Metabolic Engineering 25 (2014) 159–173

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/ymben

# Construction of robust dynamic genome-scale metabolic model structures of *Saccharomyces cerevisiae* through iterative re-parameterization

Benjamín J. Sánchez, José R. Pérez-Correa*, Eduardo Agosin*

*Department of Chemical and Bioprocess Engineering, School of Engineering, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Santiago, Chile*

ARTICLE INFO

*Article history:*
Received 7 January 2014
Received in revised form
28 May 2014
Accepted 10 July 2014
Available online 19 July 2014

*Keywords:*
Genome-scale metabolic models
Dynamic flux balance analysis
Metaheuristic optimization
Yeast
Parameter estimation
Sensitivity analysis

ABSTRACT

Dynamic flux balance analysis (dFBA) has been widely employed in metabolic engineering to predict the effect of genetic modifications and environmental conditions in the cell's metabolism during dynamic cultures. However, the importance of the model parameters used in these methodologies has not been properly addressed. Here, we present a novel and simple procedure to identify dFBA parameters that are relevant for model calibration. The procedure uses metaheuristic optimization and pre/post-regression diagnostics, fixing iteratively the model parameters that do not have a significant role. We evaluated this protocol in a *Saccharomyces cerevisiae* dFBA framework calibrated for aerobic fed-batch and anaerobic batch cultivations. The model structures achieved have only significant, sensitive and uncorrelated parameters and are able to calibrate different experimental data. We show that consumption, suboptimal growth and production rates are more useful for calibrating dynamic *S. cerevisiae* metabolic models than Boolean gene expression rules, biomass requirements and ATP maintenance.



## 1. Introduction

Mathematical modeling is a fundamental tool for metabolic engineering and the biotechnology industry since it overcomes the need for excessive experiments to validate biological hypotheses (Kitano, 2002). Among the different modeling tools, flux balance analysis (FBA) has been widely employed (Park et al., 2009). Using mass balances, under pseudo-steady state assumption and with an underlying objective function (Orth et al., 2010), FBA can predict the behavior of the whole cell metabolism. Since its first validation (Varma and Palsson, 1994), numerous efforts have been undertaken for improving its predictive performance (Copeland et al., 2012).

FBA applications have increased considerably since the introduction of genome-scale metabolic models (GSMM) (Edwards and Palsson, 2000; Osterlund et al., 2012). A GSMM consists of a detailed metabolic network, including not only most metabolites and reactions of the studied organism, but also most metabolic associated genes (Henry et al., 2010; Thiele and Palsson, 2010). Hence, predictions can be computed to assess the metabolic impact of genetic modifications, in order to overproduce a certain compound of interest, for instance. In the case of *Saccharomyces cerevisiae* (budding yeast), Förster et al. (2003) proposed the first GSMM. Later, a consensus model was developed (Herrgård et al., 2008) and further expanded to improve biochemical coverage, connectivity and knockout predictability (Dobson et al., 2010; Heavner et al., 2013, 2012).

Although the aforementioned efforts have contributed to gain knowledge of yeast's steady-state metabolism and to make accurate predictions for the overexpression of high-value metabolites in mutant strains (Oberhardt et al., 2009), the kinetics and physiology of yeast cells can be better understood in a dynamic setting, with changing environmental conditions and variable cell-density, which are standard conditions of industrial processes (Gianchandani et al., 2010; Oddone et al., 2009).

### 1.1. Dynamic flux balance analysis

Dynamic FBA (dFBA) is an extension of FBA in which, under the pseudo-steady state premise for short time steps, the evolution of the extracellular metabolite concentrations during cultivation modifies the FBA problem constraints and, in turn, the FBA solution modifies the consumption/secretion rates for the bioreactor's dynamic equations. This technique, first proposed in *Escherichia coli* by Varma and Palsson (1994), was later improved by including kinetic constraints (Mahadevan et al., 2002) and

*Abbreviations:* CC, confidence coefficient; CCC, cross-calibration coefficient; CI, confidence interval; eSS, enhanced scatter search; FIM, Fisher information matrix

* Corresponding authors.
*E-mail addresses:* bjsanche@uc.cl (B.J. Sánchez), perez@ing.puc.cl (J.R. Pérez-Correa), agosin@ing.puc.cl (E. Agosin).

http://dx.doi.org/10.1016/j.ymben.2014.07.004
1096-7176/

transcriptional regulation (Covert et al., 2008; Tepeli and Hortaçsu, 2008).

For *S. cerevisiae*, the first application of dFBA was published in 2003 (Sainz et al., 2003) and later improved to account for sugar kinetics (Hjersted and Henson, 2006; Pizarro et al., 2007) and expanded to genome-scale (Hjersted et al., 2007, Vargas et al., 2011). More recently, the shift from aerobic respiration to anaerobic fermentation was also analyzed using a genome-scale dFBA framework (Jouhten et al., 2012).

dFBA has also been applied to other microorganisms, such as *Lactoccocus lactis* (Oddone et al., 2009), *Shewanella oneidensis* (Feng et al., 2012), and to animal cell cultures, like chinese hamster ovary cells (Nolan and Lee, 2011; Provost and Bastin, 2004; Provost et al., 2006) and murine hybridoma cells (Gao et al., 2008). Moreover, co-cultivations of *E. coli* and *S. cerevisiae* (Hanly and Henson, 2011; Hanly et al., 2012; Höffner et al., 2013; Lisha and Sarkar, 2014), *S. cerevisiae* and *Scheffersomyces stipitis* (Hanly and Henson, 2013) and competition between *Geobacter sulfurreducens* and *Rhodoferax ferrireducens* (Zhuang et al., 2011) have been evaluated too.

For all the aforementioned studies, the use of several parameters, including kinetics of nutrients consumption, production rates, biomass prerequisites and ATP maintenance, is customary. Selection of the parameter's values is generally performed by manual fitting (trial and error) (Hanly and Henson, 2011; Meadows et al., 2010), parameter estimation (Nolan and Lee, 2011; Pizarro et al., 2007), or extracted from the literature (Hjersted and Henson, 2006). If the model fits the data well enough, it is considered satisfactory.

Although models with many parameters can fit experimental data accurately, they usually present problems such as lack of parameter significance (i.e. the confidence interval for the estimated parameter is larger than the estimated value itself) (Fig. 1A), low parametric sensitivity (i.e. strong variations of a parameter value results in small variations of the model output) (Smith and Missen, 2003) (Fig. 1B), non identifiability (i.e. high correlation between two parameters) (Jacquez and Greif, 1985) (Fig. 1C) and overfitting (i.e. more parameters than necessary to explain a particular behavior) (Fig. 1D). As a consequence, unrealistic model predictions arise when the model is employed under different experimental conditions, or when the calibration data has considerable noise, because multiple parameter value's combinations yield the same simulation result (Kravaris et al., 2013).

The above-mentioned problems are of special interest in dFBA modeling, in which several parameters are employed to calibrate scarce experimental data, and the simulation results are highly sensitive to the parameter values. Therefore, an important need in this field is to improve the reliability of dFBA approaches by developing methods for determining parameters that are useful for genome-scale dynamic simulations, as proposed by Park et al. (2009). These methods should include robust parameter estimation algorithms and a collection of diagnostic tools to detect insensitive and correlated fitting parameters in order to avoid over-parameterized models.

### 1.2. Iterative reparameterization

In every model, the effect of each parameter should be tested, in order to only use for estimation a certain group of parameters (estimated parameters) and have fixed the rest of them (fixed parameters) (Alberton et al., 2013). The goal for robust modeling therefore is to formulate adequate model structures (i.e., combinations of fixed and estimated parameters) that not only fit well the data, but also yield accurate predictions within a wide range of operating conditions and experimental data (Balsa-Canto et al., 2010; Chu and Hahn, 2008; Kravaris et al., 2013). This concern has not been appropriately addressed in FBA or dFBA modeling.

Sensitivity tests have been carried out to validate the usefulness of some parameters in both FBA (Nookaew et al., 2008; Varma and Palsson, 1993) and dFBA (Hjersted and Henson, 2009; Mahadevan et al., 2002; Nolan and Lee, 2011) approaches. However, results were dissimilar, most analyses did not consider identifiability and significance tests, and were performed for some isolated parameters, instead of comparing all parameters between them. Pre/post-regression diagnostics (Jaqaman and Danuser, 2006) were developed to determine which parameters cause one – or several – of the above-mentioned problems (unidentifiability, low sensitivity and no significance). Some of these parameters should afterwards be fixed (i.e., not used for calibration), and the process iterated until the estimation can be run only with the most relevant parameters.

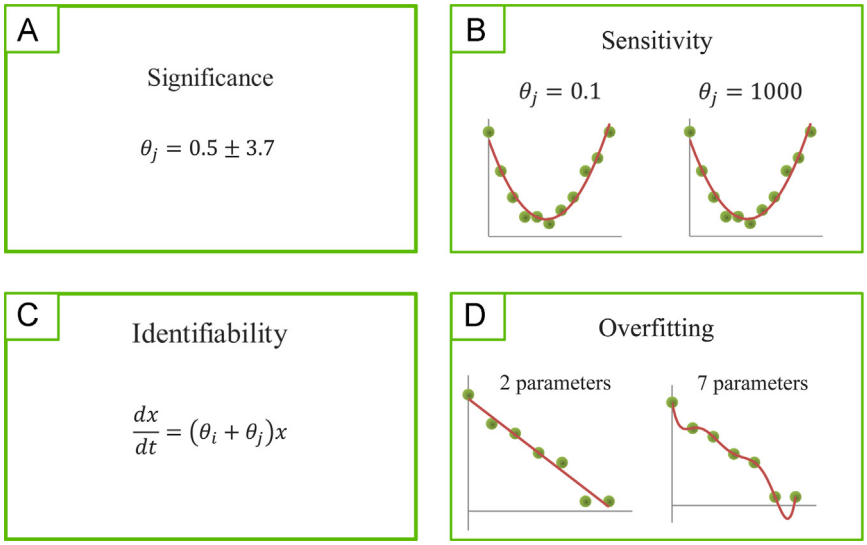


**Fig. 1.** Typical fitting problems that arise when a model has too many parameters. (A) An example of a parameter that is not significantly estimated, because the confidence interval includes the zero. (B) An example of a non-sensitive parameter, because different values yield the same result in the associated variable. (C) An example of two parameters not identifiable, because different value combinations will result in the same output. (D) An example of parameter overfitting, in which an excessive number of parameters are proposed to explain the data.

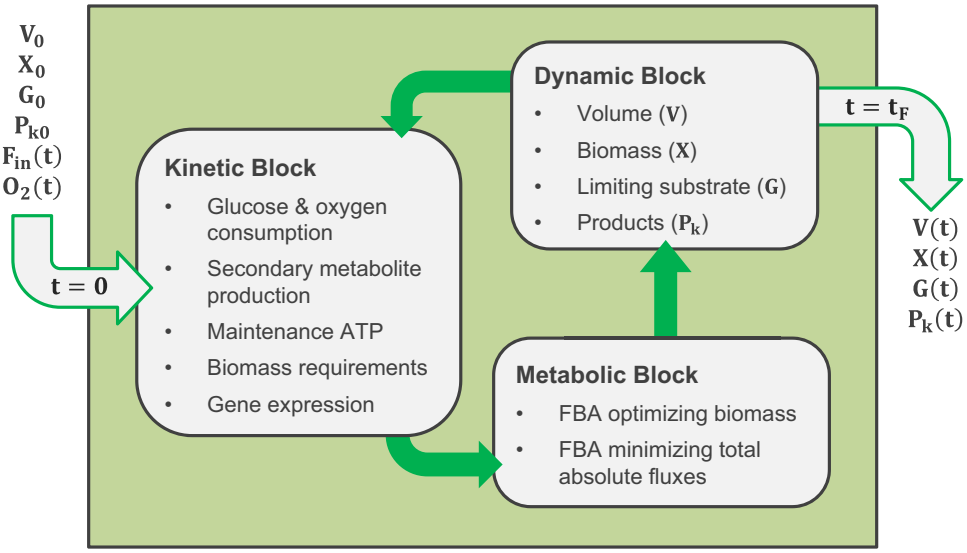


**Fig. 2.** General scheme of the dFBA framework. $V$ is volume [L], $X$ is biomass concentration [g/L], $G$ is extracellular limiting substrate (glucose) concentration [g/L], $P_k$ accounts for the different extracellular product concentrations [g/L], $F_{in}(t)$ is the feed function for the fed-batch cases [L/h], $O_2(t)$ is the predefined oxygen presence or absence, $t$ is time [h] and $t_F$ is the fermentation duration [h].

In kinetic modeling, several iterative methodologies have been proposed to find groups of estimated parameters and fixed parameters for a given model (Alberton et al., 2013; Balsa-Canto et al., 2010; Kravaris et al., 2013). However, many of these approaches require extensive symbolic manipulation, high computational power and/or deep knowledge of the dynamic model behavior in different conditions. Furthermore, none of the mentioned approaches have been designed for a dFBA scheme, in which although there is a reasonable number of parameters (15–20) and state variables (5–10), optimization problems of over 2000 variables and 1000 equations are embedded in the dynamic equations. This results in highly demanding computational times for integration, and unattainable closed-form expressions. Hence, a simple iterative methodology which can be easily integrated to any kinetic model, and in particular to any dFBA platform, is still needed.

In this work, we formulated, calibrated with experimental data and reparameterized a *S. cerevisiae* genome-scale dFBA framework, to attain model structures that have sensitive, uncorrelated and significant parameters, and at the same time are able to fit a broad range of experimental conditions. This was done here for the first time in dFBA modeling (until now the calibration results were considered sufficient if a good fit was obtained). Because we used data from both, aerobic fed-batch and anaerobic batch cultivations, the most common industrial fermentation processes, we propose two different model structures (one for each type of cultivation). To the best of our knowledge, this is the first time a model derived from a dFBA approach is calibrated for *S. cerevisiae* aerobic fed-batch cultivations. Furthermore, for reparameterization, we present a novel methodology that employs meta-heuristic optimization and pre/post-regression analysis. Fixing one parameter at a time, the procedure explores different parameter combinations and uses heuristic criteria in order to attain the mentioned model structures avoiding excessive computational time. The procedure has the following advantages: (i) it is customized for dFBA schemes, (ii) it is simple and easy to code, (iii) it does not require symbolic manipulation, (iv) it does not use too much computational time, and (v) it generates several alternative model structures, which are cross-calibrated in order to choose a solution that proves to be general enough to calibrate numerous experimental conditions.

## 2. Modeling

### 2.1. Model formulation

The dFBA framework was formulated following standard procedures (Hjersted et al., 2007; Meadows et al., 2010; Vargas et al., 2011). It was based on a pseudo-steady state assumption (Stephanopoulos et al., 1998), i.e. considering that intracellular kinetics are several orders of magnitude faster than extracellular kinetics and, therefore, the former can be disregarded if the FBA model is resolved iteratively in short integration periods.

Our model was designed as three linked blocks that are solved iteratively (Fig. 2). The inputs are the initial values for each of the state variables (volume, biomass and extracellular substrate and products), the feed function (for fed-batch cases) and the presence or absence of $O_2$ along the fermentation. This information is fed to the kinetic block, which defines the FBA optimization constraints, such as glucose uptake rate, ATP maintenance, stoichiometric requirements for biomass formation, thresholds for gene expression, byproduct production rates, etc. Then, two optimization problems are solved in the metabolic block. First, maximization of cell growth as a linear programming (LP) problem; and then, minimization of absolute flux sum as a quadratic programming (QP) problem, on a sub-optimal specific growth rate that must be fitted (for more details see Section 2.2). Next, the consumption and production rates from the solution of the FBA are transferred to the dynamic block, which integrates a set of ordinary differential equations to update the concentration state variables. Then, the kinetic block can be solved again, iterating the 3-block cycle until a predefined simulation time is achieved, or the integration turns out to be unfeasible.

The model was coded in MATLAB® 2013a (MATLAB, 2013) and implemented in 3 different machines: a Windows 7 PC with a 3.3 GHz AMD FX™ 6100 (six-core) processor, a Windows 7 PC with a 3.1 GHz Intel® Core™ 2 Duo (two-core) processor and a Linux CentOS 6.4 with a 2.3 GHz Intel® Xeon® L5640 (six-core) processor. The average computation time for a typical fermentation varied between 7 and 31 [s], depending on the fermentation characteristics and the machine used (Table S1). In the following, further details of each block and all associated equations and parameters are reported.

### 2.2. Metabolic block

FBA (Orth et al., 2010; Varma and Palsson, 1994) is based on mass balances. For $n$ reactions and $m$ metabolites a $m \times n$ stoichiometric matrix $\mathbf{S}$ can be formulated and, if we neglect the accumulation of metabolites, a mass balance for all metabolites is

$$\mathbf{S} \cdot \mathbf{v} = 0 \tag{1}$$

where $\mathbf{v}$ is the flux distribution vector [mmol/g DWh]. Additionally, lower and upper bounds for each flux can be included, based on the reversibility of each reaction, along with an objective function to be minimized or maximized, given that the problem is highly underdetermined (i.e. there are much more reactions than metabolites). As previously mentioned, in our approach the metabolic block consists of two sequential optimizations: first, a LP problem is solved by maximizing the specific growth rate (Curran et al., 2012; Hjersted and Henson, 2006; Varma and Palsson, 1994):

$$\begin{aligned} &\text{Max } \mu \\ &s.t. \quad \mathbf{S} \cdot \mathbf{v} = 0 \quad \text{(Problem 1)} \\ &\text{LB} \le \mathbf{v} \le \text{UB} \end{aligned}$$

where $\mu$ is the specific growth rate [1/h], and LB and UB are the lower and upper bounds, respectively [mmol/g DWh]. Then, a QP problem is applied to minimize the total sum of absolute fluxes based on the principle of enzyme efficiency maximization (Feng et al., 2012; Holzhütter, 2004; Schuetz et al., 2012), fixing the growth rate at a sub-optimal level

$$\begin{aligned} &\text{Min} \sum_i v_i^2 \\ &s.t. \quad \mathbf{S} \cdot \mathbf{v} = 0 \quad \text{(Problem 2)} \\ &\mu = \alpha \cdot \mu^* \\ &\text{LB} \le \mathbf{v} \le \text{UB} \end{aligned}$$

where $\mu^*$ is the value of the optimal specific growth rate obtained in Problem 1 [1/h], and $\alpha$ is a parameter that varies between 0 and 1 and is used for model calibration (see Section 2.5). For aerobic cultures, this parameter has one value ($\alpha$) during the batch stage of the cultivation and another one ($\alpha_F$) during the fed-batch stage. This is because the fermentation conditions (including the growth rate and the substrate to product yields) change dramatically upon glucose starvation.

The genome-scaled metabolic model used was a version of the consensus model of *S. cerevisiae*, *Yeast 5* (Heavner et al., 2012). We also tried *Yeast 6* (Heavner et al., 2013), the most recent version of the consensus model. However, calibrations with our experimental data showed better agreement with *Yeast 5* (Table S2). All FBA problems were solved using the Constraint-Based Reconstruction and Analysis (COBRA) toolbox (Becker et al., 2007; Schellenberger et al., 2011), which employs the programming library libSBML (Bornstein et al., 2008) and the SBML toolbox (Keating et al., 2006). Gurobi® 5.5 (Gurobi, 2013) was chosen as the optimization solver since preliminary performance tests showed it to be 4 times faster than the default solver. Finally, Gurobi Mex (Yin, 2011) was used as a Matlab-interface for calling Gurobi.

### 2.3. Dynamic block

The dynamic block consists of a set of ordinary differential equations (ODEs) that account for volume change, cell growth and metabolite consumption/production, in a batch or fed-batch culture. Considering accumulation, consumption and production, dynamic balances yield the following equations:

$$\frac{dV}{dt} = F(t) \tag{2}$$

$$\frac{d(VX)}{dt} = \mu \cdot (VX) \tag{3}$$

$$\frac{d(VG)}{dt} = F(t) \cdot G^F - v_G \cdot \text{MW}_G \cdot (VX) \tag{4}$$

$$\frac{d(VP_k)}{dt} = v_{P_k} \cdot \text{MW}_{P_k} \cdot (VX) \tag{5}$$

where $V$ is volume [L], $t$ is time [h], $F(t)$ is the feed rate [L/h] (zero for batch cases and exponential for fed-batch cases), $X$ is the biomass concentration [g/L], $\mu$ is the specific growth rate [1/h] (obtained from Problem 2 in the metabolic block), $G$ is the extracellular limiting substrate concentration [g/L], $G^F$ is the feed's substrate concentration [g/L], $P_k$ is the $k$th extracellular product concentration [g/L], $v$ is the corresponding flux exchange rate [mmol/g DWh] (obtained from Problem 2 in the metabolic block), and MW accounts for the corresponding molecular weight [g/mmol]. All fermentations were carbon-limited, with glucose as the limiting substrate, and ethanol, glycerol, citrate and lactate as the most relevant products. Therefore, Eq. (5) comprises 4 differential equations.

The integration solvers were chosen based on preliminary performance tests on all Matlab standard integrators; the best results were achieved with *ode113* for batch cultures and *ode15s* for fed-batch ones. Relative and absolute tolerances of $1e-3$ were small enough to obtain good fittings in reasonable computation times. A maximum integration step size of 0.7 h was defined, in order to avoid losing information in critical moments such as glucose starvation. Finally, all variables were forced to be non-negative.

### 2.4. Kinetic block

#### 2.4.1. Glucose consumption

The kinetic block includes glucose consumption rate ($v_G$; reaction r_1714 in *Yeast 5*), defined as a fixed flux (LB=UB) using a Michaelis–Menten kinetic with an additional term to account for ethanol inhibition (Hjersted et al., 2007; Pizarro et al., 2007)

$$v_G = \frac{v_{Gmax} G}{K_G + G} \cdot \frac{1}{1 + \frac{E}{K_E}} \tag{6}$$

where $G$ and $E$ are the glucose and ethanol concentration [g/L] respectively, $v_{Gmax}$ is the maximum glucose uptake rate [mmol/g DWh], $K_G$ is the half-saturation constant [g/L], and $K_E$ is the ethanol inhibition constant [g/L].

#### 2.4.2. Oxygen consumption

In aerobic fermentations, oxygen uptake rate ($v_{O_2}$; reaction r_1992 in *Yeast 5*) was left unconstrained (LB = −1000 [mmol/g DWh]) during the whole fermentation, because the dissolved oxygen (DO) control guaranteed enough oxygen at any time of the fermentation (Cárcamo et al., 2013). For anaerobic experiments instead, we proceeded as suggested by Heavner et al. (2012), constraining $v_{O_2}$ to zero (LB = UB = 0 [mmol/g DWh]), allowing unrestricted uptake of ergosterol (r_1757), lanosterol (r_1915), zymosterol (r_2106) and phosphatidate (r_2009), and eliminating 14-demethyllanosterol and ergosta-5,7,22,24(28)-tetraen-3beta-ol from the lipid pseudo-reaction (in the model this was achieved simply by blocking reaction r_2108 and unblocking reaction r_2109).

#### 2.4.3. Byproduct synthesis

Byproduct production was also accounted for by forcing the lower bounds of the product exchange reactions to be proportional to glucose consumption. As previously mentioned, the only 4 extracellular products detected in significant amounts were

ethanol, glycerol, citrate and lactate; therefore, we included 4 more parameters to calibrate the batch fermentations:

$$LB_i = f_i \cdot v_G\,; \quad i = 1\ldots4 \tag{7}$$

where $i$ accounts for (1) ethanol, (2) glycerol, (3) citrate and (4) lactate. $LB_i$ is the corresponding exchange reaction lower bound (reactions r_1761, r_1808, r_1687 and r_1546 in *Yeast 5*, respectively), and $f_i$ is the corresponding parameter for calibration. This approach fixes the lower bound, not the flux value; hence, if the approach was not useful for a given calibration, the lower bound would be estimated as zero by the parameter estimation routine, leaving the corresponding flux with no constraints. However, the inclusion of these parameters improved significantly the model fitting (data not shown). For the aerobic cultivations, these parameters were also allowed to change their value in the fed-batch stage:

$$LB_i = v_i; \quad i = 1\ldots4 \tag{8}$$

where $v_i$ are the corresponding extra parameters. This assumption allowed the aerobic simulations to take different yields in batch and fed-batch stages and, moreover, allowed product consumption during glucose-starvation culture conditions, a phenomenon that was observed in some conditions for ethanol and glycerol.

#### 2.4.4. ATP maintenance

ATP maintenance ($m_{ATP}$; [mmol/g DWh]) was defined as a minimum consumption flux of cytosolic ATP. Consequently, an additional exchange reaction was added to the metabolic model, which had the following lower bound:

$$LB_{ATP[cyt]_{out}} = m_{ATP} \tag{9}$$

where $LB_{ATP[cyt]_{out}}$ is the lower bound of the exchange rate of cytosolic ATP, and $m_{ATP}$ is a calibrated parameter. $m_{ATP}$ accounts for all cellular processes and functions not related to cell growth, i.e. it is the minimum non-growth associated maintenance (NGAM).

#### 2.4.5. Biomass requirements

Regarding the biomass pseudo-reaction, we considered 3 parameters for calibration that weighted each group of the major cellular requirements differently: a for aminoacids, c for carbohydrates and l for lipids:

$$a(\alpha_1 A_1 + \ldots + \alpha_{20} A_{20}) + c(\beta_1 C_1 + \ldots + \beta_4 C_4) + l(\gamma L) + \ldots \rightarrow \text{Biomass} + \ldots \tag{10}$$

with $A_1 \ldots A_{20}$ being the 20 essential aminoacids; $C_1 \ldots C_4$, the 4 sugars that are part of the biomass pseudo-reaction equation in *Yeast 5* ((1->3)-β-D-glucan, glycogen, trehalose and mannan); $L$, the lipid fraction (obtained from the lipid pseudo-reaction); and $\alpha_1 \ldots \alpha_{20}$, $\beta_1 \ldots \beta_4$ and $\gamma$, the corresponding stoichiometric coefficients. An important advantage of using this approach for studying sensitivity is that it avoids including the impact of each compound separately (i.e. every aminoacid, every sugar and every lipid), but instead the impact of each group is assessed. Therefore, one can easily conclude that if one of the 3 parameters does not significantly affect the model's output, any compound requirement of the corresponding group will neither affect it.

#### 2.4.6. Gene expression

Two additional parameters were included to account for transcriptomic (gene expression) information. Eighteen normalized microarray expression experiments (6 for aerobic and 12 for anaerobic cultivations) of different yeast strains growing on glucose were obtained (a) from a published work (Lai et al., 2006) using the query-driven search tool SPELL (Serial Pattern of Expression Levels Locator) (Hibbs et al., 2007) from the *Saccharomyces* Genome Database (SGD) (Cherry et al., 2012), and (b) from two works from our group (Aceituno et al., 2012; Orellana et al., 2014). All data was preprocessed, disregarding genes that were not included in *Yeast 5* or that were essential for biomass growth, using single-knockout gene analysis (Edwards and Palsson, 2000; Zomorrodi et al., 2012).

The methodology followed for turning off genes that were not significantly expressed was based on a published method (Åkesson et al., 2004). First, the group of aerobic or anaerobic microarrays was chosen depending on the cultivation characteristics. Afterwards, the binary variable $Y_{ij}$ was calculated for each gene $i$ and microarray $j$

$$Y_{ij} \begin{cases} 1; & \exp_{ij} \le \mu_j - t_1 \cdot \sigma_j \\ 0; & \text{otherwise} \end{cases} \tag{11}$$

where $\exp_{ij}$ is the normalized expression level of gene $i$ in microarray $j$, $\mu_j$ is the average expression in microarray $j$, $\sigma_j$ is the corresponding standard deviation, and $t_1$ is a parameter that represents an expression threshold (which will be estimated later). Once all variables $Y_{ij}$ were computed, a decision was made for each gene, according to the following rule:

$$\sum_{j=1}^{M} Y_{ij} \ge \frac{M \cdot t_2}{100} \rightarrow \text{GEN}_i \text{ off} \tag{12}$$

where $M$ is the number of microarrays used, and $t_2$ is another parameter representing a consistency threshold (a minimum percentage of the microarrays). Therefore, the expression rule can be summed up as: "the $i$th gene will be considered unexpressed if in at least $t_2$% of the microarrays the expression is $t_1$ standard deviations below the microarray average expression". If this is true for any gene, then all reactions that are catalyzed by enzymes coded by the corresponding genes will be forced to carry zero flux (LB = UB = 0 [mmol/g DWh]).

This approach assumes that each gene will be expressed (or unexpressed) along the whole cultivation. To partially overcome this limitation, we used the information of several microarrays of different yeast strains growing under different conditions; hence, it is likely that if a gene is not expressed under most of the conditions considered, it will not be expressed during the whole fermentation (aerobic or anaerobic). Also, even though gene expression is usually strain-dependent, our approach rests on the premise that if in several gene expression experiments of different yeast strains a particular gene is not expressed over a certain threshold, that gene will probably be unexpressed in most strains for that particular condition.

### 2.5. Parameter estimation

For model calibration, we formulated a nonlinear programming problem with the dFBA embedded as a constraint, and an objective function $F$ consisting of a sum of square errors between the experimental data and the simulation output, weighted by the maximum corresponding measured variable and by the number of measurements of the respective variable

$$F = \min_{\theta} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left( \frac{X_{ij}^{\text{mod}} - X_{ij}^{\text{exp}}}{n_i \times \max_j (X_{ij}^{\text{exp}})} \right)^2 \tag{13}$$

with $\theta$ representing the parameter space, $m$ the number of measured variables, $n_i$ the number of measurements for the $i$th variable, $X_{ij}^{\text{mod}}$ the dFBA output for the variable $i$ and the measurement $j$, $X_{ij}^{\text{exp}}$ the corresponding experimental value, and $\max_j(X_{ij}^{exp})$ the maximum value measured for variable $i$. With this approach, all data gets a similar weight, regardless of the number of

**Table 1**
Parameter estimation details. The symbol, name, unit, initial value, lower and upper bound for each parameter of the data fitting are shown.

| Symbol | Name | Units | LB | Initial value | UB |
|---|---|---|---|---|---|
| $v_{Gmax}$ | Maximum glucose uptake rate | mmol/g DWh | 1 | 10 | 50 |
| $K_G$ | Half-saturation constant | g/L | 0.005 | 0.05 | 10 |
| $K_E$ | Ethanol inhibition constant | g/L | 0.005 | 20 | 50 |
| $\alpha$ | Sub-optimal growth | – | 0.1 | 0.7 | 1 |
| $a$ | Aminoacid requirement | – | 0.5 | 1 | 3 |
| $c$ | Carbohydrate requirement | – | 0.5 | 1 | 3 |
| $l$ | Lipid requirement | – | 0.5 | 1 | 3 |
| $m_{ATP}$ | Maintenance ATP | mmol/g DWh | 0 | 1 | 5 |
| $t_1$ | Expression threshold 1 | – | 0 | 6 | 6 |
| $t_2$ | Expression threshold 2 | % | 75 | 100 | 100 |
| $f_E$ | Glucose–ethanol minimum yield | mmol/mmol | 0 | 1.7 | 2 |
| $f_{GL}$ | Glucose–glycerol minimum yield | mmol/mmol | 0 | 0 | 1 |
| $f_C$ | Glucose–citric acid minimum yield | mmol/mmol | 0 | 0 | 0.258 |
| $f_L$ | Glucose–lactic acid minimum yield | mmol/mmol | 0 | 0 | 1 |
| $\alpha_F$ | Sub-optimal growth (fed-batch) | – | 0.1 | 1 | 1 |
| $v_E$ | Ethanol minimum rate (fed-batch) | mmol/g DWh | −10 | 0 | 10 |
| $v_{GL}$ | Glycerol minimum rate (fed-batch) | mmol/g DWh | −10 | 0 | 0 |
| $v_C$ | Citric acid minimum rate (fed-batch) | mmol/g DWh | 0 | 0 | 10 |
| $v_L$ | Lactic acid minimum rate (fed-batch) | mmol/g DWh | 0 | 0 | 10 |

measurements or the difference of magnitudes between variables. Experimental values were measured as indicated in Section 3.

All the parameters studied, along with their units, lower and upper bounds and initial values for all optimizations are summarized in Table 1. The LB and UB of $v_{Gmax}$, $K_G$, $K_E$ and $m_{ATP}$ were chosen according to the literature (Hjersted et al., 2007; Varma and Palsson, 1994). The upper bounds of $f_E$, $f_{GL}$, $f_C$ and $f_L$ were the respective model's maximum yields from glucose. The LB of $t_2$ was set at 75%, thus forcing a minimum consistency in the microarray datasets (at least 75% of the datasets had to agree in order to delete a gene). The rest of the LB and UB were chosen to ensure that the algorithm had enough search space. Finally, initial values for parameter estimation were chosen to attain a feasible simulation.

Due to the problem complexity and the presence of multiple local minima, the parameter estimation was performed with the enhanced scatter search method eSS (Egea and Balsa-Canto, 2009), a metaheuristic global optimization algorithm for nonlinear programming problems that has been successfully used in bioprocessing (Balsa-Canto et al., 2007; Sacher et al., 2011; Sriram et al., 2012). Several options should be tuned when using eSS; in our case, the best fittings (lower objective function values in less CPU time) were achieved with a maximum of 3000 iterations and using the local optimization solver *n2fb* (Dennis et al., 1981). With these options, optimization times between 5 and 13 [h] were achieved, depending on the cultivation characteristics and the machine that ran the code (Table S1).

## 3. Materials and methods

### 3.1. Strains and culture conditions

Cultivations were carried out with 2 different strains, the industrial yeast strain *S. cerevisiae* N30 (Centrovet, Chile) and the wine yeast strain *S. cerevisiae* EC1118 (Lalvin, Switzerland). For each strain, we evaluated 4 different environmental conditions:

1. Aerobic fed-batch with slow feed rate.
2. Aerobic fed-batch with fast feed rate.
3. Anaerobic batch with low glucose initial concentration ($G_0$).
4. Anaerobic batch with large glucose initial concentration ($G_0$).

Since each experiment was carried out in duplicate, a total of 16 cultivations were performed (4 conditions × 2 strains × 2 duplicates). All media were glucose-limited and completely defined (Table S3). Conditions 1, 2 and 3 had $G_0=20$ [g/L], whereas condition 4 had a $G_0=80$ [g/L]. The glucose feed concentration for conditions 1 and 2 was 300 [g/L]. Finally, conditions 3 and 4 were also supplied with ergosterol and Tween 80, both necessary for growth under anaerobic conditions (Diderich et al., 1999), and diluted in a small amount of ethanol.

### 3.2. Experimental setup

The fermenters used for all cultivations consisted of in-house built 1 L bioreactors equipped with a condenser, a stirrer and two Rushton turbines, operating with brushless DC motors VEXTA®, AXH Series (Oriental Motor, Japan). For monitoring and control, a SIMATIC PCS7 distributed control system (Siemens, Germany) was used. Dissolved oxygen and temperature were measured with Oxymax COS22D probes (Endress Hauser, Switzerland), pH was detected with Tophit CPS471D probes (Endress Hauser, Switzerland), and off-gas composition ($CO_2$ and $O_2$) was sensed with a BlueInOne Cell gas analyzer (Bluesens, Germany). 210U and 102FS/R peristaltic pumps (Watson Marlow, USA) were used for acid, base, feed and antifoam addition; and FMA-A2407 gas flowmeters and controllers (Omega, USA) for air, $O_2$ and $N_2$ addition.

Each culture started from a 2 [mL] vial of the corresponding strain kept at −80 °C. A preculture was grown overnight at 30 °C in shake flasks with 100 [mL] of the batch medium, from which 50 [mL] for aerobic cultivations (or 60 [mL] for anaerobic cultivations) were inoculated in the 1 [L] fermenters containing 450 [mL] (or 540 [mL], respectively) of batch medium, with controlled conditions of 30 °C, pH = 5.0 and DO ≥ 2.8 [mg/L] for aerobic cultures, or DO = 0 [mg/L] for anaerobic cultures. Aerobiosis was achieved by a triple split-range control action including agitation (300–800 [RPM]), air flow (0.2–0.6 [L/min]) and pure oxygen flow (0–0.6 [L/min]) (Cárcamo et al., 2013); and anaerobiosis was achieved by sparging 0.3 [L/min] of pure nitrogen and agitating at 300 [RPM]. pH was controlled using phosphoric acid 20% [v/v] and ammonium hydroxide 15% [v/v] (the feed of the latter also gives additional nitrogen supplementation). Temperature was controlled with a mixture of hot and cold water, using a glass jacket. Lastly, foam was controlled manually using silicone antifoam 10% [v/v].

Glucose starvation was detected when a sudden decrease of the $CO_2$ composition in the off-gas occurred, and confirmed each time using Benedict's reagent (Benedict, 1909). For fed-batch experiments (conditions 1 and 2), the feed $F(t)$ was designed to track a predefined time variable growth rate and, therefore, can be

calculated from the reactor's glucose and biomass mass balances, as detailed elsewhere (Villadsen and Patil, 2007)

$$F(t) = \frac{\mu_{set}(t)}{S^F \cdot Y_{SX}} \cdot V_i X_i \cdot \exp\left\{\int_{t_i}^{t} \mu_{set}(t)dt\right\} \quad (14)$$

with $S^F$ the glucose feed concentration [g/L], $Y_{SX}$ the experimental glucose-biomass yield (fixed for Eq. 14 in 0.469 [g DW/g] (Møller et al., 2004)), $t_i$ the time at which the feed started for a given cultivation [h], $V_i$ and $X_i$ the volume [L] and biomass [g/L] values at $t_i$, respectively, and $\mu_{set}(t)$ the variable growth rate. The latter was defined as the following:

$$\mu_{set}(t) = A + B \cdot e^{-Ct} \quad (15)$$

where $A$ and $B$ are both 0.07 [1/h] for all aerobic cultivations, and $C$ is 0.14 [1/h] for condition 1, and 0.07 [1/h] for condition 2. Therefore, $\mu_{set}(t)$ decays more quickly in condition 1, which translates into a slower feed rate (Fig. S1).

### 3.3. Assay methods

Samples of ~5 [mL] were taken periodically from all cultivations. Biomass was measured in OD at 600 nm using a UV-160 UV–visible recording spectrophotometer (Shimadzu, Japan), and results were transformed to g/L using a calibration curve of 0.3797 [g DW/L/OD] for the N30 strain and of 0.3840 [g DW/L/OD] for the EC1118 strain (both determined with an infrared dryer-equipped balance (Precisa, Switzerland)).

A 2 [mL] aliquot of each sample was centrifuged for 5 min at 14,000 RPM and 4 °C, using a Mikro 22R centrifuge (Hettich, Germany). All supernatants were kept at −80 °C until the fermentation was over. Extracellular metabolites were afterwards measured by high-performance liquid chromatography (HPLC) in duplicate. 100 [μL] of a solution 27.5 [mM] $H_2SO_4$ and with 16.7 [g/L] of pivalic acid (used as internal standard) were added to 1 [mL] of each sample and each of the HPLC standards (with known concentrations of trehalose, glucose, fructose, glycerol, ethanol, citrate, malate, succinate, lactate and acetate). Afterwards, 20 [μL] of the resulting solution was injected into a LaChrom L-7000 HPLC system (Hitachi, Japan), with an Aminex HPX-87H anion-exchange column (Bio-Rad, USA) for organic acids, alcohols and sugars separation, working at 55 °C with a 0.5 [mL/min] flow of mobile phase 2.5 [mM] $H_2SO_4$ (the same concentration as the one in each sample after adding the internal standard solution). A LaChrom L-7450A diode array detector (Hitachi, Japan) was used at 210 [nm] for detecting organic acids, and a LaChrom L-7490 refraction index detector (Hitachi, Japan) for sugars and alcohols. Finally, each metabolite was quantified normalizing each area in the chromatogram by the corresponding internal standard area and using a calibration curve with the HPLC standards.

### 3.4. Reparameterization analysis

#### 3.4.1. General procedure

We employed a novel methodology to obtain model structures with identifiable and significant parameters from our dFBA framework for each of the 16 experimental cultivations. We started by calculating sensitivity and identifiability for the estimated parameters of the calibration (Fig. 3). Then, the parameters were iteratively fixed. At the end of each iteration, and depending on the result of the regression diagnostics, a parameter is eliminated (thus its value becomes fixed) for the next iteration. The following rules were considered for deciding which parameters to fix (for further details of each analysis see Section 3.4.2):

1. Identifiability: when 2 parameters had a Pearson correlation coefficient larger than 0.95, different combinations of the corresponding estimated values resulted in the same objective function in the parameter estimation procedure and, consequently, the decision was to fix at least one of them.
2. Sensitivity: when the relative sensitivity of a parameter was below 0.01 for all the variables, the parameter was considered to have no influence in the model, and was therefore fixed.

For most iterations, more than one of the above problems will arise. Therefore, exploratory branches for each of the corresponding parameters are necessary, which generates a growing exploratory tree. Our aim here was to generate several model structures, so we could choose later the one that had the best predictive capabilities. However, heuristic rules were established to reduce the search space. First, at each iteration we identified those parameters that presented identifiability or sensitivity problems. If any parameter had both problems, exploratory branches were created only for these parameters. If there were only parameters with 1 of the 2 problems, we created branches for all the problematic parameters. Finally, if no parameter had any of the mentioned problems, the model structure attained was saved for posterior analyses, and the branch no longer explored.

After finishing the exploration stage, confidence coefficients (CCs) were calculated for each non-fixed parameter in those model structures with parameter combinations that showed no identifiability or sensitivity problems (Fig. 3). If any non-fixed parameter from a model structure had a CC larger than 2, it was considered that the parameter had a value not significantly different from zero, and therefore the corresponding model structure was disregarded.

Finally, for each of the 16 experimental conditions, the model structure with the smallest mean CC was chosen as the optimal reparameterization (Fig. 3). To further improve the fitting results, each of these 16 model structures were calibrated again, but only using the corresponding non-fixed parameters; the fixed parameters had the same values than the originally estimated ones.

#### 3.4.2. Pre/post-regression analysis

In the following, we briefly explain the regression diagnostics carried out in this study, as it has been thoroughly presented elsewhere (Jaqaman and Danuser, 2006; Sacher et al., 2011). Sensitivity analysis accounts for the relative impact that each parameter has in each of the model's state variables. In our approach, we computed the relative sensitivity ($G_{ik}$), as

$$G_{ik}(t, \theta_k) = \frac{\theta_k}{X_i(t)} \frac{dX_i(t)}{d\theta_k} \quad (16)$$

where $t$ is time, $\theta_k$ is the $k$th parameter and $X_i(t)$ is the $i$th variable at time $t$. With all $G_{ik}$ values, for each time we formed a sensitivity matrix $G(t)$, in which the $k$th column denotes the sensitivity of the $k$th parameter on the state variables. In order to obtain a single normalized score (spanning all experimental times) of each parameter over each variable, we calculated average sensitivities as detailed in Hao et al. (2006). Therefore, if this score is under 0.01 in each variable for a given parameter, we chose to fix the corresponding parameter.

For identifiability calculations, the MATLAB function *corrcoef* was used to calculate the correlation coefficients between each column of the sensitivity matrices, and stored the information in a correlation coefficients matrix ($C$). If any of the off-diagonal elements of the matrix take on values over a certain threshold (in our case $|C_{ij}| \geq 0.95$), it is considered that the associated parameters are strongly correlated, and therefore one of them should be fixed.

For significance calculations, and also using the sensitivity matrices, we first calculated the Fisher Information Matrix (FIM)

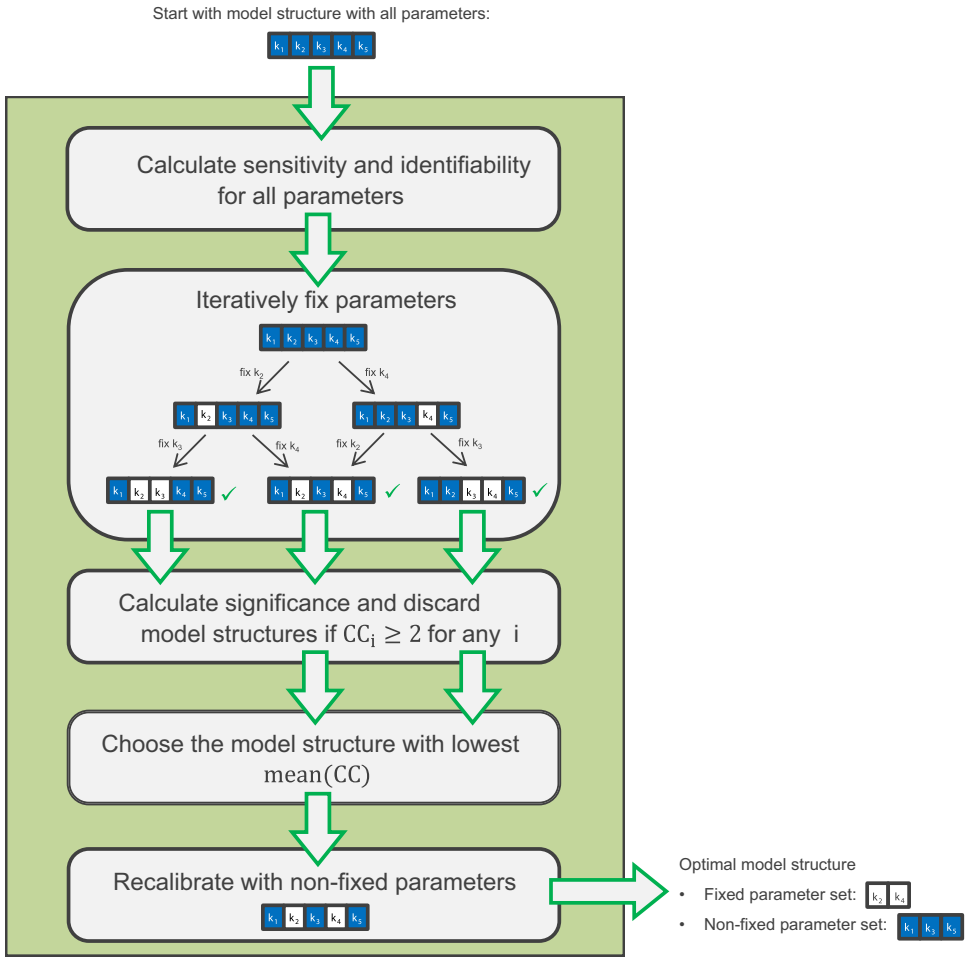


**Fig. 3.** Reparameterization methodology. As an example, a model with 5 parameters is analyzed.

(Petersen et al., 2001)

$$\mathrm{FIM} = \sum_{j=1}^{n} G_j^{\mathrm{T}} Q_j G_j \tag{17}$$

where $G_j$ is the sensitivity matrix for measurement $j$, $n$ is the number of measurements, and $Q_j$ is the inverse of the measurement error covariance matrix assuming white and uncorrelated noise, which is used as a weighting matrix. Using this matrix, the variances for each estimated parameter ($\sigma_k^2$) were calculated as (Landaw and DiStefano, 1984; Petersen et al., 2001)

$$\sigma_k^2 = \mathrm{FIM}_{kk}^{-1} \tag{18}$$

With the variances we computed the confidence interval (CI) with 5% significance for the $k$parameter as follows:

$$\mathrm{CI}_k = [\hat{\theta}_k \pm 1.96\,\sigma_k] \tag{19}$$

where $\hat{\theta}_k$ is the estimated value of the respective parameter. Finally, coefficients of confidence (CC) were calculated as follows:

$$\mathrm{CC}_k = \frac{\Delta(\mathrm{CI}_k)}{\hat{\theta}_k} = \frac{3.92\,\sigma_k}{\hat{\theta}_k} \tag{20}$$

With $\Delta(\mathrm{CI}_k)$, the CI's length. With this metric, we determined that a parameter was not significantly different from zero if the CI contained the zero, therefore if the absolute value of the CC was equal or larger than 2

$$\mathrm{CI}_k \text{ contains zero} \rightarrow \hat{\theta}_k - 1.96\sigma_k \le 0 \le \hat{\theta}_k + 1.96\sigma_k \rightarrow |\hat{\theta}_k| \le 1.96\sigma_k \rightarrow 2 \le |\mathrm{CC}_k| \tag{21}$$

The aforementioned pre/post-regression diagnostics took between 10 and 50 [min], depending on the experimental conditions and the computer employed (Table S1).

### 3.4.3. Cross-calibration

The two best model structures (the best aerobic and the best anaerobic) were chosen performing a cross-calibration between cultivations. Each model structure obtained from a given fermentation was used to calibrate the remaining 7 fermentations (aerobic or anaerobic), i.e. fixing the corresponding parameters at the values indicated by the model structure and fitting the data with the remaining parameters. Afterwards, a cross-calibration coefficient (CCC) was computed as

$$\mathrm{CCC}_{ijk} = \frac{F_{ijk}}{F_{jjk}};\ \ i,\ j = 1\ldots 8\,;\, i \neq j\,;\ k = \{0,1\} \tag{22}$$

where $F_{ijk}$ is the objective function obtained with the $i$th model structure using the experimental data of the $j$th cultivation, and $k$ is an index for distinguishing between aerobiosis ($k=1$) and anaerobiosis ($k=0$). Therefore, if $\mathrm{CCC}_{ijk}$ is close to 1, the $i$th model structure used to calibrate the $j$th experimental data is appropriate. On the other hand, if $\mathrm{CCC}_{ijk}$ is much larger than 1, a good fit was not possible and therefore the $i$th model structure is not useful for predicting different experimental conditions.

**Table 2**
Percentage of times that each parametric problem arose in (A) aerobic and (B) anaerobic calibrations. Relative sensitivity was averaged among all variables.

| (A) | \|Correlation\| ≥ 0.95 | | | | | | | | | | | | | | | | | | | Average sensitivity < 0.01 | \|CC\| ≥ 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $v_{Gmax}$ | $K_G$ | $K_E$ | $\alpha$ | $a$ | $c$ | $l$ | $m_{ATP}$ | $t_1$ | $t_2$ | $f_E$ | $f_{GL}$ | $f_C$ | $f_L$ | $\alpha_F$ | $v_E$ | $v_{GL}$ | $v_C$ | $v_L$ | | |
| $v_{Gmax}$ | – | 88% | 100% | 63% | 38% | 25% | – | 13% | – | – | 13% | 13% | – | 13% | – | – | – | – | – | – | 25% |
| $K_G$ | 88% | – | 88% | 25% | 25% | 13% | 13% | 13% | – | – | 13% | 13% | – | 13% | – | – | – | – | – | – | 88% |
| $K_E$ | 100% | 88% | – | 75% | 25% | 25% | 25% | 13% | – | – | 13% | 13% | 13% | 13% | – | – | – | – | – | – | 75% |
| $\alpha$ | 63% | 25% | 75% | – | 88% | 88% | 75% | 13% | – | – | 75% | 25% | 13% | 13% | – | – | – | – | – | – | 63% |
| $a$ | 38% | 25% | 25% | 88% | – | 100% | 100% | 13% | – | – | 50% | 25% | 13% | 13% | – | – | – | – | – | – | 88% |
| $c$ | 25% | 13% | 25% | 88% | 100% | – | 88% | 13% | – | – | 50% | 25% | 13% | 13% | – | – | – | – | – | – | 88% |
| $l$ | – | 13% | 25% | 75% | 100% | 88% | – | 25% | – | – | 38% | 25% | 13% | 13% | – | – | – | – | – | – | 100% |
| $m_{ATP}$ | 13% | 13% | 13% | 13% | 13% | 13% | 25% | – | – | – | 13% | – | – | 13% | – | – | – | – | – | 75% | 100% |
| $t_1$ | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 100% | 100% |
| $t_2$ | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 88% | 100% |
| $f_E$ | 13% | 13% | 13% | 75% | 50% | 50% | 38% | 13% | – | – | – | 25% | 13% | 13% | – | – | – | – | – | – | 25% |
| $f_{GL}$ | 13% | 13% | 13% | 25% | 25% | 25% | 25% | – | – | – | 25% | – | 13% | 13% | – | – | – | – | – | – | 25% |
| $f_C$ | – | – | 13% | 13% | 13% | 13% | 13% | – | – | – | 13% | 13% | – | – | – | – | – | – | – | – | – |
| $f_L$ | 13% | 13% | 13% | 13% | 13% | 13% | 13% | 13% | – | – | 13% | 13% | – | – | – | – | – | – | – | – | – |
| $\alpha_F$ | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 100% | 75% | – | – | – | 100% |
| $v_E$ | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 100% | – | 75% | – | – | 88% | 100% |
| $v_{GL}$ | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 75% | 75% | – | – | – | 100% | 100% |
| $v_C$ | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 25% |
| $v_L$ | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 25% |

| (B) | \|Correlation\| ≥ 0.95 | | | | | | | | | | | | | | Average sensitivity < 0.01 | \|CC\| ≥ 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $v_{Gmax}$ | $K_G$ | $K_E$ | $\alpha$ | $a$ | $c$ | $l$ | $m_{ATP}$ | $t_1$ | $t_2$ | $f_E$ | $f_{GL}$ | $f_C$ | $f_L$ | | |
| $v_{Gmax}$ | – | 88% | 100% | 75% | 88% | 100% | 88% | 63% | – | 13% | 13% | 75% | – | 13% | – | 25% |
| $K_G$ | 88% | – | 88% | 63% | 75% | 88% | 63% | 50% | – | – | 13% | 63% | 13% | 13% | 38% | 100% |
| $K_E$ | 100% | 88% | – | 75% | 88% | 100% | 88% | 63% | – | 13% | 13% | 75% | 25% | 25% | – | 63% |
| $\alpha$ | 75% | 63% | 75% | – | 63% | 75% | 75% | 63% | – | – | 13% | 88% | 25% | 25% | – | 100% |
| $a$ | 88% | 75% | 88% | 63% | – | 100% | 100% | 63% | – | – | 13% | 63% | 63% | 25% | – | 100% |
| $c$ | 100% | 88% | 100% | 75% | 100% | – | 88% | 63% | – | – | 13% | 75% | 25% | 25% | – | 100% |
| $l$ | 88% | 63% | 88% | 75% | 100% | 88% | – | 63% | – | – | 13% | 75% | 38% | 13% | 100% | 100% |
| $m_{ATP}$ | 63% | 50% | 63% | 63% | 63% | 63% | 63% | – | – | – | 13% | 63% | 38% | 13% | 75% | 100% |
| $t_1$ | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 100% | 100% |
| $t_2$ | 13% | – | 13% | – | – | – | – | – | – | – | – | – | – | – | 100% | 100% |
| $f_E$ | 13% | 13% | 13% | 13% | 13% | 13% | 13% | 13% | – | – | – | 13% | 13% | 13% | 13% | 13% |
| $f_{GL}$ | 75% | 63% | 75% | 88% | 63% | 75% | 75% | 63% | – | – | 13% | – | 25% | 13% | 38% | 88% |
| $f_C$ | – | 13% | 25% | 25% | 63% | 25% | 38% | 38% | – | – | 13% | 25% | – | 13% | – | – |
| $f_L$ | 13% | 13% | 25% | 25% | 25% | 25% | 13% | 13% | – | – | 13% | 13% | 13% | – | – | – |

## 4. Results and discussion

### 4.1. Original calibration problems

#### 4.1.1. Aerobic simulation problems

The sensitivity, identifiability and significance problems that arose in all simulations after the first calibration – with all parameters – are displayed in Table 2. In the aerobic cultivations, several issues between parameters were detected (Table 2A). Regarding identifiability, the most correlated parameters could be gathered in 3 groups. In the first group, glucose consumption parameters ($v_{Gmax}$, $K_G$ and $K_E$) were almost always structurally unidentifiable, suggesting that at most one of them should be used for calibration. The second group included all 4 parameters directly related to the biomass formation in the batch phase: $\alpha$, $a$, $c$ and $l$. Here, the correlations were high in almost all fermentations, indicating that the "suboptimal growth" effect can be modeled with just one of these parameters. The ethanol production parameter ($f_E$) in the batch phase showed significant correlations with this group; hence, in *Yeast 5*, under aerobic conditions, ethanol is more correlated than any other metabolite to biomass formation. Furthermore, $\alpha$ was most of the times correlated with $K_E$, indicating a possible connection between suboptimal growth rate and ethanol inhibition. The final group of highly intercorrelated parameters—$\alpha_F$, $v_E$ and $v_{GL}$—belongs to the fed-batch phase. The close relationship between biomass and ethanol concentrations also holds here and an additional relation with glycerol arises, probably due to the consumption of the latter at this stage.

Regarding sensitivity, the ATP maintenance and both expression thresholds ($m_{ATP}$, $t_1$ and $t_2$) were insensitive to all state variables in most of the aerobic fermentations, thus not being useful for calibration in *S. cerevisiae* dFBA (Table 2A). $v_E$ and $v_{GL}$ showed also to be insensitive and, as both were correlated with $\alpha_F$, the latter one should be preferred for calibration. Finally, CCs calculations indicated that most of the parameters have low significance. However, significance tests depend on the number of estimated parameters, therefore, several significance problems were expected at this level.

#### 4.1.2. Anaerobic simulation problems

In the anaerobic cultivations, strong correlations were found between all members of a large parameter group (Table 2B), that included $v_{Gmax}$, $K_G$, $K_E$, $\alpha$, $a$, $c$, $l$, $m_{ATP}$ and $f_{GL}$. This indicates a stronger interdependence in anaerobic cultures between glucose consumption and biomass production parameters, which was not observed for aerobic conditions. This is probably because in anaerobiosis, *Yeast 5* has fewer metabolic pathways available to produce the energetic requirements for growth (for instance, the electron transport system and other oxygen-dependent pathways are inactive); therefore consumption and production are more strongly correlated. Remarkably, the glycerol production parameter (and not the ethanol production parameter, as in aerobic cultures) is strongly correlated to the biomass parameters, probably due to the importance of glycerol in regulating the redox balance in yeast.

Analyzing sensitivity, ATP maintenance and both expression thresholds ($m_{ATP}$, $t_1$ and $t_2$) were insensitive in most of the anaerobic fermentations (the same as in the aerobic case) (Table 2B). The parameter $l$ also appeared as non-sensitive in anaerobic fermentations, possibly as a consequence that, under these conditions, ergosterol and fatty acids are supplied to the medium, i.e. there is sufficient lipid available for biomass formation. Finally, significance results showed that only a few of the parameters are significantly different from zero.

### 4.2. Reparameterization results

We performed the iterative procedure described in Section 3 to each of the 16 experimental cultivations separately. Thanks to the heuristic criteria described above, we obtained non-problematic model structures after an average of 6232 iterations for aerobic and 1858 iterations for anaerobic cultivations, which are respectively 1.2% and 11.3% of the total amount of possible combinations (Figs. S2 and S3). For each group of non-problematic model structures, we chose the one with the lowest average CC; afterwards, for each of the selected ones, we repeated the parameter estimation. Table S4 shows the best model structure for each of the 16 analyses after the second calibration. All fittings achieved have relative errors below 10% (Table S5). Finally, the results of the cross-calibration study is available in Table S6, which indicates the best aerobic and anaerobic model structures, i.e. the ones that not only have no identifiability, sensitivity or significance issues, but are also capable of calibrating adequately the rest of the corresponding cultivations.

### 4.3. Best model structures

#### 4.3.1. Aerobic model structure

The best model structure for aerobic cultivations was achieved from a *S. cerevisiae* EC1118 strain cultivation, with a fast feed in the fed-batch stage (condition 2). Estimated parameters were $\alpha$, $f_E$, $f_{GL}$, $f_C$, $f_L$, and $\alpha_F$ (Table 3). This indicates that only using suboptimal growth and minimum byproduct production rates for the batch stage, we fitted several aerobic yeast cultures, with CCs no larger than 0.46 (Table 3), no correlations between parameters and high sensitivity of the parameters in the state variables (Table S7). Hence, in order to adequately calibrate an aerobic dynamic metabolic model, information such as ATP maintenance, biomass composition, and constant Boolean expression rules do not provide additional insights if the growth and production kinetics of the system are well characterized. This is consistent with previous observations on the sensitivity of the biomass components (Varma and Palsson, 1993) and gene expression Boolean rules (Hjersted and Henson, 2009) in aerobic cultures.

We highlight here that, while all the estimated parameters have a significant effect in the model, the fixed parameters do not significantly impact the simulation output. For instance, a glucose maximum uptake rate of 22.5 [mmol/g DWh], a saturation constant of 0.88 [g/L] and an ethanol inhibition constant of 6.74 [g/L] (Table 3) are appropriate to represent the glucose consumption rate for most of the conditions explored, if the estimated

**Table 3**
Best aerobic model structure. The fixed parameters and their fixed values are shown along with the non-fixed parameters and the average CC (shown as percentage) obtained in the cross-calibration analysis.

| Fixed parameters | | Non-fixed parameters | |
|---|---|---|---|
| Symbol | Value fixed | Symbol | Average CC |
| $v_{Gmax}$ | 22.5 | $\alpha$ | 26.4% |
| $K_G$ | 0.88 | | |
| $K_E$ | 6.74 | | |
| $a$ | 1.55 | $f_E$ | 11.2% |
| $c$ | 2.23 | | |
| $l$ | 2.82 | $f_{GL}$ | 27.3% |
| $m_{ATP}$ | 1.99E−06 | | |
| $t_1$ | 4.03 | $f_C$ | 16.9% |
| $t_2$ | 75.7 | | |
| $v_E$ | −3.53 | $f_L$ | 13.7% |
| $v_{GL}$ | −0.37 | | |
| $v_C$ | 0.057 | $\alpha_F$ | 46.1% |
| $v_L$ | 0.281 | | |

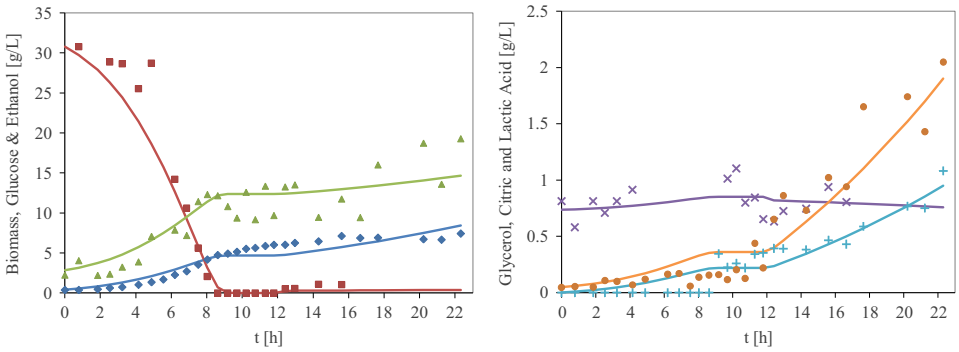


**Fig. 4.** Fitting of the best aerobic model structure attained. The experimental measurements (EC1118 strain; fast feed) for biomass (♦), glucose (■), ethanol (▲), glycerol (×), citric (+) and lactic acid (●) are shown, together with the corresponding model prediction (continuous lines).

**Table 4**
Best anaerobic model structure. The fixed parameters and their fixed values are shown along with the non-fixed parameters and the average CC (shown as percentage) obtained in the cross-calibration analysis.

| Fixed parameters | | Non-fixed parameters | |
|---|---|---|---|
| Symbol | Value fixed | Symbol | Average CC |
| $K_G$ | 6.44 | $v_{Gmax}$ | 6.9% |
| $K_E$ | 31.3 | | |
| $a$ | 1.61 | $\alpha$ | 7.2% |
| $c$ | 0.606 | | |
| $l$ | 2.29 | $f_E$ | 46.6% |
| $m_{ATP}$ | 6.76E−07 | | |
| $t_1$ | 4.37 | $f_C$ | 9.7% |
| $t_2$ | 98.1 | | |
| $f_{GL}$ | 7.04E−08 | $f_L$ | 10.9% |

parameters are appropriately fitted. Similar remarks apply to biomass requirements, ATP maintenance, gene expression thresholds and fed-batch minimum byproduct production rates.

As illustrated in Fig. 4, the aerobic model structure successfully calibrated all of the corresponding cultivation measurements. Also, cross-calibrating the results in the rest of the cultivations we obtained good fittings as well (Table S6A). One of these cross-calibrations, depicted in Fig. 6, shows that most variables are correctly predicted, with the exception of biomass, ethanol and citrate that were underpredicted in the fed-batch stage. We considered the model structure successful, considering that not only the data was appropriately fitted, but as mentioned in Section 1, we obtained a model structure with sensitive and uncorrelated parameters, that will robustly calibrate different experimenta data.

#### 4.3.2. Anaerobic model structure

For anaerobic cultivations, the best model structure was also attained from a *S. cerevisiae* EC1118 strain fermentation, with a large initial sugar concentration (condition 4), and with $v_{Gmax}$, $\alpha$, $f_E$, $f_C$ and $f_L$ as fitting parameters (Table 4). This indicates that only using these 5 parameters, we can fit several anaerobic experimental conditions of yeast growth, and also attain parameters with low CCs (Table 4), no correlations and highly sensitive (Table S8). Similarly to the aerobic case, the analysis shows that ATP maintenance, biomass composition and constant on/off expression rules are redundant if the consumption and production kinetics of the system are appropriately fitted.

Differing from the aerobic model structure, glycerol production rate is not necessary for calibration of anaerobic cultivations; instead, good estimation of glucose consumption is critical. In addition, in this model structure the ethanol inhibition constant is considerably larger (31.3 [g/L]) (Table 4) than in aerobic cultivations, indicating that under anoxia, yeast cells support higher ethanol concentrations.

Finally, Fig. 5 shows that the anaerobic model structure adequately calibrated all of the corresponding batch data. Successful fittings were attained also in the cross-calibration (Table S6B); an example of the cross-calibrations is displayed in Fig. 7, in which all variables were correctly predicted, with the exception of ethanol that was slightly underpredicted. These results confirm the usefulness of this approach for anaerobic cultivations as well, given that the calibration was successful using only uncorrelated and sensitive parameters.

### 4.4. Additional considerations

#### 4.4.1. Alternative model structures

Although the two-abovementioned model structures are highlighted as the best ones for explaining different aerobic or anaerobic conditions, other model structures also performed well in the cross-calibration (Table S6), and could be seen as alternatives for calibrating new experimental conditions. For instance, using for calibration the maximum glucose consumption rate, the ATP maintenance and some fed-batch minimum byproduct production rates, instead of the suboptimal growth rate, also allowed to achieve good fittings under various aerobic conditions. Likewise, excluding the suboptimal growth rate for estimation in anaerobic conditions can also be useful for exploring several experimental conditions. These considerations could be relevant when expanding the procedure to other conditions, such as other strains or culture media.

#### 4.4.2. Approach limitations

There are several assumptions that were made in this study that should be considered in order to understand the limitations of the model predictions. First, other parameters and variables that have been used before in dFBA modeling were disregarded from our analysis. For instance, oxygen, proton and nitrogen balances were not included in the analysis because this work focused in carbon limitation, and their supply was sufficient to avoid limitation, which made modeling unnecessary (with the exception of the anaerobic cultivations, in which there was no oxygen to model). The $CO_2$ production rate was neither evaluated, as measurements of specific productivity [mmol/g DWh] were not possible with our experimental setup.

Secondly, the growth associated maintenance (GAM), which is already included in the biomass pseudo-reaction rate (reaction r_2110 in *Yeast 5*) and has a value of 59.3 [mmol/g DWh], was not included in the analysis, given its high correlation with non-growth associated maintenance (NGAM). We choose to focus on NGAM, which can be calibrated more easily, considering that it

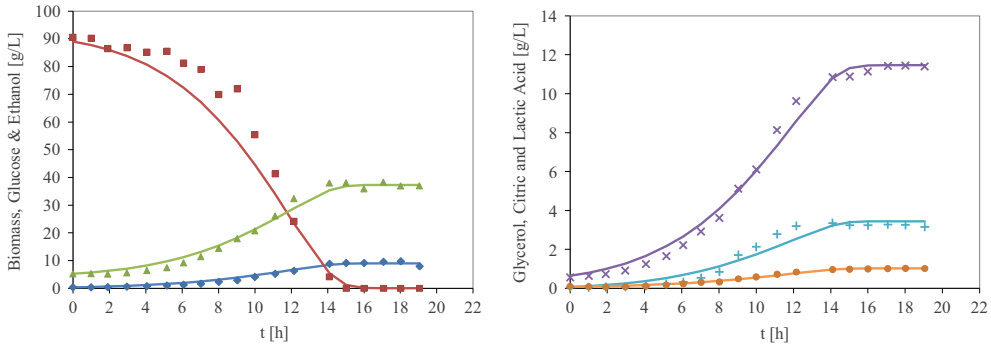


**Fig. 5.** Fitting of the best anaerobic model structure attained. The experimental measurements (EC1118 strain; large $G_0$) for biomass (♦), glucose (■), ethanol (▲), glycerol ( × ), citric ( + ) and lactic acid (●) are shown, together with the corresponding model prediction (continuous lines).

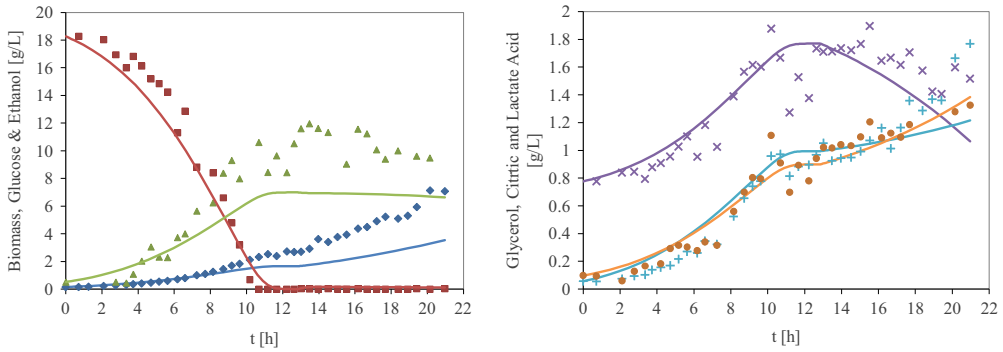


**Fig. 6.** Cross-calibration example of the aerobic model structure. The experimental measurements (N30 strain; slow feed) for biomass (♦), glucose (■), ethanol (▲), glycerol ( × ), citric ( + ) and lactic acid (●) are shown, together with the corresponding model prediction (continuous lines).

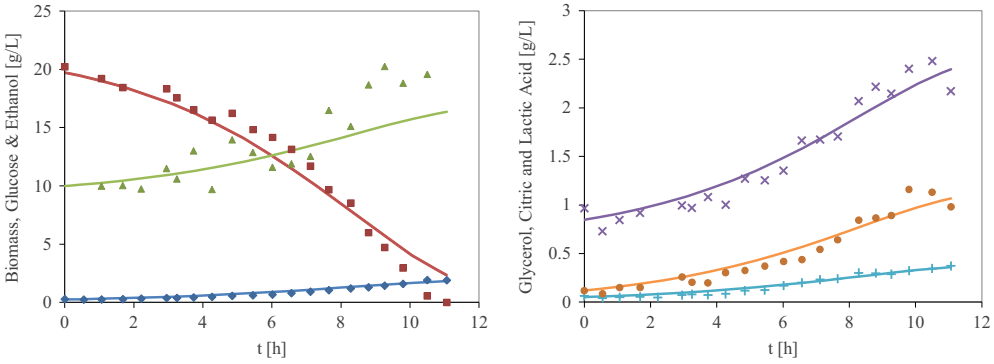


**Fig. 7.** Cross-calibration example of the anaerobic model structure. The experimental measurements (N30 strain; low $G_0$) for biomass (♦), glucose (■), ethanol (▲), glycerol ( × ), citric ( + ) and lactic acid (●) are shown, together with the corresponding model prediction (continuous lines).

does not directly impact the biomass pseudo-reaction rate. Likewise, the P/O ratio, valid only for aerobic cultures (when the electron transfer system is active), was not considered, because it is structurally unidentifiable with the ATP maintenance, and remains constant for fully aerobic conditions (Jouhten et al., 2012). Microaerobic or oxygen-limited fermentations could be performed to study appropriately this parameter, along with other oxygen-related parameters from both oxygen transfer rate (OTR) and oxygen uptake rate (OUR) kinetics (Saa et al., 2012).

Further considerations must be made regarding the gene expression rules proposed. As mentioned before, the main assumptions were that the expression was strain independent, and the rules remained constant through all fermentation. These assumptions were adopted because not enough data was collected to generate more elaborated rules. With this approach, both model structures yielded high values for the two expression thresholds ($t_1$ and $t_2$), as seen in Tables 3 and 4, which translated into no genes turned off during all cultivations. Therefore, alternative approaches (Covert et al., 2004, 2001; Mehra et al., 2003; Waldherr et al., 2013) could be considered in order to obtain relevant information from gene expression in dFBA.

It is worthy to mention that the present analysis is based on local sensitivity, identifiability and significance tools. Alternatively, one could use global tools such as structural identifiability analysis (Balsa-Canto et al., 2010), global sensitivity analysis using Monte Carlo simulations (Di Maggio et al., 2010; Kravaris et al., 2013), bifurcation theory (Kruger et al., 2007) or study higher order interrelationships between parameters (Li and Vu, 2013). However, most of these methods were designed for kinetic models, and are not applicable to dFBA modeling, in which large optimization problems are embedded in the dynamic equations. Even if the methodologies were to be adapted, extensive symbolic manipulations (Balsa-Canto et al., 2010; Li and Vu, 2013) and/or even more intensive computational simulations would be required.

We overcome these issues by showing through cross-calibration that we can attain model structures that, although locally analyzed, are useful for different conditions.

Finally, it is important to notice that the approach was developed for a reasonable number of studied parameters (15–20), which is average in dFBA modeling; however, more parameters will eventually prevent the optimization algorithm to find a good solution. Consequently, future efforts should concentrate in adapting existing methodologies to models with more parameters, for instance implementing set by set selection (Alberton et al., 2013) and/or clustering methods (Kravaris et al., 2013), in which groups of parameters are simultaneously analyzed, instead of studying one parameter at a time. These considerations are especially relevant considering novel models that account for all known cellular process and have thousands of parameters (Karr et al., 2012).

## 5. Conclusion

In this work, we developed a yeast dFBA framework that comprises several features not traditionally included in this type of models, such as a sequential optimization (biomass maximization and total flux minimization), condition-specific biomass requirements, minimum byproduct production kinetics for both batch and fed-batch stages and gene expression thresholds. The analysis presented is novel, as various culture conditions (batch/fed-batch, different strains, etc.) were used to calibrate the same framework. We also showed that several sensitivity, identifiability and significance problems arise in this kind of models, despite an adequate model fitting. The latter can severely hamper the model applicability, since simulations for different conditions are unreliable.

To solve the abovementioned issues, we proposed an iterative procedure and a cross-calibration approach; we therefore can ascertain that the model structures developed here have an adequate number of parameters, in the sense that a good fitting among several experimental conditions is obtained and, at the same time, the typical problems of too many model parameters do not arise. The procedure has some limitations (e.g., is based in local sensitivity, identifiability and significance tools) but has several advantages (e.g. has no symbolic manipulation, is not computationally expensive, and generates alternative model structures) that makes it useful for dFBA schemes.

Overall, for the aerobic model structure, suboptimal growth rate should be considered for an adequate calibration, whereas for the anaerobic model structure, glucose consumption rate is also relevant. Nonetheless, for both model structures, byproduct production rates were found to be critical for calibration. Other informations, such as gene expression constant on/off rules, biomass requirements and ATP maintenance were less relevant when calibrating these models.

One of the main applications of GSMM models – using both static and dynamic approaches – is as a guide for genetic modifications, i.e. knock-outs (gene deletions), knock-ins (additions), knock-downs (sub-expressions) or knock-ups (over-expressions), in order to overproduce a metabolite of interest (Alper et al., 2005; Bro et al., 2006; Hjersted et al., 2007). To this end, a significant amount of computational tools are currently available which, using as input a GSMM, can propose one or more of the mentioned modifications (Copeland et al., 2012; Park et al., 2009). Still, in order to have confident predictions, the models employed should be appropriately calibrated. The model structures presented in this work are expected to have important benefits when experimenting *in silico* and designing metabolic engineering strategies.

## Acknowledgments

We are grateful to Martín Cárcamo, Mariana Cepeda, Martín Concha, Jonathan Leon, Pedro Saa, Fernando Silva, Jorge Torres, Paulina Torres and Felipe Varea for their valuable technical assistance in the experimental assays; to Centrovet® for facilitating the *S. cerevisiae* N30 strain; to Dr. Claudio Gelmi for his wise recommendations and suggestions in Matlab programming and for facilitating the pre/post-regression analysis tools; to Waldo Acevedo and Ignacio Varas for their assistance in LINUX programming; to Dr. Danilo González for providing computational power from the Center for Bioinformatics and Integrative Biology (CBIB) at University Andrés Bello; and to the anonymous referees who helped with valuable feedback in order to improve the final manuscript. This work was funded by Fondecyt Grant #1130822. B.J.S. was recipient of a M.Sc. scholarship by CONICYT-PCHA/Magister Nacional/2013 – #221320015.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ymben.2014.07.004.

## References

Aceituno, F.F., Orellana, M., Torres, J., Mendoza, S., Slater, A.W., Melo, F., Agosin, E., 2012. Oxygen response of the wine yeast *Saccharomyces cerevisiae* EC1118 grown under carbon-sufficient, nitrogen-limited enological conditions. Appl. Environ. Microbiol. 78, 8340–8352.

Åkesson, M., Förster, J., Nielsen, J., 2004. Integration of gene expression data into genome-scale metabolic models. Metab. Eng. 6, 285–293.

Alberton, K.P., Alberton, A.L., Di Maggio, J.A., Díaz, M.S., Secchi, A.R., 2013. Accelerating the parameters identifiability procedure: set by set selection. Comput. Chem. Eng. 55, 181–197.

Alper, H., Jin, Y.-S., Moxley, J.F., Stephanopoulos, G., 2005. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab. Eng. 7, 155–164.

Balsa-Canto, E., Alonso, A., Banga, J., 2010. An iterative identification procedure for dynamic modeling of biochemical networks. BMC Syst. Biol. 4 (Article no. 11).

Balsa-Canto, E., Rodriguez-Fernandez, M., Banga, J., 2007. Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation. J. Food Eng. 82, 178–188.

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. Nat. Protoc. 2, 727–738.

Benedict, S.R., 1909. A reagent for the detection of reducing sugars. J. Biol. Chem. 5, 485–487.

Bornstein, B.J., Keating, S.M., Jouraku, A., Hucka, M., 2008. LibSBML: an API library for SBML. Bioinformatics 24, 880–881.

Bro, C., Regenberg, B., Förster, J., Nielsen, J., 2006. In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. Metab. Eng. 8, 102–111.

Cárcamo, M., Saa, P.A., Torres, J., Torres, S., Mandujano, P., Correa, J.R.P., Agosin, E., 2014. Effective dissolved oxygen control strategy for high cell-density cultures. IEEE Lat. Am. Trans. 12, 389–394.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hirschman, J.E., Hitz, B.C., Karra, K., Krieger, C.J., Miyasato, S.R., Nash, R.S., Park, J., Skrzypek, M.S., Simison, M., Weng, S., Wong, E.D., 2012. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 40, D700–D705.

Chu, Y., Hahn, J., 2008. Integrating parameter selection with experimental design under uncertainty for nonlinear dynamic systems. AIChE J. 54, 2310–2320.

Copeland, W.B., Bartley, B.A., Chandran, D., Galdzicki, M., Kim, K.H., Sleight, S.C., Maranas, C.D., Sauro, H.M., 2012. Computational tools for metabolic engineering. Metab. Eng. 14, 270–280.

Covert, M.W., Knight, E.M., Reed, J.L., Herrgård, M.J., Palsson, B.Ø., 2004. Integrating high-throughput and computational data elucidates bacterial networks. Nature 429, 92–96.

Covert, M.W., Schilling, C.H., Palsson, B.Ø., 2001. Regulation of gene expression in flux balance models of metabolism. J. Theor. Biol. 213, 73–88.

Covert, M.W., Xiao, N., Chen, T.J., Karr, J.R., 2008. Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. Bioinformatics 24, 2044–2050.

Curran, K.A., Crook, N.C., Alper, H.S., 2012. Using flux balance analysis to guide microbial metabolic engineering. Methods Mol. Biol. 834, 197–216.

Dennis, J.E., Gay, D.M., Walsh, R.E., 1981. An adaptive nonlinear least-squares algorithm. ACM Trans. Math. Softw. 7, 348–368.
Di Maggio, J., Diaz Ricci, J.C., Diaz, M.S., 2010. Global sensitivity analysis in dynamic metabolic networks. Comput. Chem. Eng. 34, 770–781.
Diderich, J.A., Schepper, M., Van Hoek, P., Luttik, M.A.H., Van Dijken, J.P., Pronk, J.T., Klaasen, P., Boelens, H.F.M., De Mattos, M.J.T., Van Dam, K., Kruckeberg, A.L., 1999. Glucose uptake kinetics and transcription of HXT genes in chemostat cultures of *Saccharomyces cerevisiae*. J. Biol. Chem. 274, 15350–15359.
Dobson, P.D., Smallbone, K., Jameson, D., Simeonidis, E., Lanthaler, K., Pir, P., Lu, C., Swainston, N., Dunn, W.B., Fisher, P., Hull, D., Brown, M., Oshota, O., Stanford, N. J., Kell, D.B., King, R.D., Oliver, S.G., Stevens, R.D., Mendes, P., 2010. Further developments towards a genome-scale metabolic model of yeast. BMC Syst. Biol. 4 (Article no. 145).
Edwards, J.S., Palsson, B.Ø., 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc. Natl. Acad. Sci. U. S. A. 97, 5528–5533.
Egea, J.A., Balsa-Canto, E., 2009. Dynamic optimization of nonlinear processes with an enhanced scatter search method. Ind. Eng. Chem. Res. 48, 4388–4401.
Feng, X., Xu, Y., Chen, Y., Tang, Y.J., 2012. Integrating flux balance analysis into kinetic models to decipher the dynamic metabolism of *Shewanella oneidensis* MR-1. PLoS Comput. Biol. 8 (Article no. 2).
Förster, J., Famili, I., Fu, P., Palsson, B.Ø., Nielsen, J., 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. 13, 244–253.
Gao, J., Gorenflo, V.M., Scharer, J.M., Budman, H.M., 2008. Dynamic metabolic modeling for a MAB bioprocess. Biotechnol. Prog. 23, 168–181.
Gianchandani, E.P., Chavali, A.K., Papin, J.A., 2010. The application of flux balance analysis in systems biology. Wiley Interdiscip. Rev.: Syst. Biol. Med 2, 372–382.
Gurobi, 2013. version 5.5.0. Gurobi Optimization, Inc., Houston, Texas.
Hanly, T.J., Henson, M.A., 2011. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. Biotechnol. Bioeng. 108, 376–385.
Hanly, T.J., Henson, M.A., 2013. Dynamic metabolic modeling of a microaerobic yeast co-culture: predicting and optimizing ethanol production from glucose/xylose mixtures. Biotechnol. Biofuels 6 (Article no. 44).
Hanly, T.J., Urello, M., Henson, M.A., 2012. Dynamic flux balance modeling of *S. cerevisiae* and *E. coli* co-cultures for efficient consumption of glucose/xylose mixtures. Appl. Microbiol. Biotechnol. 93, 2529–2541.
Hao, H., Zak, D., Sauter, T., Schwaber, J., Ogunnaike, B., 2006. Modeling the VPAC2-activated cAMP/PKA signaling pathway: from receptor to circadian clock gene induction. Biophys. J. 90, 1560–1571.
Heavner, B.D., Smallbone, K., Barker, B., Mendes, P., Walker, L.P., 2012. Yeast 5—an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. BMC Syst. Biol. 6 (Article no. 55).
Heavner, B.D., Smallbone, K., Price, N.D., Walker, L.P., 2013. Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. Database 2013 (Article ID bat059).
Henry, C.S., DeJongh, M., Best, A.a., Frybarger, P.M., Linsay, B., Stevens, R.L., 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat. Biotechnol. 28, 977–982.
Herrgård, M.J., Swainston, N., Dobson, P., Dunn, W.B., Arga, K.Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novère, N., Li, P., Liebermeister, W., Mo, M.L., Oliveira, A.P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasić, I., Weichart, D., Brent, R., Broomhead, D.S., Westerhoff, H.V., Kirdar, B., Penttilä, M., Klipp, E., Palsson, B.Ø., Sauer, U., Oliver, S.G., Mendes, P., Nielsen, J., Kell, D.B., 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat. Biotechnol. 26, 1155–1160.
Hibbs, M.A., Hess, D.C., Myers, C.L., Huttenhower, C., Li, K., Troyanskaya, O.G., 2007. Exploring the functional landscape of gene expression: directed search of large microarray compendia. Bioinformatics 23, 2692–2699.
Hjersted, J.L., Henson, M.A., 2006. Optimization of fed-batch *Saccharomyces cerevisiae* fermentation using dynamic flux balance models. Biotechnol. Prog. 22, 1239–1248.
Hjersted, J.L., Henson, M.A., 2009. Steady-state and dynamic flux balance analysis of ethanol production by *Saccharomyces cerevisiae*. IET Syst. Biol. 3, 167–179.
Hjersted, J.L., Henson, M.A., Mahadevan, R., 2007. Genome-scale analysis of *Saccharomyces cerevisiae* metabolism and ethanol production in fed-batch culture. Biotechnol. Bioeng. 97, 1190–1204.
Höffner, K., Harwood, S.M., Barton, P.I., 2013. A reliable simulator for dynamic flux balance analysis. Biotechnol. Bioeng. 110, 792–802.
Holzhütter, H., 2004. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. Eur. J. Biochem. 271, 2905–2922.
Jacquez, J.A., Greif, P., 1985. Numerical parameter identifiability and estimability: integrating identifiability, estimability, and optimal sampling design. Math. Biosci. 77, 201–227.
Jaqaman, K., Danuser, G., 2006. Linking data to models: data regression. Nat. Rev. Mol. Cell Biol. 7, 813–819.
Jouhten, P., Wiebe, M., Penttilä, M., 2012. Dynamic flux balance analysis of the metabolism of *Saccharomyces cerevisiae* during the shift from fully respirative or respirofermentative metabolic states to anaerobiosis. FEBS J. 279, 3338–3354.
Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Assad-Garcia, N., Glass, J.I., Covert, M.W., 2012. A whole-cell computational model predicts phenotype from genotype. Cell 150, 389–401.
Keating, S.M., Bornstein, B.J., Finney, A., Hucka, M., 2006. SBMLToolbox: an SBML toolbox for MATLAB users. Bioinformatics 22, 1275–1277.
Kitano, H., 2002. Systems biology: a brief overview. Science 295, 1662–1664.
Kravaris, C., Hahn, J., Chu, Y., 2013. Advances and selected recent developments in state and parameter estimation. Comput. Chem. Eng. 51, 111–123.
Kruger, N., Ratcliffe, G., Steuer, R., 2007. Computational approaches to the topology, stability and dynamics of metabolic networks. Phytochemistry 68, 2139–2151.
Lai, L.-C., Kosorukoff, A.L., Burke, P.V., Kwast, K.E., 2006. Metabolic-state-dependent remodeling of the transcriptome in response to anoxia and subsequent reoxygenation in *Saccharomyces cerevisiae*. Eukaryot. Cell 5, 1468–1489.
Landaw, E.M., DiStefano III, J.J., 1984. Multiexponential, multicompartmental, and noncompartmental modeling. II. Data analysis and statistical considerations. Am. J. Physiol. 246, R665–R677.
Li, P., Vu, Q.D., 2013. Identification of parameter correlations for parameter estimation in dynamic biological models. BMC Syst. Biol. 7 (Article no. 91).
Lisha, K.P., Sarkar, D., 2014. Dynamic flux balance analysis of batch fermentation: effect of genetic manipulations on ethanol production. Bioprocess Biosyst. Eng. 37, 617–627.
Mahadevan, R., Edwards, J.S., Doyle III, F.J., 2002. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys. J. 83, 1331–1340.
MATLAB, 2013. version 8.1 (R2013a). The MathWorks Inc., Natick, Massachusetts.
Meadows, A.L., Karnik, R., Lam, H., Forestell, S., Snedecor, B., 2010. Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. Metab. Eng. 12, 150–160.
Mehra, A., Lee, K.H., Hatzimanikatis, V., 2003. Insights into the relation between mRNA and protein expression patterns: I. Theoretical considerations. Biotechnol. Bioeng. 84, 822–833.
Møller, K., Sharif, M.Z., Olsson, L., 2004. Production of fungal α-amylase by *Saccharomyces kluyveri* in glucose-limited cultivations. J. Biotechnol. 111, 311–318.
Nolan, R.P., Lee, K., 2011. Dynamic model of CHO cell metabolism. Metab. Eng. 13, 108–124.
Nookaew, I., Jewett, M.C., Meechai, A., Thammarongtham, C., Laoteng, K., Cheevadhanarak, S., Nielsen, J., Bhumiratana, S., 2008. The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. BMC Syst. Biol. 2 (Article no. 71).
Oberhardt, M.A., Palsson, B.Ø., Papin, J.A., 2009. Applications of genome-scale metabolic reconstructions. Mol. Syst. Biol. 5 (Article no. 320).
Oddone, G.M., Mills, D.A., Block, D.E., 2009. A dynamic, genome-scale flux model of *Lactococcus lactis* to increase specific recombinant protein expression. Metab. Eng. 11, 367–381.
Orellana, M., Aceituno, F.F., Slater, A.W., Almonacid, L.I., Melo, F., Agosin, E., 2014. Metabolic and transcriptomic response of the wine yeast *Saccharomyces cerevisiae* strain EC1118 after an oxygen impulse under carbon-sufficient, nitrogen-limited fermentative conditions. FEMS Yeast Res. 14, 412–424.
Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248.
Osterlund, T., Nookaew, I., Nielsen, J., 2012. Fifteen years of large scale metabolic modeling of yeast: developments and impacts. Biotechnol. Adv. 30, 979–988.
Park, J.M., Kim, T.Y., Lee, S.Y., 2009. Constraints-based genome-scale metabolic simulation for systems metabolic engineering. Biotechnol. Adv. 27, 979–988.
Petersen, B., Gernaey, K., Vanrolleghem, P.A., 2001. Practical identifiability of model parameters by combined respirometric–titrimetric measurements. Water Sci. Technol. 43, 347–355.
Pizarro, F., Varela, C., Martabit, C., Bruno, C., Pérez-Correa, J.R., Agosin, E., 2007. Coupling kinetic expressions and metabolic networks for predicting wine fermentations. Biotechnol. Bioeng. 98, 986–998.
Provost, A., Bastin, G., 2004. Dynamic metabolic modelling under the balanced growth condition. J. Process Control 14, 717–728.
Provost, A., Bastin, G., Agathos, S.N., Schneider, Y.-J., 2006. Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells. Bioprocess Biosyst. Eng. 29, 349–366.
Saa, P.A., Moenne, M.I., Pérez-Correa, J.R., Agosin, E., 2012. Modeling oxygen dissolution and biological uptake during pulse oxygen additions in oenological fermentations. Bioprocess Biosyst. Eng. 35, 1167–1178.
Sacher, J., Saa, P., Cárcamo, M., López, J., Gelmi, C., Pérez-Correa, J.R., 2011. Improved calibration of a solid substrate fermentation model. Electron. J. Biotechnol. 14 (Article no. 5).
Sainz, J., Pizarro, F., Pérez-Correa, J.R., Agosin, E., 2003. Modeling of yeast metabolism and process dynamics in batch fermentation. Biotechnol. Bioeng. 81, 818–828.
Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., Kang, J., Hyduke, D.R., Palsson, B.Ø., 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat. Protoc. 6, 1290–1307.
Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., Sauer, U., 2012. Multidimensional optimality of microbial metabolism. Science 336, 601–604.
Smith, W.R., Missen, R.W., 2003. Sensitivity analysis in chemical education: part 1. Introduction and application to explicit models. Chem. Eng. Educ. 2003, 222–227.
Sriram, K., Rodriguez-Fernandez, M., Doyle, F.J., 2012. Modeling cortisol dynamics in the neuro-endocrine axis distinguishes normal, depression, and post-traumatic stress disorder (PTSD) in humans. PLoS Comput. Biol. 8 (Article no. 2).
Stephanopoulos, G., Aristidou, A.A., Nielsen, J., 1998. Metabolic Engineering: Principles and Methodologies. Academic Press, San Diego, USA.

Tepeli, A., Hortaçsu, A., 2008. A fuzzy logic approach for regulation in flux balance analysis. Biochem. Eng. J. 39, 137–148.
Thiele, I., Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 5, 93–121.
Vargas, F.A., Pizarro, F., Pérez-Correa, J.R., Agosin, E., 2011. Expanding a dynamic flux balance model of yeast fermentation to genome-scale. BMC Syst. Biol. 5 (Article no. 75).
Varma, A., Palsson, B.Ø., 1993. Metabolic capabilities of *Escherichia coli* II. Optimal growth patterns. J. Theor. Biol. 165, 503–522.
Varma, A., Palsson, B.Ø., 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl. Environ. Microbiol. 60, 3724–3731.
Villadsen, J., Patil, K.R., 2007. Optimal fed-batch cultivation when mass transfer becomes limiting. Biotechnol. Bioeng. 98, 706–710.
Waldherr, S., Oyarzún, D.A., Bockmayr, A., 2013. Dynamic optimization of metabolic networks coupled with gene expression [WWW Document]. Cornell Univ. Libr. URL ⟨http://arxiv.org/abs/1309.4936⟩ (accessed 1.2.14.).
Yin, W., 2011. Gurobi Mex: A MATLAB interface for Gurobi.
Zhuang, K., Izallalen, M., Mouser, P., Richter, H., Risso, C., Mahadevan, R., Lovley, D.R., 2011. Genome-scale dynamic modeling of the competition between Rhodoferax and Geobacter in anoxic subsurface environments. ISME J. 5, 305–316.
Zomorrodi, A.R., Suthers, P.F., Ranganathan, S., Maranas, C.D., 2012. Mathematical optimization applications in metabolic networks. Metab. Eng. 14, 672–686.