Bioprocess and Biosystems Engineering (2018) 41:641–655
https://doi.org/10.1007/s00449-018-1899-y

RESEARCH PAPER

A systematic approach for finding the objective function and active constraints for dynamic flux balance analysis

Ali Nikdel[1] · Richard D. Braatz[2] · Hector M. Budman[1]

Received: 27 July 2017 / Accepted: 16 January 2018 / Published online: 1 February 2018

Abstract
Dynamic flux balance analysis (DFBA) has become an instrumental modeling tool for describing the dynamic behavior of bioprocesses. DFBA involves the maximization of a biologically meaningful objective subject to kinetic constraints on the rate of consumption/production of metabolites. In this paper, we propose a systematic data-based approach for finding both the biological objective function and a minimum set of active constraints necessary for matching the model predictions to the experimental data. The proposed algorithm accounts for the errors in the experiments and eliminates the need for ad hoc choices of objective function and constraints as done in previous studies. The method is illustrated for two cases: (1) for in silico (simulated) data generated by a mathematical model for *Escherichia coli* and (2) for actual experimental data collected from the batch fermentation of *Bordetella Pertussis* (whooping cough).

Keywords Dynamic flux balance analysis · Flux balance analysis · Metabolic networks · Metabolic engineering · Bioprocess modeling

Abbreviations

S	Stoichiometric matrix
$\boldsymbol{v_k}$	Vector of fluxes
n	Number of reactions
k	Time instance
ψ	Concentration
$\boldsymbol{J_T}$	Fluxes that satisfy tight constraints
$\boldsymbol{J_R}$	Fluxes that satisfy relaxed constraints
$\boldsymbol{u_w}$	Weight of sum of squared errors
I	Identity matrix
w_i^{sc}	Time-varying values of the weights for all the metabolites
w_i^{u}	Weight of upper bound
W^{u}	Maximum allowable value for w_i^{sc}
w_i^{l}	Weight of lower bound
N_C	Number of the objective functions' candidates
$\boldsymbol{n_g}$	Estimated noise in the growth rate
N_{SC}	Total number of metabolites
N_m	Number of measured metabolites
$V_{i,max}$	Maximum rate
K_i	Half saturation concentration
ε	Measurement error
X_k	The biomass value at time k
w_{c_i}	The weight coefficients of the objective function candidates

✉ Hector M. Budman
hbudman@uwaterloo.ca

[1] Department of Chemical Engineering, University of Waterloo, Waterloo, ON, Canada

[2] Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Introduction

Cell metabolism is generally described by a complex network of biochemical reactions involving different metabolites interacting with each other [1]. Different levels of cellular control orchestrate the evolution of these reactions within kinetic or thermodynamic constraints. Systems biology elucidates the combinatorial role of different level of interaction between the reactions inside the cell as a system and its response to the changes in the surrounding environment of the cell [2]. This systematic understanding is instrumental for optimizing the growth and productivity of existing cell lines or for designing new cell lines.

Various metabolic models have been presented for an understanding of biological systems, including constraint-based flux balance analysis (FBA) as an important category of metabolic models. FBA models are used to describe the fluxes at steady state based on the optimality assumption

of evolutionary biology whereby cells are using resources optimally to maximize or minimize a specific objective function to survive [3].

Since most of the biological processes are dynamic in their nature, e.g., batch operations, models that represent these processes during transients are sought. Dynamic flux balance analysis (DFBA) is the dynamic extension of flux balance analysis (FBA) models to have the ability to estimate the metabolites' concentrations over time [4, 5].

Models of dynamic biological systems have traditionally consisted of systems of differential equations, each describing a dynamic balance of a main nutrient or important byproduct that is being consumed or produced generally following Michaelis–Menten reaction kinetics. These models have been often found lacking since they involve a large number of kinetic parameters that need to be calibrated to fit the experimental data. Being based on an optimization with few limiting constraints, DFBA models are advantageous over these traditional dynamic models because they typically required a smaller number of calibration parameters [6, 7].

Typically, DFBA models involve the solution at each time interval k of the optimization:

$$\max_{v_k} \boldsymbol{c}^{\mathrm{T}} \boldsymbol{v_k}, \tag{1}$$

subject to:

$$g(\psi_k) \leq \left|\boldsymbol{S}\boldsymbol{v_k}\right| \leq \boldsymbol{f}(\psi_k)$$

$$\boldsymbol{v_k} \geq 0$$

$$\psi_{k+1} \geq 0$$

where $\boldsymbol{S}$ is the stoichiometric matrix, $\boldsymbol{v_k} = (v_1, \ldots, v_n)_k$ is a vector of fluxes (moles/h·mole of biomass) at time instance k, n is the number of reactions, and the biological objective function is expressed as a specific linear combination of fluxes, $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{v_k}$. This objective function is generally related to a biological variable that is representative of the overall evolution of the culture such as growth rate and ATP production. The constraints given by functions $\boldsymbol{f}$ and $\boldsymbol{g}$ that are dependent on specific metabolites' concentrations at time k ψ_k are describing kinetic limitations associated with the consumption or production of these metabolites.

Thus, the formulation of a DFBA model involves the choice of two main elements:

1. A biological meaningful objective function.
2. A set of limiting constraints.

The stoichiometric matrix ($\boldsymbol{S}$), generally available from public sources such as the Kyoto Encyclopaedia of Genes and Genomes (KEGG), describes the metabolic reactions among species for any microorganism of interest. The idea behind the use of the constraints in this model is that only a few key metabolites are limiting in terms of their rate of consumption or production whereas all other metabolites will follow the dynamic behavior of the limiting metabolites based on static stoichiometric relations among all metabolites.

The formulation of the model as a constrained optimization is based on the assumption that cells have evolved through time to act as optimizers that allocate their resources through maximizing/minimizing an objective function subject to some kinetic constraints.

The choice of a biological meaningful objective function [8] is of key importance for formulating an accurate DFBA that will not require a large number of associated constraints and calibrating parameters. The choice of the objective function has been reported to greatly impact the model prediction accuracy [8].

Common objective functions used in previous studies, mainly for bacterial systems, are the maximization of growth rate or final biomass at the end of a batch [3]. However, several other objective functions were found, by trial and error, to more accurately predict the system behavior such as minimization of the production rate of redox potential, minimization of ATP production rate, maximization of ATP production rate, maximization and minimization of nutrient uptake rate, maximization of biomass yield per unit flux, maximization of ATP yield (maximal energy efficiency), minimization of the overall intracellular flux, maximization of ATP yield per flux unit (maximizing ATP yield while minimizing enzyme usage), maximization of biomass yield per flux unit (maximizing biomass yield while minimizing enzyme usage), minimization of glucose consumption (more efficient usage of substrate), and minimization of reaction steps (minimization of the number of reaction steps for cell growth) [8–11]. Furthermore, it has been suggested that the objective function that rules the cell behavior may be a nonlinear combination of different specific objectives such as redox minimization, growth maximization, or ATP production [12] rather than a linear combination as used in earlier DFBA models.

In terms of the choice of a suitable objective function, mammalian cells might accomplish different functionalities during the course of the fermentation comparing to bacteria. For example, a previous study on hybridoma cell central metabolism [12] studied the three choices of objective function: (1) minimizing ATP production, (2) minimizing total nutrient uptake, and (3) minimizing redox metabolism through minimizing NADH production. That study concluded that, while no single objective could solely rule the cell behavior, minimizing the NADH production can better describe the typical characteristic behavior of hybridoma cells such as their inefficient use of nutrients. This inefficient use of resources translates into

higher rates of consumption of glutamine and production of alanine [13].

Some methods have been proposed for systematically finding the biological objective function. For example, an approach referred to as ObjFind has been reported [10, 11] that is based on the minimization of the sum of squared errors between identified fluxes from experimental and simulated data. A drawback of this method is that it involves a non-convex constrained optimization for which local optima were possible [11, 12]. Furthermore, the location of these multiple optima will be sensitive to the choice of limiting constraints.

Knorr et al. [11] developed a Bayesian-based method for selecting the most suitable objective function for *E. coli* growing on succinate and producing acetate. Among five possible objective functions, the minimization of the production rate of redox potential resulted in a better model in terms of its ability to fit the data, and was the only objective function capable of prediction of the acetate production by the cells.

The biological objective solution search (BOSS) method is an additional reported bi-level optimization algorithm for inference of the objective function that does not require assumptions of candidate objectives [7, 13]. However, this method was found to lead to overfitting of experimental data and to the finding of objective functions that do not have a particular biological meaning but are rather a particular combination of certain metabolic fluxes. Furthermore, the resulting bi-level optimization is non-convex and computationally expensive [8, 14].

FBA models are often referred as being "constraint-based" since they rely on biological constraints that represent the limited ability of the cell to consume certain nutrients or produce certain byproducts [15]. These constraints can be categorized into two groups: non-adjustable and adjustable. Non-adjustable constraints such as stoichiometric constraints, enzyme, and transporter capacities are time independent and describe intrinsic characteristics of the cells under consideration. On the other hand, kinetic and regulatory constraints may change as a function of environmental conditions and, therefore, are reasonable to adjust to match the model predictions to the experimental observations [15].

In an earlier work by the authors [16], the limiting constraints were found by inspecting the values of the Lagrange multipliers and assessing from them whether particular constraints are active or inactive at the solution. The disadvantage of that approach is that we had to decide on threshold values for the Lagrange multipliers as these values could not be systematically related to actual uncertainty/noise in the measurements.

An additional challenge in our earlier study was related to multiplicity of solutions that is also common in FBA models due to a large number of reactions that are considered versus the limited number of measured variables that are used to constrain the problem. In the literature, the addition of both ad hoc-chosen capacity constraints and/or thermodynamic constraints have been proposed to limit the solution space [17]. In our previous work, it was necessary to choose by trial and error a set of loose constraints to limit the solution space to address multiplicity.

In this work, we propose a novel optimization-based algorithm to systematically and simultaneously identify both the limiting adjustable constraints (that is, an additional set of constraints to limit the solution space) and the biological objective function. For the limiting constraints, the objective is to identify the metabolites for which a change in their uptake or production rates significantly affects the value of the objective function and as a result on the estimated flux values. Constraints that are used to limit the solution space are also systematically identified from the proposed optimization. The significance of all the constraints is directly related in this work to measurement error. For the objective function, we investigate various probable objective functions by considering them simultaneously within the proposed optimization to find the solution of this problem that results in the best fit for describing the data.

Then, after identifying the necessary constraints and biological objective function, a predictive model is formulated where the identified constraints are related to the corresponding metabolite concentrations and the future metabolite concentrations can be predicted using the fluxes resulting from the optimization into dynamic mass balances of all metabolites.

The paper is organized as follows. The mathematical algorithm for identifying the limiting constraints and objective function is presented in "Identification algorithm" and the following section formulates a predictive model of metabolites based on the identified objective function and constraints. The subsequent section presents experimental methods used to collect data for the *B. Pertussis* case study and how to generate the in silico data used for the *E. coli* study following which the results are explained. The final section concludes the paper.

Theoretical and experimental methods

The development of the DFBA model in this study involves two steps:

1. An identification algorithm of limiting constraints and the biological objective function.
2. The development of a predictive DFBA model based on the identified constraints and biological objective function.

These two steps are discussed separately below.

Identification algorithm

The proposed identification algorithm was formulated to meet three main goals:

1. Formulate the identification problem as a one-level optimization.
2. Find limiting constraints such as the level of fit that can be directly related to the measurement error.
3. Integrate into the algorithm a search for a biologically motivated objective function that results in best fit between data and model predictions.

Regarding Goal 1, as mentioned in the Introduction, some approaches that were proposed for identifying DFBA models are based on bi-level optimization formulations which can be numerically challenging due to the existence of local minima. This problem is especially important for the objective of simultaneously identifying the objective function and the set of limiting constraints. Bi-level formulations derive from the need to minimize some measure of the quality of fit of model predictions and experiments while maximizing a biological objective as required in DFBA formulations as shown in Eq. (1). The measure of the fit, e.g., the sum of squared errors, is calculated with the input data which generally consist of the concentrations of extracellular metabolites at different time intervals. To avoid a bi-level formulation, a key idea for the approach proposed in the current study is that the data (concentrations of extracellular metabolites) are represented by convex sets whereby, at each time interval, data points are limited by upper and lower bounds. For example, Fig. 1 shows the evolution of glucose concentration during a batch where at each time interval the glucose level lies between an upper and lower limit.

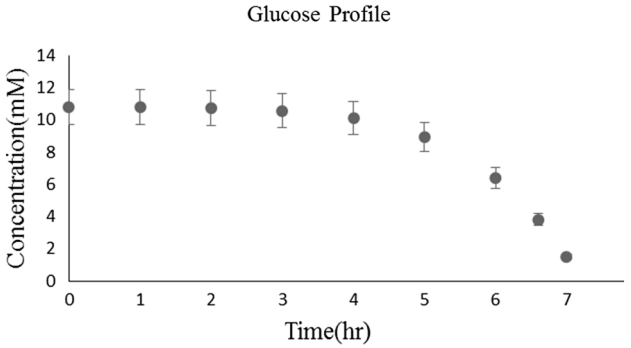


Fig. 1 Convex set constraints for glucose concentrations of *E. coli* diauxic metabolism with 10% error bound

The use of convex sets to represent biological experimental data has been recently reported in several studies dealing with identification of biological models [16, 18]. In this approach, the experimental data within the convex set are generally assigned the same probability to occur. Experimental data in biological experiments are highly variable due to measurement noise or lack of exact repeatability due to unmeasured disturbances such as variability in initial conditions. Then, if the number of experimental repeats is not large, it is reasonable to ascribe the same probability to the values contained between the bounds at each time interval as shown in Fig. 1.

Since the convex sets can be formulated as inequality constraints within an optimization, there is no need in our proposed approach to use a dedicated optimization level for minimizing the errors in fitting between data and model predictions. Hence, using set constraints, a bi-level optimization formulation is avoided.

Regarding Goal 2, to identify a model that is robust with respect to uncertainty in measurements, the key idea is to perform an optimization with respect to two different sets of fluxes: one set of fluxes ($\boldsymbol{J}_{\mathbf{T}}$) that satisfies the convex sets' inequalities without measurement uncertainty and a second set of fluxes ($\boldsymbol{J}_{\mathbf{R}}$) that satisfies these inequalities with an accuracy related to the measurement error. This means that the original convex sets (used to solve $\boldsymbol{J}_{\mathbf{T}}$) represent lack of repeatability of batches due to unmeasured disturbances, e.g., unmeasured changes in inoculum amount and population content, unmeasured changes in minor species, etc.

In addition to these unmeasured disturbances there is measurement noise of magnitude epsilon. The flux vector $\boldsymbol{J}_{\mathbf{R}}$ is solved by satisfying the original convex sets with the added noise.

Then, we seek a minimal set of constraints by adding to the optimization objective one term that it is related to the error between $\boldsymbol{J}_{\mathbf{T}}$ and $\boldsymbol{J}_{\mathbf{R}}$ and additional terms that penalize the activation of the limiting constraint. By minimizing the resulting objective function, we are minimizing the number of constraints that will be active at the solution while at the same time minimizing the error between the ideal set of fluxes $\boldsymbol{J}_{\mathbf{T}}$ and the uncertain set of fluxes $\boldsymbol{J}_{\mathbf{R}}$.

Regarding Goal 3, the search for the most suitable biologically meaningful objective function for solving the DFBA is conducted by adding different candidate functions, e.g., growth rate, redox potential etc., into the overall objective. Then, to force the solution towards one specific candidate, the cross-products of the candidates' biological functions are also added to the objective function such as, at the solution of the minimization, these cross-products will be driven to zero and only one particular candidate will be left.

Following the above considerations, the resulting optimization is

$$\min_{\boldsymbol{J_R},\boldsymbol{J_T},\boldsymbol{w^L},\boldsymbol{w^U},\boldsymbol{w_c}} (\boldsymbol{J_T}-\boldsymbol{J_R})^{\mathrm{T}} u_{\mathrm{w}} (\boldsymbol{J_T}-\boldsymbol{J_R}) - \sum_{i=1}^{N_{\mathrm{C}}} (w_{\mathrm{c}_i}) \boldsymbol{c_i}^{\mathrm{T}} \boldsymbol{J_T} - \sum_{i=1}^{N_{\mathrm{C}}} (w_{\mathrm{c}_i}) \boldsymbol{c_i}^{\mathrm{T}} \boldsymbol{J_R} + \frac{1}{2} \sum_{\substack{i=1 \\ j=1 \\ i \neq j}}^{N_{\mathrm{C}}} (w_{c_i})(w_{c_j}) + \left(\sum_{i=1}^{N_{\mathrm{m}}} w_i^L - \sum_{i=1}^{N_{\mathrm{m}}} w_i^{\mathrm{U}} \right) \frac{n_{\mathrm{g}}}{\varepsilon N_{\mathrm{m}}} \quad (2a)$$

subject to:

$$\left.\frac{\mathrm{d}\psi_i}{\mathrm{d}t}\right|^{\mathrm{l}} w_i^{\mathrm{l}} \leq \boldsymbol{S_i}\,\boldsymbol{J_R} X_k \leq \left.\frac{\mathrm{d}\psi_i}{\mathrm{d}t}\right|^{\mathrm{u}} w_i^{\mathrm{u}} \quad (2b)$$

$$\left.\frac{\mathrm{d}\psi_i}{\mathrm{d}t}\right|^{\mathrm{l}} \leq \boldsymbol{S_i}\,\boldsymbol{J_T} X_k \leq \left.\frac{\mathrm{d}\psi_i}{\mathrm{d}t}\right|^{\mathrm{u}} \quad (2c)$$

$$1 \leq w_i^{\mathrm{u}} \leq 1 + \varepsilon \quad (2d)$$

$$1 - \varepsilon \leq w_i^{\mathrm{l}} \leq 1 \quad (2e)$$

$$\boldsymbol{J_R} \geq 0 \quad (2f)$$

$$\boldsymbol{J_T} \geq 0 \quad (2g)$$

$$\psi_i^{k+1} \geq 0 \quad (2h)$$

$$0 \leq w_{\mathrm{c}_i} \leq 1 \quad (2i)$$

$$\sum_{i=1}^{N_{\mathrm{c}}} w_{\mathrm{c}_i} = 1 \quad (2j)$$

We defined $\left.\frac{\mathrm{d}\psi_i}{\mathrm{d}t}\right|^{\mathrm{l}} = \frac{\psi_i^{\mathrm{l},k+1} - \psi_i^k}{\psi \mathrm{T}}$ and $\left.\frac{\mathrm{d}\psi_i}{\mathrm{d}t}\right|^{\mathrm{u}} = \frac{\psi_i^{\mathrm{u},k+1} - \psi_i^k}{\psi \mathrm{T}}$. The first term in the objective function (2a) is a quadratic term that minimizes the sum of square errors between the assumed two sets of fluxes $\boldsymbol{J_T}$ and $\boldsymbol{J_R}$, $\boldsymbol{J_T}$ represents the fluxes that satisfy tight constraints on metabolites' concentrations or consumption and production rates of these metabolites as observed from the limited data available, and $\boldsymbol{J_R}$ are the fluxes that satisfy relaxed constraints for which it is assumed that the data are uncertain. The $\boldsymbol{S_i}$ is the row of the stoichiometric matrix representing metabolite i, and X_k is the biomass value at time k.

The sum of square errors between the elements of the vector $\boldsymbol{J_T}$ and the elements of the vector $\boldsymbol{J_R}$ weighted by the scalar u_{w} that is selected sufficiently large so as for at the solution the first term in (2a) is smaller than a tolerance (10^{-4}) that is of the order of the noise n_{g} in the biological objective (e.g., the measurement noise of the growth rate if the latter is used as the biological objective). If u_{w} is too large the difference between $\boldsymbol{J_R}$ and $\boldsymbol{J_R}$ is small but the results are conservative since the model outputs do not span the allowable set based bounds with noise ε requiring the use of many active constraints. If we select the u_{w} too small, the difference between the models results and the data was too large and the simulation becomes unfeasible at some time interval.

To account for the uncertainty in consumption/production rates, the bounds on the relaxed constraints are obtained by multiplying the bounds on the tight constraints by weights w_i^{u} and w_i^{l} in Eq. 2a. These weights change between 1 and $1+\varepsilon$ or $1-\varepsilon$ and 1 for the upper or lower constraints, respectively (Eqs. 2d, 2e) where ε is chosen to reflect the expected uncertainty in the measurements of consumption/production rates of metabolites.

The second and third terms in the cost function (2a) describe the biological objective functions to be maximized in a DFBA model computed as a function of either the $\boldsymbol{J_T}$ or the $\boldsymbol{J_R}$ vectors of fluxes that satisfy either the tight and relaxed constraints, respectively. N_{C} is the number of the objective functions' candidates for the problem and the w_{c_i} are the weight coefficients corresponding to the candidates of the objective function.

The fourth term in Eq. (2a) is cross-product of the corresponding weights of any two candidate biological objective functions. The minimization in (2a) is expected to force these cross-products close to zero values thus only leaving one nonzero value of the weights w_{c_i}. This nonzero value of a w_{c_i} will indicate which one of the candidate biological objectives is the one to be chosen for the problem, i.e., the goal is that at each time interval only one of the weight coefficients which has the highest impact will be driven to one and the rest will tend to zero. Mathematically, there may be a situation that the resulting bilinear optimization will give fractional values, especially in the case that the sum of the cross-products of the weights is not significant as compared to the other terms in the objective function of the problem. If this happens it could be resolved by weighing the cross-products with a higher scalar weight.

The fifth term in Eq. (2a) is to drive as many constraints as possible to non-active status, i.e., the redundant ones so as to identify only the necessary ones. It should be noticed that $1 \leq w_i^{\mathrm{u}} \leq 1+\varepsilon$ and $1-\varepsilon \leq w_i^{\mathrm{l}} \leq 1$ and these inequalities imply that the extreme values of the weights are $w_i^{\mathrm{u}} = 1 + \varepsilon$ and the smallest $w_i^{\mathrm{l}} = 1-\varepsilon$. Thus, the minimization of the 5th term $\left(\sum_{i=1}^{N_{\mathrm{m}}} w_i^{\mathrm{L}} - \sum_{i=1}^{N_{\mathrm{m}}} w_i^{\mathrm{U}}\right)$ will drive

as many as possible w_i^{l} to $1-\varepsilon$, or for w_i^{u} to $1+\varepsilon$, respectively, and thus, when they are active at these extreme values, the corresponding constraints are not necessary (not active) for the given noise. Hence, if one can expand the bounds on the consumption/production rates up to the limit of the noise without affecting the error $\boldsymbol{J_T}-\boldsymbol{J_R}$ in a significant way, this means that the corresponding constraints are not necessary to explain the data. Since the fifth term is of the order of $\varepsilon N_{\mathrm{m}}$ (the ones are cancelling out with each other), we must introduce the factor, $\frac{n_{\mathrm{g}}}{\varepsilon N_{\mathrm{m}}}$, which ultimately renders this term of the order of the assumed noise in the growth rate of magnitude n_{g}.

In fact, the first and the fifth terms are introduced in the objective function of the optimization (Eq. 1) to achieve a tradeoff between the error $\left(\boldsymbol{J_T}-\boldsymbol{J_R}\right)$ and the limiting constraints active at the solution of $\boldsymbol{J_T}$ and $\boldsymbol{J_R}$.

Constraints (2b) and (2c) are the key elements for the development of a predictive DFBA model presented in the next subsection. The idea is to find for which amino acid constraints (2b) and (2c) become active and to eventually replace these constraints by their corresponding kinetics functions to be able to predict future concentrations. More details of this process has been explained in "Predictive DFBA model".

The objective function in (2a) is nonlinear involving a bilinear term and a quadratic term, and the constraints are all linear. Since the bilinear term involves products of weights that are each bounded between 0 and 1, the term is convex. Thus, overall this formulation results in a nonlinear convex optimization that can be solved by common optimization solvers such as *fmincon* in *MATLAB*. Accordingly, the optimization search must be conducted for different initial guesses.

For the special case that the biological objective function is known or assumed a priori and only the active constraints need to be identified, the nonlinear optimization of Eq. (2) that results from the presence of bilinear terms [terms 2 and 3 in Eq. (2a)] converts into the simple quadratic optimization:

$$\min_{\boldsymbol{J_R},\,\boldsymbol{J_T},\boldsymbol{w^L},\boldsymbol{w^U}} (\boldsymbol{J_T}-\boldsymbol{J_R})^{\mathrm{T}} u_{\mathrm{w}} \left(\boldsymbol{J_T}-\boldsymbol{J_R}\right) - \boldsymbol{c}^{\mathrm{T}}(\boldsymbol{J_T}+\boldsymbol{J_R}) + \left(\sum_{i=1}^{N_m} w_i^{\mathrm{L}} - \sum_{i=1}^{N_m} w_i^{\mathrm{U}}\right)\frac{n_{\mathrm{g}}}{\varepsilon N_{\mathrm{m}}} \tag{3a}$$

subject to:

$$\left.\frac{\mathrm{d}\psi_i}{\mathrm{d}t}\right|^{\mathrm{l}} w_i^{\mathrm{l}} \le \boldsymbol{S_i}\,\boldsymbol{J_R} X_k \le \left.\frac{\mathrm{d}\psi_i}{\mathrm{d}t}\right|^{\mathrm{u}} w_i^{\mathrm{u}} \tag{3b}$$

$$\left.\frac{\mathrm{d}\psi_i}{\mathrm{d}t}\right|^{\mathrm{l}} \le \boldsymbol{S_i}\,\boldsymbol{J_T} X_k \le \left.\frac{\mathrm{d}\psi_i}{\mathrm{d}t}\right|^{\mathrm{u}} \tag{3c}$$

$$1 \le w_i^{\mathrm{u}} \le 1+\varepsilon \tag{3d}$$

$$1-\varepsilon \le w_i^{\mathrm{l}} \le 1 \tag{3e}$$

$$\boldsymbol{J_R} \ge 0 \tag{3f}$$

$$\boldsymbol{J_T} \ge 0 \tag{3g}$$

$$\psi_i^{k+1} \ge 0 \tag{3h}$$

This quadratic optimization can be solved, for example, using *cplexqp* (**IBM CPLEX**) in a MATLAB-compatible code. If the fluxes $\boldsymbol{J_R}=\boldsymbol{J_T}$ at the solution of the QP then the quadratic term in (3a) vanishes and the solution is identical to the solution of an LP that corresponds to the original formulation of a DFBA model (Eq. 1) augmented by the second term that represents added uncertainty to the biological objective as explained before in this section.

A main negative consequence of the multiplicity of solutions is that, for some problems such as the *B. Pertussis* case study in this work, some of the rate constraints in Eqs. (2b) and (2c) become active only on isolated time intervals along a batch fermentation and this sporadic activation of constraints depends highly on the initial guesses assumed for the optimization. On the other hand, it should be remembered that one of the expected benefits of the DFBA model is to predict concentrations using only a few kinetic constraints so a small number of kinetic parameters would be needed. The sporadic activation of many constraints defeats this purpose since it would require constraining many amino acids and calibration of many corresponding kinetic expressions.

A way to address this situation is by limiting the solution space for the optimization through the imposition of upper bounds on consumption or production rates of metabolites. Sometimes, these bounds are available from the literature. In the case that these bounds are not available, we propose to find the bounds from a modification of the optimization in (2a):

$$\min_{\boldsymbol{J_R},\boldsymbol{J_T},\boldsymbol{w^L},\boldsymbol{w^U},\boldsymbol{w_c},\boldsymbol{w^{sc}}} (\boldsymbol{J_T}-\boldsymbol{J_R})^{\mathrm{T}} u_{\mathrm{w}} (\boldsymbol{J_T}-\boldsymbol{J_R}) - \sum_{i=1}^{N_C} (w_{c_i}) \boldsymbol{c_i}^{\mathrm{T}} \boldsymbol{J_T} - \sum_{i=1}^{N_C} (w_{c_i}) \boldsymbol{c_i}^{\mathrm{T}} \boldsymbol{J_R} + \frac{1}{2} \sum_{\substack{i=1 \\ j=1 \\ i \neq j}}^{N_C} (w_{c_i})(w_{c_j}) + \left(\sum_{i=1}^{N_m} w_i^{\mathrm{L}} - \sum_{i=1}^{N_m} w_i^{\mathrm{U}} \right) \frac{n_g}{\varepsilon N_m} - \left(\sum_{i=1}^{N_{sc}} w_i^{sc} \right) \frac{n_g}{W^{\mathrm{U}} N_{sc}} \tag{4a}$$

subject to:

$$\left| \frac{d\psi_i}{dt} \right|^{l} w_i^{l} \leq \boldsymbol{S_i} \boldsymbol{J_R} X_k \leq \left| \frac{d\psi_i}{dt} \right|^{u} w_i^{u} \tag{4b}$$

$$\left| \frac{d\psi_i}{dt} \right|^{l} \leq \boldsymbol{S_i} \boldsymbol{J_T} X_k \leq \left| \frac{d\psi_i}{dt} \right|^{u} \tag{4c}$$

$$1 \leq w_i^{u} \leq 1 + \varepsilon \tag{4d}$$

$$1 - \varepsilon \leq w_i^{l} \leq 1 \tag{4e}$$

$$\boldsymbol{J_R} \geq 0 \tag{4f}$$

$$\boldsymbol{J_T} \geq 0 \tag{4g}$$

$$\psi_i^{k+1} \geq 0 \tag{4h}$$

$$0 \leq w_{c_i} \leq 1 \tag{4i}$$

$$\sum_{i=1}^{N_c} w_{c_i} = 1 \tag{4j}$$

$$\left| \boldsymbol{S_i} \boldsymbol{J_T} X_k \right| \leq w_i^{sc}(k) \tag{4k}$$

$$0 \leq w_i^{sc}(k) \leq W^{u} \tag{4l}$$

The optimization (4a), in contrast with (2a), includes a sixth term in the objective function whose purpose is to maximize at each instance the absolute consumption/production rate of each metabolite (defined in 4k) towards an upper bound W^{u} as defined by (4l). It can be easily verified that the additional term in (4a) is also of the order of the noise, n_g since the term is normalized by the maximum allowable value W^{u} (Eq. 4l) and by the total number of metabolites N_{SC}. To constrain the solution space, it has been often proposed in the FBA literature to bound all the fluxes by a value of the order of the largest known consumption rate of the limiting nutrients [17]. In this work, we have used a similar criterion to choose the value of W^{u}. In principle, this value can be iteratively adjusted by executing the optimization in (4a)–(4e) for different values of W^{u} until a satisfactory fitting between the vectors of fluxes $\boldsymbol{J_T}$ and $\boldsymbol{J_R}$ is obtained.

The fundamental difference between constraints (4b)–(4c) with respect to the constraints (4k)–(4l) is that the former are tight constraints of the order of the noise (ε) in the metabolite consumption/production rates that are frequently active along a batch whereas the latter are loose rate constraints related to the noise n_g in the biological objective that are only important in isolated time intervals along the batch. In principle, the additional term in (4a) together with constraints (4i)–(4l) results in time-varying values of the weights $w_i^{sc}(k)$ for each metabolite i. For the purpose of limiting the solution space of a predictive model, one time-independent upper bound is found for each metabolite $\boldsymbol{i}$ from

$$w_i^{sc,max} = \max \left[w_i^{sc}(k) \right], \tag{5}$$

where k ranges from 0 to T. Selecting the maximum bound is justified since, for the worst case that all $\boldsymbol{w_i^{sc}} = W^{u}$, the effect of the sixth term in (4a) is guaranteed to be no larger than n_g, i.e., the noise in the biological objective function. Then, the bound $w_i^{sc,max}$ for each metabolite $\boldsymbol{i}$ is used as an upper bound for the rate of that metabolite in the predictive model that is presented in the next section.

Predictive DFBA model

The predictive DFBA model is given by

At each time instance k ($k = 1, \ldots, T$)

$$\max_{\boldsymbol{v_k}} \boldsymbol{c}^{\mathrm{T}} \boldsymbol{v_k}, \tag{6a}$$

subject to:

$$\left| \boldsymbol{S_i} \boldsymbol{v_k} \right| \leq \frac{V_i, \max . \psi_i^k}{K_i + \psi_i^k} \frac{\mathrm{mmol}}{\mathrm{g\ dw\ h}} (i = 1, 2, \ldots, \mathrm{N_m};\ N_m \text{ is the number of metabolites}) \tag{6b}$$

$$\boldsymbol{v_k} \geq 0 \tag{6c}$$

$$\left|\frac{d\psi_i}{dt}\right| \leq w_i^{sc,max} \tag{6d}$$

$$\psi_i^{k+1} \geq 0 \tag{6e}$$

$$\psi_i^{k+1} = \psi_i^k + \boldsymbol{S_i}\mathbf{v_k}\boldsymbol{X_k}\boldsymbol{\Delta T} \geq 0 \tag{6f}$$

where $\boldsymbol{S}$ is the stoichiometric matrix and $\boldsymbol{S_i}$ is the row of stoichiometric matrix associated with the metabolite i. Equations (6a)–(6f) are a modified version of the optimization (4a)–(4k) with respect to four elements:

1. The objective function (6a) includes only the biological objective (second term in 4a). This objective is based on the assumption that, at the optimal solution of (4a), the first term vanishes since, at $\boldsymbol{J_T} \approx \boldsymbol{J_R}$, the third term becomes identical to the second term, the fourth and sixth terms are neglected since they are of the order of the noise n_g, and the fifth term also vanishes since only one $w_{c_1} = 1$ and the rest are zero.
2. The limiting constraints identified, as explained above, the upper and lower bounds in (4b) and (4c), are replaced by explicit functions (6b) of the corresponding metabolite concentrations at time k.
3. Constraints (4i) and (4l) are replaced by upper bound constraints (6d) using the values $w_i^{sc,max}$ calculated from (5).
4. Equations to predict the concentrations of metabolites over time (Eq. 6f) are formulated based on mass balances, that is, Eq. (6f) is time-discretized mass balances for all metabolites that are used to integrate the concentrations over time.

The optimization (6a)–(6f) is an LP that is solved with the CPLEX optimization software. A fundamental advantage of the proposed model (6a)–(6f) is that the parameters in the kinetic expressions corresponding to the limiting constraints (6b) can be separately identified for each metabolite. This is particularly advantageous as compared to previously reported algorithms where the parameters corresponding to all the kinetic expressions involved in the constraints of the DFBA problem must be simultaneously identified from bi-level optimization formulations.

The ability to identify the kinetic expression separately is due to the fact that, at each time instance, the solution of the optimization (4a) provides which constraint i is active; it is possible to find the corresponding values of the gradients, either $\frac{d\psi_i}{dt}\Big|^u$ or $\frac{d\psi_i}{dt}\Big|^l$, that are active at the constraint. Then, Michaelis–Menten expression

$$\frac{d\psi_i}{dt} = \frac{V_{i,\max}\cdot\psi_i}{K_i + \psi_i}X \tag{7a}$$

or alternative types of kinetic expressions can be found to correlate the bounds on the uptakes or production rates $\left(\frac{d\psi_i}{dt}\right)$ of the metabolite as a function of its concentration(ψ_i) at every time interval, where $V_{i,\max}$, K_i are the kinetic parameters that should be identified. For example, if it is found that glucose is a limiting constraint, then a kinetic expression can be used to represent the relation:

$$\frac{dGLC}{dt} = \frac{V_{GLC,\max} \times GLC}{K_{GLC} + GLC}X \tag{7b}$$

Then, the two parameters $V_{GLC,max}$ and K_{GLC} can be found from time data of glucose alone since the gradient of glucose on the left-hand side of Eq. (5) and the corresponding concentration of glucose at each time interval are available.

If Michaelis–Menten kinetics are assumed, graphical methods such as the Lineweaver–Burk are used to identify the parameters in the kinetic expressions (7a) or (7b) [19]. For other types of kinetic expressions, nonlinear regression is done using the Curve Fitting Toolbox in the MATLAB environment.

Experimental and in silico data used in the case studies

The proposed method is applied in two case studies for two model microorganisms. The first study is the batch fermentation of *E. coli* system for which a DFBA model has been reported by Mahadevan et al. [7]. The model was used to generate in silico data consisting of simulated data with superimposed white noise. Since the DFBA model with the required biological objective function and constraints was available a priori, this case study serves to verify whether our methodology is able to identify the correct objective function and limiting constraints.

For the second case study, we studied the fermentation of *Bordetella pertussis (B. Pertussis)* used to produce the antigens of the whooping cough vaccine. The metabolic network for this organism involves 49 reactions. This process was assumed to operate initially in batch and, after depletion of the main nutrient, in fed-batch mode. Glutamate was the main nutrient for this microorganism. For this example, we use actual measurements of sixteen amino acids' concentrations measured by HPLC.

Results

Case study 1: *E. coli*

The DFBA model describing the diauxic growth of an *E. coli* system involves four species: glucose, acetate, oxygen, and the biomass in its network. All species except the biomass are considered as potentially active constraints.

The model proposed by Mahadevan et al. [7] is based on a simplified metabolic network of four reactions involving the four aforementioned species shown in "Appendix I". Mahadevan's model assumes that the growth rate is maximized at each time interval subject to two kinetic limiting constraints on glucose and oxygen consumptions.

At all time instances k:

$$\max_{\boldsymbol{v}} \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v} \tag{8}$$

$$\left|\mathbf{S}_{\mathbf{Glc}}\boldsymbol{v}\right| \leq \frac{V_{\mathrm{Glc,max.}}\psi_{\mathrm{Glc},k}}{K_{\mathrm{Glc}}+\psi_{\mathrm{Glc},k}}\frac{\mathrm{mmol}}{\mathrm{gdw\ h}}$$

$$\left|\mathbf{S}_{\mathbf{O}_2}\boldsymbol{v}\right| \leq \mathrm{OUR}\ \frac{\mathrm{mmol}}{\mathrm{gdw\ h}}$$

This model was simulated for a total of 10 h to simulate experimental conditions. Gaussian noise was added to the metabolite responses at a level of 5% of the total range of change of each metabolite during the batch. These simulated values were used as in silico data to test the proposed identification methodology as given by Eqs. (4a)–(4e). The sampling interval used for discretization was 0.1 h and the oxygen uptake rate was OUR = 10. Initial condition for metabolites and biomass at time zero was glucose (10.4 mM), acetate (0.4 mM), oxygen (0.21 mM) and biomass (0.001 g). Based on this level of noise, ε was set equal to 0.1 and the upper bound in constraint (4 k), W^{u}, was set to 10 mM/h gdw. The weight u_{w} as per Eq. (2j) was set to 0.01. As mentioned in the methodology section, this weight is selected large enough such as the first term in (4a) at the optimum is of the order of n_{g} (estimated noise in the growth rate). We used interior-point algorithms with cplexqp and cplexlp function in CPLEX for MATLAB toolbox.

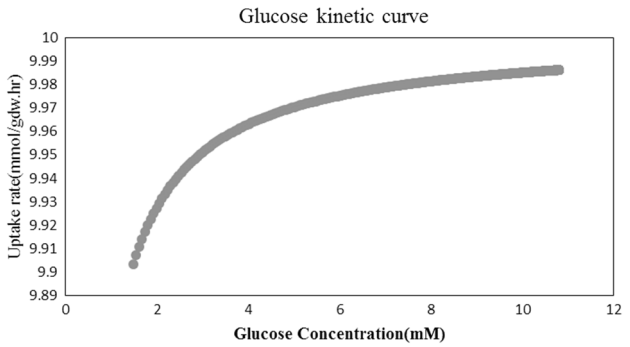


Fig. 2 Kinetic curve of glucose

To test our methodology, we assume that either one of the nutrients (glucose, acetate, or oxygen) can be potentially limiting. In addition, we assume two possible candidate biological objectives for maximization: (1) growth rate as given by Eq. (8), i.e., $\boldsymbol{c}_1{}^{\mathrm{T}} = [1\ 1\ 1\ 1]$ (see Eq. 4a) (2) maximal consumption rate of acetate, which is represented by $\boldsymbol{c}_2{}^{\mathrm{T}} = [39.43\quad 0\quad -1.24\quad -12.12]$ in Eq. (4a).

The solution over time of the weights w_i^{u} and w_i^{l} for all the species is identified. These values indicate which rate constraints are considered at each time interval according to Eqs. (4b) and (4c). Following the solution of problem (4a), it was found that the upper bound of oxygen depletion rate is active from $t = 0$ until the end of the batch whereas the upper bound in glucose becomes active after $t = 4.2$ h and thereafter. These results coincide with Mahadevan's model where the limiting rate constraints are glucose and oxygen while acetate is not limiting, thus corroborating the ability of the method to identify the constraints.

The weights involved in constraints (4k) and (4l) which are introduced to limit the solution space were found to be equal to their upper bound, i.e., $w_i^{\mathrm{sc}} = W^{\mathrm{u}}$, for all the four species and for the entire duration of the batch. The weights w_{c_1} and w_{c_2} that correspond to the growth rate and to the acetate objectives in Eq. (4a) are 1 and 0 for the entire duration of the batch, thus correctly identifying the growth rate as the objective to be maximized. This again verifies the ability of the proposed approach to identify the correct biological objective function. For the purpose of formulating the predictive model according to Eqs. (6a)–(6f), it is necessary to identify kinetic expressions of the two identified limiting rate constraints as a function of the corresponding metabolite concentrations, i.e., the rate constraints in glucose and oxygen are expressed as a function of the corresponding glucose and oxygen concentrations respectively. Figure 2 shows the consumption rate of glucose obtained at each time interval as a function of the corresponding glucose concentration. For this plot, a Michaelis–Menten kinetic expression was identified using the Lineweaver–Burk graphical method. The resulting kinetic parameters for the glucose are as in the following table. It should be noted that these results are almost identical to the non-linear parametric estimation.

Parameter	Original value	Estimated value
$V_{\mathrm{Glc,max}}$	10	11.11
k_{Glc}	0.015	0.02

The procedure identifies, in accordance with Mahadevan's model used to generate the in silico data, Michaelis–Menten

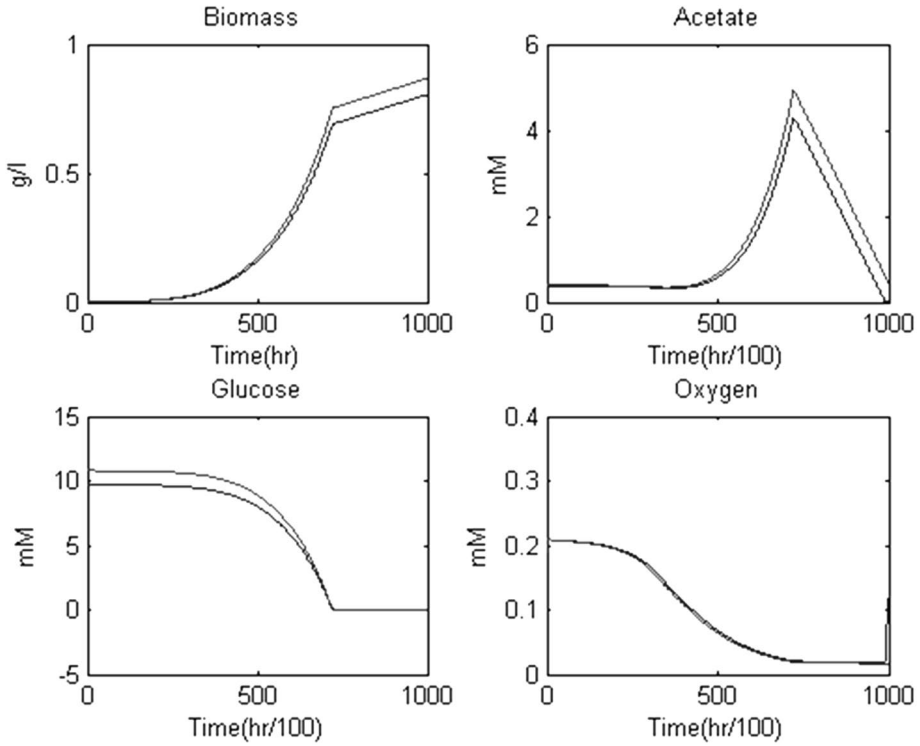


Fig. 3 Metabolite concentration profiles of *E. coli* case study red line is the simulated original model and the blue line is our identified model

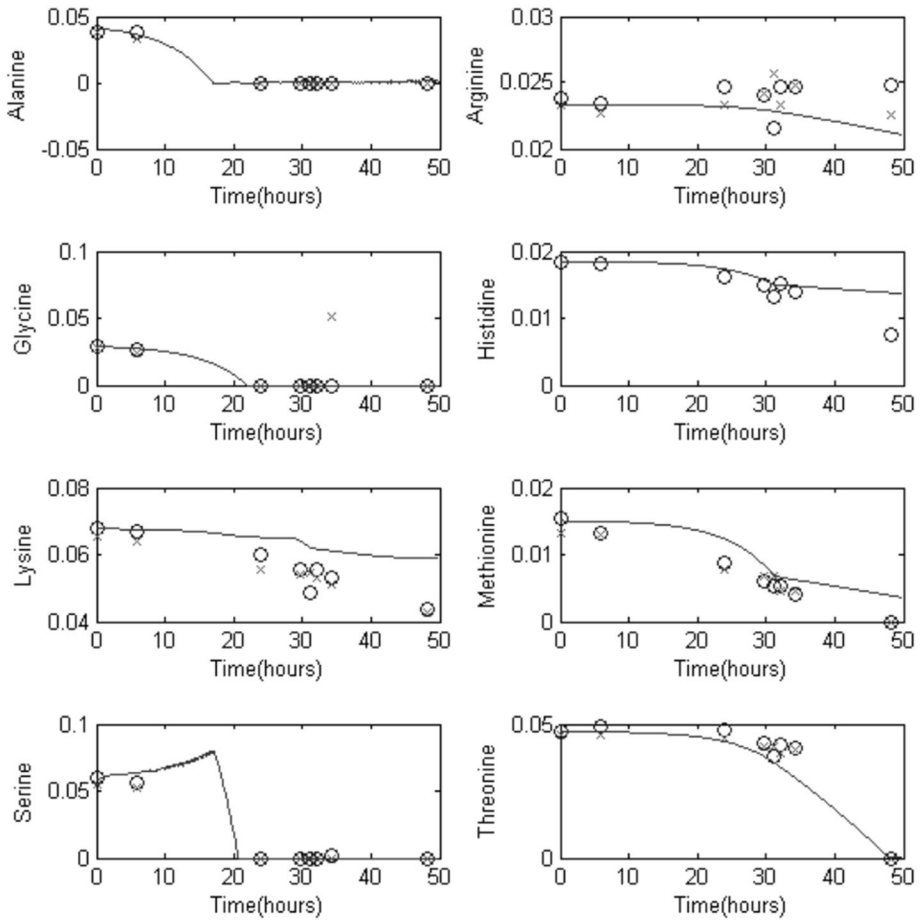
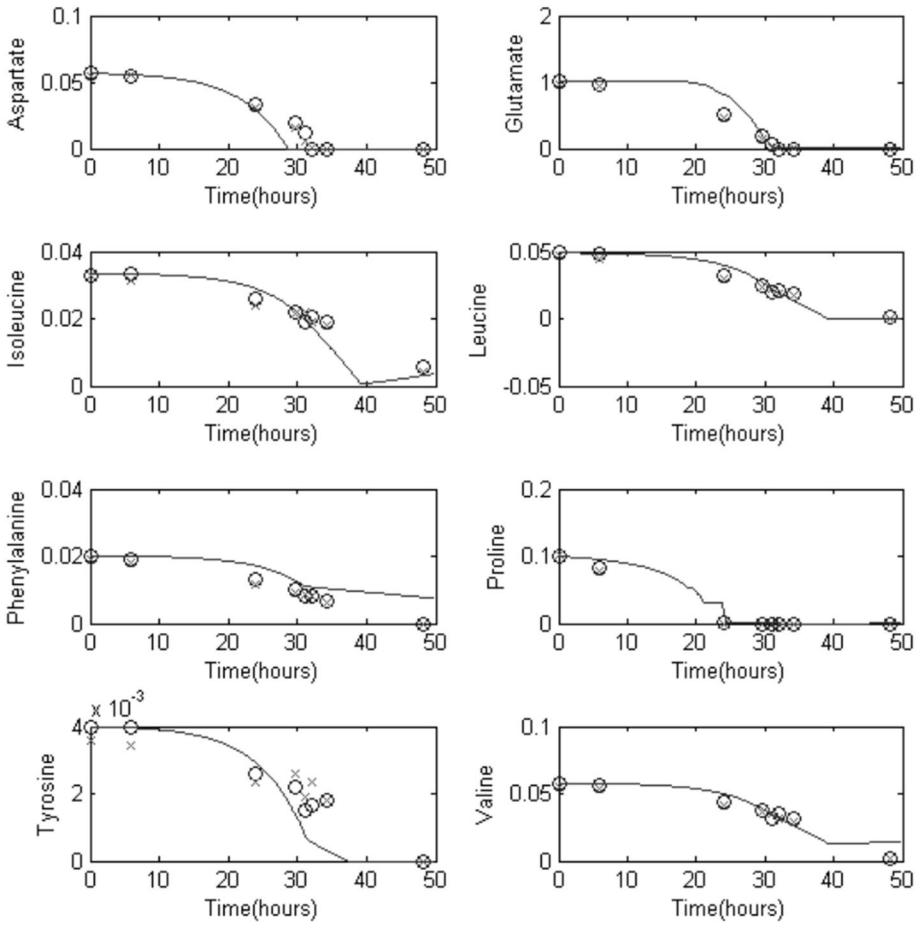


Fig. 4 Metabolite concentration profiles of *B. Pertussis* case study resulted from the DFBA

Fig. 4 (continued)



kinetics for the glucose consumption rate bound whereas a constant upper bound for the consumption of oxygen is identified. The numerical values of the kinetic constants are slightly different from Mahadevan's model because they were identified by the procedure in (4a)–(4l) from simulations of the original model with superimposed noise.

Figure 3 shows the comparison between the time evolutions of the four species as a function of batch time obtained from the original model without the noise (red line) and the identified model (blue line). The simulated results are slightly different between the original and identified models due to the differences in the kinetic rate expressions for the limiting constraints that resulted from the assumed noise.

Case study 2: *B. pertussis*

Bordetella Pertussis is a Gram-negative, aerobic, and pathogenic bacteria causing Pertussis or whooping cough [20]. In this case, our algorithm is used for systematic identification from experimental data of limiting constraints and a suitable biological objective function of a *B. Pertussis* fed-batch culture. Glutamate is the main carbon source used for the feeding of this culture, which is started at time 0 in batch operation until depletion of glutamate occurs and then fed-batch operation is started with a constant glutamate feed rate of 4.3 g/h. A metabolic network involving 49 metabolites and 50 reactions was available from a previous study by Budman et al. [21]. The reactions that were considered are listed in "Appendix II". HPLC measurements of 16 amino acids were available for a 50 h fermentation process at 8 time instances: 0, 6, 24, 29, 31.1, 32, 34, 48.3 h. The data are shown by circles in Fig. 4. Due to confidentially reasons, all the presented data in this figure have been normalized based on the initial concentration of glutamate which is the main nutrient in this fermentation process.

A nonlinear polynomial interpolation function was applied to interpolate the data between the available measured values. The sampling interval used for the identification procedure given by equations (4a)–(4l) was 0.1 h. The level of noise in the HPLC measurements was estimated as 10% of the full range of change of each amino acid and this value was used to estimate the $\varepsilon = 0.1$ used in the rate constraints (4b) and (4c). The upper bound W^{u} in constraint (4l) was set to 10 mM/h mM as in the previous case study. In principle, this value can be iteratively further adjusted as explained in

"Theoretical and experimental methods". The weight u_w in Eq. (2j) was set to 0.01.

Two candidates were considered as possible biological objectives for maximization: maximization of the growth rate and maximization of the lactate production rate. The growth rate is defined by the stoichiometric relations derived from the reactions listed in "Appendix II".

The identification procedure given by Eqs. (4a)–(4l) was applied to the data to identify the objective function and limiting constraints. Among constraints (4b, 4c), only one constraint is active for the entire duration of the batch corresponding to the upper bound in the consumption rate of phenylalanine. The biological reason for phenylalanine being the limiting active constraint might be due to the fact that phenylalanine is not produced by *B. Pertussis* and should be externally supplied in the medium.

A key difference between the current case study and the previous case study is that here only a fraction of the species involved in the network is measured whereas in the previous study all species were measured. Thus, the current case study is highly under-determined and consequently, constraints (4k)–(4l) that are used to limit the solution space become important. These constraints were found to be important for limiting the solution space for both consuming and producing species that are not being measured.

The weights w_i^{sc} (constraint 4k) were identified at each time interval and accordingly $w_i^{sc,max}$ was calculated using Eq. (5).

Six of the amino acids can be produced according to the metabolic reactions listed in "Appendix II": alanine, aspartate, glutamate, isoleucine, arginine, and proline. However, it was found from the data that the concentration of these metabolites did not exhibit any noticeable increases in time. The reason is that these metabolites start to be synthesized only after they are completely depleted so as to further generate biomass, i.e., the production of these amino acids is exactly balanced by their consumption towards biomass especially during the fed-batch phase. In principle, an upper bound of 0 could have been assumed for the production rates but it was found that due to Euler discretization that it was difficult to satisfy the positivity constraint (6e) when the produced amino acid concentrations are very close to zero. For these six amino acids, having a constant upper bound with the value of 10^{-4} was necessary. The preferable objective function was found to be the growth rate. This coincides with the objective function that has been prevalent if previous bacterial models and also corresponds to the objective function used in a *B. Pertussis* model reported in Budman et al. [21].

Using the growth rate as the biological objective, it is possible to formulate the optimization problem as a QP as explained in Eq. (3a). Solving the problem as a QP serves to verify the nonlinear optimization used to solve the problem in (3a)–(3h). The *cplexqp* solver in MATLAB was used to solve this quadratic optimization. The results were very similar to the results of the nonlinear optimization except for small differences in the values of the weights, w_i^{sc}.

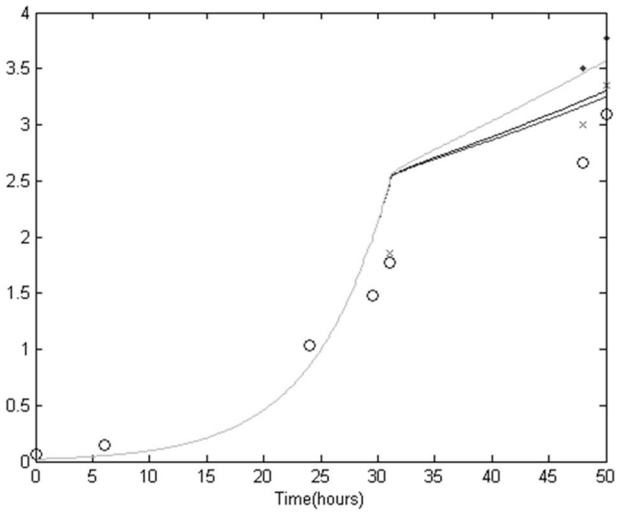


Fig. 5 Biomass profile of *B. Pertussis* in fed-batch mode: red, blue, and green lines are model predictions with fed-batch rate of 4.3, 6, and 7.4 g/h; circle, ×, and pink dots mark experimental data points for the feed rate of 4.3, 6, and 7.4 g/h

To formulate a predictive model as given by Eqs. (6a)–(6f), a kinetic expression was sought relating the maximum consumption rate in phenylalanine as a function of its concentration. From the relation between the measured consumption rates and the corresponding concentrations, it was concluded that a Michaelis–Menten relation is not suitable. Instead, as assumed in our earlier work [16], a Hill-type function [22] was used to describe the maximal uptake rate of the phenylalanine by

$$r_{\mathrm{Phe}} = \frac{0.0066\ \psi_{\mathrm{Phe}}{}^{3.2}}{0.022^{3.2}\ \psi_{\mathrm{Phe}}{}^{3.2}} \tag{9}$$

The values of the parameters in this expression were found using the *lsqcurvefit* of MATLAB.

The bounds $w_i^{sc,max}$ used in constraints (6d) are obtained from the calculated values of the corresponding w_i^{sc} according to Eq. (5); their values are presented in "Appendix II".

The simulated results for the 16 measured amino acids using Eqs. (6a–6f) using the phenylalanine expression in Eq. (9) and the constraints (6b) are shown in Fig. 4. The sum of squared errors between simulations and measured values is SSE = 6.25. This error is significantly lower as compared to our earlier model of SSE = 15.42 [21] where the limiting constraints in developing the DFBA model

were based on trial and error in contrast with the systematic approach proposed in this work. The reason that our model is not able to give a perfectly prediction for some of the metabolites might be due to the inaccuracies in the stoichiometric matrix and in the calculation of biomass amino acids' composition.

To test further the predictive ability of the model, no constraints were included for biomass. The goal was to test whether the model without any biomass-related constraints could correctly predict the biomass evolution over time. This test is significant since all amino acids contribute to the composition of biomass and thus predicting correctly the evolution of biomass implies also a certain level of fidelity in the prediction of the amino acids. Biomass measurements were available for two different feeding rates of glutamate 4.3 and 4.7 g/h .The model predictions of biomass for both feeding rates show very good agreement between measured and predicted values (Fig. 5).

It also should be noted that the accuracy of the model could be increased by including more measurements in the identification step.

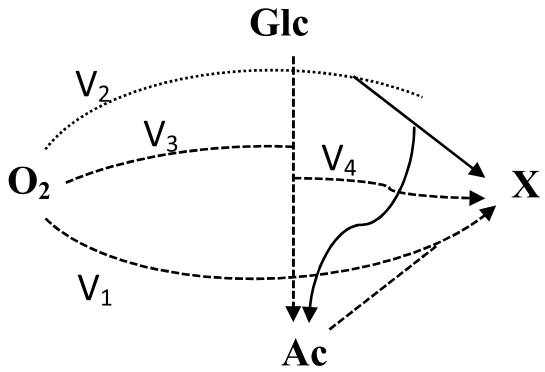
Conclusion

The formulation of a dynamic metabolic flux balance model requires the identification of a suitable biological objective function to be maximized or minimized and active limiting constraints. In this contribution, we propose a systematic approach to identify both of these attributes for developing DFBA models. A key advantage of the proposed approach is that its formulation as a one-level optimization that simultaneously maximizes the biological objective together with terms that are related to the limiting constraints on consumption and production rates of amino acids. Proper weighting of these additional terms allows the constraint to be related to the relative errors incurred in measuring both amino acid concentrations as well as the biological objective variable. The use of set-based constraints ensures fitting between the data and the model predictions while avoiding the need for two-level optimization formulations. Two case studies were investigated. In the first case study for *E. coli* for which a priori known model was used to generate in silico data, the proposed approach successfully recognized the correct objective function and constraint of the model. In the second case study that involved experimental data of *B. Pertussis* fermentation, the scarcity of data and the complexity of the metabolic network rendered the problem underdetermined. Additional constraints were required to only allow realistic solutions. In this case, our approach systematically identified a single kinetic rate bound on one amino acid (phenylalanine) and a set of constant upper bounds on consumption and production rates where the latter served to limit the solution space.

The resulting model was able to predict the evolution of biomass concentration with time whereas biomass values were not used for model calibration.

Acknowledgements The authors would like to thank Natural Science and Engineering Research Council (NSERC).

Appendix I



J_1 $39.43\ Ac + 35O_2 \rightarrow X$
J_2 $9.46Glc + 12.92O_2 \rightarrow X$
J_3 $9.84\ Glc + 12.73O_2 \rightarrow 1.24Ac + X$
J_4 $19.23Glc \rightarrow 12.12Ac + X$

Appendix II

J_1 Pyruvate → PEP
J_2 Pyruvate + CoA → Acetyl-coA + CO_2
J_3 Acetyl-CoA + H_2O + oxaloacetate → citrate + CoA
J_4 Citrate → α-ketoglutarate + CO_2
J_5 α-ketoglutarate + enz-N6 → succinyl transf. + CO_2
J_6 Succinyl transf. + CoA → succinyl-CoA + enz-N6
J_7 Succinyl-CoA + phosphate → CoA + succinate
J_8 Succinate + acceptor → fumarate + reduction acceptor
J_9 Fumarate + H_2O → malate
J_{10} Malate → oxaloacetate
J_{11} Glutamate + NH_3 → phosphate + glutamine
J_{12} 2 Glutamate ← glutamine + α-ketoglutarate
J_{13} Glutamate + H_2O → α-ketoglutarate + NH_3
J_{14} Proline + $2H_2O$ → glutamate
J_{15} Oxaloacetate + glutamate → aspartate + α-ketoglutarate
J_{16} Aspartate + NH_3 ← aspargine
J_{17} PEP + HCO_3^- → phosphate + oxaloacetate*

J_{18} Lactate → pyruvate
J_{19} Acetate + CoA ← acetyl-CoA + phosphate
J_{20} 2 Acetyl-CoA → CoA + acetoacetyl-CoA
J_{21} Acetoacetyl-CoA → PHB
J_{22} Glucose-6-phosphate + 3 glyceraldehyde-3-phosphate → 3 ribose-5-phosphate
J_{23} Acetyl-CoA + carrier-protein → acetoacetyl-carrier + CoA + CO_2
J_{24} Threonine + 3 pyruvate + 2 glutamate + acetyl-CoA + H_2O → NH_3 + $3CO_2$ + $2H_2O$ + isoleucine + α-ketoglutarate + valine + CoA + leucine
J_{25} Serine + tetrahydrofolate ←→ 5,10-methylenetetrahydrofolate + glycine + H_2O
J_{26} Serine ←→ pyruvate + NH_3
J_{27} Threonine ←→ glycine + acetaldehyde
J_{28} Aspartate → threonine + phosphate
J_{29} Hydrogen sulfide + acetyl-CoA + serine → CoA + cysteine + acetate
J_{30} Glutamate + pyruvate ← α-ketoglutarate + alanine
J_{31} Aspartate + pyruvate + glutamate + succynyl-CoA → phosphate + α-ketogultarate + succinate + lysine + CO_2 + CoA
J_{32} Malate → pyruvate + CO_2
J_{33} Amino acids → biomass
J_{34} Amino acids → pertactin
J_{35} PEP → glyceraldehyde 3-P + phosphate
J_{36} 2GAP + H_2O → glucose 6-P + phosphate
J_{37} **J_3**
J_{38} Amino acids → Pertussis toxin
J_{39} Amino acids → fimbria
J_{40} Amino acids → FHA
J_{41} Inverse of J14
J_{42} Inverse of J27†
J_{43} Inverse of J30
J_{44} 2 Glutamate + aspartate → fumarate + α-ketoglutarate + arginine
J_{45} Glucose + 6-P GAP + 2 PEP + glutamate → tyrosine + D-xylose + α-ketoglutarate
J_{46} Inverse of J26
J_{47} Valine + α-ketoglutarate → glutamate + 4 methyl-2-oxopentanoate
J_{48} 2 pyruvate + serine + aspartate + cysteine → CoA + 3 CO_2 + glycine
J_{49} Inverse of J18

Stoichiometry of purines

Adenine: pyruvate + 2 glutamine + 2 aspartate + glycine → adenine + 2 glutamate + fumarate

Guanine: pyruvate + 3 glutamine + aspartate + glycine → guanine + 3 glutamate + fumarate

Stoichiometry of pyrimidines

UMP (Uridylic acid): glutamine + HCO_3^- + aspartate + ribose-5-phosphate → UMP + CO_2 + glutamate

CMP (Cytidylic acid): glutamine + HCO_3^- þ aspartate + ribose-5-phosphate + NH_3 → CMP + CO_2 + glutamate

TMP (Thymidylic acid): HCO_3^- + serine + glutamine + aspartate + ribose-5-phosphate → glycine + TMP + glutamate + CO_2

Metabolite	Value
Ala	− 1.028497997
Arg	− 0.023
Asp	− 0.209047
Asn	− 0.029
Gln	0.6234
Glu	− 3.623564
Gly	− 0.635784
His	− 0.173987138
Ileu	− 0.253078
Leu	− 0.26459
Lys	− 0.232495
Meth	− 0.45017
Phe	− 0.028945
Pro	− 1.532302
Ser	− 2.423
Thr	− 0.02846
Try	0
Tyr	− 0.00823
Val	− 0.92
Cys	0
Pyr	0
AcCoA	0
GAP	0
Lac	7.125
Ammonia	6.2675
Co2	5.243
A-Ketoglutarate	0
Biomass	200
Pertactin	0.4171
Citrate	0.50043
Succ-transferase	0
Succ CoA	0
Succinate	0
Fumarate	0
Malate	0
Oxaloacetate	0.005
Ribose	8
G6P	0.4184
PEP	0
AcAcCoA	0

Metabolite	Value
PHB	0
FFA	0.234
Pt toxin	0.33
Fimbria	0.31243
FHA	0.334125
Xylose	0
CoA	0
ATP	0

References

1. Schilling CH, Letscher D, Palsson BO (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J Theor Biol 203(3):229–248
2. Jaqaman K, Danuser G (2006) Linking data to models: data regression. Nat Rev Mol Cell Biol 7(11):813–819
3. Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis?. Nat Biotechnol 28(3):245–248
4. Foguet C et al (2016) HepatoDyn: a dynamic model of hepatocyte metabolism that integrates C-13 isotopomer data. PLoS Comput Biol 12(4):e1004899
5. Llaneras F, Sala A, Picó J (2012) Dynamic estimations of metabolic fluxes with constraint-based models and possibility theory. J Process Control 22(10):1946–1955
6. Varma A, Palsson BO (1995) Parametric sensitivity of stoichiometric flux balance models applied to wild-type *Escherichia coli* metabolism. Biotechnol Bioeng 45(1):69–79
7. Mahadevan R, Edwards JS, Doyle FJ (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys J 83(3):1331–1340
8. Sanchez CEG, Saez RGT (2014) Comparison and analysis of objective functions in flux balance analysis. Biotechnol Prog 30(5):985–991
9. Pramanik J, Keasling JD (1997) Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. Biotechnol Bioeng 56(4):398–421
10. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol Syst Biol 3:119
11. Knorr AL, Jain R, Srivastava R (2007) Bayesian-based selection of metabolic objective functions. Bioinformatics 23(3):351–357
12. Burgard AP, Maranas CD (2003) Optimization-based framework for inferring and testing hypothesized metabolic objective functions. Biotechnol Bioeng 82(6):670–677
13. Savinell JM, Palsson BO (1992) Network analysis of intermediary metabolism using linear optimization.1. Development of mathematical formalism. J Theor Biol 154(4):421–454
14. Gianchandani EP et al (2008) Predicting biological system objectives de novo from internal state measurements. BMC Bioinform 9:43
15. Llaneras F, Picó J (2008) Stoichiometric modelling of cell metabolism. J Biosci Bioeng 105(1):1–11
16. Nikdel A, Budman H (2016) Identification of active constraints in dynamic flux balance analysis. Biotechnol Prog 33(1):26–36
17. Maranas CD, Zomorrodi AR, Wiley Online Library (2016) Optimization methods in metabolic networks, Wiley, Hoboken **(1 online resource)**
18. Borchers S et al (2013) Identification of growth phases and influencing factors in cultivations with AGE1.HN cells using set-based methods. PLoS One 8(8):e68124
19. Shuler ML, Kargi F (1992) Bioprocess engineering: basic concepts. Prentice Hall international series in the physical and chemical engineering sciences, Prentice Hall, Englewood Cliffs, pp 16+479 s
20. Thalen M et al (2006) Fed-batch cultivation of *Bordetella pertussis*: metabolism and Pertussis Toxin production. Biologicals 34(4):289–297
21. Budman H et al (2013) A dynamic metabolic flux balance based model of fed-batch fermentation of *Bordetella pertussis*. Biotechnol Prog 29(2):520–531
22. Weiss JN (1997) The Hill equation revisited: uses and misuses. Faseb J 11(11):835–841