Computers and Chemical Engineering 72 (2015) 325–333

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Gobal optimization of hybrid kinetic/FBA models via outer-approximation

Carlos Pozo[a], Antoni Miró[a], Gonzalo Guillén-Gosálbez[a,b,*], Albert Sorribas[c], Rui Alves[c], Laureano Jiménez[a]

[a] Departament d'Enginyeria Química, Universitat Rovira i Virgili, Av. Països Catalans 26, 43007 Tarragona, Spain
[b] School of Chemical Engineering and Analytical Science, University of Manchester, Mill, Sackville Street, Manchester M13 9PL, United Kingdom
[c] Departament de Ciències Mèdiques Bàsiques, Institut de Recerca Biomèdica de Lleida (IRBLLEIDA), Universitat de Lleida, Avinguda Alcalde Rovira Roure 80, 25198 Lleida, Spain

ARTICLE INFO

Article history:
Received 16 January 2014
Received in revised form 16 June 2014
Accepted 18 June 2014
Available online 28 June 2014

Keywords:
Metabolic engineering
FBA
GMA
Global optimization
Kinetic models

ABSTRACT

Flux balance analysis (FBA) is a linear programming-based framework widely used to predict the behavior, in terms of the resulting flux distribution, of cellular organisms in different media. FBA models are constructed using only stoichiometric information, and for this reason they sometimes fail in predicting fluxes precisely. In this work, we formally define the concept of hybrid FBA/kinetic models, in which kinetic information of key processes is used to tighten the search space of standalone FBA formulations, thereby enhancing their predictive capabilities. This approach leads to non-linear non-convex models that may exhibit multiple local optima. To solve them to global optimality, we use a customized outer-approximation algorithm that exploits the structure of the kinetic equations. Numerical results show that our method enhances the quality of standalone FBA models, providing more accurate predictions.

1. Introduction

Metabolic networks are controlled by means of non-obvious combinations of metabolic and hierarchical regulations which are hard to understand just by visual inspection of the system. Mathematical models can be used in combination with other techniques such as optimization in order to elucidate the gears behind these biological mechanisms.

Methods available for the analysis of metabolic systems are generally based on (i) kinetic data or (ii) structural and stoichiometric modeling (Tomar and De, 2013). Kinetic models differ in the enzyme kinetic rate law they adopt, which is typically a function of certain kinetic parameters and reactant concentrations (both of which determine how fast the reaction will progress). Michaelis–Menten and General Mass Action (GMA) models are among the most popular kinetic formulations (Alves et al., 2008a,b).

Under the structural and stoichiometric modeling, we find the so-called network-based and constraint-based approaches. The latter include the most popular approach used in the analysis of metabolic networks (i.e., flux balance analysis, FBA). FBA, which was pioneered by Papoutsakis (1984) and later expanded by Palsson and co-workers (Varma and Palsson, 1994a,b; Schilling et al., 1999, 2000), is based on principles of constraints analysis. FBA makes quantitative predictions about fluxes of metabolic networks based on steady-state mass balance constraints and without relying on any type of kinetic information.

FBA models are typically under-determined, that is, the number of variables exceeds the number of equations, so there are an infinite number of feasible solutions satisfying the mass balance constraints. The key idea in FBA is to solve an optimization problem in which a biological objective function is optimized subject to the network stoichiometry (i.e., mass balances). Hence, among the infinite number of feasible solutions to the mass balance constraints, the goal is to identify the one that optimizes a biological criterion that reflects the cell's behavior. This leads to a linear programming (LP) formulation, which can be efficiently solved by well-known methods. Hence, FBA works under the hypothesis that evolution results in selecting the cell's machinery that provides a (quasi)optimum response for one or several physiological constraints. Typical objective functions include maximizing the biomass growth (Orth et al., 2010), the synthesis rate of a particular product (Kauffman et al., 2003), or minimizing the ATP consumption. In practice, rather than calculating and predicting precisely the metabolic network behavior, FBA aims to narrow down the range of

* Corresponding author. Tel.: +34 977558618; fax: +34 977559621.
E-mail address: gonzalo.guillen@urv.cat (G. Guillén-Gosálbez).

http://dx.doi.org/10.1016/j.compchemeng.2014.06.011
0098-1354/

all possible phenotypes (i.e., feasible space) that a metabolic system can display (Palsson, 2000).

The simplicity given by the linear nature of FBA is one of its main advantages, as it facilitates its application to genome-wide scale models of microorganisms (Edwards and Palsson, 2000), and more recently also to plants (Poolman et al., 2009; de Oliveira Dal'Molin et al., 2010). As an example, FBA has been used for elucidating the flux modifications that are required in *Escherichia coli* strains in order to increase the ethanol production under different conditions. Recent work has employed as well FBA models to address cutting edge research questions ranging from understanding the dynamics of microbial communities to predicting perturbations according to complex metabolic engineering objectives (Chaganti et al., 2011).

Despite these advances, there are still cases in which FBA provides poor in silico predictions of the cellular behavior observed in vivo (Edwards et al., 2002). This is partly due to the fact that FBA focuses only on metabolic fluxes and neglects both, the regulatory effects within a metabolic network and the concentrations of key metabolites. Several attempts have been made recently for improving the quality of FBA models. Some authors have proposed to incorporate additional knowledge based on flux capacities and/or thermodynamics into FBA models (Price et al., 2004; Orth et al., 2010). Constraint-based formulations that add transcriptional regulatory constraints for certain conditions have also been proposed to achieve the same goal (Covert et al., 2004; Chandrasekaran and Price, 2010; Shlomi et al., 2007). In addition, physico-chemical constraints have been included in FBA models with the aim of improving their predicting capabilities (Price et al., 2003).

A promising path to enhance the quality of FBA predictions consists of incorporating kinetic information of key velocities into the model. Unfortunately, kinetic models of metabolism require intensive data that are difficult to obtain. Following this approach, Lee and Voit (2010) combined FBA models with GMA formulations for analyzing the monolignol biosynthetic pathway in *Populus xylem*. The resulting models were solved by means of the Indifrect Optimization Method (IOM, see Torres et al., 1997), which relies on transforming a given GMA formulation into an approximated S-System model with the aim of carrying a linear optimization of the problem. As stated by the authors, the main drawback of this strategy is that the solution to the S-system model may not be globally optimal in the original one. Because of this, unstable solutions might be detected as optimal. More recently, Cotten and Reed (2013) estimated kinetic parameters in a simplified kinetic model of central *E. coli* metabolism integrating fluxomic, proteomic, and metabolomics data. Then this model was used to obtain lower and upper bounds which were imposed on fluxes in a FBA model. None of these works addresses the global optimization of FBA models where kinetic rate laws are explicitly included in the formulation. Note that the introduction of kinetic expressions leads to non-linear non-convex models that are more difficult to handle. Losing the nice linear properties of LP is therefore the price to pay for increasing the accuracy of the predictions.

In this work, we formally introduce the concept of hybrid kinetic/FBA models that combine stoichiometric and kinetic equations of selected key processes. These kinetic constraints help brush up stoichiometric predictions by incorporating key regulatory information. Hybrid models lead to non-convex formulations that can show several local optima in which standard optimizers might get trapped during the search. Hence, besides formally defining the concept of hybrid model, we introduce a tailored outer-approximation algorithm to solve them that guarantees convergence to the global optimum within an epsilon tolerance. We illustrate the capabilities of our strategy through its application to a FBA model of *Saccharomyces cerevisiae*, in which we progressively tighten the search space by adding kinetic information of key velocities.

2. Standalone FBA models

FBA predicts intracellular fluxes using stoichiometric information. In a metabolic network, the concentration of the internal metabolites X can be obtained from the fluxes v producing/depleting them and their corresponding stoichiometric matrix S, as described by Eq. (1).

$$\dot{X} = S \cdot v = 0 \tag{1}$$

The assumption of steady state conditions in Eq. (1) is relevant for most intracellular reactions since they are typically much faster than the rate of change in the resultant cellular phenotypes, such as cell growth and process dynamics (Varma and Palsson, 1994a,b).

The system of equations in Eq. (1) is in general underdetermined, that is, there is an infinite number of solutions satisfying it. Among them, FBA identifies the solution of an optimization problem whose constraints are given by Eq. (1), and whose objective function (f) depends on the particular application addressed. This leads to an LP problem of the following form:

$$\begin{aligned} (\text{FBA}) \quad \min \quad & f \\ s.t. \quad & \dot{X} = S \cdot v = 0 \\ & v^{LB} \le v \le v^{UB} \\ & v \in \Re^{+} \end{aligned} \tag{2}$$

where v^{LB} and v^{UB} are lower and upper bounds imposed on the velocities, respectively. The problem given by Eq. (2) can be efficiently solved by means of well-known LP techniques. At this point, it is important to distinguish between the standard FBA approach, which aims to predict fluxes assuming a biological objective function driving the cell's machinery, and FBA models employed for computing optimal enzymatic modulations for microbial strain optimization. This latter application of FBA leads to MILP formulations, for instance when a limit is imposed on the number of enzymes which can be simultaneously modified. Other cases in which MILP formulations arise are those in which binary variables are used to model knock-out alternatives. In some cases, these MILPs are formulated in the context of a bi-level optimization problem, with outer and inner problems denoting the bioengineering objective and the biological objective, respectively. Hence, in this work, we will refer to the basic FBA formulation. Note, however, that our general approach can be easily extended in order to deal with FBA-based MILP models as well.

The polyhedral feasible region defined by the mass balance constraints tends to be very large, and this might result in inaccurate predictions. Hence, after solving the model optimizing a plausible biological objective (typically, biomass growth rate), we expect that the fluxes calculated in silico through linear programming will be as close as possible to the true fluxes found in vivo. In practice, both sets of values will differ, and several objective functions and/or additional constraints can then be tested for obtaining better predictions.

Particularly, FBA-based predictions can be enhanced through additional constraints that remove solutions violating basic physical principles. Thermodynamic constraints, for instance, ensure that a reaction is consistent with metabolite concentrations and Gibbs free energy of reaction (Pharkya and Maranas, 2006). In this paper we use instead kinetic and regulatory information to tighten the search space of FBA models (Pharkya and Maranas, 2006).

3. Hybrid FBA/kinetic models

As a first step toward the construction of genome-scale kinetic models, we present here a methodology that allows incorporating kinetic information into FBA models using the power-law

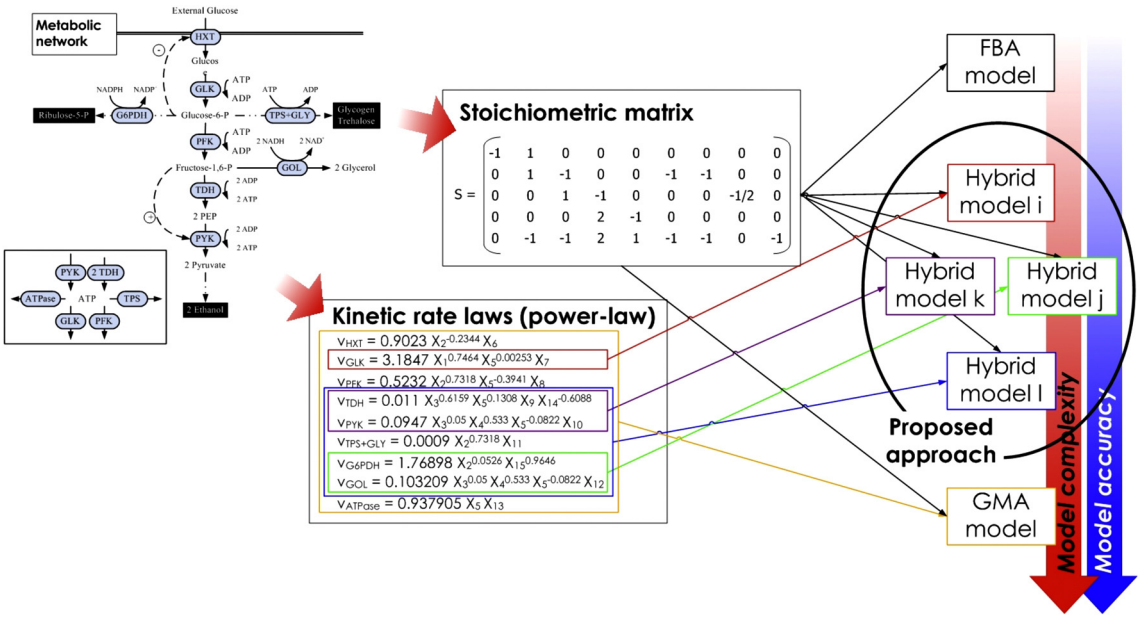

Fig. 1. The standalone consideration of the stoichiometric matrix of the metabolic network gives rise to a standard FBA model. Hybrid models of increasing predicting capabilities and complexity can be built by adding from one to several kinetic equations to the problem formulation (for instance one kinetic equation in Hybrid model *i*, two in Hybrid models *k* and *j*, and five in Hybrid model *l*). In the extreme case, all the kinetics are embedded and the model is equivalent to a full kinetic model (GMA model in this particular case where power-law equations are used).

formalism and global optimization methods. We describe first how to build a hybrid FBA model that contains kinetic equations based on the power-law formalism. We then discuss how we can globally optimize this model via a customized outer-approximation algorithm. Hence, we start from a standard FBA model of the following form:

$$\begin{aligned} (\text{MFBA}) \quad \min \quad & f \\ s.t. \quad & \dot{X} = S \cdot v = 0 \\ & v^{LB} \le v \le v^{UB} \\ & v \in \Re^{+} \end{aligned} \tag{3}$$

where f is an objective function (e.g. biomass growth, the synthesis rate of a particular product, ATP consumption) and where the feasible space is defined by the steady-state mass balance equations and a set of lower and upper bounds imposed on the velocities. The next step is to introduce kinetic equations that allow determining the velocity v_r of flux r from the concentration of those metabolites affecting such velocity. Without loss of generality, we use here the so-called power-law formalism, which is a general kinetic approach to model reaction rates that facilitates the modeling of metabolic networks (Alves et al., 2008b). The use of this formalism is rather convenient, since it adequately captures the non-linear behavior of metabolic regulations while yet exhibiting useful mathematical properties such as linearity in the logarithmic space. In this formalism, each velocity is described as follows:

$$v_r = \gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}} \quad r = 1, \ldots, p \tag{4}$$

In this equation, γ_r is the apparent rate-constant of the enzyme ruling process r and f_{rj} is the kinetic order of metabolite j in this process.

The inclusion of kinetic equations into the standalone FBA leads to the following non-convex NLP formulation:

$$\begin{aligned} \min \quad & f \\ s.t. \quad & S \cdot v = 0 \\ & v_r = \gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}} \quad r = 1, \ldots, p \\ & v^{LB} \le v \le v^{UB} \\ & X^{LB} \le X \le X^{UB} \\ & v, X \in \Re^{+} \end{aligned} \tag{5}$$

Note that this formulation can be extended to address realistic metabolic engineering problems by introducing continuous and binary variables, thus leading to MINLP models (see Appendix A).

The main difference between this model and a traditional FBA model is the incorporation of kinetic equations based on the power-law formalism. Besides, this hybrid model allows constraining the concentrations of metabolites within realistic lower and upper limits X^{LB} and X^{UB}, respectively. These constraints, in turn, introduce flux limitations through the kinetic equations. The specific effects attached to saturation, cooperativity, and the kinetic properties are here made explicit through the kinetic equation and its parameters. In a more elaborated extension of our approach, parameter boundaries may be introduced. However, this would require sophisticated computational improvements due to the non-linearities associated to the kinetic function.

Complete hybrid models would require a bunch of kinetic data that is typically hard to obtain. In addition, they lead to large-scale non-convex NLPs with potential multiple local optima. To overcome these limitations, we propose to include kinetic equations for a subset of reactions only. As we will show later in the article, by doing so we can generate a set of hybrid models of increasing complexity and with different predictive capabilities (see Fig. 1). In practice, introduction of few kinetic equations may dramatically improve predictions.

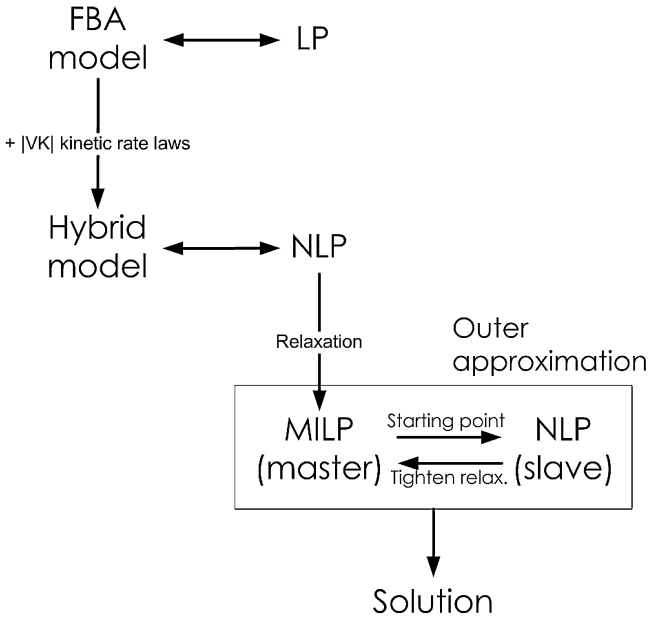

Fig. 2. Solution approach based on the generation of a set of hybrid models containing |*VK*| velocities which are later globally optimized by means of a customized outer approximation. Each problem is decomposed into two subproblems: a master MILP, constructed by relaxing the original model using piecewise envelopes and hyperplanes, that provides a lower bound, and a slave NLP that yields an upper bound. The algorithm iterates between these two levels until a termination criterion is satisfied.

Thus, following this approach, we define kinetic equations for only a subset *VK* of velocities. The resulting optimization problem is as follows:

$$\begin{aligned} (\text{MHFBA}) \quad \min \quad & f \\ s.t. \quad & S \cdot v = 0 \\ & v_r = \gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}} \qquad v_r \in VK \\ & v^{LB} \le v \le v^{UB} \\ & X^{LB} \le X \le X^{UB} \\ & v, X \in \Re^{+} \end{aligned} \tag{6}$$

The resulting model is a non-convex NLP where non-convexities are given by the power-law equations (i.e., non-linear equality constraints). To solve this non-convex NLP, we use a customized outer approximation-based algorithm introduced in a previous work (Pozo et al., 2010).

Fig. 2 illustrates the main idea of this algorithm, which decomposes the original problem (Eq. (6)) into two sub-problems at different hierarchical levels: an upper level master mixed-integer linear programming (MILP) problem, and a lower level slave non-linear (NLP) problem. The master level solves an MILP that is a relaxation of the original non-convex NLP. This MILP predicts a valid lower bound on the global optimum (when minimizing, the solution of the relaxation will be, at least, as good as that of the original problem). In the lower level, the original problem is locally optimized, thereby providing a valid upper bound (when minimizing) on the global optimum. These two subproblems are solved iteratively until the optimality gap is reduced below a given tolerance. Further details on this approach can be found in (Pozo et al., 2010).

Note that our outer approximation algorithm allows building linear relaxations of the original model of increasing quality and complexity. This leads to a set of hybrid linear FBA/kinetic models in which the kinetic equations are replaced by linear underestimators. Starting from the FBA, we can generate a set of hybrid non-linear models, in which some velocities are modeled using kinetic equations. Each of these models can be solved to global optimality using our customized outer-approximation algorithm.

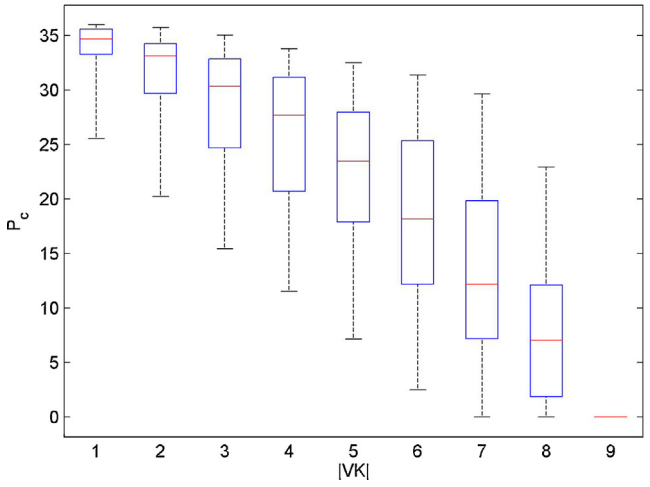

Fig. 3. Box plot of P_c for the different cases each with a given number of kinetic equations incorporated.

Note that the outer-approximation algorithm exploits the particular structure of the kinetic formalism (i.e., GMA equations), and for this reason outperforms standard global optimization software packages (i.e., BARON), as was shown in a previous publication (Pozo et al., 2010). We emphasize here that hybrid models can be solved by any type of suitable global optimization package. Hence, our approach is not restricted to the use of the outer approximation algorithm. Global optimization is a growing field that is evolving rapidly (Misener and Floudas, 2012; Zorn and Sahinidis, 2014). Hence, our general approach will benefit from the future theoretical and computational developments made in the area.

4. Quality of in silico predictions using hybrid models

Adding kinetic equations into a FBA model will likely result in an improvement of its predictive capabilities. We introduce a metric to quantify the quality of a prediction made by a model *c*. This metric, that we call predictive indicator and denote by P_c, measures the size of the search space within which a feasible solution must fall. Larger search spaces will result in worst predictions, so we look for very tight search spaces, which in the limit correspond to one single point obtained by solving the fully kinetic model.

To calculate this quality metric, we proceed as follows. We first maximize and minimize each velocity of the hybrid model subject to the mass balance equations and to a subset of kinetic constraints (i.e., those that are modeled via the power-law formalism in the hybrid model). These calculations allow identifying the boundary values for each velocity that are compatible with all the constraints. With the information obtained from these runs of the hybrid model, we next determine the predictive parameter P_c as follows:

$$P_c = \sum_{r=1}^{p} \left(\frac{v_r^{*UB} - v_r^{*LB}}{(v_r^{*UB} + v_r^{*LB})/2} \right)^2 \quad \forall c \tag{7}$$

where v_r^{*LB} and v_r^{*UB} are the upper and lower bounds for each velocity, respectively, which are obtained from the corresponding maximization and minimization sub-problems. A small value of P_c implies that the intervals within which the velocities fall are

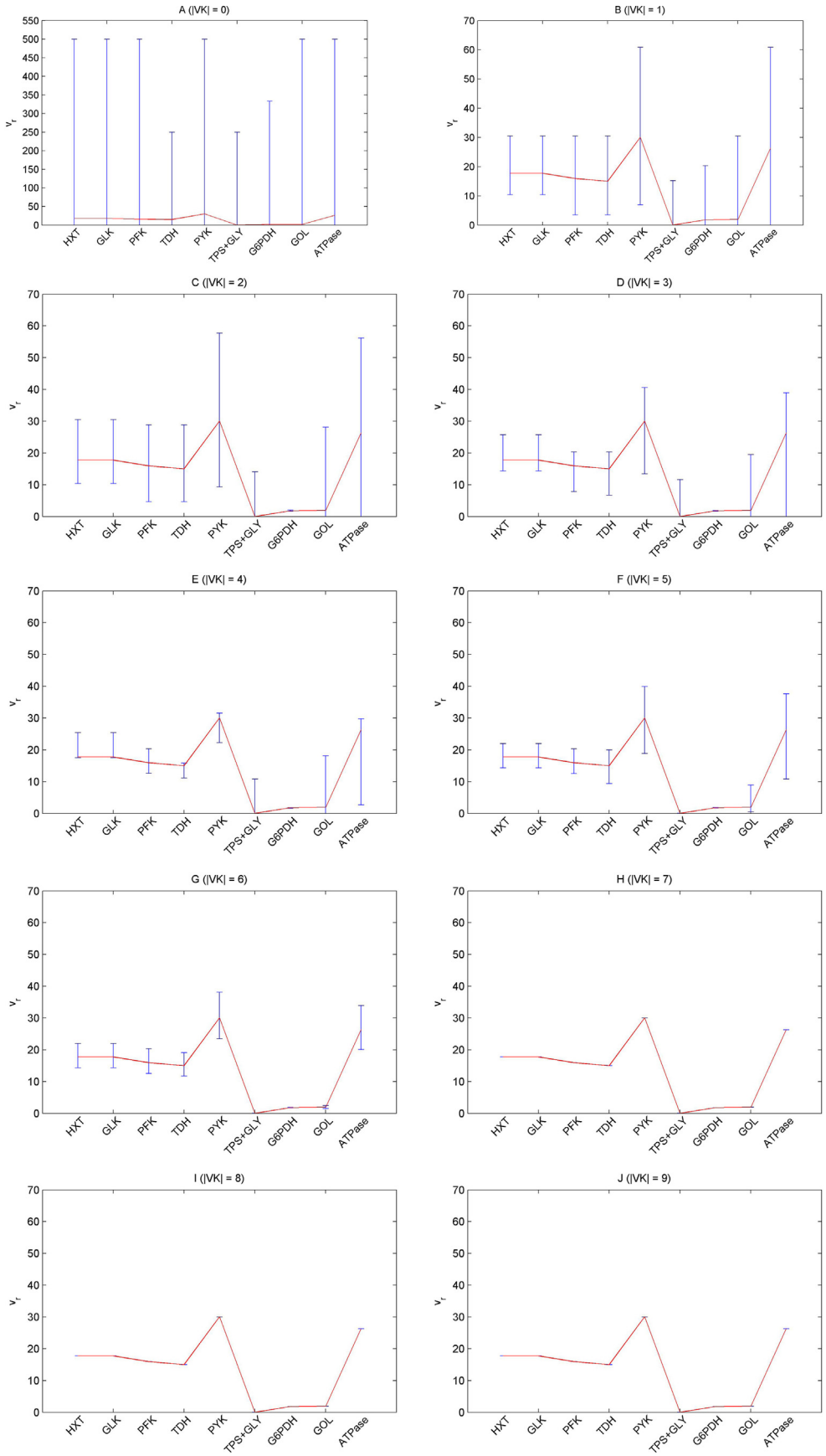

Fig. 4. Graphical representation of the search space of the subset of velocities in Table 1. (For interpretation of the references to color in text, the reader is referred to the web version of this article.)

small, which indicates a high predictive capability.[1] A large P_c value implies the converse (i.e., the confidence intervals for the velocities are large, and therefore the predictive capabilities are poor). Note that, in the limit, when the model is fully constrained (i.e., kinetic model), the solution is in general unique and the value of P_c is 0.

One key point in our approach concerns the manner in which we select the velocities that will be modeled through kinetic equations. As will be discussed later, this issue has a large impact on the predictive capabilities of the hybrid model. We will deal with this aspect of the approach in more detail in Section 5.

5. Case study

We illustrate the capabilities of our approach through its application to a FBA model defined for *S. cerevisie* (Fig. 1, model parameters are available in Vilaprinyo et al., 2006). The hybrid models were all implemented in GAMS and solved with CPLEX 12.3.0.0 (master subproblems) and CONOPT 3.15A (slave subproblems) on an AMD Athlon II 2.99 GHz computer.

We start with the basic FBA model and then solve a set of hybrid models of different size in each of which we model with kinetic power-law equations only a subset of the reactions. Following this approach, we solve a total of 511 models, 9 hybrid models where only one single reaction is modeled via the power-law formalism, 36 hybrid models with two power-law kinetic equations, and so on, until we reach the fully kinetic model in which all the reactions are modeled with power-law equations (i.e., a GMA kinetic model). Each hybrid model is run 18 times, 9 to maximize the single velocities and 9 to minimize them. In total, we thus solve 9216 non-convex NLPs. Recall that each of these models is globally optimized using our customized outer-approximation algorithm.

The results of the aforementioned runs are depicted in Figs. 3 and 4 and Table 1. Fig. 3 is a box plot showing the P_c values for a given number of kinetic equations ($|VK_c|$). Note that due to the combinatorial nature of the problem, the number of hybrid models varies according to the number of velocities modeled via the power-law formalism.

As expected, the value of P_c decreases on average as we increase the number of power-law equations embedded in the model, yet this tendency does not hold always (i.e., models with more kinetic equations do not necessarily lead to better P_c values). As an example, the best model with a single kinetic equation yields a $P_c = 25.55$, outperforming half of the models containing 4 kinetic equations ($|VK| = 4$), and even some models with 8 power-law equations ($|VK| = 8$). Thus, for a given number of kinetic equations, it is clear that there are some combinations that provide much better results than others. This is clearly the case even when many equations are considered.

To further investigate this issue, we compare the models with better performance for every given number of kinetic equations (Table 1). As observed, the performance of the best hybrid model for a given number of kinetic equations increases with the number of velocities modeled via the GMA formalism.

In Fig. 4 we analyze the intervals calculated for each velocity in the best hybrid model for each case. As shown, as we add more kinetic constrains, the intervals shrink. The basal solution (red line) has also been depicted in the figure for comparison purposes. Note that, given a solution, if there is at least one velocity outside the feasible range, then the solution is guaranteed to be unfeasible. Conversely, if the value of a velocity lies within the lower and upper bounds predicted in silico, then the associated solution will only be feasible if and only if the remaining velocities also lie within their corresponding bounds. Ideally, all the experimental velocities should fall within the predicted intervals provided the model contains the most relevant metabolic reactions.

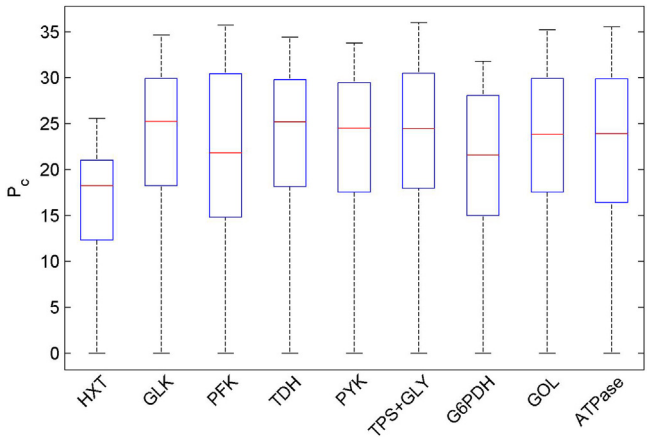

Fig. 5. Box plot of P_c for the different cases in which a given velocity is being modeled by a power-law kinetic equation.

Models A and J are the extreme cases. Model A corresponds to a standard FBA model, in which the difference between the upper and lower bound is significantly large for all the velocities (Fig. 4). On the other extreme, we find the full kinetic model J, in which all the velocities are described with power-law equations (i.e., the corresponding GMA model). Model J is therefore fully determined, as it features the same number of equations and constraints. Hence, as long as the kinetic information used to build the model is accurate, it will show the best possible predictive capabilities. Note that due to the non-convex nature of the equations, the full kinetic model might contain more than one feasible solution. In practice, however, we found this situation quite unlikely, so when we maximize and minimize the velocities we obtain a single point satisfying the mass balance and kinetic equations. Hence, the full kinetic model identifies always the same basal state solution regardless of the kind of optimization (i.e., minimization or maximization) carried out.

Between the two extreme cases, we find intermediate hybrid models yielding predictions of different quality. Note that 7 kinetic expressions (out of 9) suffice to produce the same results as the full GMA model (see subcases H and J). Subcase I containing 8 kinetic expressions produces as well the same results as the full kinetic model. These results suggest that the selection of key velocities to be modeled with kinetic expressions is crucial for generating high-quality results. For instance, by inspecting Table 1, it can be seen that, as $|VK_c|$ increases, successful cases are mainly built upon adding a new kinetic equation to those in the previous case ($|VK_c| - 1$). In other words, most of the cases satisfy the following condition: velocities included in the best model containing $|VK_c|$ kinetic equations, appear also in the best model containing $|VK_c| + 1$ kinetics.

In order to further inspect this tendency, we compute the box plot of P_c for all the cases modeling the same velocity v_r and regardless of whether the remaining velocities are modeled with kinetic equations or not, that is, we show the performance of all the models modeling $v_r \in VK_c$ (see Fig. 5). For instance, the box above HXT, corresponds to all the hybrid models in which a kinetic equation is included for v_{HXT}.

The figure shows that models including the kinetics of v_{HXT} have the best predicting capabilities on average. Remarkably, this is the velocity identified in case A, when only a single velocity was

[1] Remarks: the proposed model will provide accurate predictions if and only if the kinetic expressions employed are indeed a good representation of the real system, and if the model is noise free. Otherwise there is no guarantee that real experimental observations will be contained by the intervals predicted by the model.

Table 1
Subset of velocities for which kinetic constraints are included. "X" means that there is a power-law kinetic equation for this velocity, while a blank space implies the converse. We show only the best hybrid model (the one yielding the lowest P_c value) for each case being studied.

Subcase	P_C	$\|VK_c\|$	v_{HXT}	v_{GLK}	v_{PFK}	v_{TDH}	v_{PYK}	$v_{TPS+GLY}$	v_{G6PDH}	v_{GOL}	v_{ATPase}
A	36.00	0									
B	25.55	1	X								
C	20.25	2	X						X		
D	15.46	3	X		X				X		
E	11.53	4	X		X				X		X
F	7.16	5	X		X			X	X	X	
G	2.53	6	X		X		X	X	X	X	
H	0	7	X		X		X	X	X	X	X
I	0	8	X		X	X	X	X	X	X	X
J	0	9	X	X	X	X	X	X	X	X	X

selected, and also appears in the remaining cases of Table 1 (that is, in the best combination of velocities for any given $|VK_c|$). A possible explanation for this is that this velocity models the input flow to the network which acts as a rate limiting step (i.e., limiting its value is equivalent to limiting the remaining fluxes).

Apart from v_{HXT}, there are two velocities yielding significantly better results than the rest: v_{PFK} and v_{G6PDH}. Again, this result is consistent with what is shown in Table 1, since v_{G6PDH} is selected in all the cases with $|VK_c| > 1$, and v_{PFK} is selected in all the cases with $|VK_c| > 2$. We argue that the reason why these are the next two processes to be included in the model lies in the fact that fixing their values contributes to reduce the degrees of freedom of the network to a large extent. Recall that if the network was completely linear, limiting the input flow (which is something the model already does by including the power-law equation for v_{HXT}) would be enough to limit the remaining fluxes of the network. Since this pathway is not linear (i.e., it contains branching points), this condition does not suffice, and additional limits for fluxes at branching points must be included. In this particular case, the best results are achieved by adding kinetic equations for v_{PFK} and v_{G6PDH}, which are two of the three fluxes depleting glucose-6-P (i.e., in a branching point of the network). Note that the input flow to this branching point (v_{GLK}) is also limited once v_{HXT} is, since at steady state $v_{HXT} = v_{GLK}$. Hence, by imposing additional limits on v_{PFK} and v_{G6PDH} we further reduce the degrees of freedom of the network.

6. Conclusions

This work explored the use of hybrid kinetic/FBA models as a manner to enhance the quality of standalone FBA models of metabolic networks. This is accomplished by tightening the search space via addition of kinetic constraints (i.e., power-law equations). This approach leads to non-convex problems that can be solved by means of global optimization techniques.

The capabilities of the proposed approach were illustrated using a FBA model of *S. cerevisiae*. Results show that the predictive capabilities of the standalone FBA model can be greatly improved by adding kinetic data. In building hybrid models, we need to assess the inherent trade-off between model complexity and accuracy. Our method constitutes a promising alternative for genome-wide metabolic networks for which it is hard to acquire the information required by a complete kinetic model, but for which some kinetic constraints can actually be derived from existing data.

Our approach can solve successfully kinetic models of small and medium size involving up to 30 metabolites and 60 reactions. The computational burden of the method presented, however, might grow rapidly with the size of the metabolic network. The main advantage of the hybrid approach is that it allows defining hybrid models with different levels of complexity depending on the number of velocities that are modeled via kinetic equations.

Future work will focus on developing smart strategies to select key velocities whose kinetic modeling results in improved predictions at the expense of a marginal increase in complexity. Our final goal is to generate a reliable representation of the system while keeping it in a manageable size in terms of kinetic complexity. Results suggest that the most promising candidates are those idenpendent velocities representing key degrees of freedom of the network, mainly the input flow and reactions at branching points.

Acknowledgements

The authors acknowledge the financial support of the following institutions: Spanish Ministry of Education and Science (CTQ2009-14420-C02, CTQ2012-37039-C02, DPI2012-37154-C02-02 and ENE 2011-28269-CO3-03), Spanish Ministry of External Affairs (projects PHB 2008-0090-PC).

Appendix A. MINLP formulation for metabolic engineering applications

In metabolic engineering, one seeks the enzymatic profile (optimal phenotype) that maximizes the production rate of a given metabolite. We can modify the standard metabolic model to reflect this situation by rewriting the enzyme activity in Eq. (4) as a product of the basal enzyme activity γ_r and its fold-change K_r:

$$v_r = \gamma_r K_r \prod_{j=1}^{n+m} X_j^{f_{rj}} \quad v_r \in VK \tag{8}$$

Solutions involving a large number of genetic manipulations might be impractical, so the model needs to acknowledge whether a specific enzyme is modified or not. To this end, we define the following disjunction:

$$\begin{bmatrix} Y_{r1} \\ K_r^{LB} \le K_r \le 1-\delta \end{bmatrix} \vee \begin{bmatrix} Y_{r2} \\ 1-\delta \le K_r \le 1+\delta \end{bmatrix} \vee \begin{bmatrix} Y_{r3} \\ 1+\delta \le K_r \le K_r^{UB} \end{bmatrix} \quad r : v_r \in VK \tag{9}$$

Here δ is a sufficiently small parameter, $K_r{}^{LB}$ and $K_r{}^{UB}$ are lower and upper bounds on the modulation level of the activity of the enzyme controlling process r, and Y_r is a Boolean variable that is true if the associated term of the disjunction is satisfied and false otherwise. This disjunction can be reformulated into equations by means of the Big-M or the convex hull reformulations (Vecchietti et al., 2003). In particular, the Big-M formulation is as follows:

$$K_r - (1-\delta) \le M(1-y_{r1}) \quad r : v_r \in VK \tag{10}$$

$$1 - \delta - K_r \le M(1-y_{r2}) \quad r : v_r \in VK \tag{11}$$

$$K_r - (1+\delta) \le M(1-y_{r2}) \quad r : v_r \in VK \tag{12}$$

$$1+\delta-K_r \le M(1-y_{r3}) \quad r:v_r \in VK \tag{13}$$

$$K_r^{LB} \le K_r \le K_r^{UB} \quad r:v_r \in VK \tag{14}$$

$$y_{r1}+y_{r2}+y_{r3}=1 \quad r:v_r \in VK \tag{15}$$

Whereas the convex hull encompasses the following equations:

$$K_r = k_{r1}+k_{r2}+k_{r3} \quad r:v_r \in VK \tag{16}$$

$$K_r^{LB} y_{r1} \le k_{r1} \le (1-\delta)y_{r1} \quad r:v_r \in VK \tag{17}$$

$$(1-\delta)y_{r2} \le k_{r2} \le (1+\delta)y_{r2} \quad r:v_r \in VK \tag{18}$$

$$(1+\delta)y_{r3} \le k_{r3} \le K_r^{UB} y_{r3} \quad r:v_r \in VK \tag{19}$$

$$y_{r1}+y_{r2}+y_{r3}=1 \quad r:v_r \in VK \tag{20}$$

The above equations require the definition of binary variable y_r, which takes a value of 1 if the associated term of the disjunction holds true and zero otherwise.

Enzymes controlling fluxes for which no kinetic equation is defined are also susceptible of being either over-expressed or down-regulated. This is accounted for in the model using variable v_r, which will show a higher or lower value when compared to the basal state according to whether the corresponding enzyme is over-expressed or inhibited, respectively. In order to account for this effect, we must consider the following disjunction:

$$\begin{bmatrix} Y_{r1} \\ v_r^{LB} \le v_r \le v_r^{bas}-\delta \end{bmatrix} \vee \begin{bmatrix} Y_{r2} \\ v_r^{bas}-\delta \le v_r \le v_r^{bas}+\delta \end{bmatrix} \vee \begin{bmatrix} Y_{r3} \\ v_r^{bas}+\delta \le v_r \le v_r^{UB} \end{bmatrix} \quad r:v_r \notin VK \tag{21}$$

where ${v_r}^{bas}$ is the value of velocity v_r in the basal state. We can transform this disjunction into equations by means of either the BigM or convex hull reformulation, with the former leading to the following equations:

$$v_r-(v_r^{bas}-\delta) \le M(1-y_{r1}) \quad r:v_r \notin VK \tag{22}$$

$$v_r^{bas}-\delta-v_r \le M(1-y_{r2}) \quad r:v_r \notin VK \tag{23}$$

$$v_r-(v_r^{bas}+\delta) \le M(1-y_{r2}) \quad r:v_r \notin VK \tag{24}$$

$$v_r^{bas}+\delta-v_r \le M(1-y_{r3}) \quad r:v_r \notin VK \tag{25}$$

$$v_r^{LB} \le v_r \le v_r^{UB} \quad r:v_r \notin VK \tag{26}$$

$$y_{r1}+y_{r2}+y_{r3}=1 \quad r:v_r \notin VK \tag{27}$$

and the later to the following ones:

$$v_r = v_{r1}+v_{r2}+v_{r3} \quad r:v_r \notin VK \tag{28}$$

$$v_r^{LB} y_{r1} \le v_{r1} \le (v_r^{bas}-\delta)y_{r1} \quad r:v_r \notin VK \tag{29}$$

$$(v_r^{bas}-\delta)y_{r2} \le v_{r2} \le (v_r^{bas}+\delta)y_{r2} \quad r:v_r \notin VK \tag{30}$$

$$(v_r^{bas}+\delta)y_{r3} \le v_{r3} \le v_r^{UB} y_{r3} \quad r:v_r \notin VK \tag{31}$$

$$y_{r1}+y_{r2}+y_{r3}=1 \quad r:v_r \notin VK \tag{32}$$

Note that Eqs. (20) and (32) are indeed the same equation, but defined over complementary sets. Therefore, we only need to include one of them in the model and define it over the complete set of velocities, that is, $r = 1,\ldots,p$.

Finally, the binary variable y_r is used to impose an upper limit ME on the number of enzymes that can be modified simultaneously:

$$\sum_{r:v_r \in VK} y_{r1} + \sum_{r:v_r \in VK} y_{r3} \le ME \tag{33}$$

The resulting non-convex MINLP embedding a stoichiometric–kinetic hybrid model for metabolic engineering applications is as follows:

$$\begin{aligned}
&\min \quad f \\
&s.t. \quad S \cdot v = 0 \\
&\left.\begin{aligned} &v_r = \gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}} \\ &K_r = k_{r1}+k_{r2}+k_{r3} \\ &K_r^{LB} y_{r1} \le k_{r1} \le (1-\delta)y_{r1} \\ &(1-\delta)y_{r2} \le k_{r2} \le (1+\delta)y_{r2} \\ &(1+\delta)y_{r3} \le k_{r3} \le K_r^{UB} y_{r3} \end{aligned}\right\} \quad r:v_r \in VK \\
&\left.\begin{aligned} &v_r = v_{r1}+v_{r2}+v_{r3} \\ &v_r^{LB} y_{r1} \le v_{r1} \le (v_r^{bas}-\delta)y_{r1} \\ &(v_r^{bas}-\delta)y_{r2} \le v_{r2} \le (v_r^{bas}+\delta)y_{r2} \\ &(v_r^{bas}+\delta)y_{r3} \le v_{r3} \le v_r^{UB} y_{r3} \end{aligned}\right\} \quad r:v_r \notin VK \\
&y_{r1}+y_{r2}+y_{r3}=1 \\
&v^{LB} \le v \le v^{UB} \\
&X^{LB} \le X \le X^{LB} \quad r=1,\ldots,p \\
&v, v_1, v_2, v_3 X, K, k_1, k_2, k_3 \in \Re^+
\end{aligned} \tag{34}$$

This model can be efficiently solved via the tailored outer approximation-based algorithm developed by the authors in a previous work (Pozo et al., 2010).

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compchemeng.2014.06.011.

References

Alves R, Vilaprinyo E, Hernandez-Bermejo B, Sorribas A. Mathematical formalisms based on approximated kinetic representations for modeling genetic and metabolic pathways. Biotechnol Genet Eng Rev 2008a;25:1–40.

Alves R, Vilaprinyo E, Sorribas A. Integrating bioinformatics and computational biology: perspectives and possibilities for in silico network reconstruction in molecular systems biology. Curr Bioinform 2008b;3:98–129.

Chaganti SR, Kim DH, Lalman JA. Flux balance analysis of mixed anaerobic microbial communities: effects of linoleic acid (LA) and pH on biohydrogen production. Int J Hydrogen Energy 2011;36:14141–52.

Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA 2010;107(41):17845–50.

Cotten C, Reed JL. Mechanistic analysis of multi-omics datasets to generate kinetic parameters for constraint-based metabolic models. BMC Bioinform 2013;14:32.

Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high throughput and computational data elucidates bacterial networks. Nature 2004;429(6987):92–6.

de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. AraGEM, a genome-scale reconstruction of the primary metabolic network in *Arabidopsis*. Plant Physiol 2010;152:579–89.

Edwards JS, Covert M, Palsson B. Metabolic modelling of microbes: the flux-balance approach. Environ Microbiol 2002;4(3):133–40.

Edwards JS, Palsson BO. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci USA 2000;97:5528–33.

Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. Curr Opin Biotechnol 2003;14:491–6.

Lee Y, Voit EO. Mathematical modeling of monolignol biosynthesis in Populus xylem. Math Biosci 2010;228(1):78–89.

Misener R, Floudas CA. Global optimization of mixed-integer quadratically-constrained quadratic programs (MIQCQP) through piecewise linear and edge-concave relaxations. Math Progr 2012;136(1):155–82.

Orth JD, Thiele I, Palsson BO. What is flux balance analysis? Nat Biotechnol 2010;28(3):245–8.

Palsson B. The challenges of in silico biology. Nat Biotechnol 2000;18:1147–50.

Papoutsakis ET. Equations and calculations for fermentations of butyric acid bacteria. Biotechnol Bioeng 1984;26:174–87.
Pharkya P, Maranas CD. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng 2006;8:1–13.
Poolman MG, Miguet L, Sweetlove LJ, Fell DA. A genomescale metabolic model of *Arabidopsis* and some of its properties. Plant Physiol 2009;151:1570–81.
Pozo C, Guillén-Gosélbez G, Sorribas A, Jiménez L. Outer approximation-based algorithm for biotechnology studies in systems biology. Comp Chem Eng 2010;34(10):1719–30.
Price ND, Reed JL, Palsson BO. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2004;2(11):886–97.
Price ND, Papin JA, Schilling CH, Palsson BO. Genomescale microbial in silico models: the constraints-based approach. Trends Biotechnol 2003;21:162–9.
Schilling CH, Edwards JS, Palsson BO. Toward metabolic phenomics: analysis of genomic data using flux balances. Biotechnol Progr 1999;15:288–95.
Schilling CH, Edwards JS, Letscher D, Palsson BO. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. Biotechnol Bioeng 2000;71:286–306.
Shlomi T, Eisenberg Y, Sharan R, Ruppin E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. Mol Syst Biol 2007;3:101.
Tomar N, De RK. Comparing methods for metabolic network analysis and an application to metabolic engineering. Gene 2013;521:1–14.
Torres NV, Voit EO, Glez-Alcón C, Rodríguez F. An indirect optimization method for biochemical systems: Description of method and application to the maximization of the rate of ethanol, glycerol, and carbohydrate production in Saccharomyces cerevisiae. Biotechnol Bioeng 1997;55(5):758–72.
Varma A, Palsson BO. Metabolic flux balancing-basic concepts, scientific and practical use. Nat Biotechnol 1994a;12(10):994–8.
Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli*. Appl Environ Microbiol 1994b;60:3724–31.
Vecchietti A, Sangbum L, Grossmann I. Modeling of discrete/continuous optimization problems: characterization and formulation of disjunctions and their relaxations. Comput Chem Eng 2003;27:433–48.
Vilaprinyo E, Alves R, Sorribas A. Use of physiological constraints to identify quantitative design principles for gene expression in yeast adaptation to heat shock. BMC Bioinform 2006;7:184.
Zorn K, Sahinidis NV. Global optimization of general non-convex problems with intermediate bilinear substructures. Optim Methods Softw 2014;29(3):442–62.