science* 2012, **336:**601–604.
18. Sendín JO, Alonso AA, Banga JR: **Multi-Objective Optimization of Biological Networks for Prediction of Intracellular Fluxes.** In *2nd International Workshop on Practical Applications of Computational Biology and Bioinformatics (IWPACBB).* Berlin Heidelberg: Springer; 2008:197–205.
19. Varma A, Boesch BW, Palsson BO: **Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates.** *Appl Environ Microbiol* 1993, **59**(8):2465–2473.
20. Edwards J, Ibarra R, Palsson B: **In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data.** *Nat Biotechnol* 2001, **19**(2):125–130.
21. Jin Y, Jeffries T: **Stoichiometric network constraints on xylose metabolism by recombinant Saccharomyces cerevisiae.** *Metab Eng* 2004, **6**(3):229–238.
22. Varma A, Palsson BO: **Metabolic capabilities of Escherichia coli: I. Synthesis of biosynthetic precursors and cofactors.** *J Theor Biol* 1993, **165**(4):477–502.
23. David H, Åkesson M, Nielsen J: **Reconstruction of the central carbon metabolism of Aspergillus niger.** *Eur J Biochem* 2003, **270**(21):4243–4253.
24. Tortajada M, Llaneras F, Picó J: **Validation of a constraint-based model of *Pichia pastoris* growth under data scarcity.** *BMC Syst Biol* 2010, **4**(115):1–11.
25. Tortajada M, Llaneras F, Ramón D, Picó J: **Estimation of recombinant protein production in *Pichia pastoris* based on a constraint-based model.** *J Process Control* 2012, **22**(6):1139–1151.
26. Llaneras F: **Interval and Possibilistic Methods for Constraint-Based Metabolic Models.** In *PhD Thesis.* Universidad Politécnica de Valencia: Departamento de Ingeniería de Sistemas y Automática; 2011.
27. Lofberg J: **YALMIP. A Toolbox for Modeling and Optimization in MATLAB, In Flux-Balance Approach.** In *IEEE International Symposium on Computer Aided Control Systems Design.* Edited by IEEE. Taipei: IEEE; 2004:284–289.
28. Dragosits M, Stadlmann J, Albiol J, Baumann K, Maurer M, Gasser B, Sauer M, Altmann F, Ferrer P, Mattanovich D: **The effect of temperature on the proteome of recombinant *Pichia pastoris*.** *J Proteome Res* 2009, **8**(3):1380–1392.
29. Caspeta L, Shoaie S, Agren R, Nookaew I, Nielsen J: **Genome-scale metabolic reconstructions of Pichia stipitis and *Pichia pastoris* and in silico evaluation of their potentials.** *BMC Syst Biol* 2012, **6**(1):24.
30. Inan M, Meagher MM: **Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of *Pichia pastoris*.** *J Biosci Bioeng* 2001, **92**(6):585–589.
31. Zhang W, Bevins MA, Plantz BA, Smith LA, Meagher MM: **Modeling *Pichia pastoris* growth on methanol and optimizing the production of a recombinant protein, the heavy-chain fragment C of botulinum neurotoxin, serotype A.** *Biotechnol Bioeng* 2000, **70:**1–8.
32. Kim S, Warburton S, Boldogh I, Svensson C, Pon L, d'Anjou M, Choi BK: **Regulation of alcohol oxidase 1 (AOX1) promoter and peroxisome biogenesis in different fermentation processes in *Pichia pastoris*.** *J Biotechnol* 2013, **166**(4):174–181.
33. Santos S: **Análisis cuantitativo y modelización del metabolismo de la levadura *Pichia pastoris*.** In *PhD Thesis.* Universitat Autónoma de Barcelona: Departamento de Ingeniería Quimica; 2008.
34. Heyland J, Fu J, Blank LM, Schmid A: **Carbon metabolism limits recombinant protein production in *Pichia pastoris*.** *Biotechnol Bioeng* 2011, **108**(8):1942–1953.
35. Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D: **Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins.** *Biotechnol Bioeng* 2008, **100**(1):177–183.
36. Charoenrat T, Ketudat-Cairns M, Stendahl-Andersen H, Jahic M, Enfors SO: **Oxygen-limited fed-batch process: an alternative control for *Pichia pastoris* recombinant protein processes.** *Bioprocess Biosyst Eng* 2005, **27**(6):399–406.

37. Carnicer M, Baumann K, Töplitz I, Sanchez F, Mattanovich D, Ferrer P, Albiol J: **Macromolecular and elemental composition analysis and extracellular metabolite balances of Pichia pastoris growing at different oxygen levels.** *Microb Cell Fact* 2009, **8:**65.
38. Cereghino GP, Cereghino JL, Ilgen C, Cregg JM: **Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*.** *Curr Opin Biotechnol* 2002, **13**(4):329–332.
39. Baumann K, Carnicer M, Dragosits M, Graf A, Stadlmann J, Jouhten P, Maaheimo H, Gasser B, Albiol J, Mattanovich D, Ferrer P: **A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions.** *BMC Syst Biol* 2010, **4:**141.
40. Garcia X, Ferrer P, Montesinos J, Valero F: **Fed-batch operational strategies for recombinant Fab production with *Pichia pastoris* using the constitutive GAP promoter.** *Biochem Eng J* 2013, **79:**172–181.
41. Inan M, Meagher MM: **The effect of ethanol and acetate on protein expression in *Pichia pastoris*.** *J Biosci Bioeng* 2001, **92**(4):337–341.
42. Chiruvolu V, Eskridge KM, Cregg JM, Meagher MM: **Effects of glycerol concentration and pH on growth of recombinant Pichia pastoris.** *Appl Biochem Biotechnol* 1998, **75:**163–173.
43. Solà A, Jouhten 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.** *Microbiology* 2007, **153:**281–290.
44. Jordà J, Suarez C, Carnicer M, Pierick A, Heijnen J, Van Gulik W, Ferrer P, Albiol J, Aljoscha W: **Glucose-methanol co-utilization in Pichia pastoris studied by metabolomics and instationary 13C flux analysis.** *BMC Sys Biol* 2013, **7**(17):2–16.
45. Solà A: **Estudi del metabolisme central del carboni de *Pichia pastoris*.** In *PhD Thesis.* Universitat Autònoma de Barceloana: Escola Tècnica Superior d'Enginyeria; 2004.
46. Jungo C, Marison I, Stockar U: **Mixed feeds of glycerol and methanol can improve the performance of Pichia pastoris cultures: a quantitative study based on concentration gradients in transient continuous cultures.** *J Biotechnol* 2007, **128:**824–837.
47. d'Anjou MC, Daugulis AJ: **A rational approach to improving productivity in recombinant Pichia pastoris fermentation.** *Biotechnol Bioeng* 2001, **72:**1–11.
48. Zhang WH, Liu CP, Inan M, Meagher MM: **Optimization of cell density and dilution rate in *Pichia pastoris* continuous fermentations for production of recombinant proteins.** *J Ind Microbiol Biotechnol* 2004, **31:**330–334.
49. 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 13C-labelling data from proteinogenic amino acids.** *New Biotecnol* 2014, **31**(1):120–132.
50. Chung B, Selvarasu S, Camattari A, Ryu J, Lee H, Ahn J, Lee D: **Research Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement.** *Microb Cell Fact* 2010, **9:**50.
51. 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:**936–944.
52. Tortajada M: **Process Development for the Obtention and use of Recombinant Glycosidases: Expression, Modelling and Immobilization.** In *PhD Thesis.* Universidad Politécnica de Valencia: Departamento de Ingeniería de Sistemas y Automática; 2012.

SUPLEMENTARY MATERIAL

Additional file 1: *P. pastoris* Metabolic Network.



Reactions number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
Reversibilities	0	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0
GLCcyt	-1																																							1								0,00000
G6Pcyt	1	-1																			-1																										-0,03635
F6Pcyt		1	-1																						1	1																				-0,03635
FBPcyt			1	-1																							1																				0,00000
DHAPcyt				1	-1																														1												0,00000
GAPcyt				1	1	-1																		1	-1	1								1												0,00000
PG3cyt						1	-1																	1																							-0,01654
PEPcyt							1	-1																												1											-0,02525
PYRcyt								1	-1	-1																				-1													-1			0,00000
GOLcyt																					-1																									0,00000
NADPHcyt												1																													1						-0,19947
iCO2										-1	1			1		1	1	1			2											1				1			-1							0,01770
RU5Pcyt																					1	-1	-1					-1																		0,00000
R5Pcyt																						1	-1	-1					1																		-0,01066
XU5Pcyt																						1		-1		-1							-1													0,00000
S7Pcyt																								1	-1																					0,00000
E4Pcyt																									1	-1																				-0,01464
OACcyt									1																			-1								-1										-0,02424
PYRmit															-1														1																		-0,02939
ACCOAmit															1	-1																					1										-0,00798
OACmit																-1				1								1																		-0,00079
ICITmit															1	-1	-1																														0,00000
NADH							1				-1				1		1	1	1								1	-2				2											-1				0,06268
AKGmit																1	1	-1												-1																0,00000
NADPHmit																	1														1																-0,05605
AKGcyt																		1	-1																												-0,02666
SUCmit																			1	-1																											0,00000
MALmit																				1	-1																											0,00000
ACDcyt										1	-1	-1																																			0,00000
ACEcyt												1	-1																																		0,00000
ACCOAcyt													1																																			-0,00333
iO2																												-1				-1					-1		1								-0,00002
EtOH cyt											1																																-1					0,00000
MeOHcyt																																-1														1		0,00000
HCHOcyt																																1	-1	-1														0,00000
DHAcyt																																		1	-1													0,00000
ATP	-1		-1			1		1	-1				-2				1										-1	3							-1	-1		-1									-1,70830

Additional file 2: experimental datasets.

	μ	QGlu	QGlyc	QMet	QPyr	QCit	QEtOH	OUR	CPR	Qp	Source
	Cmmol/g/h	mmol/g/h	mmol/g/h	mmol/g/h	mmol/g/h	mmol/g/h	mmol/g/h	mmol/g/h	mmol/g/h	mg/g/h	
1	3.72	0.96	0.00	0.00	0.00	0.01	0.016	2.002	2.088	0.020	Dragosits 2008
2	3.72	0.94	0.00	0.00	0.00	0.002	0.035	1.776	1.868	0.040	Dragosits 2008
3	3.74	0.92	0.00	0.00	0.00	0.00	0.00	1.693	1.746	0.060	Dragosits 2008
4	3.01	0.99	0.00	0.00	0.00	0.03	0.00	2.120	2.370	0.055	Dragosits 2008
5	3.71	1.33	0.00	0.00	0.06	0.02	0.33	1.570	2.030	0.056	Dragosits 2008
6	3.73	1.74	0.00	0.00	0.09	0.04	1.00	0.540	1.650	0.089	Dragosits 2008
7	1.88	0.00	1.09	0.00	0.00	0.00	0.00	2.16	1.56	0.000	Sola 2007
8	2.07	0.00	0.95	0.63	0.00	0.00	0.00	2.70	1.70	0.001	Sola 2007
9	1.72	0.00	0.74	1.48	0.00	0.00	0.00	3.90	2.10	0.014	Sola 2007
10	2.02	0.00	0.57	2.33	0.00	0.00	0.00	4.85	2.21	0.024	Sola 2007
11	6.17	0.00	2.75	0.00	0.00	0.00	0.00	3.62	2.35	0.000	Sola 2007
12	6.18	0.00	2.77	1.87	0.00	0.00	0.00	7.19	4.18	0.001	Sola 2007
13	6.24	0.00	2.23	2.73	0.00	0.00	0.00	7.20	3.60	0.012	Sola 2007
14	5.74	1.51	0.00	0.00	0.00	0.00	0.00	2.71	3.18	0.000	Sola 2004
15	1.60	0.00	0.00	6.31	0.00	0.00	0.00	7.56	3.44	0.613	Sola 2004
16	3.27	0.81	0.00	1.09	0.00	0.00	0.00	4.02	2.68	0.000	Sola 2004
17	3.25	0.78	0.00	1.00	0.00	0.00	0.00	3.65	2.45		Sola 2004
18	3.33	0.89	0.00	1.32	0.00	0.00	0.00	4.72	3.35		Sola 2004
19	2.32	0.00	0.67	2.01	0.00	0.00	0.00	3.21	1.77	0.012	Jungo 2007
20	2.32	0.00	0.51	2.49	0.00	0.00	0.00	3.46	1.89	0.020	Jungo 2007
21	2.32	0.00	0.43	2.73	0.00	0.00	0.00	3.58	1.97	0.019	Jungo 2007
22	2.32	0.00	0.31	3.09	0.00	0.00	0.00	3.76	2.09	0.021	Jungo 2007
23	2.32	0.00	0.28	3.18	0.00	0.00	0.00	3.79	2.09	0.021	Jungo 2007
24	2.32	0.00	0.18	3.49	0.00	0.00	0.00	3.96	2.17	0.021	Jungo 2007
25	2.32	0.00	0.13	3.62	0.00	0.00	0.00	4.02	2.21	0.022	Jungo 2007
26	2.32	0.00	0.11	3.69	0.00	0.00	0.00	4.06	2.25	0.020	Jungo 2007
27	2.32	0.00	0.09	3.74	0.00	0.00	0.00	4.08	2.25	0.021	Jungo 2007
28	2.32	0.00	0.00	4.02	0.00	0.00	0.00	4.22	2.33	0.022	Jungo 2007
29	0.39	0.00	0.27	0.38	0.00	0.00				0.015	d'Anjou 2001
30	0.77	0.00	0.54	0.50	0.00	0.00				0.020	d'Anjou 2001
31	1.16	0.00	0.82	0.63	0.00	0.00				0.040	d'Anjou 2001
32	1.93	0.00	1.09	0.66	0.00	0.00				0.025	d'Anjou 2001
33	2.71	0.00	1.36	0.94	0.00	0.00				0.050	d'Anjou 2001
34	3.09	0.00	1.90	0.55	0.00	0.00				0.060	d'Anjou 2001
35	3.48	0.00	2.45	0.44	0.00	0.00				0.065	d'Anjou 2001
36	0.31	0.00	0.00	0.99	0.00	0.001				0.006	Zhang,W2004
37	1.39	0.00	0.00	4.66	0.00	0.003				0.037	Zhang,W2004

38	1.62	0.00	0.00	5.64	0.00	0.005				0.036	Zhang,W2004
39	1.04	0.00	0.00	3.02	0.00	0.001				0.019	Zhang,W2004
40	1.20	0.00	0.00	3.39	0.00	0.003				0.030	Zhang,W2004
41	1.93	0.00	0.00	4.34	0.00	0.004				0.021	Zhang,W2004
42	1.31	0.00	0.00	3.62	0.00	0.006				0.032	Zhang,W2004
43	1.66	0.00	0.00	4.02	0.00	0.006				0.025	Zhang,W2004
44	0.54	0.00	0.00	1.73	0.00	0.001				0.025	Zhang,W2004
45	0.66	0.00	0.00	1.67	0.00	0.003				0.012	Zhang,W2004
46	4.54	0.00	2.53	0.18	0.00	0.00	0.00	4.78	3.25	0.000	Jorda J 2013
47	5.63	0.00	2.61	1.76	0.00	0.00	0.00	5.35	3.09	0.001	Jorda, J 2013
48	5.44	0.00	2.22	2.58	0.00	0.00	0.00	5.73	3.33	0.012	Jorda, J 2013
49	1.19	0.42	0.00	0.00				0.86	0.81		Chung 2010
50	2.77	0.72	0.00	0.00				1.23	1.29		Chung 2010
51	3.91	1	0.00	0.00				1.49	1.72		Chung 2010
52	5.57	1.39	0.00	0.00				1.85	2.24		Chung 2010
53	1.97	0.00	0.00	4.06							Caspeta 2012
54	2.96	0.00	0.00	5.93							Caspeta 2012
55	3.54	0.00	0.00	7.18							Caspeta 2012
56	1.22	0.00	0.00	2.72				2.931	1.57	0.016	Jungo, C 2006
57	2.12	0.00	0.00	4.44				4.697	2.48	0.020	Jungo, C 2006
58	2.31	0.00	0.00	4.79				5.052	2.72	0.022	Jungo, C 2006
59	2.34	0.00	0.00	4.85				5.083	2.68	0.020	Jungo, C 2006
60	3.53	0.00	0.00	7.12				7.218	3.76	0.024	Jungo, C 2006
61	4.47	0.00	0.00	8.90				8.670	4.46	0.028	Jungo, C 2006
62	0.90	0.00	0.52	0.00				0.82	0.62	0.00	Jungo, C 2006
63	1.11	0.00	0.62	0.00				0.87	0.65	0.00	Jungo, C 2006
64	2.38	0.00	1.21	0.00				1.65	1.22	0.00	Jungo, C 2006
65	4.89	0.00	2.40	0.00				3.12	2.29	0.00	Jungo, C 2006
66	8.37	0.00	4.04	0.00				4.77	3.40	0.00	Jungo, C2006
67	0.90	0.00	0.00	1.89				1.26	0.99	0.00	Tortajada 2012
68	1.40	0.00	0.00	2.55				2.16	1.15	0.00	Tortajada 2012
69	0.94	0.00	0.00	1.87				1.67	0.93	0.00	Tortajada 2012
70	8.07	0.00	3.68	0.00				3.99	2.96	0.00	Tortajada 2012
71	8.79	0.00	4.50	0.00				4.71	4.70	0.00	Tortajada 2012
72	8.66	0.00	4.06	0.00				4.19	3.53	0.00	Tortajada 2012

Carbon balance and possibilistic evaluation results.

Code	substrate	Error C- balance	Π*	E**
1	Glu	1%	1.00	1%
2	Glu	1%	1.00	2%
3	Glu	1%	1.00	5%
4	Glu	9%	0.75	6%
5	Glu	19%	0.13	12%
6	Glu	28%	0.00	38%
7	Gly	5%	1.00	3%
8	Gly-MeoH	8%	0.74	6%
9	Gly-MeoH	4%	0.25	10%
10	Gly-MeoH	5%	0.08	14%
11[a]	Gly	3%	1.00	2%
12	Gly-MeoH	2%	0.82	6%
13	Gly-MeoH	5%	0.32	9%
14	Glu	2%	0.65	7%
15	MeoH	10%	0.05	12%
16[a]	Glu-MeoH	0%	0.16	12%
17	Glu-MeoH	0%	0.18	11%
18	Glu-MeoH	0%	0.36	9%
19	Gly-MeoH	2%	1.00	2%
20	Gly-MeoH	5%	1.00	3%
21	Gly-MeoH	7%	1.00	3%
22	Gly-MeoH	10%	1.00	5%
23	Gly-MeoH	10%	1.00	5%
24	Gly-MeoH	12%	0.91	5%
25	Gly-MeoH	14%	0.71	6%
26	Gly-MeoH	14%	0.64	6%
27	Gly-MeoH	14%	0.61	6%
28	MeoH	16%	0.50	7%
29[b]	Gly-MeoH	n.a	1.00	1%
30[b]	Gly-MeoH	n.a	1.00	1%
31[b]	Gly-MeoH	n.a	1.00	1%
32[b]	Gly-MeoH	n.a	1.00	1%
33[b]	Gly-MeoH	n.a	1.00	1%
34[b]	Gly-MeoH	n.a	1.00	1%
35[b]	Gly-MeoH	n.a	1.00	1%
36[b]	MeoH	n.a	1.00	1%
37[b]	MeoH	n.a	1.00	1%

38[b]	MeoH	n.a	1.00	1%
39[b]	MeoH	n.a	1.00	1%
40[b]	MeoH	n.a	1.00	1%
41[b]	MeoH	n.a	1.00	1%
42[b]	MeoH	n.a	1	1%
43[b]	MeoH	n.a	1	1%
44[b]	MeoH	n.a	1.00	1%
45[b]	MeoH	n.a	1.00	1%
46[a]	Gly-MeoH	0%	1.00	3%
47	Gly-MeoH	9%	1.00	5%
48	Gly-MeoH	5%	1.00	3%
49	Glu	31%	1.00	5%
50	Glu	11%	0.07	15%
51	Glu	11%	0.02	20%
52	Glu	12%	0.02	21%
53[b]	MeoH	51%	1.00	1%
54[b]	MeoH	50%	1.00	1%
55[b]	MeoH	51%	1.00	1%
56	MeoH	3%	1.00	1%
57	MeoH	4%	1.00	1%
58	MeoH	5%	1.00	1%
59	MeoH	4%	1.00	1%
60	MeoH	3%	1.00	1%
61	MeoH	1%	1.00	1%
62[a]	Gly	2%	1.00	1%
63[a]	Gly	5%	1.00	1%
64[a]	Gly	1%	1.00	1%
65[a]	Gly	0%	1.00	1%
66[a]	Gly	3%	1.00	1%
67[a]	MeoH	0%	1.00	2%
68[a]	MeoH	0%	1.00	1%
69[a]	MeoH	0%	1.00	1%
70[a]	Gly	0%	1.00	1%
71[a]	Gly	0%	1.00	1%
72[a]	Gly	0%	1.00	2%

*Possibility of the most possible flux distribution.

** Degree of measurements uncertainty necessary to find a fully possible flux distribution (π=1).

(a) Dataset that does not produce heterologous protein.

(b) Dataset without CO2 and O2 measurements.

ARTICLE II

AN FBA MODEL OF *PICHIA PASTORIS* TO PREDICT PROTEIN PRODUCTION AND GROWTH OVER METHANOL AND GLYCEROL, BASED ON ENERGETIC/ATP ALLOCATION

Submitted to *Biotechnology and Bioengineering. (JCR quartile:* Q1; impact factor: in 2014; 24 among 163 in Biotechnology & applied microbiology).

Morales, Y., Tortajada, M., Picó, J., Vehí, J., Llaneras, F. “An FBA model of Pichia pastoris to predict protein production and growth over methanol and glycerol, based on energetic/ATP allocation”. Manuscript submitted for publication

Abstract

Maximizing recombinant protein production is a key feature of Pichia pastoris cultures. In this work, we present a Flux Balance Analysis (FBA) model with growth maximization assumption, able to predict recombinant protein production simultaneously with growth rate in chemostat cultures of genetically modified organisms (GMOs) of P. pastoris. The model has been developed for pure methanol and mixed feeds (methanol-glycerol) as substrates. In a previous work, we predicted the behavior of wild type P. pastoris strains with a small constraint-based metabolic model and FBA. Herein, we add a mass balance to this small model to represent ATP consumption, and a hypothesis about how modified cells distribute their ATP resources –which are used only for growth in wild type cells- for growth and recombinant protein synthesis. The predictions were validated with several experimental scenarios from literature. The results show the model is able to predict with reasonable accuracy protein production and growth rate. Some features of the model are the following: i) is remarkably accurate despite cells were expressing different heterologous proteins in a variety of operating conditions; ii) requires little information to perform its predictions – just the availability of substrates in the environment–, and iii) is a small and simple model with only a few parameters that have been tuned from experimental data. The approach followed in this work is of interest in industrial environments and pilot laboratories where experimental data are not abundant, as it provides valuable predictions using few data and a small and simple metabolic model.

Keywords

Constraint-based model, flux balance analysis, protein expression.

AN FBA MODEL OF *PICHIA PASTORIS* TO PREDICT PROTEIN PRODUCTION AND GROWTH OVER METHANOL AND GLYCEROL, BASED ON ATP ALLOCATION

Yeimy Morales[1§], Marta Tortajada[2], Jesús Picó[3], Josep Vehí, Francisco Llaneras[1]

[1]MICElab, IIIA, Universitat de Girona, Campus Montilivi, P4, Girona, 17071, Spain

[2]BIÓPOLIS S.L., Parc Científic Universitat de València, C/ Catedrático Agustín Escardino Benlloch, 9, 46980, Paterna, Valencia, Spain.

[3] Institut Universitari d'Automàtica i Informàtica Industrial, Universitat Politècnica de València, Camino de Vera s/n, Edificio 5C, 46022 Valencia, Spain

[§]Corresponding author

Email addresses:

YM: yeimy.morales@outlook.com

MT: marta.tortajada@biopolis.es

JP: jpico@ai2.upv.es

FLL: francisco.llaneras@udg.edu

JV: josep.vehi@udg.edu

Highlights

- We present an FBA model of *Pichia pastoris* able to predict recombinant protein production.
- We validate the model with several experimental data sets.
- We introduce a simple hypothesis to predict how protein production impacts growth of mutant strains.

ABSTRACT

Maximizing recombinant protein production is a key feature of *Pichia pastoris* cultures. In this work, we present a Flux Balance Analysis (FBA) model with growth maximization assumption, able to predict recombinant protein production simultaneously with growth rate in chemostat cultures of genetically modified organisms (GMOs) of *P. pastoris*. The model has been developed for pure methanol and mixed feeds (methanol-glycerol) as

ARTICLE III

PFA TOOLBOX: A MATLAB TOOL FOR METABOLIC FLUX ANALYSIS

Accepted in *BMC Systems Biology*. (JCR quartile: Q1; Impact factor: 2.435 in 2014; 13 among 57 journals in Mathematical & computational Biology).

PFA TOOLBOX: A MATLAB TOOL FOR METABOLIC FLUX ANALYSIS

Yeimy Morales[1§], Gabriel Bosque[2], Josep Vehí[1], Jesús Picó[2], Francisco Llaneras[1]

[1] MICElab, IIIA, Universitat de Girona, Campus Montilivi, P4, Girona, 17071, Spain.

[2] Institut Universitari d'Automàtica i Informàtica Industrial, Universitat Politècnica de València, Camino de Vera s/n, Edificio 5C, 46022 Valencia, Spain.

[§] Corresponding author.

Email addresses:

YM: yeimy.morales@udg.edu

GB: gabbosch@upv.es

JV: josep.vehi@udg.edu

JP: jpico@ai2.upv.es;

FLL: francisco.llaneras@udg.edu

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 13C-MFA which uses stable isotopically labeled substrates (e.g., 13C-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-13C-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-suited 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 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 **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, $\mathbf{N} \in \mathbf{R}^{\{nxr\}}$ and $\mathbf{D} \in \mathbf{R}^{\{rxr\}}$ is a diagonal matrix with 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 **v** that are solution of (1) define a flux distribution.

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

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

where $\mathbf{v}_{\boldsymbol{m}}^{\boldsymbol{m}}$ and $\mathbf{v}_{\boldsymbol{m}}^{\boldsymbol{M}}$ are vectors with the minimum and maximum possible values that the measured fluxes $\mathbf{v}_{\boldsymbol{m}}$ 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 **v** can be obtained solving two Linear Programming (LP) problems, as follows:

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

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:

$$MEC = \begin{cases} \boldsymbol{w_m} = \mathbf{v_m} + \boldsymbol{\varepsilon_1} - \boldsymbol{\mu_1} + \boldsymbol{\varepsilon_2} - \boldsymbol{\mu_2} \\ \boldsymbol{\varepsilon_1}, \boldsymbol{\mu_1} \geq 0 \\ 0 \leq \boldsymbol{\varepsilon_2} \leq \boldsymbol{\varepsilon_2^{max}} \\ 0 \leq \boldsymbol{\mu_2} \leq \boldsymbol{\mu_2^{max}} \end{cases} \quad (4)$$

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 π 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:

$$J = \boldsymbol{\alpha} \cdot \boldsymbol{\varepsilon}_1 + \boldsymbol{\beta} \cdot \boldsymbol{\mu}_1 \quad (5)$$

The possibility of each solution is defined as:

$$\pi(\delta) = \exp\big(-J(\delta)\big) \;\; \delta \, \epsilon \, MEC \;\cap MOC \quad (6)$$

Where **α** and **β** 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», **α** and **β** 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:

$$J_{\min} = \min_{\boldsymbol{\varepsilon},\boldsymbol{\mu},\boldsymbol{v}} J = \boldsymbol{\alpha} \cdot \boldsymbol{\varepsilon}_1 + \boldsymbol{\beta} \cdot \boldsymbol{\mu}_1 \; s.t \; \{MOC \cap MEC\} \quad (7)$$

The solution flux vector **v**, that we call $\mathbf{v}_{mp}$, 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 $\left[\mathrm{v}_\gamma^{\mathrm{m}}, \mathrm{v}_\gamma^{\mathrm{M}}\right]$ for a flux v, with a conditional possibility higher than γ, can be computed solving two extra LP's:

$$\mathbf{v}_\gamma^m = \min_{\boldsymbol{\varepsilon},\boldsymbol{\mu},\mathbf{v}} \mathrm{v} \; s.t \begin{cases} MOC \cap MEC \\ J - J_{\min} < -\ln \gamma \end{cases} \quad (8)$$

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.

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

Table 1. List of functions in the PFA Toolbox.

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

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.

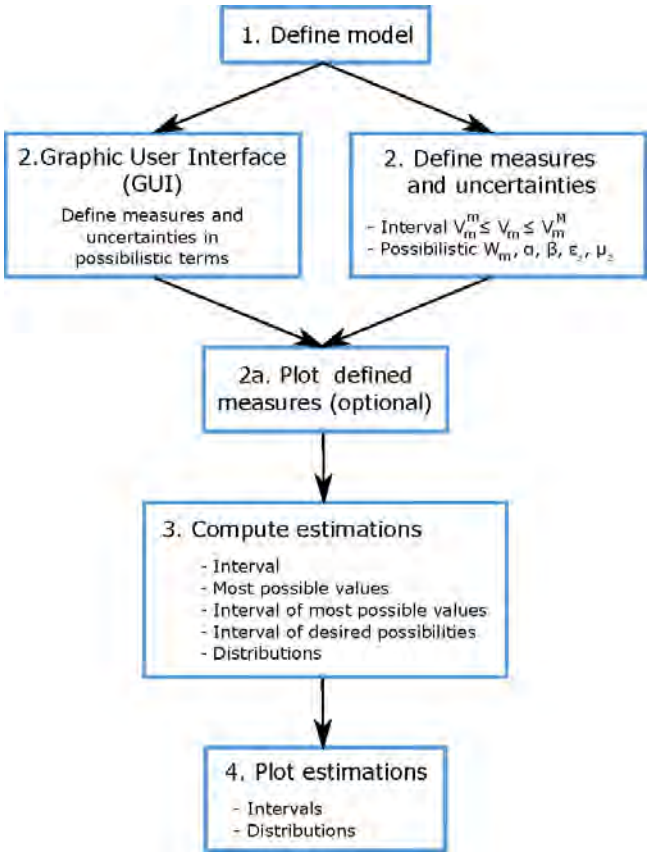


Figure 1. Protocol to use the PFA toolbox. A step by step to use the PFA toolbox. Protocol is the same to solve the MFA problems with Interval and possibilistic MFA. Possibilistic has two additional steps, which are optional, a User Guide Interface (GUI) to represent graphically the measures in possibilistic terms and a function to check if the measures and their uncertainties are well-defined.

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.

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 Figure 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.

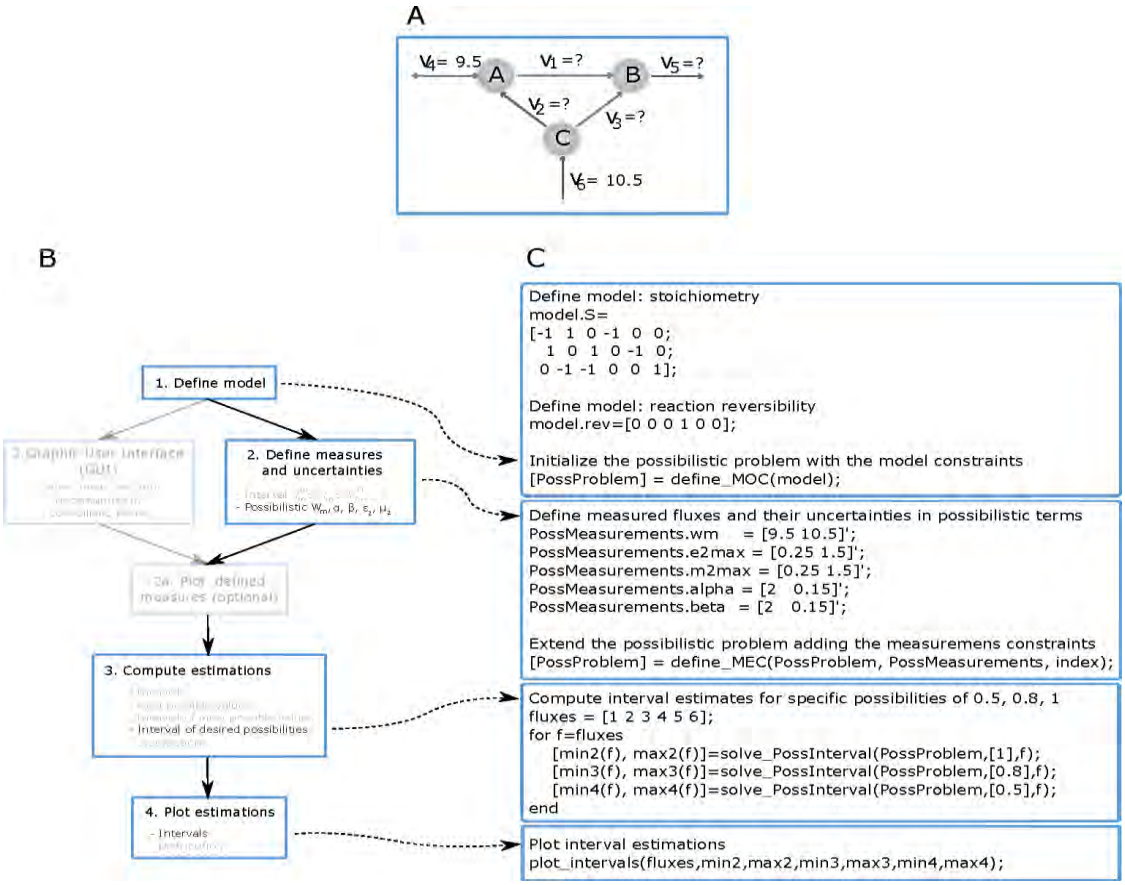


Figure 2. PFA toolbox methodology to solve example of flux estimation under data scarcity. (A) Upper panel present a simple metabolic network. Metabolites are in capital letters, each Vj represent a flux and the double arrows indicate a reversible reaction. (B) The step-by-step procedure follows to solve the MFA problem where only two measures are known. (C) Right panel shows a piece of MATLAB code used to perform the computations.

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 μ_2 and ε_2, and the weights α and β (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 Figure 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 fuction *plot_distribution*. As an example, Figure 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 supplementary material (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.

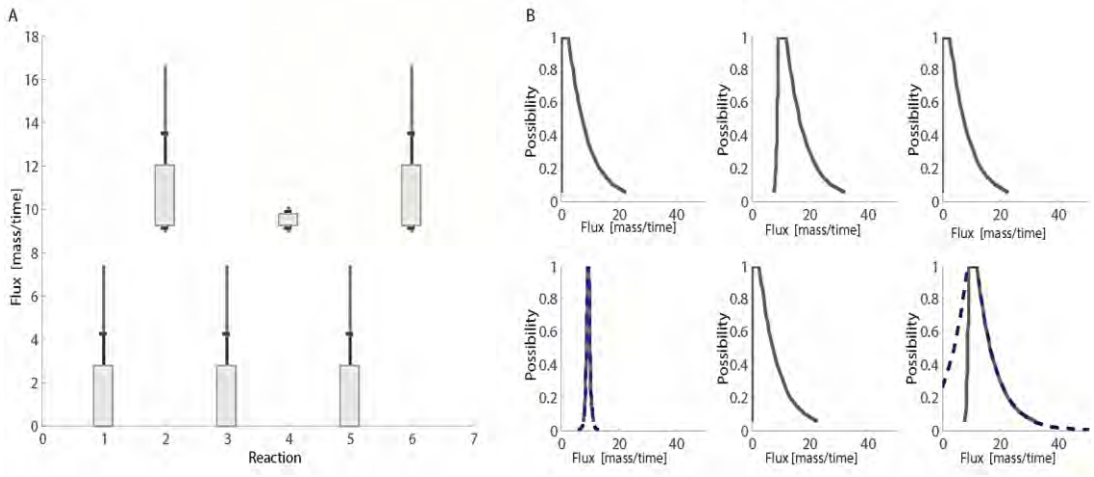


Figure 3. Flux estimation. Estimations for every flux were obtained with the PFA Toolbox. (A) Three interval estimates are given, for maximum conditional possibility (box), possibility of 0.8 (black line), and 0.5 (gray line). (B) Possibility distributions are depicted with solid lines and dashed lines represent measured values.

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 File 2 in Supplementary Material). 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 Supplementary Material (File 3).

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 Figure 4A.

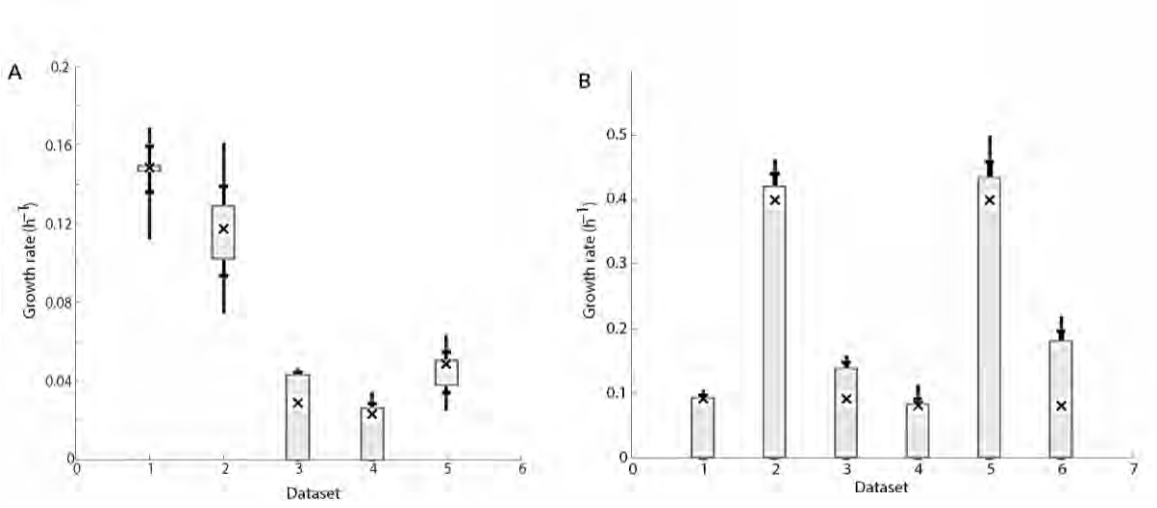


Figure 4. Growth estimations with possibilistic MFA for P. pastoris and E. coli. (A) Example with six *P. pastoris* experiments. **(B)** Example with *E. coli* experiments. In both cases, three interval estimates are represented, for conditional possibilities equal to 0.99 (box), 0.5 (bar) and 0.1 (line). The crosses represent the actual experimental values.

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 Supplementary Material (File 1b).

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

Here we use a well-known model of *E. coli*, taken from [25] and illustrated in the Supplementary Material (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_2 production and acetate and pyruvate secretion). The model and the measurements can be found in the Supplementary Material (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 Supplementary Material (File 1c).

The flux estimates computed with the toolbox are compatible with the actual growth rate in all scenarios (Figure 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_maxPoss* function. This provides the maximum possibility flux vector and the associated degree of possibility (π_{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).

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 Supplementary Material (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.

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

Reference	μ	Q_{Glu}	Q_{Glyc}	Q_{Met}	Q_{Pyr}	Q_{Cit}	Q_{EtOH}	OUR	CPR	π_{mp}**
	Cmmol* /g/h	mmol /g/h	mmol /g/h	mmol /g/h	mmol /g/h	mmol /g/h	mmol /g/h	mmol /g/h	mmol /g/h	
[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**
[34]	2.38	0.00	1.21	0.00	N.A.	N.A.	N.A.	1.65	1.22	**1.00**
[34]	4.89	0.00	2.40	0.00	N.A.	N.A	N.A.	3.12	2.29	**1.00**
[35]	1.40	0.00	0.00	2.55	N.A.	N.A.	N.A.	2.16	1.15	**1.00**
[35]	0.94	0.00	0.00	1.87	N.A.	N.A.	N.A.	1.67	0.93	**1.00**

*Cmmol= Carbon mmol **Dimensionless value of the possibility of the most possible flux distribution.

All the computations of these four examples were performed with the PFA Toolbox. The computations take approximately 13 seconds 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 minutes in an AMD A10-5800K 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.

DECLARATIONS

List of 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

MFA: Metabolic Flux Analysis

MOC: model-based constraints

MEC: Measurement constraints

PFA: Possibilistic Flux Analysis

YALMIP: Modelling language for advanced modeling and solution of optimization problems

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Availability of data and materials:

Supplementary Material 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 (File 1a), *P. pastoris* (File 1b) and *E. coli* (File 1c).

Supplementary Material 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.

Supplementary Material 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), CO2 production (CPR), and formation of byproducts (ethanol, citrate, and pyruvate).

Supplementary Material File 4 – Metabolic network of *Escherichia Coli.* Metabolic network for the *Escherichia coli* model.

Supplementary Material 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.

Competing interests

The authors declare that they have no competing interests.

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).

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.

Acknowledgements

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

References

1. Sauer U, Hatzimanikatis V, Bailey J, Hochuli M, Szyperski T, Wuethrich K. Metabolic fluxes in riboflavin-producing Bacillus subtilis. Nature biotechnology. 1997;15(5):448-452.
2. Wittmann C. Metabolic flux analysis using mass spectrometry. In: Tools and Applications of Biochemical Engineering Science. Berlin: Springer; 2002;39-64.
3. Antoniewicz M. Methods and advances in metabolic flux analysis: a mini-review. J Ind Microbiol Biot. 2015;42(3):317-325.
4. Araúzo-Bravo MR, Shimizu JK. An improved method for statistical analysis of metabolic flux analysis using isotopomer-mapping matrices with analytical expressions. J Biotech. 2003;05:117-133.
5. 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):734-751.
6. Llaneras F. Interval and possibilistic methods for constraint-based metabolic models. *PhD Thesis.* Universidad Politécnica de Valencia, Departamento de Ingeniería de Sistemas y Automática; 2011.
7. 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.
8. Llaneras F, Sala A, Picó J. A possibilistic framework for constraint-based metabolic flux analysis. BMC Syst Biol. 2009;3(1):79.

9. 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.
10. 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.
11. Iyer, V. V., Ovacik, M. A., Androulakis, I. P., Roth, C. M., & Ierapetritou, M. G. Transcriptional and metabolic flux profiling of triadimefon effects on cultured hepatocytes. Toxicology and applied pharmacology, 2010;248(3):165-177.
12. Zamorano F, Wouwer A, Bastin, G. Detailed metabolic flux analysis of an underdetermined network of CHO cells. J Biotechnol. 2010;150(4):497-508.
13. Iyer V, Yang H, Ierapetritou M, Roth, C. Effects of glucose and insulin on HepG2-C3A cell metabolism. Biotechnol Bioeng. 2010;107(2):347-356.
14. 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. 2010:118-123.
15. Orman, M. A., Arai, K., Yarmush, M. L., Androulakis, I. P., Berthiaume, F., & Ierapetritou, M. G. Metabolic flux determination in perfused livers by mass balance analysis: effect of fasting. Biotechnology and bioengineering. 2010;107(5):825-835.
16. 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.
17. 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.
18. 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.
19. Stephanopoulos GN, Aristidou AA, Nielsen J: Metabolic Engineering: Principles and Methodologies. San Diego, USA: Academic Press; 1998.
20. 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.
21. Lofberg J: YALMIP: A toolbox for modeling and optimization in MATLAB. IEEE International Symposium on Computer Aided Control Systems Design .2004:284-289.
22. YALMIP Home Page [http://users.isy.liu.se/johanl/yalmip/]. Accessed 11 May 2016.

23. IBM ILOG CPLEX- High-performance mathematical programming engine. [http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/]. Accessed 11 May 2016.
24. GLPK (GNU Linear programming kit) [http://www.gnu.org/software/glpk/]. Accessed 11 May 2016.
25. 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).
26. 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):152-164.
27. 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.
28. Bonarius H, Schmid G, Tramper J. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. Trends in Biotechnology. 1997;15(8):308–314.
29. Palsson B Ø. Systems biology: properties of reconstructed networks. New York: Cambridge University Press; 2006.
30. 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.
31. Solà A, Jouhten 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;153:281-290.
32. Solà A: Estudi del metabolisme central del carboni de *Pichia pastoris. PhD Thesis.* Universitat Autònoma de Barceloana, Escola Tècnica Superior d'Enginyeria. 2004.
33. 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 13C-labelling data from proteinogenic amino acids. New Biotecnol. 2014;31(1):120–132.
34. 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;39:936-944.

35. 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.

SUPPLEMENTARY MATERIAL

Supplementary Material 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 (File 1a), *P. pastoris* (File 1b) and E. coli (File 1c). Code files are presented in additional

Supplementary Material 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.

Supplementary Material File 3 – Stoichiometric matrix and experimental data *for Pichia pastoris*.

An excel 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), CO2 production (CPR), and formation of byproducts (ethanol, citrate, and pyruvate).

Supplementary Material File 4 – Metabolic network of Escherichia Coli. Metabolic network for the Escherichia coli model.

Supplementary Material File 5 – Stoichiometric matrix and experimental data for *Escherichia coli*.

An excel 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.

Considering the environment, only the experimental data sets from Supplementary material File 3, and File 4 are presented. The complete files of supplementary material are included in an additional folder.

P. pastoris experimental datasets taken from the literature.

μ Cmmol/ 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	Reference
1.88	0.00	1.09	0.00	0.00	0.00	0.00	2.16	1.56	Solà et al., 2007
5.74	1.51	0.00	0.00	0.00	0.00	0.00	2.71	3.18	Solá, 2004
4.54	0.00	2.53	0.18	0.00	0.00	0.00	4.78	3.25	Jordà, et al., 2014
1.11	0.00	0.62	0.00	N.A	N.A	N.A	0.87	0.65	Jungo et al., 2006
0.90	0.00	0.00	1.89	N.A	N.A	N.A	1.26	0.99	Tortajada, 2012

Results evaluation degree of consistency

μ Cmmol/ 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	poss*	Reference
6.17	0.00	2.75	0.00	0.00	0.00	0.00	3.62	2.35	**0.16**	Solà et al., 2007
3.27	0.81	0.00	1.09	0.00	0.00	0.00	4.02	2.68	**1.00**	Solá, 2004
2.38	0.00	1.21	0.00	N.A	N.A	N.A	1.65	1.22	**1.00**	Jungo et al., 2006
4.89	0.00	2.40	0.00	N.A	N.A	N.A	3.12	2.29	**1.00**	Jungo et al., 2006
1.40	0.00	0.00	2.55	N.A	N.A	N.A	2.16	1.15	**1.00**	Tortajada, 2012
0.94	0.00	0.00	1.87	N.A	N.A	N.A	1.67	0.93	**1.00**	Tortajada, 2012

* Possibility of the most possible flux distribution.

E. coli experimental datasets taken from the literature. This includes measurements of biomass, glycerol, OUR, CPR and pyruvate, and results evaluation degree of consistency

μ h^{-1}	Q_{Glu} Mmol /g/h	Q_{Ace} Mmol /g/h	Q_{Pyr} Mmol /g/h	OUR Mmol /g/h	CPR Mmol /g/h	Poss*	Reference
0.09	1.4	0	0	4.6	4.9	1.00	Emmerling et al., 2002
0.4	4.8	0	0	11.8	12.4	1.00	Emmerling et al., 2002
0.09	2.2	1.3	0.1	5.1	5.2	1.00	Emmerling et al., 2002
0.08	1.4	0	0	5.5	5.6	1.00	Emmerling et al., 2002
0.4	5	0	0	12.8	13.7	1.00	Emmerling et al., 2002
0.08	2.7	1.2	0	7	6.3	1.00	Emmerling et al., 2002

* Possibility of the most possible flux distribution.

SUPPLEMENTARY INFORMATION

POSSIBILISTIC FLUX ANALYSIS TOOLBOX V1 USER'S MANUAL

Possibilistic Flux Analysis Toolbox v1
User's manual

Contents

INTRODUCTION

Metabolic flux analysis (MFA) is a widely used procedure to estimate the metabolic fluxes within living cells. All MFA-wise methods combine a model with a set of measured fluxes to estimate those that are unknown.

However, traditional MFA methods have limitations. Particularly in scenarios of uncertainty – which are indeed common–, where measurements are scarce and imprecise. The methods implemented in this toolbox, Interval MFA and Possibilistic MFA, have been developed to face those scenarios.

Interval MFA represents the measurements as intervals and provides interval estimates. This way we can deal with imprecise measurements and provide reliable estimates even if our knowledge is incomplete (Llaneras & Picó, 2007; 2008). The method is simple and reliable. Interval MFA has been used by several groups in the last years (D'Huys, 2010; Iyer, 2010; Tortajada, 2010; Zamorano, 2010; Hope, 2011; Lorh, 2014).

Possibilistic MFA is an extension of Interval MFA. Instead of using intervals, the known fluxes can be represented with a possibilistic distribution. And it provides interval and distributions as estimates. These possibilistic estimates are also reliable in uncertain scenarios, as interval ones, but richer and more informative (Llaneras & Pico, 2009; Tortajada, 2012, Morales, 2014).

Herein, we present the PFA Toolbox for MATLAB, which provide a set of functions to easily apply Interval MFA and Possibilistic MFA. The main features of PFA Toolbox are the following:

» Provides reliable MFA estimations in uncertain (or underdetermined) scenarios, where only a few fluxes can be measured.

» Provides MFA estimations accounting for measurements imprecision.

» Makes easy to plot the results as interval estimates or flux distributions.

» Is composed of simple functions that MATLAB users can use in flexible ways and modify if needed.

» Includes a user graphic interface that helps the user to represent its measurements and their uncertainty.

The toolbox is compatible with COBRA, a well-known Toolbox to work with constraint-based metabolic models (Schellenberger, 2011; Agren, 2013). Any COBRA-compatible model can be used in PFA.

LIST OF FUNCTIONS

The PFA Toolbox for MATLAB provides a set of functions to easily apply Interval MFA and Possibilistic MFA. Here we list the functions and their purpose. Each function will be explained below.

Initialization

`initPFAtoolbox`	Start the PFA Toolbox

1. MFA problem formulation

`define_MOC`	Define the constraint-based model and PFA problem.
`define_PossMeasurements`	Define a set of measurements in possibilistic terms.
`define_MEC`	Add the measurements as constraints.

2. Computing estimations

`solve_maxPoss`	One most possible set of flux values.
`solve_maxPossIntervals`	The interval of most possible flux values.
`solve_PossInterval`	The interval of flux values with the desired poss.

3. Plot

`plot_PossMeasurements`	Plot measurements in possibilistic terms.
`plot_distribution`	Plot the distribution of a given flux.
`plot_intervals`	Plot interval estimates of a given flux.

Other

`Solve_possintervalYMP`	(advanced function)
`solve_Interval`	To solve an Interval MFA problem.

SETTING UP THE PFA TOOLBOX

The core methods implemented in PFA toolbox require solving linear programming problems (LP). To solve these problems, we use a flexible and efficient external optimizer: YALMIP (Löfberg, 2004). With the toolbox we provide a copy of YALMIP, but further information about this excellent piece of code can be found in http://users.isy.liu.se/johanl/yalmip/.

YALMIP, and therefore the PFA Toolbox, is able to use different LP solvers. Here is a list of those that we have tested:

- IBM ILOG CPLEX Optimizer. An efficient and reliable solver from IBM. A 90-day evaluation version can be downloading for free.
- GLPK solver. An efficient, reliable and widely used LP solver. Free.
- Linprog. The standard LP solver in MATLAB is not particularly efficient nor reliable, but it can works fine with small MFA problems.

If you want to use a different solver, check the YALMIP documentation.

Setting up the toolbox

These are the steps to install the PFA Toolbox: (1) copy the PFA Toolbox folder where you want it to be installed, (2) start MATLAB, (3) navigate to the PFA toolbox path, and (4) run the following script:

```
>> initPFAtoolbox
```

This action will add the PFA Toolbox directories to your MATLAB path and run the initialization script of YALMIP.

To check which LP solver will be used, run **yalmiptest**:

```
>> yalmiptest
```

The LP solver can be changed. Please read YALMIP help for further details.

Tested systems

The PFA Toolbox has been tested both on Mac OS (v. 10.01, Yosemite) and Windows 7 (64bit). The MATLAB version that have been tested are: R2012a (Windows), R2013b (Mac OS), and R2014a (Windows).

BACKGROUNDS ON MFA

Constraint-based model is an extensively used approach in metabolic modeling of living cells (Palsson, 2006; Stephanoupulos, 1998; Llaneras & Picó, 2008). In this context, MFA is a popular approach to, exploiting these models and some available measurements, estimate all the metabolic fluxes within cells.

Traditional MFA typically combines a linear stoichiometric model with a set of measurements of uptake and production rates or fluxes (which are those easy to obtain). See (Stephanopoulos, 1998) or (Heijden, 1994) for details. However, traditional MFA has some limitations that can be tackled with simple extensions, such as Interval MFA or Possibilistic MFA (Llaneras, 2011). These the two methodologies implemented in PFA Toolbox.

Firstly, let as summarize how traditional MFA is performed.

Consider a metabolic network represented by a stoichiometric matrix **N**. If we add a set of irreversibility constraints, we define a space of feasible steady-state flux distribution. These Model-based Constraints are denoted as MOC:

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

Where **D** is a diagonal matrix with $\boldsymbol{D}_{ii} = 1$ if the flux is reversible (0 otherwise).

The second step of traditional MFA is to introduce a set of measurements fluxes in $\mathbf{v}$ as constraints. If we acknowledge that measurements are imprecise, these measurements could be represented as follows:

$$\mathbf{w_m} = \mathbf{v_m} + \mathbf{e_m} \quad (2)$$

Where $\mathbf{e_m}$ is a vector that represent the intrinsic uncertainty of each of the experimental measured fluxes in the vector $\mathbf{w_m}$.

Thus, Traditional MFA can be defined as the exercise of determining the complete flux vector $\mathbf{v}$ that satisfies (1-2), for a ''reasonably small'' imprecision of the measurements in $\mathbf{w_m}$. Traditional MFA is often formulated as two steps procedure: 1) analyze the consistency of the measurements to detect gross errors – something which is only possible in overdetermined problems– , and 2) solve a Least Squares Problem to estimate $\mathbf{v}$. A nice explanation of this approach can be found in (Klamt, 2002).

These traditional formulations have several problems, particularly in scenarios where data is scarce. 1) They do not account for inequality constraints, which are a useful way of improve the estimation with an information that is often available – reactions irreversibilities– . 2) They only provide point wise estimates, which are uninformative and unreliable when uncertainty is significant. 3) For the previous reason, Traditional MFA cannot be used in scenarios of uncertainty, for example, where only a few measurements are available or those we have are imprecise. Interval MFA and Possibilistic MFA, the methods implemented here, tackle those limitations.

Interval MFA

The interval MFA is a simple yet powerful extension of Traditional MFA. Basically, it is based on representing each measured flux as an interval, $[v_{m,i}^{min}, v_{m,i}^{Max}]$, which can be described as a set of inequality constraints.

$$\mathbf{v_m^m} \leq \mathbf{v_m} \leq \mathbf{v_m^M} \qquad (3)$$

This way, the equations (1-3) define a constraint base, **CB**. These set of constraints define the space of feasible fluxes, according to our model and the constraints imposed by the measurements.

From this **CB**, interval estimates can be achieved for each measured and non-measured flux. The interval of feasible (possible) values for a flux $\boldsymbol{v}$ can be obtained solving two LP problems, as follows:

$$\forall\, \boldsymbol{v}, \begin{cases} \boldsymbol{v_i^m} = \min \boldsymbol{v} \;\; s.t. \boldsymbol{v} \in \mathbf{CB} \\ \boldsymbol{v_i^M} = \max \boldsymbol{v} \;\; s.t. \boldsymbol{v} \in \mathbf{CB} \end{cases} \qquad (4)$$

This provides an interval estimate for each flux of interest.

These estimates are particularly useful in two common situations: when measurements are highly imprecise, and when only a few fluxes are measured. The main benefit of interval estimates is that we can perform MFA even if data is scarce, because the estimates will be only as precise as allowed by the actual uncertainty. Details about Interval MFA can be found in (Llaneras, 2007a; 2007b; 2011).

Example 1: Interval MFA

Now we will illustrate the use of PFA Toolbox with a simple example of Interval MFA. Consider the toy metabolic model shown in figure 1. The network has six fluxes and three metabolites, which impose three independent stoichiometric relationships. Consider also that two fluxes, v_4 and v_6, have been measured. The MFA problem will consist on estimating all other fluxes, and improve our estimates of v_4 and v_6, if possible.

Now we will use the PFA Toolbox to estimate all fluxes in the network, based on the model and the measurements.

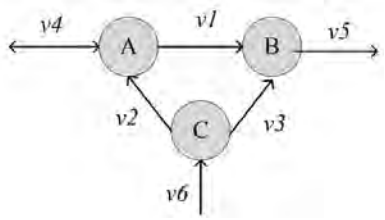


Figure 1. A toy metabolic model.

First, to define the constraint-based model we create the structure *model.* This structure contains the stoichiometric matrix and the information regarding reactions reversibility, as follows:

```
model.S= [-1 1 0 -1 0 0;
    1 0 1 0 -1 0;
    0 -1 -1 0 0 1];
model.rev= [0 0 0 1 0 0];
```

Then, we define flux variables using YALMIP syntax.

```
v = sdpvar(6,1);
```

Now we can generate the constraint base, CB, with all constraints defining the MFA problem so far.

```
CB = [model.S*v==0];
CB = [CB, diag(not(model.rev))*v>=0];
```

We can also bound any flux (if desired), as follows:

```
CB = [CB, v<=1000];
CB = [CB, v>=0];
```

The model has been defined.

Now we add the measurements, represented as intervals to capture their uncertainty. Let us assume, for example, that this uncertainty is ± 3 (flux units) for v_6 and ±0.5 for v_4. In this case, the measured fluxes can be established as follows:

```
CB = [CB, 9 <= v(6) <= 12];
CB = [CB, 9.25 <= v(4) <= 9.75];
```

This is all. The Interval MFA problem has been formulated.

Now, to compute interval estimates we use *solve_Interval*. For example, we can get interval estimate for flux v, as follows:

```
[vmin, vmax] = solve_Interval (CB, v(3));
```

In addition, we can get interval estimates for all six fluxes, as follows:

```
for i=1:6
 [vmin(i,:),vmax(i,:)]=solve_Interval(CB,v(i));
end
```

Finally, The PFA Toolbox also provides functions to plot the estimates. For example, we can use *plot_intervals* to depict interval estimates.

```
plot_intervals([1 2 3 4 5 6],vmin,vmax)
axis square
xlabel('flux'),ylabel('Values')
```

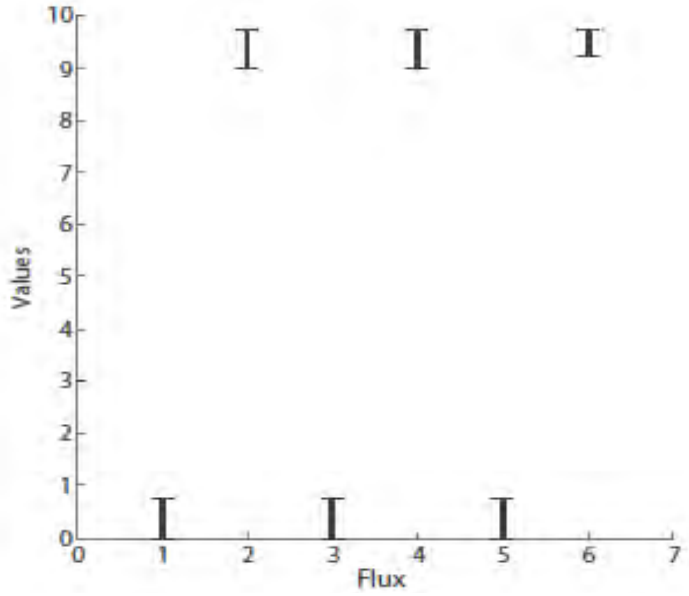


Figure 2. Fluxes estimation with our toolbox using Interval MFA approach.

Possibilistic MFA

In this section we describe Possibilistic MFA. This method can be seen as an extension of Interval MFA, more flexible and able to provide richer estimates. Possibilistic MFA can represent the measured fluxes in a more flexible way – as distributions– and provides interval flux estimates of any desired degree of possibility and also possibility distributions. And in addition to perform flux estimations, Possibilistic MFA can also be used evaluate the consistency between a model and a set of measurements.

The possibilistic framework exploits the notion of "possibilistic constraint satisfaction problems", which was introduced in (Dubois, 1996). Similar optimization approaches to logic reasoning were previously explored in (Sala, 2001; Sala, 2008). The framework is based on two ideas: (1) represent 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 (2) use computationally efficient optimization-based methods to query for the "most possible" solutions. This framework provides a simple and powerful way to deal with uncertainty both in the measurements and

the model (e.g., imprecision and lack of knowledge), which is a typical difficulty in flux estimation problems.

Herein we summarize the formulation of Possibilistic MFA, but further details can be found in (Llaneras, 2009; Llaneras, 2011). First, let as summarize how Possibilistic MFA is performed.

Constraints: model and measurements. Let us start considering the constraints forming the model (MOC) that were given in equation (1). Then, we incorporate measurements of (some) extracellular fluxes as additional linear constraints, the measurement-based constraints (MEC).

Consider a constraint-based model, defined as *MOC*, as in (1). Now we will add the measurements, in a similar as in (2), but we will represent them in possibilistic terms by means of linear constraints and two non-negative slack variables that will represent measurements errors or uncertainty. These constraints are called measurement constraints, or *MEC*, as follows:

$$MEC = \begin{cases} \boldsymbol{w_m} = \mathbf{v_m} + \boldsymbol{\varepsilon_1} - \boldsymbol{\mu}_1 + \boldsymbol{\varepsilon}_2 - \boldsymbol{\mu}_2 \\ \boldsymbol{\varepsilon}_1, \boldsymbol{\mu}_1 \geq 0 \\ 0 \leq \boldsymbol{\varepsilon}_2 \leq \boldsymbol{\varepsilon}_2^{max} \\ 0 \leq \boldsymbol{\mu_2} \leq \boldsymbol{\mu}_2^{max} \end{cases} \quad (5)$$

Where $\mathbf{v_m}$ is the vector of actual fluxes for each measured flux, and $\mathbf{w}$ is the vector of measured values. Both differ due to errors and imprecision. This uncertainty will be represented by the slack variables $\boldsymbol{\varepsilon}$, $\boldsymbol{\mu}$, $\boldsymbol{\varepsilon_2}$ and $\boldsymbol{\mu_2}$. The slack variables $\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon_2}$ represent errors in the measurement $\mathbf{w_m}$ in one direction (measure smaller than actual value), whereas $\boldsymbol{\mu}$ and $\boldsymbol{\mu_2}$ represent errors in the opposite direction. As explained below, error components $\boldsymbol{\varepsilon_1}$ and $\boldsymbol{\mu_1}$ will be penalized in a cost index (6) to assign a decreasing possibility to increasing errors, while $\boldsymbol{\varepsilon}_2$ and $\boldsymbol{\mu_2}$ will be remain non-penalized, thus defining a band of fully possible $\mathbf{v}_m$ values around the measured $\mathbf{w_m}$, in particular [$\mathbf{w} + \boldsymbol{\varepsilon}_2^M$, $\mathbf{w} - \boldsymbol{\mu}_2^M$]. Once $\mathbf{v}_m$ surpass these bounds, $\boldsymbol{\varepsilon_1}$ and $\boldsymbol{\mu_1}$ must be nonzero to fulfill (5) and the associated "cost" (6) will indicate some disagreement with the model (i.e., lower possibility, as later defined).

Each candidate solution of (1) and (5) can be denoted as $\delta=\{\mathbf{v}, \boldsymbol{\varepsilon_1}, \boldsymbol{\mu_1}, \boldsymbol{\varepsilon_2}, \boldsymbol{\mu_2}\}$. Now, we define a function, $\pi(\delta):\Delta\rightarrow[0,1]$, that assigns possibility in [0, 1] to each solution, ranging between impossible and fully possible. A simple yet sensible way to build this function is

using a linear cost index J to penalize large deviations between actual fluxes and their measured values:

$$J = \boldsymbol{\alpha} \cdot \boldsymbol{\varepsilon_1} + \boldsymbol{\beta} \cdot \boldsymbol{\mu_1} \quad (6)$$

And the possibility of each solution is defined as follows:

$$\pi(\delta) = \exp\big(-J(\delta)\big) \;\; \delta \, \epsilon \, MEC \;\cap MOC \quad (7)$$

Where α and β are row vectors of sensor accuracy coefficients.

This way equations (5-7) can be interpreted as representing the statement: «given a measured value $\boldsymbol{w_m}$, the assertion $\boldsymbol{v}_m$=$\boldsymbol{w_m}$ is fully possible, and the more $\boldsymbol{v}_m$ and $\boldsymbol{w_m}$ differ, the less possible such situation is».

The actual possibility for each value of $\boldsymbol{v_m}$ depends on how the user defines the "sensor" accuracy coefficients in (5) and (6): the maximum bounds for $\boldsymbol{\varepsilon}_2^M$ and $\boldsymbol{\mu}_2^M$ define an interval of fully possible values (π=1), and the possibility of $\boldsymbol{v}_m$ being out of this interval depends on the user-selected weights $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in (5). Notice that measurements uncertainty can be non-symmetric, and that very complex descriptions can be achieved by adding slack variables.

Poss-MFA: estimating fluxes

The simplest flux estimate $\mathbf{v_{mp}}$ in $\boldsymbol{\delta}_{mp}$={$\mathbf{v}_{mp}$, $\boldsymbol{\varepsilon}_{1,mp}$, $\boldsymbol{\mu}_{1,mp}$, $\boldsymbol{\varepsilon}_{2,mp}$, $\boldsymbol{\mu}_{2,mp}$} is given by the maximum possibility (minimum-cost) solution of the constraint satisfaction problem (1)-(5), which can be obtained solving a linear programming problem (LP):

$$J^{min} = \min_{\boldsymbol{\varepsilon},\boldsymbol{\mu},\boldsymbol{v}} J \; s.t \; \{MOC \cap MEC \quad (8)$$

Notice that point-wise estimates would be unreliable if multiple solutions were reasonably possible, so it is advisable to get interval estimates and distributions. Interval estimates for desired marginal (π) and conditional or *a posteriori* (γ) possibilities can be found, again, solving efficient LP optimizations (Llaneras, 2011).

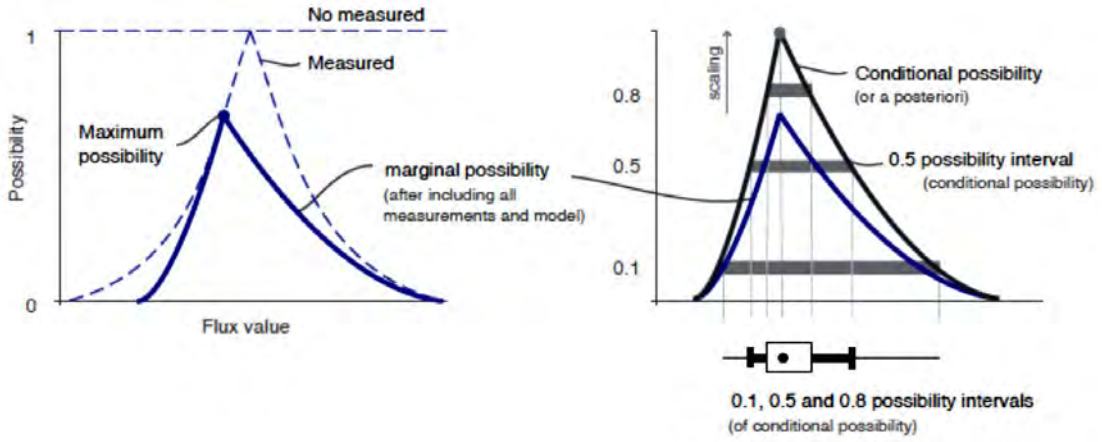


Figure 3. Possibilistic flux estimations. (Left) the figure shows the possibilistic distribution representing the original measurement, the marginal distribution and the maximum possibility flux estimation. (Right) The figure shows the marginal and conditional (a posteriori) possibility of π=0.8, 0.5 and 0.1, and the maximum possibility estimation, are depicted in a box-plot chart.

Interval estimates. The interval of values with conditional possibility higher than γ for a given flux $\left[v_{i,\gamma}^{m}, v_{i,\gamma}^{M}\right]$ can be computed solving two extra LPs:

$$\boldsymbol{v}_{i,\gamma}^{m} = \min_{\boldsymbol{\varepsilon},\boldsymbol{\mu},\boldsymbol{v}} \boldsymbol{V}_i \; s.t \begin{cases} MOC \cap M\varepsilon C \\ J - J_{min} < -\log \gamma \end{cases} \tag{9}$$

The upper bound is conformed replacing minimum by maximum.

These possibilistic intervals have a similar interpretation to confidence intervals (*credible intervals*) in Bayesian statistics. In practice, getting estimates with conditional possibilities γ is equivalent to normalize the marginal possibility distributions to a maximum equal to one. We typically compute the interval of most possible values (γ=1) and other less possible intervals (γ=0.8, 0.5 and 0.1) to capture uncertainty.

Marginal and conditional possibility distribution. Additionally to interval estimates, PFA toolbox allow estimate conditional possibility distributions. Each interval is estimate as was previously explained for different possibilities to build a distribution. Figure 3 shown a conditional possibility distribution, and marginal distribution is also presented

The marginal distribution can be interpreted as the ''distribution of the possible values for each flux in the network given the measurements'' and the conditional possibility is a normalization of marginal possibility, where we assume that the model and measurements are correct (Llaneras, 2009).

Poss-MFA: evaluating data consistency

Poss-MFA can also be applied to evaluate the degree of consistency between given model (1) and a set of experimental measurements (5). Notice that the most possible solution of the constraint satisfaction problem (1) and (2) computed with (6) has an associated degree of possibility that grades consistency:

$$\pi^{mp} = \exp(-j^{min}) \qquad (10)$$

This value, π^{mp} in [0, 1], grades the consistency between model and measurements. Possibility equal to one must be interpreted as complete agreement between the model and the measurements, whereas lower values imply that there is certain degree of error in the measurements, the model or both. The evaluation of consistency can be used to (a) conciliate a set of experimental measurements, (b) serve as basis for process monitoring and fault detection systems, or (c) validate a constraint-based model (Llaneras, 2009; 2011; Tortajada, 2010).

Example 2: introduction to Possibilistic MFA

We will consider the same network of example 1, showed in figure 4. The network has six fluxes and three metabolites, which impose three independent stoichiometric relationships. Consider also that two fluxes v_4 and v_6, have been measured. The MFA problem will consist on estimating all other fluxes, and improve our estimates of v_4 and v_6, if possible.

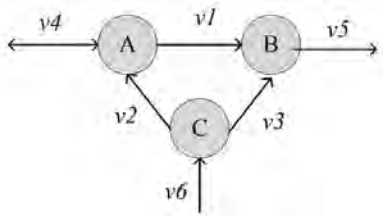


Figure 4. A toy metabolic model.

Again, we start by defining the constraint-based model. The structure *model* contains the stoichiometric matrix and the information regarding reactions reversibility, as follows:

```
model.S= [-1 1 0 -1 0 0;
  1 0 1 0 -1 0;
  0 -1 -1 0 0 1];
model.rev= [0 0 0 1 0 0];
```

The model defines the model constraints (MOC), that we use to create the Constraint Base for our Possibilistic MFA problem:

```
[PossProblem] = define_MOC(model);
```

Now we add the measurements. Let us assume, for example, that we have measurements for fluxes 4 and 6, with $w_4 = 9.5$ and $w_6 = 10.5$, and that the first measurement is very accurate and last one is unreliable. We can choose $\boldsymbol{\varepsilon}_2^{max}, \boldsymbol{\mu}_2^{max}, \boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, accordingly. (In subsequent examples, we will give more details about how to choose these variables.)

```
PossMeasurements.wm  = [9.5 10.5]';
PossMeasurements.e2max = [0.25 1.5]';
PossMeasurements.m2max = [0.25 1.5]';
PossMeasurements.alpha  = [2  0.15]';
PossMeasurements.beta  = [2  0.15]';
```

At this point, we can check how our measurements look.

```
% To plot one single measured flux, for flux 4.
plot_PossMeasurements(possmeas,1);

% To plot all measured flux
index = [4 6]; % Indexes in model of the measured fluxes
for f=[1:6]
 subplot(2,3,f), xlim([0 50]), hold on
 m=find(index==f);
 if(not(isempty(m)))
  [x,y] = plot_PossMeasurements(PossMeasurements, m);
   plot(x,y,'b--')
 end
end
```

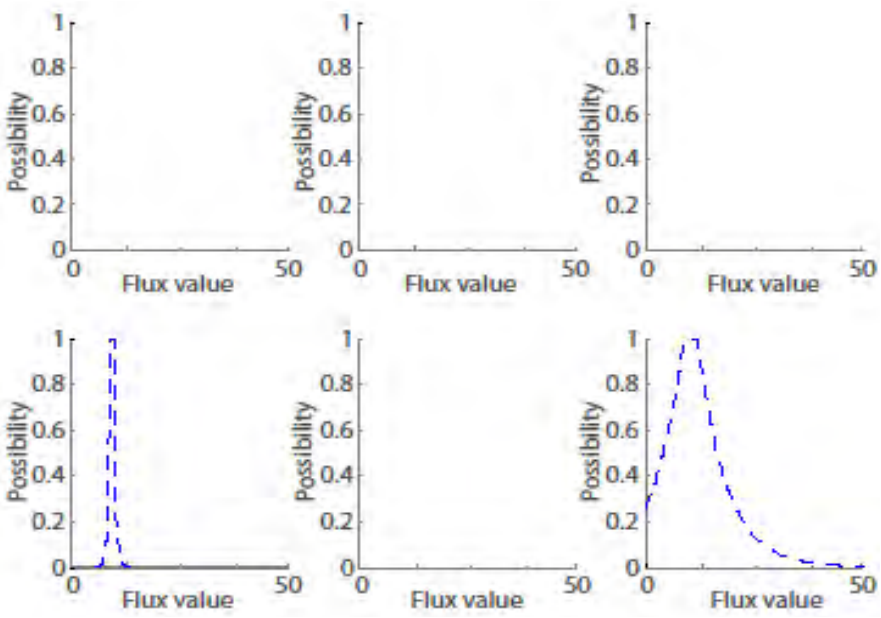


Figure 5. Measured fluxes and their uncertainty, represented as a possibilistic distribution.

Once the measurements have been defined, we can add the as measurement constraints (MEC) to our Possibilistic MFA problem:

```
index = [4 6]; % Indexes in model of the measured fluxes
[PossProblem] = define_MEC(PossProblem, PossMeasurements, index);
```

At this point, once we have added the model and the measurements, the Possibilistic MFA problem is completely defined. Now we can get the flux estimations. Let us present the three different options.

You can perform a **point-wise** estimation:

```
[v,poss]=solve_maxPoss(PossProblem);
```

It returns two outputs, v, the estimate for each flux in the network, and *poss*, the possibility of this most possible solution.

However, you can also get a more reliable estimate, the **interval** of values with maximum possibility:

```
[vmin,vmax]=solve_maxPossIntervals(PossProblem);
```

The outputs *vmin* and v*max* are two vectors that contain the minimum and maximum values for each flux with maximum possibility. This way, if there are multiple flux values with maximum possibility, you will get al.,l of them.

Similarly, you can also calculate any **interval estimation with specific possibility**. For example, you can get the interval of values for v_6 with a conditional possibility of 0.8 (i.e., that are "quite" possible).

```
[min6, max6]=solve_PossInterval(PossProblem,0.8,6);
```

Alternatively, you can get three intervals estimates for every flux, with a loop:

```
fluxes = [1 2 3 4 5 6];
for f=fluxes
 [mn2(f), mx2(f)]=solve_PossInterval(PossProblem,[1],f);
 [mn3(f), mx3(f)]=solve_PossInterval(PossProblem,[0.8],f);
 [mn4(f), mx4(f)]=solve_PossInterval(PossProblem,[0.5],f);
end
```

We provide a function to plot these rich and compact estimates:

```
plot_intervals(fluxes,min2,max2,min3,max3,min4,max4);
```

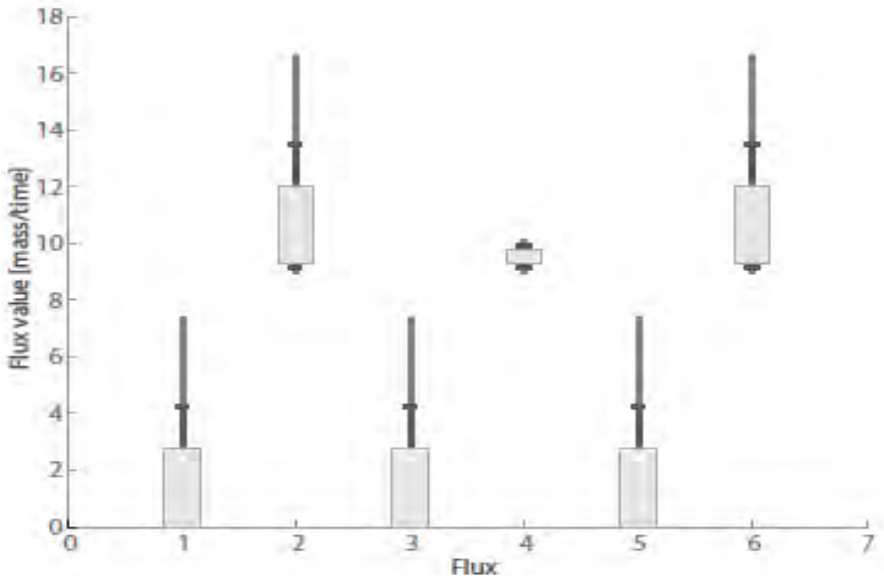


Figure 6. Estimates for every flux. Three interval estimates are given, for maximum conditional possibility (box), possibility of 0.8 (black line), and 0.5 (gray line).

Finally, you can compute a complete possibility distribution for any flux. To plot the results, a plot functions is provided. The following example computes and plots the possibility distribution for every flux in our example.

```
possibilities = [0.05:0.05:1];        % granularity
fluxes = [1 2 3 4 5 6];               % fluxes to plot
for f=fluxes
 [min_p(:,f), max_p(:,f)]= solve_PossInterval(PossProblem,possibilities,f);
end
```

```
figure
for f=fluxes
     subplot(2,3,f), hold on
     [x,y]= plot_distribution(min_p(:,f),max_p(:,f),possibilities);
     plot(x,y,'k','lineWidth',2)
 xlim([0 50])
     if(find(index==f))
  [x,y]=plot_PossMeasurements(meas,find(f==index));
  plot(x,y,'b--')
 end
end
```

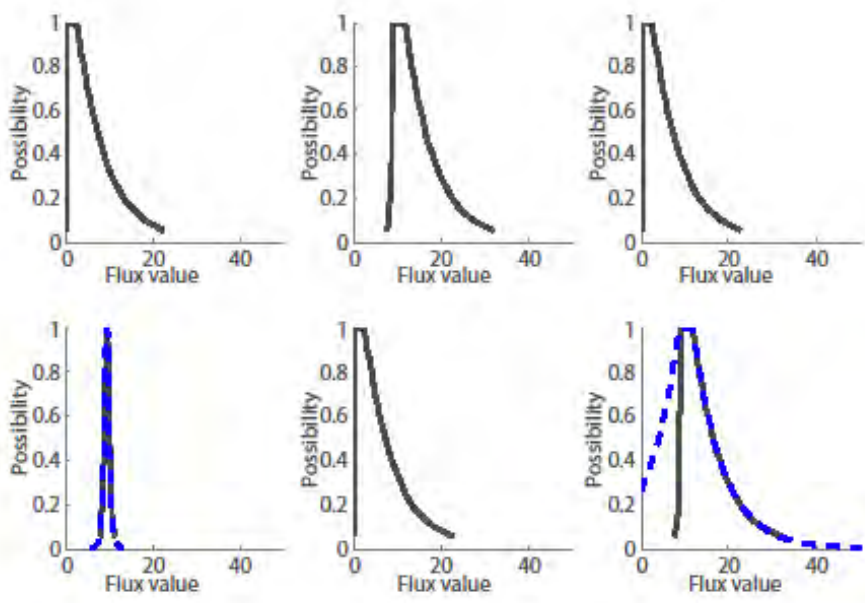


Figure 7. Distribution of possible values for every flux in the network. The original measured values are depicted in dashes lines. Solid lines represent the estimates provided by Possibilistic MFA. Notice that measured fluxes can also be estimated.

Note: all the computations for specific degrees of possibility have been computed for conditional possibilities. This is the default. If you want to compute marginal distribu-

tion, you should add a parameter when calling the functions (check their help for details). An explanation about marginal vs. conditional possibility can be found in (Llaneras, 2009; 2011).

Example 3: how to represent the measurements

In any MFA problem, there is one keystone step: decide how uncertainty your measurements are. Let us discuss this issue with a new example. Consider, for instance, that we have measurements of three uptake and production rates, for oxygen, ethanol and glucose. The measurements are: $w_{glu} = 40.6$, $w_{EtOH} = 15.96$, $w_{O_2} = 61.9$. The glucose measurement is very accurate, the ethanol one is moderately accurate, and the measurement of oxygen is quite unreliable. Now, we, as user, have to translate this information about the measurements and their uncertainty into a possibilistic representation, by choosing the weights α and β and the limits $\boldsymbol{\varepsilon_2^{max}}, \boldsymbol{\mu_2^{max}}$.

The PFA Toolbox provides two routes to achieve this.

The straightforward approach: let the user define these variables. First, we define the limits $\boldsymbol{\varepsilon_2^{max}}, \boldsymbol{\mu_2^{max}}$, which define an interval of values around the measured value with full possibility. Then, we define the weights α and β, which are associated with the slack variables ε_1 and μ_1– to penalize the values out of full possibility interval and make them "less possible". If uncertainty is assumed to symmetric, $\alpha = \beta$, and $\boldsymbol{\varepsilon_2^{max}} = \boldsymbol{\mu_2^{max}}$. These parameters need to be defined to every flux.

For example, this is a representation of our measurements:

```
PossMeasurements.wm=[40.6 15.96 61.9];
PossMeasurements.e2max=[0.5 0 4];
PossMeasurements.m2max=[0.5 0 4];
PossMeasurements.alpha=[2 0.5 0.11];
PossMeasurements.beta=[2 0.5 0.11];
```

As before, you can plot your measurements to check if they look, as you –-the user–, want them to look.

```
figure
for f=[1:3]
 subplot1,3,f), hold on
 [x,y] = plot_PossMeasurements(possmeas,f);
  plot(x,y,'b--')
  xlim([0 120]);
end
```

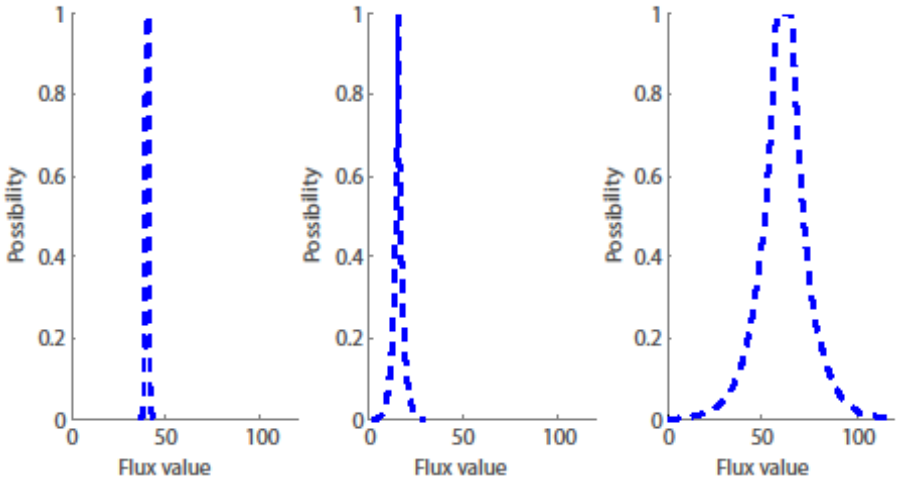


Figure 8. Our measurements, and its uncertainty, represented in possibilistic terms.

A simplest approach: use a function to define the measurements. The PFA Toolbox provides a function to assign the uncertainty of the measurements in an easier way that is suitable in many occasions. Basically, user only defines two intervals of high possibility (`intFP`) and low possibility (`intLP`), and parameters **α, β**, $\boldsymbol{\varepsilon_2^{max}}$ & $\boldsymbol{\mu_2^{max}}$ are defined accordingly:

- Full possibility (π=1) is assigned to de interval $\boldsymbol{w_m} \pm$ `intFP`.
- Larger deviations are penalized so that values equal to $\boldsymbol{w_m} \pm$ `intLP` have a possibility of π=0.1.

Coming back to our example, we could do the following:

```
wm=[40.6 15.96 61.9]; %define measures and its full and lower possibility
intFP=[0.5 0 4];
intLP=[1.65 4.6 24.9];

[PossMeasures]=define_PossMeasurements(wm,intFP,intLP);
```

And plot:

```
figure
for f=[1:3]
 subplot(1,3,f), hold on
 [x,y] = plot_PossMeasurements(PossMeasures,f);
  plot(x,y,'r--')
  xlim([0 120]);
end
```

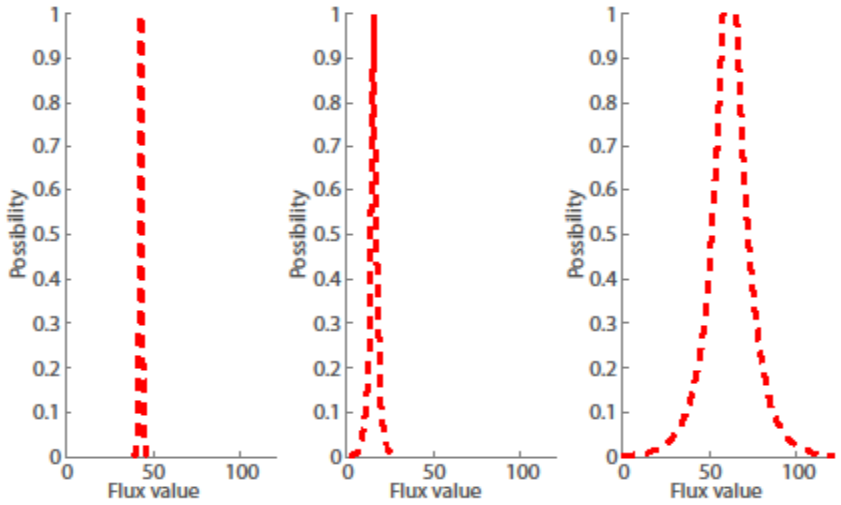


Figure 9. Our measurements, and its uncertainty, represented in possibilistic terms.

When using the PFA toolbox, we have noticed that there are often advantages in defining the interval of low and high possibility in relative terms. For example, one clear and simple way of define the uncertainty of O2 measurements is to state somethings like: «a 5% deviation from the measured values is fully possible, but a deviation larger than 20% is an event of low possibility».

If we reason like this for every measurement, they can be represented as follows:

```
wm=[40.6 15.96 61.9];
intFP = max(0.01,abs(wm.*0.05));
intLP = max(0.02,abs(wm.*0.20));

[PossMeasures]=define_PossMeasurements(wm,intFP,intLP);
```

Notice that the intervals are defined with a minum size of 0.01 and 0.02 flux units. We do this to avoid problems in case a measurement in $\mathbf{w_m}$ is zero or near zero. In general, when defining measurements uncertainty in absolute terms, is useful to consider also a minimal, absolute band of uncertainty. Otherwise, the uncertainty of near zero measurements tends to be underrepresented.

Example 4. Evaluating consistency

As mentioned above, Possibilistic MFA and the PFA Toolbox can be also useful to evaluate the degree of consistency between a given model and a set of experimental measurement. You can find real applications in (Tortajada, 2010; Morales, 2014), herein we present a simple example with the toy model of figure 1.

Let us consider the following situation: we have six different datasets, corresponding – for example– to six chemostat experiments, with four measured fluxes in each case (w_2, w_3, w_4 & w_6) and we want to check if there are larger error than expected in the measurements, by evaluating the consistency between the model and each dataset. (As we mention in the backgrounds, consistency analysis can be used both to validate a model or a set of data. In this example, we assume that the model has been shown to be precise and we use it to find error in the measurements.)

The first step was defining the measurements uncertainty; in this case, we considered the same uncertainty for all the measures. We consider that the fluxes are fully possible with deviations ±0.5 units of flux, and deviations larger than ±1.2 units will have a lower possibility.

The second one is computing the possibility of the most possible solution, with *solve_maxPoss*, which provides a simplest indication of how consistent model and measurements are. The results are provided in Table 1.

Dataset	Flux 2	Flux 3	Flux 4	Flux 6	π in [0,1]
1	1.88	1.09	0.0	2.16	1
2	2.07	0.95	0.63	2.70	1
3	1.72	1	1.48	2.90	1
4	2.02	0.57	1.33	3.55	1
5	2.32	1.13	2.62	3.52	1
6	0.73	2.04	1.98	2.55	**0.439**

Table 1. Experimental data and consistency analysis.

The results are that five in six datasets are fully possible, whereas the dataset 6 has a lower possibility (0.44). This implies that one of measured fluxes in dataset six is highly deviated, or more than one is sensibly deviated.

The procedure to perform all these calculations with the PFA Toolbox is presented below.

We start by defining the model.

```
model.S= [-1 1 0 -1 0 0;
  1 0 1 0 -1 0;
  0 -1 -1 0 0 1];
model.rev=[0 0 0 1 0 0];
[PossProblem] = define_MOC(model);
```

Then we are going to define the measurements and its uncertainties, which were defined previously as 0.5 for full possibility, and 1.2 for lower possibility.

```
index  =[ 2  3  4   6 ];
exp{1}.wm=[ 1.88 1.09 0.0 2.16];
exp{2}.wm=[ 2.07 0.95 0.63 2.70];
exp{3}.wm=[ 1.72  1  1.48 2.90];
exp{4}.wm=[ 2.02 0.57 1.33 3.55];
exp{5}.wm=[ 2.32 1.13 2.62 3.52];
exp{6}.wm=[ 0.73 2.04 1.98 2.55];

intFP=[0.5];
intLP=[1.2];
```

Then we compute the possibility of the most possible solution for each dataset.

```
for i=1:length(exp)
 [measures]=define_PossMeasurements(exp{i}.wm,intFP, intLP);
 [PossProblem]=define_MEC(PossProblem, measures, index);
 [v, poss]=solve_maxPoss(PossProblem);
 poss_all(i)=[poss];
  i
end
```

These results are those provided in Table 1.

Investigating the inconsistencies. To better understand where the inconsistencies come from, we can compute the marginal distributions for every flux in the dataset 6 (the inconsistent one).

First, we define the problem:

```
[PossProblem] = define_MOC(model);
[PossMeasures]=define_PossMeasurements(exp{6}.wm,intFP, intLP);
[PossProblem]= define_MEC(PossProblem, PossMeasures, index);
```

Then we compute the complete marginal possibility distribution for every flux and and plot the results.

```
poss = [0.01:0.01:0.439];
flux = [1 2 3 4 5 6];

for f=flux
     [minp(:,f),maxp(:,f)]=solve_PossInterval(PossProblem,poss,f,'none');
end

figure
for f=fluxes
 subplot(2,3,f), hold on
 [x,y] = plot_distribution(minp(:,f),maxp(:,f),poss);
 plot(x,y,'k','lineWidth',2)
 xlim([0 5])
 if(find(index==f))
  [x,y] = plot_PossMeasurements(PossMeasures, find(f==index));
   plot(x,y,'b--')
   xlim([0 5])
 end
end
```

These results shown that, taken the model into account, there are larger than expected errors in the measurements. Further investigations can be performed to find which measurements may be provoking the inconsistency. See (Llaneras, 2009; Llaneras, 2011) for more elaborate examples of these procedures.

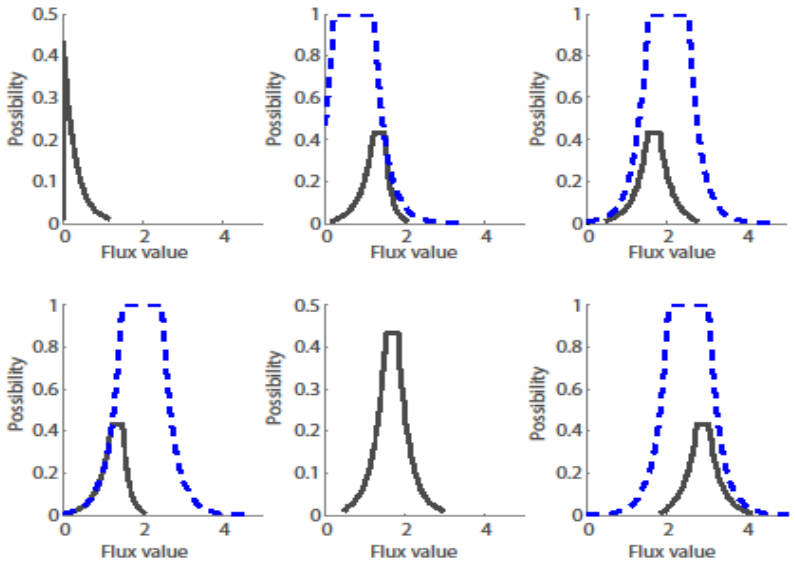


Figure 10. Possibilistic MFA performed to detect error in a set of measurements. Solid lines show the marginal possibility distribution for each flux (in the dataset 6). Dashed lines represent the original measurements, or a priori, that are of course fully possible. The results shown that, taken the model into account, it seems to be larger than expected errors in the measurements.

FUNCTIONS DESCRIPTION

The following section describes each function of the PFA Toolbox. We show its syntax, a brief description, and the lists of inputs and outputs. This information can also be consulted within MATLAB.

Initialization

`initPFAtoolbox`	Start the PFA Toolbox

1. MFA problem formulation

`define_MOC`	Define the constraint-based model and PFA problem.
`define_PossMeasurements`	Define a set of measurements in possibilistic terms.
`define_MEC`	Add the measurements as constraints.

2. Computing estimations

`solve_maxPoss`	One most possible set of flux values.
`solve_maxPossIntervals`	The interval of most possible flux values.
`solve_PossInterval`	The interval of flux values with the desired poss.

3. Plot

`plot_PossMeasurements`	Plot measurements in possibilistic terms.
`plot_distribution`	Plot the distribution of a given flux.
`plot_intervals`	Plot interval estimates of a given flux.

Other

`Solve_possintervalYMP`	(advanced function)
`solve_Interval`	To solve an Interval MFA problem.

initPossToolbox

The function initiates the PFA Toolbox.

Syntax

```
>> initPFAtoolbox
```

Description

The function initiates the PFA Toolbox. It adds the toolbox folder to the MATLAB path. If YALMIP is not already installed, a copy is also added to the path. The optimization solver GLPK is selected, if it is installed and is detected. Otherwise, YALMIP is initialized with the available solver.

defineMOC

Generate the constraint-based model structure.

Syntax

```
[PossProblem] = define_MOC(model)
```

Description

[PossProblem] = define_MOC(model) returns a struct that defines the Possibilistic MFA problem. Initially, it contains some symbolic decision variables (the fluxes v) and the first constraints into the CB (the stoichiometry and the irrerversibilities). The function receives as input another struct, model. This struct can be a COBRA model – it has the same fields– or be created by the user. The struct is a simple one, containing the following:

model.S	The stoichiometric matrix.
model.rev	A vector indicating which reactions are reversible with '1' for those reversible and '0' otherwise.
model.lb	(optional) Lower bound for each flux.
model.ub	(optional) Upper bound for each flux.

define_PossMeasurements

It generates a set of possibilistic measurements.

Syntax

```
[PossMeasurements]=define_PossMeasurements(wm, intFP, intLP);
[PossMeasurements]=define_PossMeasurements(meas);
```

Description

[PossMeasurements]=define_PossMeasurements(wm, intFP, intLP) returns a *PossMeasurements* struct that contains the measured fluxes `wm` the limits `e2max` and `m2max`, and the weights `alpha` and `beta`, required to represent a set of measurements in possibilistic terms. Vector *wm* is a vector with the measured values for each flux. *intFP* represent the interval around *wm* in which flux values have maximum possibility. It can be a single value, equal for all the measurements, or a vector with a specific value for each measurement. *intLP* represent the interval around *wm* in which measurements have a low possibility. It can be a single value or a vector.

[PossMeasurements]=define_PossMeasurements(meas) does exactly the same, but receives a struct as input, with fields: `wm`, *intFP* and *intLP*.

wm	Measures values vector
intFP	Interval with maximum possibility
intLP	Interval with low possibility

Recall that *define_PossMeasurements* defines the measurements and their uncertainty based on two intervals. Basically, the user only defines two intervals of high possibility (*intFP*) and low possibility (*intLP*), and parameters **α, β**, $\boldsymbol{\varepsilon_2^{max}}$ & $\boldsymbol{\mu_2^{max}}$ are defined accordingly:

- Full possibility (π=1) is assigned to the interval $w \pm intFP$.
- Larger deviations are penalized so that values equal to $w \pm intLP$ have a possibility of π=0.1.

As an alternative, advanced user can define directly the variables in the *PossMeasurements* struct.

defineMEC

Add the measurement-constraints to a *PossProblem* structure.

Syntax

```
[PossProblem]=define_MEC(PossProblem, PossMeasurements, index)
```

Description

[PossProblem]=define_MEC(PossProblem, PossMeasurements, index) returns the *PossProblem* received as input adding the measurements as constraints. The output struct has the following fields:

.v	The vector of fluxes (as a YALMIP decision variable)
.CB	The constraints-base, i.e., the set of all constraints (in YALMIP syntax).
.e1	The vector of slack variables (as a YALMIP decision variable)
.m1	The vector of slack variables (as a YALMIP decision variable)
.e2	The vector of slack variables (as a YALMIP decision variable)
.m2	The vector of slack variables (as a YALMIP decision variable)
.J	The objective function penalizing the deviations between fluxes and measured values, accordingly to alpha and beta.

The input `PossProblem` is a struct generated by `defineMOC` function, and defines the model constraints of the MFA problem. The input `PossMeasurements` is a structure generated by *define_PossMeasurements* or manually, and defines the measurements in possibilistic terms. Finally, the vector *index* indicates the indexes of the measured fluxes in the model.

solve_maxPoss

Returns one flux vector estimate of maximum possibility.

Syntax

```
[v,poss] = solve_maxPoss(PossProblem);
[v, poss, diagnostic] = solve_maxPoss(PossProblem, options);
```

Description

[v, poss]=solve_maxPoss(PossProblem) returns the column vector **v** with a set of fluxes with maximum possibility. Notice, however, that it could more than one flux vector with maximum possibility, and the solver returns only one of them. To know if there are multiple candidates, use *solve_maxPossInterval*. The only mandatory input is *PossProblem*, a structure defining the Possibilistic MFA problem (see *define_MEC* and *define_MOC*).

The optional input "*options*" specifies the YALMIP options (see '*help yalmip'*).

The optional output "*diagnostic*" returns information about the solver status (see '*help yalmiperror*'). "*diagnostic.error*" indicates if the problem was successfully solved (it return '0' if the problem is successfully solved, '1' if the problem is infeasible, etc. See '*yalmiperror*'). "*diagnostic.details*" provides all the info returned by the optimization solver.

solve_maxPossIntervals

It returns the interval estimate of fluxes with maximum possibility.

Syntax

```
[vmin,vmax]=solve_maxPossIntervals(PossProblem);
[vmin, vmax, diagnostic] = solve_maxPossIntervals(PossProblem, options);
```

Description

[vmin, vmax]=solve_maxPossIntervals(PossProblem) returns the interval estimate with maximum possibility in vectors v*min*, and *vmax*, which contain the lower and upper limits for each flux. The only mandatory input is *PossProblem*, a structure defining the Possibilistic MFA problem (see *define_MEC* and *define_MOC*).

The optional input "*options*" specifies the YALMIP solver options (see *'help yalmip'*).

The optional output "*diagnostic*" returns information about the solver status (see *'help yalmiperror'*). "*diagnostic.error*" indicates if the problem was successfully solved (it return '0' if the problem is successfully solved, '1' if the problem is infeasible, etc. See '*yalmiperror*'). "*diagnostic.details*" provides all the info returned by the optimization solver.

solve_PossInterval

It returns an interval estimate for a flux for the desired degree of possibility.

Syntax

```
[vmin, vmax]=solve_PossInterval(PossProblem, poss, flux);
[vmin, vmax]=solve_PossInterval(PossProblem, poss, flux, mode);
[vmin, vmax, diagnostic]=solve_PossInterval(PossProblem, poss, flux, mode, options);
```

Description

[vmin, vmax, diagnostic]=solve_PossInterval(PossProblem, poss, flux) returns an interval estimate for a flux, for the desired degree of conditional possibility. The vectors v*min* and v*max* contain the upper and lower limits of the flux of interest for the degrees of possibility specified as input.

The input *PossProblem* is a structure defining the Possibilistic MFA problem (see ***define_MEC*** and ***define_MOC***). The vector *poss* indicates the degree of possibility for the intervals that you want to compute (e.g., 0.8, or [0.99 0.8 0.1], etc.). Flux indicates the index of the flux to be estimated.

> Example. *[vmin, vmax]= solve_PossInterval (ProblemA, [0.99 0.5 0.1], 7)* computes three interval estimates for the flux 7 in the ProblemA, for conditional of possibilities 0.99, 0.5 and 0.1.

The optional input "*mode*" can be used to get estimates of marginal possibility, instead of conditional possibility. If mode is not provided, the function provides conditional possibilities as a default.

The optional input "*options*" specifies the YALMIP solver options (see '*help yalmip'*).

The optional output "*diagnostic*" returns information about the solver status (see '*help yalmiperror*'). "*diagnostic.error*" indicates if the problem was successfully solved (it return '0' if the problem is successfully solved, '1' if the problem is infeasible, etc. See '*yalmiperror*'). "*diagnostic.details*" provides all the info returned by the optimization solver.

plot_intervals

Plot interval estimates for a set of fluxes.

Syntax

```
plot_intervals(flux, vmin_c1, vmax_c1);
plot_intervals(flux, vmin_c1, vmax_c1 ,vmin_c2, vmax_c2);
plot_intervals(…, …, …, …, vmin_c3, vmax_c3);
[h_out] = plot_intervals(…);
```

Description

plot_intervals(flux, vmin_c1, vmax_c1) creates a 2D plot to show a set of interval estimates. The function receives as input a vector of x coordinates, ***flux*** (e.g., [1:5] or [1 7 8]). Vectors ***vmin_c1, vmax_c1*** define the lower and upper limits of the intervals to be plotted, which typically would have been computed with ***solve_PossInterval***.

plot_intervals(flux, vmin_c1, vmax_c1, vmin_c2, vmax_c2) and *plot_intervals(flux, vmin_c1, max_c1, vmin_c2, vmax_c2, vmin_c3, vmax_c3)* allow to plot two and three pairs of interval estimates in a single, compact graph.

If an output variable is indicated, "*h_out*", the functions will return a structure with the handles of every object in the figure. This way, it can be customized.

plot_distribution

Plot a complete the possibilistic distribution of a flux.

Syntax

```
plot_distribution(vmin,vmax,poss)
[meas_out,poss_out] = plot_distribution(vmin,vmax,poss)
```

Description

plot_distribution(vmin,vmax,poss) plots the possibilistic distribution for a specific flux. The function receives three inputs, vectors *vmin*, and v*max*, with the lower and higher limits of a set of interval estimates, each interval corresponding to the degrees of possibility indicated in *poss*. v*min* and *vmax* are the output of the *solve_PossInterval(…, poss, …).*

If output variables are given, as in *[meas_out, poss_out]=…*, instead of drawing a graph, the data is returned as two output variables for x and y coordinates. This way, the user can use plot function to plot the data with a custom style.

Example: *[x, y] = plot_distribution(vmin, vmax, poss); plot(x, y, 'r').*

plot_PossMeasurements

Plot measurements defined in possibilistic terms.

Syntax

```
plot_PossMeasurements(PossMeasurements,flux);
plot_PossMeasurements(possmeas,flux, resolution, pos_min);
[meas_out,poss_out]= plot_PossMeasurements(…);
```

Description

plot_PossMeasurements(PossMeasurements, flux) creates a 2D plot with the possibilistic distribution of one measured flux. This way the user can check if the measurements and its uncertainty is well-defined. The function receives as input a set of measurements in a structure, *PossMeasurements*. The measured flux to be plotted is indicated with *flux*.

The input *PossMeasurements* is a struct with the measured fluxes `wm` the limits `e2max` and *m2max*, and the weights `alpha` and `beta`. It can be defined manually by the user or via function *define_PossMeasurements*.

The optional inputs "*resolution*" and "*pos_min*" are used to specify the number of points used to create the plots (default, 20) and the minimum possibility to be plotted (default, 0.001).

If output variables are given, as in *[meas_out, poss_out]=…*, instead of drawing a graph, the data is returned as two output variables for x and y coordinates. This way, the user can use plot function to plot the data with a custom style.

Example: *[x, y] = plot_PossMeasurements(A, 3); plot(x, y, '*k').*

Solve_possintervalYMP (for advanced users only)

It returns an interval estimate for a flux for the desired degree of possibility. This function uses a different and rudimentary syntax. It is of use only for advanced users wanting to do non-standard computations.

Syntax

```
[vmin,vmax,diagnostic]=solve_PossIntervalYMP(F,J,poss,var);
[__,__,__] = solve_PossIntervalYMP (__,__,__,__,mode);
[__,__,__] = solve_PossIntervalYMP (__,__,__,__,options);
```

Description

This function uses a different and rudimentary syntax. It is of use only for advanced users wanting to do non-standard computations.

[vmin, vmax]=solve_PossIntervalYMP(F, J, poss, var) returns an interval estimate for a flux, for the desired degree of conditional possibility. The vectors v*min* and v*max* contain the upper and lower limits of the flux of interest for the degrees of possibility specified as input.

The input *F* is a YALMIP structure defining a set of constraints. *J* is a YALMIP object function defining the possibility of each candidate solution of *F*. The vector *poss* indicates the degree of possibility of the intervals that you want to compute (e.g., 0.8, or [0.99 0.8 0.1], etc.). Finally, *var* is the variable – typically a flux– that you want to estimate.

The optional input "*mode*" can be used to get estimates of marginal possibility, instead of conditional possibility. If mode is not provided, the function provides conditional possibilities as a default.

The optional input "*options*" specifies the YALMIP solver options (see 'help yalmip').

solve_interval

Solve an interval MFA problem.

Syntax

```
[vmin, vmax] = solve_interval(constraints,flux);
[__, __ , diagnostic]=solve_interval(__, __, options);
```

Description

[vmin, vmax]=solve_interval(constraints, flux, solver_options) returns the column vectors *vmin* and *vmax*, which define the interval estimates for the fluxes that have been asked. Regarding the inputs, the *constraints* are a YALMIP struct with a set of constraints for interval MFA problem. *Flux* is a vector with the indexes of the fluxes to be estimated. *options* is an optional input that allows to specify the YALMIP solver options (use 'help yalmip' for details).

GRAPHIC USER INTERFACE

New user of Possibilistic MFA find difficult to represent the measurements in possibilistic terms. In order to facilitate this task, the PFA toolbox includes a Graphic User Interface (GUI) that allows to easily representing the measurements.

Initialize

To initiate the user interface, write in the *command window* of MATLAB:

```
>>guiPossMeasurements
```

The GUI contains four panels: one to create new or upload a set of flux measurements, a second one to add and remove measurements to the set, a third one to edit each single flux, and the last one to save the work.

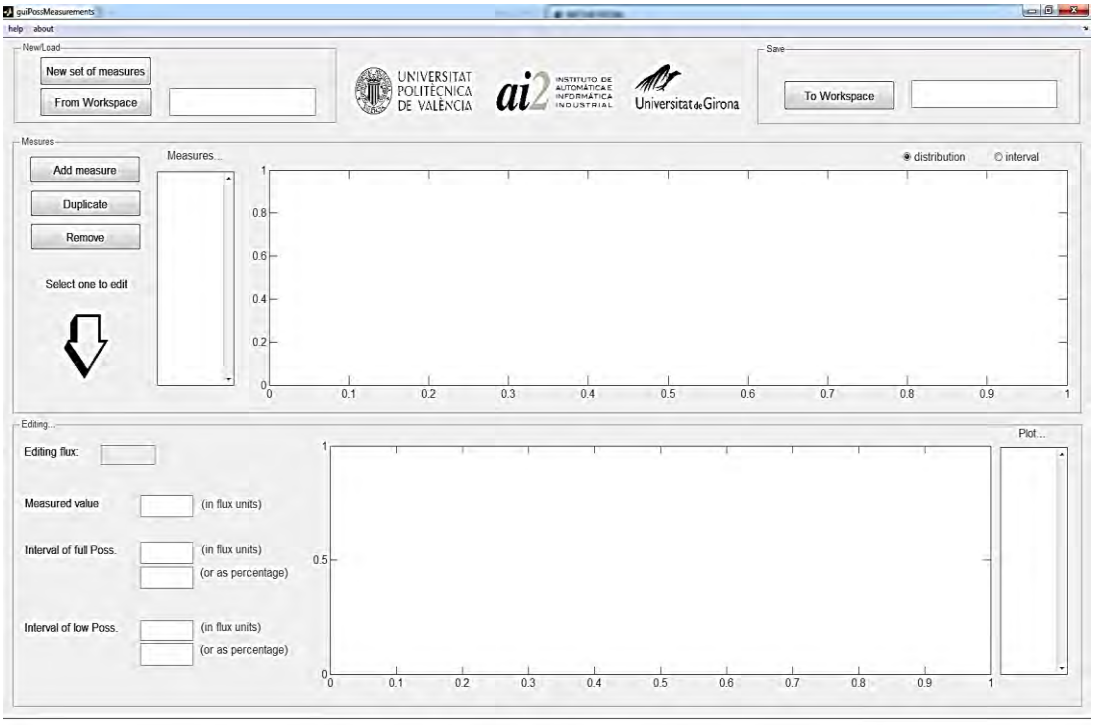


Figure 11. User interface of PFA toolbox for represent possibilistic measures.

(1) New/load: create or upload a set of measures

The initial step to use the GUI is to generate or upload a set of measurements. To create a new set of measures just click in the button *«new set of measures»*. Alternatively, you can import a set of measurements from the workspace to continue a previous work. In this case, write the name of the variable of the previously saved data and click *«from workspace»*.

(2) Measures: Addition, duplicate or remove measures

Once a set of measurements has been create or uploaded, each measure will be listed and ploted in the box below. Here you can click button to add, remove or duplicate the selected flux. In the axes on the right, all measurements will be plotted simultaneously. You can chose if measurements are plotted as possibilistic distributions (more detailed) or intervals (more compact).

(3) Editing flux measurements

To edit a flux measurement, first you must select it in the list of measurements. Then you can change the measured value and its ucertainty (by means of the intervals of flux and half possibility). To do this, simply modify the values in the corresponding boxes.

The measurements uncertainty can be defined both in absolute (flux units) and relative terms (as a percentage of the measured value).

The axes on the right allow you to visualize the measurments being edited and any other of your interest (selecting them in the box on the right).

(4) Save the measurements

After adding all the measurements and defining their uncertainty, the GUI can generate a matlab struct with the results. This struct can then be used with PFA Toolbox.

To save this structure to the MATLAB workspace, just write a name for the variable and click «*to workspace»*. A struct will be saved to the workspace with the given name and three fields, `wm` (with the measured values), `intFP` and `intLP` (with the intervals defining the uncertainty of each measurement). This structure is the input for *define_PossMeasurements*, one of the functions to perform Possibilistic MFA.

Note: If you want to save the structure into a file, use the standard MATLAB command, save.

References

The following are the references cited in the text, within others useful readings.

About constraint-based models

Llaneras, F., Picó, J. (2008). Stoichiometric modelling of cell metabolism.*Journal of Bioscience and Bioengineering*, *105*(1), 1-11.

Bernhard, P. (2006). Systems biology: properties of reconstructed networks.

About COBRA Toolbox

Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. Ø., Herrgard, M. J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols*, 2(3), 727-738.

Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M.,Palsson, B. Ø. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. *Nature protocols*, 6(9), 1290-1307.

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., Nielsen, J. (2013). The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. PLoS Comput Biol, 9(3), e1002980.

Interval MFA methodology

Llaneras, F., Picó, J. (2007). An interval approach for dealing with flux distributions and elementary modes activity patterns. Journal of theoretical biology, 246(2), 290-308.

Llaneras F, Bastin G, Picó J (2007a). On metabolic flux analysis when measurements are insufficient and/or uncertain. IAP Dysco workshop.

Llaneras, F. (2011). Interval and possibilistic methods for constraint-based metabolic models. PhD Thesis. Universidad Politécnica de Valencia.

Interval MFA Applications

Llaneras, F., Picó, J. (2007b). A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient. BMC bioinformatics, 8(1), 421.

Iyer, V. V., Yang, H., Ierapetritou, M. G., Roth, C. M. (2010). Effects of glucose and insulin on HepG2-C3A cell metabolism. Biotechnology and bioengineering, 107(2), 347-356.

D'Huys, P. J., Lule, I., Van Hove, S., Vercammen, D., Anné, J., Bernaerts, K.,Van Impe, J. (2010). Flux balance analysis and flux spectrum analysis for Streptomyces lividans batch fermentations. Book of Abstracts-Study Day Interuniversity Attraction Pole IAP VI/4.

Zamorano, F., Wouwer, A. V.,Bastin, G. (2010). A detailed metabolic flux analysis of an underdetermined network of CHO cells. Journal of biotechnology, 150(4), 497-508.

Hoppe, A., Hoffmann, S., Gerasch, A., Gille, C., Holzhütter, H. G. (2011). FASIMU: flexible software for flux-balance computation series in large metabolic networks. BMC bioinformatics, 12(1), 28.

Lohr, V., Hädicke, O., Genzel, Y., Jordan, I., Büntemeyer, H., Klamt, S., Reichl, U. (2014). The avian cell line AGE1. CR. pIX characterized by metabolic flux analysis. BMC biotechnology, 14(1), 72.

Possibilistic MFA methodology

Llaneras, F., Sala, A., Picó, J. (2009). A possibilistic framework for constraint-based metabolic flux analysis. BMC systems biology, 3(1), 79.

Llaneras, F. (2011). Interval and possibilistic methods for constraint-based metabolic models. PhD Thesis. Universidad Politécnica de Valencia.

Sala, A., Albertos, P. (2001). Inference error minimization: fuzzy modelling of ambiguous functions. Fuzzy Sets and Systems, 121(1), 95-111.

Sala, A. (2008). Encoding fuzzy possibilistic diagnostics as a constrained optimization problem. Information Sciences, 178(22), 4246-4263.

Dubois, D., Fargier, H.,Prade, H. (1996). Possibility theory in constraint satisfaction problems: Handling priority, preference and uncertainty. Applied Intelligence, 6(4), 287-309.

Possibilistic MFA applications

Llaneras, F., Sala, A., Picó, J. (2009). Applications of possibilistic reasoning to intelligent system monitoring: a case study. In Control Applications, (CCA) & Intelligent Control, (ISIC), IEEE. 171-176

Tortajada, M., Llaneras, F., Picó, J. (2010). Possibilistic Validation of a Constraint-Based Model under Data Scarcity: Application to Pichia Pastoris Cultures. In Computer Applications in Biotechnology (Vol. 11, No. 1, pp. 19-23).

Folch-Fortuny A, Tortajada M, Prats-Montalbán JM, Llaneras F, Picó J, Ferrer A (2014). MCR-ALS on metabolic networks: Obtaining more meaningful pathways. Chemometrics and Intelligent Laboratory Systems.

Folch-Fortuny, A., Tortajada, M., Prats-Montalbán, J. M., Llaneras, F., Picó, J., Ferrer, A. (2015). MCR-ALS on metabolic networks: Obtaining more meaningful pathways. Chemometrics and Intelligent Laboratory Systems, 142, 293-303.

Morales, Y., Tortajada, M., Picó, J., Vehí, J., Llaneras, F. (2014). Validation of an FBA model for Pichia pastoris in chemostat cultures. BMC systems biology, 8(1), 142.

About Traditional MFA methods

Klapa, M. I., Stephanopoulos, G. (2000). Metabolic flux analysis. In Bioreaction Engineering (pp. 106-124). Springer Berlin Heidelberg.

Klamt, S., Schuster, S., Gilles, E. D. (2002). Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. Biotechnology and Bioengineering, 77(7), 734-751.

(Chapter II) Llaneras, F. (2011). Interval and possibilistic methods for constraint-based metabolic models. PhD Thesis. Universidad Politécnica de Valencia.

Stephanopoulos, G. N, Aristidou, A. A. (1998). Metabolic Engineering: Principles and Methodologies. San Diego, USA: Academic Press.

Heijden, R. T., Romein, B., Heijnen, J. J., Hellinga, C., Luyben, K. C. (1994). Linear Constraint Relations in Biochemical Reaction Systems: I & II. Biotechnology & Bioengineering, 43(1):3–10.

About YALMIP and solvers

Löfberg, J. (2004, September). YALMIP: A toolbox for modeling and optimization in MATLAB. In Computer Aided Control Systems Design, 2004 IEEE International Symposium on (pp. 284-289). IEEE.

Löfberg, J. (2008, July). Modeling and solving uncertain optimization problems in YALMIP. In Proceedings of the 17th IFAC World Congress (pp. 1337-1341).

IBM ILOG CPLEX- High-performance mathematical programming engine. [http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/]

GLPK (GNU Linear programing kit) [http://www.gnu.org/software/glpk/]

Yeimy Liceth Morales Pérez