Stoichiometric capacitance reveals the theoretical capabilities of metabolic networks

Abdelhalim Larhlimi[1,2,*], Georg Basler[3], Sergio Grimbs[1], Joachim Selbig[1,2] and Zoran Nikoloski[3,*]

[1]Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, D-14476 Potsdam, Germany
[2]Bioinformatics, Max-Planck Institute for Molecular Plant Physiology, Am Mühlenberg 1, D-14476 Potsdam, Germany
[3]Systems Biology and Mathematical Modelling, Max-Planck Institute for Molecular Plant Physiology, Am Mühlenberg 1, D-14476 Potsdam, Germany

ABSTRACT

Motivation: Metabolic engineering aims at modulating the capabilities of metabolic networks by changing the activity of biochemical reactions. The existing constraint-based approaches for metabolic engineering have proven useful, but are limited only to reactions catalogued in various pathway databases.

Results: We consider the alternative of designing synthetic strategies which can be used not only to characterize the maximum theoretically possible product yield but also to engineer networks with optimal conversion capability by using a suitable biochemically feasible reaction called 'stoichiometric capacitance'. In addition, we provide a theoretical solution for decomposing a given stoichiometric capacitance over a set of known enzymatic reactions. We determine the stoichiometric capacitance for genome-scale metabolic networks of 10 organisms from different kingdoms of life and examine its implications for the alterations in flux variability patterns. Our empirical findings suggest that the theoretical capacity of metabolic networks comes at a cost of dramatic system's changes.

Contact: larhlimi@mpimp-golm.mpg.de, or nikoloski@mpimp-golm.mpg.de

Supplementary Information: Supplementary tables are available at *Bioinformatics* online.

1 INTRODUCTION

Metabolic reactions play a fundamental role in sustaining cell growth through the import of nutrients from the environment and their conversion into molecules rendering a viable organism. Metabolic reactions do not operate in isolation; they are in fact related through shared metabolites and form large-scale metabolic networks (Klamt and Stelling, 2003; Schilling *et al.*, 2000; Schuster *et al.*, 2000). Altering the structure of metabolic networks provides the means for modulating their capabilities. To this end, metabolic engineering aims at re-programming cellular networks by controlling the activity of biochemical reactions in order to obtain a desirable output and has already found numerous application in biotechnology and medicine (Feist *et al.*, 2009; Lee *et al.*, 2011; Müller and Hausmann, 2011). From the early attempts of engineering bacterial strains, it had become apparent that the biosynthetic capabilities of organisms are limited not only by the input in a form of nutrients but also by

the reaction fluxes obeying the basic physico-chemical constraints. Therefore, constraint-based approaches have gained considerable attention in the metabolic engineering community (Price *et al.*, 2003).

Based on the stoichiometry of the considered reactions, flux balance analysis (FBA) has provided the basic framework for investigating reaction fluxes in a metabolic network (Edwards and Palsson, 2000; Kauffman *et al.*, 2004; Lee *et al.*, 2006; Varma and Palsson, 1994). Extensions of this framework have subsequently allowed systematic investigations of network modifications directed at enhancing the production of selected targets. For instance, OptKnock (Burgard *et al.*, 2003), OptStrain (Pharkya *et al.*, 2004), OptReg (Pharkya and Maranas, 2006), OptForce (Ranganathan *et al.*, 2010) and CASOP (Haedicke and Klamt, 2010) as well as EMILiO (Yang *et al.*, 2011) have facilitated prediction and *in silico* verification of network modifications, including insertion and deletion of reactions from other species or under- and over-expression of gene products. All of these alternatives for network modifications rely on the assumption that gene expression is proportional to reaction flux. The findings of these optimization-based approaches suggest alternative pathways and/or flux redirections necessary to achieve the objective of interest. We note that the enumerated approaches, although very useful in designing strategies for metabolic engineering, are limited to the reactions already catalogued in various databases (Karp *et al.*, 2000; Kanehisa and Goto, 2000; Kumar *et al.*, 2012). The performance of the suggested strategies is in turn determined by the maximum achievable yield of a desired target which is calculated with the help of various optimization techniques.

Herein, we argue that quantifying the performance of metabolic engineering strategies in a given metabolic network should rely on the maximum 'theoretically' possible product yield in the network, given by the optimal inter-conversion of metabolites by using any biochemically feasible reaction. The notion of theoretical capacity refers to the estimated maximum production that can be delivered during a given period, provided a set of inputs. To emphasize the analogy, in the case of metabolic networks, the inputs are represented by the nutrients together with the set of metabolic intermediates and reactions, while the maximum production to be estimated is given by the suitably chosen objective function. Clearly, the inputs are given by the existing biochemical knowledge structured in a form of cellular networks together with the corresponding environmental constraints.

---

[*]To whom correspondence should be addressed.

Estimation of the theoretical capacity for large-scale metabolic networks is all the more pressing issue due to the incompleteness of biological knowledge. The gaps in the current genome-scale metabolic networks arising as a result of the uncharacterized function of 20–60% of gene products (Hanson *et al.*, 2010) can have profound effect on the outcome of optimization-based approaches (Feist *et al.*, 2009; Marashi and Bockmayr, 2011). This provides the first justification for using the theoretical capacity as a golden standard. Therefore, the maximum theoretically possible yield lends itself as a standard analogous to the 'theoretical capacity' used in finance, transportation/logistics and batteries research (Todorov *et al.*, 1999; Transportation Research Board, 1997).

In addition, given the large fraction of unknown chemical reactions in biological systems and the immense space of macromolecules potentially catalyzing chemical reactions (Dobson, 2004), the most promising targets for metabolic engineering may not be found among the already characterized enzymes. Instead, synthetic alternatives obtained through systematic screening for chemically feasible, but so far uncharacterized, reactions may offer novel promising options for metabolic engineering (Basler *et al.*, 2012). This becomes particularly interesting in cases where the synthetic reactions can be readily decomposed into pathways of equivalent net conversion capability, involving only known enzymes.

The article is organized as follows: we first provide a brief overview of optimization-based approaches for metabolic network analysis. We then give a detailed and formal description of our novel concept, called stoichiometric capacitance. Our approach is designed to determine the theoretical capacity of a given metabolic network and relies on determining a feasible chemical reaction whose addition to the network results in maximum product yield. Therefore, it can be used in designing synthetic metabolic engineering strategies aimed at improving the yield of any product of interest. We illustrate the proposed concept on the Krebs cycle with the idea of verifying if glucose can be synthesized from fatty acids (de Figueiredo *et al.*, 2009). In addition, we apply the proposed method to improve biomass yield in genome-scale metabolic networks of 10 organisms from different kingdoms of life. Finally, we examine the implications of stoichiometric capacitance for the alterations in flux variability patterns.

2 METHODS

2.1 Optimization-based approaches

Metabolic networks are composed of metabolites inter-converted by a set of biochemical reactions. Based on the chosen system boundary, the metabolites in a given metabolic network can be categorized as 'internal' and 'external' (Heinrich and Schuster, 1996). The structure of a metabolic network is fully described by its stoichiometric matrix $S \in \mathbb{R}^{m \times n}$ which is equivalent to the incidence matrix of the underlying weighted directed hypergraph (Klamt *et al.*, 2009). The rows of the stoichiometric matrix correspond to the internal metabolites and each column represents a reaction. At steady state, the rate of formation of every internal metabolite equals the rate of its consumption. This is expressed by the flux balance equation

$$Sv = 0, \qquad (1)$$

where $v \in \mathbb{R}^n$ denotes the vector of fluxes, also termed 'flux distribution', for the reactions included in the metabolic network. Aside from the flux balance conditions, flux distributions are also subject to other constraints related to

thermodynamics, environmental conditions and flux capacities, which are often modeled as linear inequalities

$$v^{\min} \leq v \leq v^{\max}, \qquad (2)$$

where $v^{\min}$ and $v^{\max}$ stand for the lower and upper bounds of the flux capacities of the considered reactions, respectively.

In addition, the law of mass conservation implies that all internal reactions must be stoichiometrically 'mass balanced', i.e. each stoichiometric column corresponding to an internal reaction must lie in the kernel of the mass matrix, M, which contains the molecular sum formulas of the internal metabolites. More formally,

$$MS_{*j} = 0 \text{ for all internal reactions } j \in \{1, \ldots, n\}, \qquad (3)$$

where $M_{k,l}$ is the k-th atomic species (i.e. chemical element) in the sum formula of the l-th metabolite and S_{*j} stands for the j-th column in the stoichiometric matrix corresponding to the internal reaction j. Interestingly, this fundamental physical constraint is rarely considered in metabolic network analyses, although it provides a necessary precondition for the flux balance equation (1). Indeed, violation of equation (3) for any metabolite would allow for the removal or creation of atomic species out of nothing, thus rendering the flux balance equation meaningless.

The linear constraints given in equations (1) and (2) define a polyhedron

$$P = \{v \in \mathbb{R}^n \mid Sv = 0, \ v^{min} \leq v \leq v^{max}\}, \qquad (4)$$

which contains all possible steady-state flux distributions. Here, we are interested in those flux distributions that maximize a predefined objective corresponding to a metabolic network function, such as growth or product yield. We assume that the network objective of interest is a linear function τ defined by a vector $c \in \mathbb{R}^n$, such that $c^T v \geq 0$ for all $v \in P$. As in FBA, the optimal value, τ^*, for a metabolic network with stoichiometric matrix, S, is obtained by solving the following linear programming (LP) problem:

$$\tau^* = \max \{c^T v : \ Sv = 0, \ v^{min} \leq v \leq v^{max}\}. \qquad (5)$$

Clearly, the set of feasible fluxes for the problem in equation (5) is narrowed down by the constraints in equations (1) and (2). Namely, under these constraints, certain reversible reactions proceed only in one direction, while other reactions, termed 'blocked', are unable of carrying flux. A reaction carrying non-zero flux is termed 'unblocked'. Furthermore, due to these constraints, a non-zero flux of certain reactions implies a non-zero flux through other ones (Burgard *et al.*, 2004). In particular, several reactions can be essential for performing the considered metabolic capability τ.

In analogy to essential reactions for the growth of a micro-organism (Edwards and Palsson, 2000), if $\tau^* > 0$, an unblocked reaction i is termed *essential* for τ if there exists $\alpha^i > 0$ such that $c^T v \leq \alpha^i v_i$ for all $v \in P$. Let the essential reactions for τ be denoted by Ψ. If we have $\alpha^i = \min \{\frac{c^T v}{v_i} : \ v \in P\}$ for all essential reactions $i \in \Psi$, then the optimal (i.e. maximum) capability τ^* is given by

$$\tau^* = \min \{\alpha^i v_i : \ v \in P \text{ and } i \in \Psi\}. \qquad (6)$$

Therefore, increasing the maximum value of the objective τ^* requires the relaxation of the upper bound imposed by the essential reactions on the objective τ. Such a relaxation either allows essential reactions to carry more flux or reduces the set of essential reactions, i.e. certain reactions are no longer essential for τ.

In the following, we introduce the concept of stoichiometric capacitance, which inherently depends on the aforementioned relaxation, and illustrate its applications in genome-scale metabolic networks as well as well-studied pathways.

2.2 Stoichiometric capacitance

The relaxation of some of the linear inequalities (2) may lead to an increase of the maximal objective τ^* in the original network. This can be achieved by placing the living system in an environment which is in favor of

the maximization of the objective τ; for instance, the living system can import more nutrients. The change in the objective value τ due to the relaxation of environmental conditions can be assessed using shadow prices in linear programming (Varma and Palsson, 1993). Alternatively, a significant improvement of τ can be obtained by determining a sparse vector $r \in \mathbb{R}^m$ such that $\tau^*(r) > \tau^*$ with

$$\tau^*(r) = \max \{c^T v : \ Sv + r = 0, \ v^{\min} \leq v \leq v^{\max}\}. \tag{7}$$

Let $I(r) = \{i \in \{1, \ldots, m\} \mid r_i \neq 0\}$. The vector r can be chosen to simulate a boundary reaction exporting (respectively importing) the accumulated (respectively depleted) internal metabolites in $I(r)$. More importantly, if the vector r lies in the null space of the mass matrix M, defined in equation (3), then r represents the stoichiometry of a feasible chemical conversion of the internal metabolites $I(r)$, up to multiplication by scalars. This may include any set of feasible chemical reactions with net conversion given by r. In the following, we refer to r as stoichiometric capacitance. We note that if the net conversion of the stoichiometric capacitance is thermodynamically feasible, its inclusion in the considered metabolic model, specified by equation (4), may increase the network objective, τ, given the same amount of nutrients.

In physics and engineering, the term capacitance is classically used to denote the change that must be added to a system to raise the electrical potential by one unit. Analogously, the concept of stoichiometric capacitance will denote the change in the metabolic structure given by the feasible chemical reaction whose inclusion in the network increases the objective of interest. The reference point for the increase is provided by the maximum value of the objective in the original metabolic network.

To investigate the extent to which we can improve the objective τ by adding a stoichiometric capacitance r to an investigated metabolic network, for each internal metabolite $i \in \{1, \ldots, m\}$, we first introduce a binary variable x_i such that $i \in I(r)$ implies $x_i = 1$. Let μ denote a given maximum number of non-zero components of r and λ be a predefined upper bound on the flux through the added reaction. The maximum capability $\tau^*(r)$ and the corresponding stoichiometric capacitance r can be obtained by solving the following mixed integer linear programming (MILP) problem

$$\tau^*(r) = \max \ c^T v$$
$$\text{subject to:} \qquad Sv + r = 0,$$
$$v^{\min} \leq v \leq v^{\max},$$
$$Mr = 0, \tag{i}$$
$$Tr \leq 0, \tag{ii}$$
$$-\lambda x \leq r \leq \lambda x, \tag{iii}$$
$$\sum_{i=1}^{m} x_i \leq \mu, \tag{iv}$$
$$r_i = x_i = 0, \text{for all } i \notin E, \tag{v}$$
$$v \in \mathbb{R}^n, \ r \in \mathbb{R}^m, \ x \in \{0,1\}^m.$$

Condition (i) ensures that the stoichiometric capacitance, i.e. the added reaction, is chemically feasible. In other words, it ensures that r represents a mass-balanced net conversion that preserves the number of consumed and produced atoms. The inequality in (ii) guarantees that the stoichiometric capacitance is thermodynamically feasible, with T denoting the vector of the standard Gibbs free energy of formation of the internal metabolites. Constraints (iii) and (iv) ensure that r involves a limited number of metabolites from $I(r)$, i.e. $|I(r)| \leq \mu$. Finally, the optional condition (v) allows modeling a stoichiometric capacitance that involves only a predefined subset $E \subseteq \{1, \ldots, m\}$ of metabolites. For instance, E may be the set of metabolites in a particular subcellular compartment or tissue. It is clear that further constraints can be included in the above MILP problem to determine a stoichiometric capacitance corresponding to a net conversion of only reactions that are more likely to occur in nature.

Although the stoichiometric capacitance represents a chemically feasible reaction, it may not correspond to any of the characterized enzymatic reactions from investigated organisms. Therefore, it is interesting to address the problem of determining if the stoichiometric capacitance, determined by solving the proposed MILP problem, can be expressed as a combination of a given set $\mathcal{R}$ of known enzymatic reactions. The set $\mathcal{R}$ can be defined by

using the reactions catalogued in several existing databases of biochemical reactions (Kanehisa and Goto, 2000; Karp *et al.*, 2000; Kumar *et al.*, 2012).

Here, we define a linear programming problem whose solution resolves the raised issue. Suppose, the stoichiometry of the reactions in $\mathcal{R}$ is given by the matrix

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix},$$

where $A \in \mathbb{R}^{m \times p}$, $B \in \mathbb{R}^{m \times q}$ and $C \in \mathbb{R}^{m' \times q}$ such that A is a stoichiometric matrix defined in the space of m internal metabolites and $\begin{pmatrix} B \\ C \end{pmatrix}$ describes the stoichiometry of the reactions in $\mathcal{R}$ which involve additional m' metabolites. An identified stoichiometric capacitance r can be considered as a net conversion of the reactions in $\mathcal{R}$ if there exist two vectors $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$ such that

$$A\alpha + B\beta = r,$$
$$C\beta = 0,$$
$$\alpha_i(TA_{*i}) \leq 0, \text{for all } i \in \{1, \ldots, p\},$$
$$\beta_i(TB_{*i} + T'C_{*i}) \leq 0, \text{for all } i \in \{1, \ldots, q\},$$

with T' denoting the vector of the standard Gibbs free energy of formation of the additional m' metabolites.
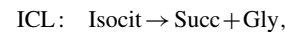
3 RESULTS AND DISCUSSION

We implemented the proposed method for determining the stoichiometric capacitance within the TOMLAB environment (Holmström, 1999).

To illustrate our novel concept, we first consider the Krebs cycle whose mass-balanced reactions are depicted in Figure 1. Although a net synthesis of glucose from fatty acids may be feasible in the entire network of human metabolism (Kaleta *et al.*, 2011), a recent theoretical analysis has demonstrated that such a conversion cannot be achieved by using the Krebs cycle (de Figueiredo *et al.*, 2009). Here, we investigate whether this conversion becomes possible when a stoichiometric capacitance is added to the model of the Krebs cycle. To this end, we focus on finding a stoichiometric capacitance r with the objective of synthesizing glucose, where the set $I(r)$ of involved metabolites consists only of the following six metabolites: acetyl coenzyme A (AcCoA), coenzyme A (CoA), isocitrate (Isocit), (S)-malate (Mal), succinate (Succ) and water (H_2O). The rows of the matrices M and T corresponding to these metabolites are included in Supplementary Table S1.

Solving the proposed MILP problem, with S given by the stoichiometric matrix of the glycolysis/gluconeogenesis and Krebs cycle (de Figueiredo *et al.*, 2009), the matrices M and T for all the involved metabolites, and the synthesis of glucose as a metabolic function τ, results in the stoichiometric capacitance r whose equation is

$$\text{Isocit} + \text{AcCoA} + H_2O \rightarrow \text{Mal} + \text{Succ} + \text{CoA}.$$
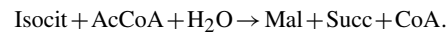
Interestingly, the stoichiometric capacitance r is the sum of the enzymatic reactions catalyzed by the isocitrate lyase (ICL) and the malate synthase (MAS) from the glyoxylate cycle:

$$\text{ICL}: \ \text{Isocit} \rightarrow \text{Succ} + \text{Gly},$$
$$\text{MAS}: \ \text{Gly} + \text{AcCoA} + H_2O \rightarrow \text{Mal} + \text{CoA},$$

where 'Gly' stands for Glyoxylate. This result is in agreement with the findings in de Figueiredo *et al.* (2009) which state that glucose can be produced out of AcCoA using the Krebs cycle when the
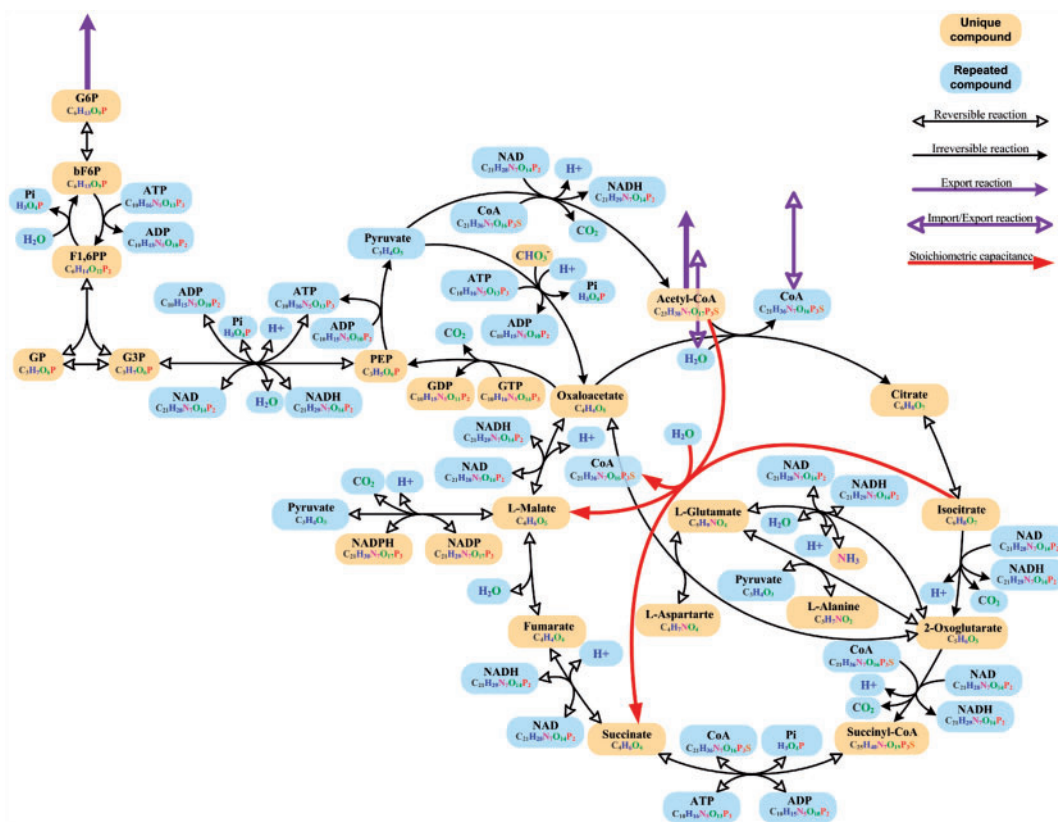
Fig. 1. Metabolic model of glycolysis/gluconeogenesis and Krebs cycle in humans adopted from de Figueiredo *et al.* (2009). It has been shown that a net synthesis of glucose from fatty acids cannot be achieved by using the Krebs cycle. A stoichiometric capacitance with the synthesis of glucose as a metabolic function τ is colored in red. Adding this chemical reaction to the network allows for a net conversion of fatty acids into glucose

Table 1. Comparison of the increases in biomass yield by using the stoichiometric capacitance (SC)

Network	FBA	SC	Reaction equation
*M. barkeri, iAF*692	0.03	160847%	6 H_2O + 2 nac ($C_6H_4NO_2$) = 2 glycogen ($C_6H_{10}O_5$) + N_2
*S. cerevisiae, iND*750	0.10	72%	24 CO_2 + 10 sbt_L ($C_6H_{14}O_6$) = 14 glycogen ($C_6H_{10}O_5$) + 19 O_2
*M. tuberculosis, iNJ*661	0.05	486%	11 H + 58 ppa ($C_3H_5O_2$) = 22 glycogen ($C_6H_{10}O_5$) + 3 ttdca ($C_{14}H_{27}O_2$)
*E. coli, iJR*904	0.92	830%	12 CO_2 + 20 H = 2 glycogen ($C_6H_{10}O_5$) + 7 O_2
*E. coli, iAF*1260	0.74	3202%	6 CO_2 + 5 glc_D ($C_6H_{12}O_6$) = 6 glycogen ($C_6H_{10}O_5$) + 6 O_2
*E. coli, iJO*1366	0.98	1714%	12 CO_2 + 9 succ ($C_4H_4O_4$) = 8 acon ($C_6H_3O_6$) + 6 H_2O_2
*S. aureus, iSB*619	0.07	44%	5aizc ($C_9H_{11}N_3O_9P$) + arg_L ($C_6H_{15}N_4O_2$) = aicar ($C_9H_{13}N_4O_8P$) + citr_L ($C_6H_{13}N_3O_3$)
*H. pylori, iIT*341	0.69	370%	2 hom_L ($C_4H_9NO_3$) = 2obut ($C_4H_5O_3$) + acac ($C_4H_5O_3$) + 2 NH_4
H. vulgare seeds	2.76	0.75%	2 mal ($C_4H_6O_5$) + 3 phpyr ($C_3H_5O_7P$) = 5 CO_2 + 3 e4p ($C_4H_9O_7P$)
B. subtilis	0.40	58%	43 dha ($C_3H_6O_3$) + 6 3dhsk ($C_7H_7O_5$) + 9 pime ($C_7H_{10}O_4$) = 39 glycogen ($C_6H_{10}O_5$)

FBA stands for the optimal biomass yield using flux balance analysis, SC denotes the increases in the optimal biomass yield when a stoichiometric capacitance, determined by solving the proposed MILP problem, is added to the corresponding network. The last column shows the reaction equation of the corresponding stoichiometric capacitances.

enzymes involved in the glyoxylate shunt are added. We note that the proposed MILP problem often does not have a unique solution, i.e. there may be more than one stoichiometric capacitance satisfying the imposed constraints. For instance, additional stoichiometric capacitances for the synthesis of glucose are given in Supplementary Table S2.

To benchmark the proposed method, in the following we explore the possibility of increasing biomass production in the following metabolic networks: *Saccharomyces cerevisiae*, *iND*750 (Duarte *et al.*, 2004), *Methanosarcina barkeri*, *iAF*692 (Feist *et al.*, 2006), *Mycobacterium tuberculosis*, *iNJ*661 (Jamshidi and Palsson, 2007), *Escherichia coli*, *iJR*904, *iAF*1260, *iJO*1366 (Feist *et al.*, 2007; Orth *et al.*, 2011; Reed *et al.*, 2003), *Staphylococcus aureus*, *iSB*619 (Becker and Palsson, 2005), *Helicobacter pylori*, *iIT*341 (Thiele *et al.*, 2005), *Hordeum vulgare* seed (Grafahrend-Belau *et al.*, 2009) and *Bacillus subtilis* (Oh *et al.*, 2007).

We constrain the stoichiometric capacitance to include up to four metabolites localized in the cytoplasm of the 10 considered networks. The results for the increase of biomass yield due to the inclusion of the stoichiometric capacitance are included in Table 1. It is evident that the stoichiometric capacitance allows for a very small increase of 0.75% in the case of *H. vulgare*'s relatively small and tightly constrained network. However, in the cases of *M. barkeri*, *iAF*692 and *E. coli*, *iAF*1260, the maximum theoretically possible yield is 160 847% and 3202%, respectively.

To further investigate the effect of adding the stoichiometric capacitance to a given metabolic network, we used flux variability analysis (Mahadevan and Schilling, 2003). To this end, we computed the flux ranges with and without adding the corresponding stoichiometric capacitance to the networks of *E. coli*, *iJO*1366, *S. cerevisiae*, *iND*750 and *M. tuberculosis*, iNJ661. Reactions whose flux is zero in all optimal FBA pathways are called 'excluded reactions' as they are not involved in optimizing the biomass yield. Other reactions, termed 'indispensable' for growth, carry non-zero flux in all optimal FBA pathways. Figure 2 shows the changes in the types of reactions due the inclusion of the stoichiometric capacitance in the three considered network models.

By inspecting the panels in Figure 2, the following trends, obtained from the flux variability analysis, become apparent: (i) the number of indispensable reactions specific to the network with stoichiometric capacitance (in orange) is smaller compared to that of indispensable reactions specific to the original network (i.e. without the stoichiometric capacitance, in red), (ii) the number of excluded reactions specific to the network with stoichiometric capacitance (in green) is smaller compared to that of excluded reactions specific to the original network (in blue) and (iii) the number of reactions that switch from excluded in the original network to indispensable in the altered (in blue/yellow) is smaller than that of reactions that switch from indispensable in the original network to excluded in the network altered with the stoichiometric capacitance (in red/green); however, this holds valid only for *E. coli* and *S. cerevisiae*.

In addition, we observe in Figure 2 that only few reactions become indispensable (numbers in orange) or excluded (numbers in green) and only few reactions change between these types when adding the stoichiometric capacitance in the three networks. On the other hand, some of the indispensable (respectively excluded) reactions in the original network become no longer indispensable (respectively excluded). Consequently, the flux of these reactions may vary dramatically and so affect the optimal biomass yield, which corresponds to the relaxation of the constraints imposed on the feasible flux distributions when adding the stoichiometric capacitance. However, we point out that by far most reactions remain indispensable or excluded (red/orange and blue/green intersections), which demonstrates that the stoichiometric capacitance may still strongly rely on the fluxes of reactions from the original network, while allowing for a significant increase in biomass yield (Table 1). On one hand, this suggests that the identified stoichiometric capacitance allows for improving the performance of the original network without strongly redirecting the activity of the commonly used pathways, thus yielding a realistic scenario for metabolic engineering approaches. Nevertheless, since the changes happen across inspected types of reactions with and without stoichiometric capacitance, we conclude that the inclusion of the corresponding stoichiometric capacitance in each of the considered models may result in dramatic system's changes.
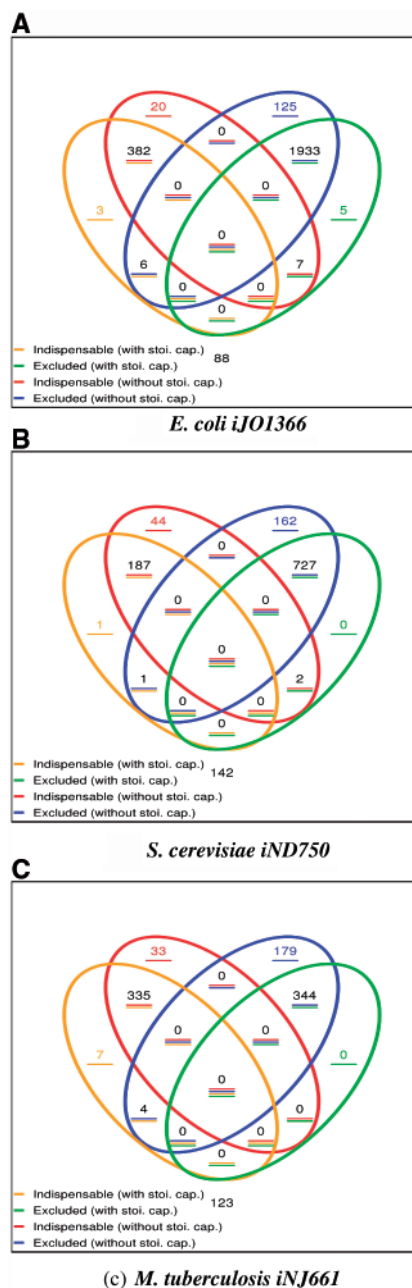
Fig. 2. Venn diagram of reaction types in the metabolic networks of *E. coli*, *iJO*1366, *S. cerevisiae*, *iND*750 and *M. tuberculosis*, *iNJ*661. The reaction types are determined by flux variability analysis using either the corresponding original FBA model (without stoi. cap.) or the altered FBA model (with stoi. cap.) which includes the corresponding stoichiometric capacitance given in Table 1. Excluded reactions are those which are not involved in any optimal FBA pathway, whereas reactions termed indispensable for growth carry non-zero flux in all optimal FBA pathways

4 CONCLUSION

Our novel concept of stoichiometric capacitance allows not only for quantifying the theoretical capacity of metabolic networks with respect to a given objective but also suggests verifiable interventions (i.e. feasible chemical reactions) which provide this capacity, thus

facilitating the design of metabolic engineering strategies. The used optimization-based formulation takes into account physico-chemical constraints to restrict the search space in a biochemically meaningful manner. To this end, we used a mixed integer linear program that can be efficiently resolved with existing optimization software. It becomes apparent that the proposed approach can easily be extended to consider non-linear objective functions discussed in Schuetz *et al.* (2007). Further constraints can be included to determine a stoichiometric capacitance corresponding to a net conversion of only reactions that are more likely to occur in nature. Our theoretical framework was used in addressing a concrete question that has long been debated in biochemistry, and concerns the possibility to produce sugars from fatty acids. The identified stoichiometric capacitance nicely confirms existing experimental results. This example also illustrates how the stoichiometric capacitance, taken as an overall reaction, can be decomposed into single enzymatic reactions. We also provided a linear program that addresses the issue of determining whether the proposed stoichiometric capacitance can be expressed as an overall conversion of a given set of enzymatic reactions. Furthermore, the analysis of genome-scale metabolic networks from various organisms showed that the stoichiometric capacitance could dramatically increase biomass production. Taken together, we believe that the presented approach will prove valuable not only in evaluating and classifying different metabolic networks according to their theoretical efficiency but also as a method of choice for designing synthetic metabolic engineering strategies.

Conflict of Interest: none declared.

REFERENCES

Basler,G. *et al.* (2012) Optimizing metabolic pathways by screening for feasible synthetic reactions. *Biosystems*, **109**, 186–191.

Becker,S.A. and Palsson,B.Ø. (2005) Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation. *BMC Microbiol.*, **5**, 8.

Burgard,A.P. *et al.* (2003) OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.*, **84**, 647–657.

Burgard,A.P. *et al.* (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.*, **14**, 301–312.

de Figueiredo,L.F. *et al.* (2009) Can sugars be produced from fatty acids? A test case for pathway analysis tools. *Bioinformatics*, **25**, 152–158.

Dobson,C.M. (2004) Chemical space and biology. *Nature*, **432**, 824–828.

Duarte,N.C. *et al.* (2004) Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.*, **14**, 1298–1309.

Edwards,J.S. and Palsson,B.Ø. (2000) Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics*, **1**, 1.

Feist,A.M. *et al.* (2006) Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri. *Mol. Syst. Biol.*, **2**, 2006.0004.

Feist,A.M. *et al.* (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.

Feist,A.M. *et al.* (2009) Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.*, **7**, 129–143.

Grafahrend-Belau,E. *et al.* (2009) Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism. *Plant Physiol.*, **149**, 585–598.

Hädicke,O. and Klamt,S. (2010) CASOP: a computational approach for strain optimization aiming at high productivity. *J. Biotechnol.*, **147**, 88–101.

Hanson,A.D. *et al.* (2010) Unknown proteins and 'orphan' enzymes: the missing half of the engineering parts list–and how to find it. *Biochem. J.*, **425**, 1–11.

Heinrich,R. and Schuster,S. (1996) *The Regulation of Cellular Systems*. Chapman and Hall, NY, USA.

Holmström,K. (1999) The TOMLAB Optimization Environment in Matlab. *Adv. Model. Optim.*, **1**, 47–69.

Jamshidi,N. and Palsson,B.Ø. (2007) Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC Syst. Biol.*, **1**, 26.

Kaleta,C. *et al.* (2011) *In Silico* evidence for gluconeogenesis from fatty acids in humans. *PLoS Comput. Biol.*, **7**, e1002116.

Kanehisa,M. and Goto,S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.

Karp,P.D. *et al.* (2000) The EcoCyc and MetaCyc databases. *Nucleic Acids Res.*, **28**, 56–59.

Kauffman,K.J. *et al.* (2004) Advances in flux balance analysis. *Curr. Opin. Biotechnol.*, **14**, 491–496.

Klamt,S. and Stelling,J. (2003) Two approaches for metabolic pathway analysis? *Trends Biotechnol.*, **21**, 64–69.

Klamt,S. *et al.* (2009) Hypergraphs and cellular networks. *PLoS Comput. Biol.*, **5**, e1000385.

Kumar,A. *et al.* (2012) MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC Bioinformatics*, **13**, 6.

Lee,J.M. *et al.* (2006) Flux balance analysis in the era of metabolomics. *Brief. Bioinformatics*, **7**, 140–150.

Lee,J.W. *et al.* (2011) Systems metabolic engineering for chemicals and materials. *Trends Biotechnol.*, **29**, 370–378.

Mahadevan,R. and Schilling,C.H. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.*, **5**, 264–276.

Marashi,S.-A. and Bockmayr,A. (2011) Flux coupling analysis of metabolic networks is sensitive to missing reactions. *BioSystems*, **103**, 57–66.

Müller,M.M. and Hausmann,R. (2011) Regulatory and metabolic network of rhamnolipid biosynthesis: traditional and advanced engineering towards biotechnological production. *Appl. Microbiol. Biotechnol.*, **91**, 251–264.

Oh,Y.-K. *et al.* (2007) Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.*, **282**, 28791–28799.

Orth,J.D. *et al.* (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011. *Mol. Syst. Biol.*, **7**, 535.

Pharkya,P. and Maranas,C.D. (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab. Eng.*, **8**, 1–13.

Pharkya,P. *et al.* (2004) OptStrain: A computational framework for redesign of microbial production systems. *Genome Res.*, **14**, 2367–2376.

Price,N.D. *et al.* (2003) Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol.*, **21**, 162–169.

Ranganathan,S. *et al.* (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.*, **6**, e1000744.

Reed,J.L. *et al.* (2003) An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol.*, **4**, R54.

Schilling,C.H. *et al.* (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**, 229–248.

Schuetz,R. *et al.* (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol. Syst. Biol.*, **3**, 119.

Schuster,S. *et al.* (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.*, **18**, 326–332.

Thiele,I. *et al.* (2005) Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. *J. Bacteriol.*, **187**, 5818–5830.

Todorov,Y.R. *et al.* (1999) Determination of theoretical capacity of metal ion-doped LiMn2O4 as the positive electrode in Li-ion batteries. *J. Power Sources*, **77**, 198–201.

Transportation Research Board. (1997) *Highway capacity manual*. Technical report, National Research Council, Washington, DC.

Varma,A. and Palsson,B.Ø. (1993) Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors. *J. Theor. Biol.*, **165**, 477–502.

Varma,A. and Palsson,B.Ø. (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.*, **60**, 3724–3731.

Yang,L. *et al.* (2011) EMILiO: a fast algorithm for genome-scale strain design. *Metab. Eng.*, **13**, 272–281.