Predicting Metabolic Fluxes Using Gene Expression Differences As Constraints

Rogier J.P. van Berlo, Dick de Ridder, Jean-Marc Daran, Pascale A.S. Daran-Lapujade, Bas Teusink, and Marcel J.T. Reinders

Abstract—A standard approach to estimate intracellular fluxes on a genome-wide scale is flux-balance analysis (FBA), which optimizes an objective function subject to constraints on (relations between) fluxes. The performance of FBA models heavily depends on the relevance of the formulated objective function and the completeness of the defined constraints. Previous studies indicated that FBA predictions can be improved by adding regulatory *on/off* constraints. These constraints were imposed based on either absolute [21], [3] or relative [20] gene expression values. We provide a new algorithm that directly uses regulatory *up/down* constraints based on gene expression data in FBA optimization (tFBA). Our assumption is that if the activity of a gene drastically changes from one condition to the other, the flux through the reaction controlled by that gene will change accordingly. We allow these constraints to be violated, to account for posttranscriptional control and noise in the data. These *up/down* constraints are less stringent than the *on/off* constraints as previously proposed. Nevertheless, we obtain promising predictions, since many *up/down* constraints can be enforced. The potential of the proposed method, tFBA, is demonstrated through the analysis of fluxes in yeast under nine different cultivation conditions, between which approximately 5,000 regulatory *up/down* constraints can be defined. We show that changes in gene expression are predictive for changes in fluxes. Additionally, we illustrate that flux distributions obtained with tFBA better fit transcriptomics data than previous methods. Finally, we compare tFBA and FBA predictions to show that our approach yields more biologically relevant results.

Index Terms—Gene expression data, mixed-integer linear programming, optimization function, FBA.

1 Introduction

In recent years, metabolic flux analysis (MFA) has become one of the major tools in metabolic engineering [26], [29]. Knowledge of metabolic pathways, describing at least the essential central reactions of the cell, is combined with measurements of net conversion rates to predict the fluxes through all reactions of the cell's metabolic network, exploiting the fact that all metabolites should be balanced in steady state.

However, several questions around metabolic flux analysis remain unsolved, one of the most significant ones being the identification of fluxes through parallel routes and cycles [30]. These are often estimated by employing flux-balance analysis (FBA). FBA assumes that the operation of a metabolic network can be described by optimality principles [18], [23]. Many optimality principles have been proposed, such as biomass, ATP, and product maximization. In this case, we employ flux minimization, because the product, biomass, and ATP (required for maintenance) rates have been measured and can directly be incorporated (as constraints) into the FBA optimization.[1] Flux minimization postulates that cells convert the available nutrients into products in enzymatically the most efficient way and that the shortest path is the most enzymatically efficient path. However, it is questionable whether cells really operate as efficiently as possible. Moreover, it is unlikely that the shortest path is always the most efficient path, since efficiency also depends on the kinetics of the reactions. This suggests that unreliable flux estimates may be obtained using FBA. Therefore, we search for additional experimental data to improve these estimates. A valuable data source for this purpose might be gene expression data.

We have a compendium of chemostat-based transcriptome data sets, comprised of 170 microarrays together with extracellular flux measurements [10]. In this study, our aim is to utilize the gene expression data to improve intracellular flux predictions. We make use of the assumption that if the expression of a gene changes from one condition to the other, the flux through the reaction controlled by that gene will change accordingly.

The possibility of improving FBA predictions using gene expression data has been studied before. Most of these studies use *absolute* gene expression values for this purpose. For example, Covert and Palsson [4] proposed to add

- *R.J.P. van Berlo, D. de Ridder, and M.J.T. Reinders are with The Delft Bioinformatics Lab, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands, and the Kluyver Centre for Genomics of Industrial Fermentation, Julianalaan 67, 2628 BC Delft, The Netherlands. E-mail: {r.j.p.vanberlo, d.deridder, m.j.t.reinders}@tudelft.nl.*
- *J.-M. Daran and P.A.S. Daran-Lapujade are with the Industrial Microbiology Group, Department of Biotechnology, Delft University of Technology, Julianalaan 67, 2628 BC Delft, The Netherlands, and the Kluyver Centre for Genomics of Industrial Fermentation, Julianalaan 67, 2628 BC Delft, The Netherlands. E-mail: {j.g.daran, p.a.s.daran-lapujade}@tudelft.nl.*
- *B. Teusink is with the Systems Bioinformatics IBIVU, Faculty of Earth and Life Sciences, Vrije Universiteit, De Boelelaan 1081A, 1081 HV Amsterdam, The Netherlands, and the Kluyver Centre for Genomics of Industrial Fermentation, Julianalaan 67, 2628 BC Delft, The Netherlands. E-mail: bas.teusink@falw.vu.nl.*

Manuscript received 18 Nov. 2008; revised 27 Feb. 2009; accepted 23 Apr. 2009; published online 29 May 2009.
For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBB-2008-11-0200. Digital Object Identifier no. 10.1109/TCBB.2009.55.

1. As a result, it does not make any sense anymore to optimize for them. Note that even if one incorporates the measured growth, product, and ATP rates in FBA optimization, multiple solutions still exist.

regulatory *on/off* constraints based on gene expression data, a method they named rFBA. Their constraints forced reaction rates to zero, if the genes encoding the enzymes for those reactions were not transcribed at all [3]. A Boolean formalism was used to describe the effect of isoenzymes (OR) and protein complexes (AND) on reaction rates. They utilized these *on/off* constraints to model *dynamic* experiments. Akesson et al. [1] suggested to employ a similar approach for modeling *steady state* situations. They still forced fluxes to zero if genes were not present, but did not explicitly exploit the Boolean formalism. They used their approach in modeling aerobic and anaerobic glucose-limited chemostat cultivations, as well as batch cultivations with high glucose levels. They showed that, particularly in this last situation, standard FBA fails and gene-expression-based constraints helped in improving predictions. Shlomi et al. [20] proposed to use the same Boolean formalism as Covert and Palsson [4] for modeling static situations, which they called SR-FBA. In this model, the Boolean rules were transformed into linear constraints that could directly be incorporated into FBA optimization. For example, a rule like "$a = b$ AND c" was formulated as "$-1 \geq 2b + 2c - 4a \geq 3$." Since these variables (a, b, and c) were forced to be binary, the resulting optimization problem became a mixed-integer optimization problem.

Although models using absolute expression values in FBA optimization can be valuable, we expect them to be less powerful for modeling flux distributions in *S. cerevisiae*, since enzymatic genes in yeast are rarely not transcribed at all. Recently, Shlomi et al. [19] introduced an approach using *relative* rather than absolute expression values for inferring regulatory constraints. However, they used these to define *absolute on/off* constraints. If a gene was down-regulated in a particular condition, it was assumed that the corresponding flux in that condition was zero. They imposed this using soft constraints. As a result, some of these regulatory constraints may be violated. They discretized the continuous gene expression data to assess whether genes were down-regulated (-1), baseline (0), or up-regulated (1). However, this approach failed on our compendium data set as the regulatory constraints defined turned out to be too heavily dependent on the cultivation conditions compared. To illustrate this, let us assume that we study three conditions, and that a particular gene is highly expressed in the first, moderately expressed in the second, and lowly expressed in the third condition. If a compendium would only comprise these three conditions, the gene is clearly the least active in condition 3. As a result, condition 1 will be assigned the discrete label $+1$, condition 2 the label 0, and condition 3 the label -1. The model of Shlomi et al. [20] would now predict the flux through the reaction controlled by this gene to be zero in condition 3.

Let us now assume that at a certain moment, expression data for a *fourth* condition become available. This additional experimental data indicate that the expression of our gene is even lower in condition 4 than in condition 3. As a result, condition 4 will be assigned the discrete label -1 and the label of condition 3 will change from -1 to 0. This not only means that the flux in condition 4 is forced to zero, but also that the constraint forcing the flux to zero in condition 3 disappears. This illustrates that the constraints for a particular condition can change drastically as additional experimental data become available.

A method is required that uses *relative* gene expression values between all pairs of conditions to predict *relative* intracellular fluxes. If we expand our compendium with a new condition, then all the regulatory constraints between the old conditions should remain the same. Only additional constraints between the new condition and each of the old ones should be incorporated. Therefore, we suggest to add regulatory *up/down* constraints to FBA, which we will call transcriptional controlled FBA (tFBA). To account for posttranscriptional control and noise in the gene expression data, we formulate these constraints as *soft constraints*, such that several of them may be violated (see Methods). Consequently, our objective will be to minimize the number of violated regulatory *up/down* constraints. Since binary variables are needed to define these constraints (see Section 2), our approach entails a mixed-integer linear optimization problem (MILP) instead of a linear optimization as in (r)FBA. The regulatory constraints are defined *between* all pairs of conditions (as opposed to Shlomi et al. [20] and Covert and Palsson [4]), such that multiple FBA models (one per condition) are solved simultaneously.

We applied our approach to nine different cultivation conditions of this yeast microarray compendium. These microarray experiments enable us to define approximately 5,000 regulatory constraints between pairs of cultivation conditions. We show that changes in gene expression are indeed predictive for changes in flux. Furthermore, we demonstrate that gene expression data point toward the utilization of more pathways than expected based on flux minimization, especially in conditions where carbon is available in excess. In these conditions, tFBA predicts C1-metabolism to be much more active than (r)FBA, which is in agreement with results of previous *dynamic* experiments.

2 Methods

2.1 Flux Variability Analysis (FVA)

In steady state, the rate of production of each intracellular metabolite i should equal the rate of consumption, leading to the following mass-balance constraints [17], [16], [24]:

$$\sum_{j \in R} S_{ij} v_j^c = 0, \quad \forall i \in M. \tag{1}$$

Here, S_{ij} denotes the stoichiometric coefficient corresponding to metabolite i in reaction j. v_j^c represents the flux through reaction j in condition c at steady state. M and R denote the sets of metabolites and reactions, respectively.

Upper and lower limits can be applied to each flux by using the bounds α_j^c and β_j^c in the following inequality:

$$\alpha_j^c \leq v_j^c \leq \beta_j^c. \tag{2}$$

Since it is hard to say anything about the upper limits of reactions, we set β_j^c to 1,000 (mmol.gDW^{-1}.h^{-1}), which is almost 200 times the carbon uptake rate in each of the examined cultivation conditions (see Table 1). For reversible reactions, α_j^c is set to $-\beta_j^c$, whereas for irreversible reactions α_j^c is forced to zero. In a similar way, we enforce the (measured) external fluxes (including the biomass and "ATP for maintenance" flux [7]) to be equal to their measured value, e_j^c:

$$v_j^c = e_j^c. \tag{3}$$

TABLE 1
The Nine Different Cultivation Conditions Considered in This Paper Differing in Aeration Type, Limiting Nutrient, and Carbon Source

Cultivation condition parameters				Calculated (Measured) conversion rates										
				Input fluxes ($\frac{mmol}{gDW*h}$)					Output fluxes ($\frac{mmol}{gDW*h}$)					Growth ($\frac{1}{h}$)
Abbreviation	Aeration	Lim. Nutr.	C-Source	Acetate	Ethanol	Galactose	Glucose	O_2	Acetate	Ethanol	Glycerol	CO_2		Biomass
Ac Ae	Aerobic	Carbon	Acetate	5.5(5.8)	−(−)	−(−)	−(−)	7.4(7.3)	−(−)	−(−)	−(−)	7.5(7.4)		0.1(0.1)
Eth Ae	Aerobic	Carbon	Ethanol	−(−)	3.5(3.7)	−(−)	−(−)	6.8(6.8)	−(−)	−(−)	−(−)	3.4(3.2)		0.1(0.1)
Gal Ae	Aerobic	Carbon	Galactose	−(−)	−(−)	1.2(1.1)	−(−)	3.4(2.8)	−(−)	−(−)	−(−)	3.5(2.9)		0.1(0.1)
Glc Ae	Aerobic	Carbon	Glucose	−(−)	−(−)	−(−)	1.1(1.1)	3.0(2.7)	−(−)	−(−)	−(−)	3.1(2.8)		0.1(0.1)
N-lim Ae	Aerobic	Nitrogen	Glucose	−(−)	−(−)	−(−)	5.5(5.4)	3.6(3.3)	−(−)	8.5(8.3)	0.1(0.1)	12.2(12.4)		0.1(0.1)
P-lim Ae	Aerobic	Phosphorus	Glucose	−(−)	−(−)	−(−)	5.9(6.1)	4.0(3.7)	0.4(0.3)	8.4(8.0)	0.4(0.2)	12.7(13.1)		0.1(0.1)
Glc An	Anaerobic	Carbon	Glucose	−(−)	−(−)	−(−)	6.0(5.5)	−(−)	−(−)	9.7(10.0)	0.8(0.8)	10.1(10.0)		0.1(0.1)
N-lim An	Anaerobic	Nitrogen	Glucose	−(−)	−(−)	−(−)	8.2(8.3)	−(−)	−(−)	14.3(14.0)	0.6(0.3)	14.6(14.8)		0.1(0.1)
P-lim An	Anaerobic	Phosphorus	Glucose	−(−)	−(−)	−(−)	8.7(8.6)	−(−)	−(−)	14.9(14.8)	1.0(0.9)	15.4(15.5)		0.1(0.1)

Columns 5-14 represent the estimated (e_j^c) and measured (m_j^c) conversion rates (measured values in parentheses) for the substrates and products, respectively. The estimated values where calculated by minimizing the mean squared error between the estimated and measured conversion rates (see (26)).

FVA can now be performed to calculate the minimal ($v_j^{c,min}$) and maximal ($v_j^{c,max}$) flux through each reaction j subject to the governing stoichiometric, thermodynamic, and extracellular flux constraints:

$$\begin{aligned} \min \textit{or} \max \quad & v_j^c \\ \text{subject to} \quad & \text{I. Stoichiometric constraints:} \\ & \sum_{j\in R} S_{ij} v_j^c = 0 \qquad \forall i \in M. \\ & \text{II. Thermodynamic constraints:} \\ & \alpha_j^c \le v_j^c \le \beta_j^c \qquad \forall j \in R. \\ & \text{III. Extracellular measurement constraints:} \\ & v_j^c = e_j^c \qquad \forall j \in E. \end{aligned} \tag{4}$$

Here, E denotes the set of measured extracellular fluxes. The flux range for each reaction can then be defined as $v_j^c \in [v_j^{c,min}, v_j^{c,max}]$.

2.2 Flux Minimization (FBA)

FBA is a constraint-based modeling approach employed to locate an optimal flux distribution within the solution space defined by the governing constraints. If flux minimization is used as optimality criterion, the FBA problem can be formulated as:

$$\begin{aligned} z_c = \min \quad & \sum_{j\in R} \left| v_j^c \right| \\ \text{subject to} \quad & \text{I. Stoichiometric constraints:} \\ & \sum_{j\in R} S_{ij} v_j^c = 0 \qquad \forall i \in M. \\ & \text{II. Thermodynamic constraints:} \\ & \alpha_j^c \le v_j^c \le \beta_j^c \qquad \forall j \in R. \\ & \text{III. Extracellular measurement constraints:} \\ & v_j^c = e_j^c \qquad \forall j \in E. \end{aligned} \tag{5}$$

Because several reactions are reversible, reaction rates can be negative. Therefore, we add up the *absolute* values of the fluxes, $\sum |v_j^c|$. This causes the objective in (5) to be nonlinear. To transform it into a linear objective, we represent each reversible reaction, v_j^c, by a forward and a backward reaction, v_{jf}^c and v_{jb}^c:

$$\begin{aligned} v_j^c &= v_{jf}^c - v_{jb}^c, \\ \max\left(0, \alpha_j^c\right) &\le v_{jf}^c \le \max\left(0, \beta_j^c\right), \\ \max\left(0, -\beta_j^c\right) &\le v_{jb}^c \le \max\left(0, -\alpha_j^c\right). \end{aligned} \tag{6}$$

If we constrain only the forward or the backward reaction rate to be nonzero, then the absolute value of $|v_j^c|$ can be written as:

$$|v_j^c| = |v_{jf}^c - v_{jb}^c| = v_{jf}^c + v_{jb}^c. \tag{7}$$

Note that this constraint is automatically fulfilled if we perform flux minimization using $\sum_{j\in R}(v_{jf}^c + v_{jb}^c)$ as objective function, since this objective function will only be minimal if either the forward or the backward reaction rate is zero (or both). This means that (5) becomes:

$$\begin{aligned} z_c = \min \quad & \sum_{j\in R} \left(v_{jf}^c + v_{jb}^c\right) \\ \text{subject to} \quad & \text{I. Stoichiometric constraints} \\ & \text{II. Thermodynamic constraints:} \\ & v_j^c = v_{jf}^c - v_{jb}^c \qquad \forall j \in R, \\ & \max\left(0, \alpha_j^c\right) \le v_{jf}^c \le \max\left(0, \beta_j^c\right) \qquad \forall j \in R, \\ & \max\left(0, -\beta_j^c\right) \le v_{jb}^c \le \max\left(0, -\alpha_j^c\right) \qquad \forall j \in R. \\ & \text{III. Extracellular measurement constraints} \end{aligned} \tag{8}$$

2.3 Regulatory FBA (rFBA, [5])

Covert et al. [5] proposed to add regulatory *on/off* constraints to FBA. These constraints forces fluxes to zero if the genes controlling those fluxes are low expressed:

$$\begin{aligned} w_c = \min \quad & \sum_{j\in R} \left(v_{jf}^c + v_{jb}^c\right) \\ \text{subject to} \quad & \text{I. Stoichiometric constraints} \\ & \text{II. Thermodynamic constraints} \\ & \text{III. Extracellular measurement constraints} \\ & \text{IV. Regulatory } on/off \text{ constraints:} \\ & v_j^c = 0 \qquad \forall j \in A^c, \end{aligned} \tag{9}$$

where A^c denotes the set of reactions for which the corresponding genes are low expressed in condition c ($g_j^c \le T_0$, in which g_j^c represents the expression in

condition c of the gene which encodes the enzyme for catalyzing reaction j). Unfortunately, rFBA models often do not have a solution (even if T_0 is low), since the fluxes through the reactions in set A^c cannot be forced to zero without violating other constraints (i.e., stoichiometric, thermodynamic, and extracellular measurement constraints). Covert et al. [5] did not encounter this problem, since they only fixed the limiting nutrient to its measured value (and then optimized for biomass yield). Therefore, we propose to adjust for this phenomenon and reformulate the regulatory constraints as *soft constraints*. This is accomplished by using a two-step procedure. In the first step, the fluxes through the low-expressed genes are minimized (instead of set to zero) subject to all non-regulatory constraints:

$$\begin{aligned} &\textit{step I}: \\ &w_c = \min \quad \sum_{j \in A^c} \left(v_{jf}^c + v_{jb}^c\right) \\ &\text{subject to} \quad \text{I. Stoichiometric constraints} \\ &\qquad\qquad \text{II. Thermodynamic constraints} \\ &\qquad\qquad \text{III. Extracellular measurement constraints} \end{aligned} \tag{10}$$

As a result, we obtain the minimal allowed flux through the reactions in set A^c denoted by w^c. Note that this may contain solutions in which some of these fluxes are set to zero. In the second step, standard flux minimization is performed over the complete flux set R, but with the additional regulatory constraint: $\sum_{j \in A^c}(v_{jf}^c + v_{jb}^c) = w^c$.

$$\begin{aligned} &\textit{step II}: \\ &\min \quad \sum_{j \in R} \left(v_{jf}^c + v_{jb}^c\right) \\ &\text{subject to} \quad \text{I. Stoichiometric constraints} \\ &\qquad\qquad \text{II. Thermodynamic constraints} \\ &\qquad\qquad \text{III. Extracellular measurement constraints} \\ &\qquad\qquad \text{IV. Regulatory } \textit{on/off} \text{ constraints:} \\ &\qquad\qquad \sum_{j \in A^c} \left(v_{jf}^c + v_{jb}^c\right) = w^c. \end{aligned} \tag{11}$$

If there are multiple solutions fulfilling the regulatory constraint $\sum_{j \in A^c}(v_{jf}^c + v_{jb}^c) = w^c$, then (11) will seek the one with the lowest total flux, $\sum_{j \in R}(v_{jf}^c + v_{jb}^c)$.

2.4 Transcriptional Controlled FBA (tFBA)

2.4.1 Regulatory Up/Down Constraints

In contrast to Covert et al. [5], we propose to add regulatory *up/down* constraints to FBA. Assume that gene g_j, which encodes the enzyme for catalyzing reaction j, increases in expression if we shift from condition k to l ($k, l \in C$, $k \neq l$, where C represents the set of considered cultivation conditions). Then, we impose that the flux through reaction j in condition l should exceed the flux in condition k, resulting in the constraint:

$$|v_j^l| \geq |v_j^k| \quad \text{iff } g_j^l > g_j^k. \tag{12}$$

To cope with gene expression measurement noise, genes are considered as being changed only if the fold change (i.e., $\log_2$ of the relative change in expression) as well as the absolute change in expression exceeds a certain threshold. As a result, (12) is replaced by:

$$\log_2\left(\frac{g_j^l}{g_j^k}\right) \geq T_1 \text{ and } \left(g_j^l - g_j^k\right) \geq T_2. \tag{13}$$

We employ both measures, since fold changes turn out to be more predictive for flux data when the absolute changes are larger and vice versa (see Section 4).

To account for posttranscriptional control, we soften the constraint in (12) by introducing the slack variable $\gamma_j^{k\to l}$ ($\gamma_j^{k\to l} \geq 0$):

$$|v_j^l| + \gamma_j^{k\to l} \geq |v_j^k|. \tag{14}$$

We want $\gamma_j^{k\to l}$ to be zero if $|v_j^l|$ exceeds $|v_j^k|$, and to be $|v_j^k| - |v_j^l|$ if $|v_j^l|$ is smaller than $|v_j^k|$, since then $\gamma_j^{k\to l}$ measures how much flux does not change according to the gene expression data (i.e., how much our assumption is violated). This can be accomplished by introducing the variable $\delta_j^{k\to l}$ ($\delta_j^{k\to l} \geq 0$), such that (14) becomes an equality constraint:

$$\left|v_j^l\right| + \gamma_j^{k\to l} = \left|v_j^k\right| + \delta_j^{k\to l}, \tag{15}$$

and by imposing either $\gamma_j^{k\to l}$ or $\delta_j^{k\to l}$ to be zero:

$$\begin{aligned} \epsilon * \Gamma_j^{k\to l} \leq \gamma_j^{k\to l} \leq \epsilon + \Gamma_j^{k\to l} * \beta_j^k, \\ \epsilon * \Delta_j^{k\to l} \leq \delta_j^{k\to l} \leq \epsilon + \Delta_j^{k\to l} * \beta_j^l, \\ \Gamma_j^{k\to l} + \Delta_j^{k\to l} \leq 1, \\ \Gamma_j^{k\to l} \in \{0,1\}, \\ \Delta_j^{k\to l} \in \{0,1\}. \end{aligned} \tag{16}$$

Here, $\Gamma_j^{k\to l}$ and $\Delta_j^{k\to l}$ are binary variables indicating whether or not $\gamma_j^{k\to l}$ and $\delta_j^{k\to l}$ are nonzero. Note that the third constraint, $\Gamma_j^{k\to l} + \Delta_j^{k\to l} \leq 1$, enforces only $\gamma_j^{k\to l}$ or $\delta_j^{k\to l}$ to be nonzero. ϵ prevents counting small γs and δs ($\gamma, \delta < \epsilon$), which may be caused by round-off errors in the stoichiometric matrix and noise in the gene expression and extracellular flux data.[2] This means that both $\Gamma_j^{k\to l}$ and $\Delta_j^{k\to l}$ are zero, if $|v_j^l| \pm |v_j^k| < \epsilon$. We set ϵ equal to 0.01 in all experiments.

2.4.2 Regulatory Objective Function

Equations (15) and (16) encompass the regulatory *up/down* constraints that we want to add to our FBA framework. Just as in rFBA, the regulatory constraints in tFBA are defined as soft constraints. Therefore, we also adopt a two-step procedure for solving tFBA problems. In the first step, we try to fulfill as many regulatory *up/down* constraints as possible. Since we introduced four variables $(\gamma, \delta, \Gamma, \Delta)$ for measuring how well regulatory *up/down* constraints are fulfilled, our objective function can be formulated as:

$$\max \sum_{k \in C} \sum_{l \in C} \sum_{j \in G^{k\to l}} \left(\alpha_1 \delta_j^{k\to l} - \alpha_2 \gamma_j^{k\to l} + \alpha_3 \Delta_j^{k\to l} - \alpha_4 \Gamma_j^{k\to l}\right). \tag{17}$$

2. Note that the upper bound either restricts $\gamma_j^{k\to l}$ to ϵ (in case $\Gamma_j^{k\to l} = 0$) or is unbounded (equal to β_j^k in case $\Gamma_j^{k\to l} = 1$).

Here, $G^{k\to l}$ denotes the set of reactions that are controlled by genes that *increase* in expression from condition k to l (according to (13)). Note that if a pairwise comparison between conditions I and II is performed, two reaction sets are obtained, $G^{I\to II}$ and $G^{II\to I}$. $G^{k\to l}$ is empty if k equals l.

We can optimize for each of the four regulatory variables. $\sum \delta_j^{k\to l}$ measures how *much* flux changes in concordance with gene expression data, whereas $\sum \gamma_j^{k\to l}$ measures how *much* flux violates this behavior. In a similar way, $\sum \Delta_j^{k\to l}$ measures how *many* fluxes change in concordance with gene expression data, whereas $\sum \Gamma_j^{k\to l}$ measures how *many* fluxes violate this behavior. By using different parameter settings $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_4]$, different aspects of the optimization function can be strengthened.

We use the parameter setting $\boldsymbol{\alpha} = [0, 1, 0, 0]$, as this brings tFBA flux distribution predictions closest to standard FBA predictions. This means that the differences between FBA and tFBA predictions can fully be ascribed to changes in gene expression and not to the choice of the objective function. Accordingly, the following optimization problem is solved in the first step of tFBA:

$$
\begin{aligned}
&\textit{step } I\textit{:}\\
&y = \min \quad \sum_{k\in C}\sum_{l\in C}\sum_{j\in G^{k\to l}} \gamma_j^{k\to l}\\
&\text{subject to}\\
&\quad \text{I. Stoichiometric constraints:}\\
&\quad \sum_{j\in R} S_{ij} v_j^c = 0 \qquad \forall i \in M, \forall c \in C.\\
&\quad \text{II. Thermodynamic constraints:}\\
&\quad \alpha_j^c \le v_j^c \le \beta_j^c \qquad \forall j \in R, \quad \forall c \in C.\\
&\quad \text{III. Extracellular measurement constraints:}\\
&\quad v_j^c = e_j^c \qquad \forall j \in E, \forall c \in C.\\
&\quad \text{IV. Regulatory } \textit{up/down} \text{ constraints:}\\
&\quad |v_j^l| + \gamma_j^{k\to l} = |v_j^k| + \delta_j^{k\to l} \qquad \forall j \in G^{k\to l}, \forall k \in C, \forall l \in C,\\
&\quad \epsilon * \Gamma_j^{k\to l} \le \gamma_j^{k\to l} \le \epsilon + \Gamma_j^{k\to l} * \beta_j^k \qquad \forall j \in G^{k\to l}, \forall k \in C, \forall l \in C,\\
&\quad \epsilon * \Delta_j^{k\to l} \le \delta_j^{k\to l} \le \epsilon + \Delta_j^{k\to l} * \beta_j^l \qquad \forall j \in G^{k\to l}, \forall k \in C, \forall l \in C,\\
&\quad \Gamma_j^{k\to l} + \Delta_j^{k\to l} \le 1 \qquad \forall j \in G^{k\to l}, \forall k \in C, \forall l \in C,\\
&\quad \gamma_j^{k\to l} \ge 0 \qquad \forall j \in G^{k\to l}, \forall k \in C, \forall l \in C,\\
&\quad \delta_j^{k\to l} \ge 0 \qquad \forall j \in G^{k\to l}, \forall k \in C, \forall l \in C,\\
&\quad \Gamma_j^{k\to l} \in \{0,1\} \qquad \forall j \in G^{k\to l}, \forall k \in C, \forall l \in C,\\
&\quad \Delta_j^{k\to l} \in \{0,1\} \qquad \forall j \in G^{k\to l}, \forall k \in C, \forall l \in C.
\end{aligned}
\tag{18}
$$

Note that the regulatory *up/down* constraints are constraints between pairs of conditions. As a result, we solve FBA models for multiple conditions ($\forall c$) *simultaneously*. Let y denote the minimal flux that does not change according to the expression data, obtained by solving (18). In the second step of tFBA, standard flux minimization is employed (for all conditions together), but using the additional linear constraint $\sum_{k\in C}\sum_{l\in C}\sum_{j\in G^{k\to l}} \gamma_j^{k\to l} = y$:

$$
\begin{aligned}
&\textit{step } I\textit{:}\\
&x = \min \quad \sum_{j\in R}\sum_{c\in C} |v_j^c|\\
&\text{subject to}\quad \text{I. Stoichiometric constraints}\\
&\qquad \text{II. Thermodynamic constraints}\\
&\qquad \text{III. Extracellular measurement constraints}\\
&\qquad \text{IV. Regulatory } \textit{up/down} \text{ constraints}\\
&\qquad \text{V. Additional regulatory up/down constraint:}\\
&\qquad \sum_{k\in C}\sum_{l\in C}\sum_{j\in G^{k\to l}} \gamma_j^{k\to l} = y.
\end{aligned}
\tag{19}
$$

2.4.3 Reversible Reactions

If either the forward or the backward reaction rate is zero for each reaction j in each condition c, then (15) reduces to:

$$v_{jf}^l + v_{jb}^l + \gamma_j^{k\to l} = v_{jf}^k + v_{jb}^k + \delta_j^{k\to l}, \tag{20}$$

and the problem in (18) becomes linear. This constraint is automatically fulfilled in FBA and rFBA, but not in tFBA, since the first step of tFBA does not minimize fluxes, but differences between fluxes. We need to impose this, otherwise the regulatory constraints in (20) can be fulfilled by making both v_{jf}^l and v_{jf}^l nonzero (undesirable cycles). Therefore, we enforce it by introducing the binary variables J_{jf}^c and J_{jb}^c. The reaction rates v_{jf}^c and v_{jb}^c can only be nonzero if their corresponding binary variables J_{jf}^c and J_{jb}^c are nonzero. By imposing the additional constraint:

$$J_{jf}^c + J_{jb}^c \le 1 \tag{21}$$

only the forward or the backward reaction rate can be nonzero. As a result, the nonlinear optimization problem in (18) (and (19)) reduces to a MILP:

$$
\begin{aligned}
&\textit{step } I\textit{:}\\
&y = \min \quad \sum_{k\in C}\sum_{l\in C}\sum_{j\in G^{k\to l}} \gamma_j^{k\to l}\\
&\text{subject to}\quad \text{I. Stoichiometric constraints}\\
&\qquad \text{II. Thermodynamic constraints:}\\
&\qquad v_j^c = v_{jf}^c - v_{jb}^c \qquad \forall j \in R, \forall c \in C,\\
&\qquad J_{jf}^c \times \max(0, \alpha_j^c) \le v_{jf}^c \le J_{jf}^c \times \max(0, \beta_j^c) \qquad \forall j \in R, \forall c \in C,\\
&\qquad J_{jb}^c \times \max(0, -\beta_j^c) \le v_{jf}^c \le J_{jb}^c \times \max(0, -\alpha_j^c) \qquad \forall j \in R, \forall c \in C,\\
&\qquad J_{jf}^c + J_{jb}^c \le 1 \qquad \forall j \in R, \forall c \in C,\\
&\qquad J_{jf}^c \in \{0,1\} \qquad \forall j \in R, \forall c \in C,\\
&\qquad J_{jb}^c \in \{0,1\} \qquad \forall j \in R, \forall c \in C.\\
&\qquad \text{III. Extracellular measurement constraints}\\
&\qquad \text{IV. Regulatory } \textit{up/down} \text{ constraints:}\\
&\qquad v_{jf}^l + v_{jb}^l + \gamma_j^{k\to l} = v_{jf}^k + v_{jb}^k + \delta_j^{k\to l}\\
&\qquad \forall j \in G^{k\to l}, \forall k \in C, \forall l \in C,
\end{aligned}
\tag{22}
$$

in which all other regulatory constraints from (18) remain as they are.

2.5 Performance in Terms of Gene Expression Fit

We employed the following performance criterion for comparing the three FBA strategies:

$$P = \frac{\sum \Delta_j^{k\to l}}{\sum \Gamma_j^{k\to l} + \sum \Delta_j^{k\to l}}. \tag{23}$$

The Δs and Γs for FBA are obtained by solving step I of tFBA with the additional constraint that the total amount of flux should be equal to $\sum_{c=1}^{|C|}\sum_{j=1}^{|R|} v_j^c = \sum_{c=1}^{|C|} z^c$ (see (5)). Similarly, the Δs and Γs for rFBA are calculated by incorporating both the constraint $\sum_{c=1}^{|C|}\sum_{j=1}^{|A^c|} v_j^c = \sum_{c=1}^{|C|} w^c$ and the constraint $\sum_{c=1}^{|C|}\sum_{j=1}^{|R|} v_j^c = \sum_{c=1}^{|C|} z^c$ in step I of tFBA (see (10)):

$$\begin{aligned}
&\text{step I:}\\
&\min \quad \sum_{k\in C}\sum_{l\in C}\sum_{j\in G^{k\to l}} \gamma_j^{k\to l}\\
&\text{s.t.} \quad \text{I. Stoichiometric constraint}\\
&\qquad \text{II. Thermodynamic constraints}\\
&\qquad \text{III. Extracellular measurement constraint}\\
&\qquad \text{IV. Regulatory } \textit{up/down} \text{ constraints}\\
&\qquad \text{V. Flux minimization constraint:}\\
&\qquad\quad \sum_{j\in R} v_j^c = z^c \qquad \forall c \in C.\\
&\qquad (\textit{or } \text{VI. Regulatory } \textit{on/off} \text{ constraint:}\\
&\qquad\quad \sum_{j\in A^c} v_j^c = w^c \qquad \forall c \in C).
\end{aligned} \tag{24}$$

This means that if multiple solutions exist for both the rFBA and FBA problem, the ones that best fulfill the regulatory *up/down* constraints are selected. In this way, we always obtain a unique solution. Let $\mathbf{L}_0$ denote this solution. This was verified by employing flux variability analysis, using all optimized criterion values as equality constraints. For example, if our objective was to minimize the total amount of regulatory violation $\sum_{k\in C}\sum_{l\in C}\sum_{j\in G^{k\to l}} \gamma_j^{k\to l}$, and we found an optimal value y, then we added the constraint $\sum_{k\in C}\sum_{l\in C}\sum_{j\in G^{k\to l}} \gamma_j^{k\to l} = y$.

For the FVA, we had to solve a large number of MILP problems, which is highly computationally demanding. We, therefore, used the following strategy to reduce the convergence time. First, we employed FVA including only the stoichiometric and thermodynamic constraints and the constraint on the minimal flux, ignoring the regulatory *up/down* constraints. This means that we only have a few binary variables (for possibly reversible reactions). For example, for the tFBA problem, this implies that we use the constraint $\sum_{j\in R}\sum_{c\in C} |v_j^c| = x$ (see (19)) for minimal flux. By employing this FVA, we could already reveal the direction of many of the reversible reactions. In the second step, we solved the FVA incorporating the constraints on minimal flux as well as the regulatory *up/down* constraints. Since we already found the direction of almost all reversible reactions in the first step, this MILP problem can be solved much faster than the original one, as far fewer binary variables are needed in the optimization. Moreover, this MILP problem is much easier to solve than the MILP problem as described in (22), since we already have a good initial solution ($= \mathbf{L}_0$) and the solution space is much smaller as a result of the minimal flux constraint. As a result, each FVA for one reaction could be solved within a minute. Since we had to solve approximately 5,400 (= number of reactions) of these FVA problems for one type of analysis (tFBA, rFBA, or FBA), performing the entire FVA took approximately a week.

2.6 Functional Enrichment Analysis

We performed a hypergeometric test to assess over-representation of KEGG pathways [9] and MIPS functional categories [13] in the set of reactions N^c through which the flux changes (up or down) if we shift from using FBA to tFBA:

$$\begin{aligned}
\text{up:} \quad & j \in N^c, \text{ if } v_j^{c,tFBA} - v_j^{c,FBA} \geq \epsilon,\\
\text{down:} \quad & j \in N^c, \text{ if } v_j^{c,tFBA} - v_j^{c,FBA} \leq \epsilon.
\end{aligned} \tag{25}$$

We used the functional annotations of the genes to annotate the reactions. This means that if a reaction is catalyzed by multiple genes, then the reaction is assigned to each functional category to which any of its catalyzing genes belong.

3 Experimental Setup

3.1 Cultivation Conditions

We applied our methodology to the metabolic network of the yeast *S. cerevisiae* using a microarray compendium consisting of nine different cultivation conditions (see Table 1). Since the yeast was grown in chemostat cultures, several relevant physical parameters could be accurately controlled, such as nutrient concentrations and growth rate [22], [2]. The specific growth rate was equal to 0.1 h^{-1} for all conditions. Furthermore, an estimate of the ATP needed for growth-associated (69.2 mmol of ATP/gDW) and non-growth-associated (1 mmol of ATP/gDW per h) activities was derived from literature [7], [28], [22]. This implies that besides the measured extracellular fluxes, also the biomass and ATP rate could be specified.

Since the extracellular measurements are noisy, we could not fix the extracellular fluxes to their measured value. This would result in unsolvable linear programs, as it may not be possible to fulfill some of these constraints. Therefore, we first inferred the input and output fluxes, $v_j^c \in E$, from their measured values, m_j^c by minimizing the mean squared error:

$$\begin{aligned}
\min \quad & \sqrt{\sum_{j=1}^{|E|} \left(v_j^c - m_j^c\right)^2}\\
\text{subject to} \quad & \sum_{j=1}^{|R|} S_{ij} v_j^c = 0 \quad i \in M\\
& \alpha_j^c \leq v_j \leq \beta_j^c \quad j \in R.
\end{aligned} \tag{26}$$

Here, m_j^c represents the measured conversion rate for reaction j in cultivation condition c. As a result, we obtained a flux value, e_j^c, for each extracellular flux v_j^c, $j \in E$. Equation (26) entails a constrained linear least squares optimization problem. Table 1 shows that in most cases the calculated conversion rates (e_j^c) do not differ much from the measured conversion rates (m_j^c).

3.2 Metabolic Network Model

We used the compartmentalized stoichiometric model iND750, as presented by Duarte et al. [6]. It is a genome-scale model containing 1,150 reactions. It includes a reaction for biomass production, which is based on research of

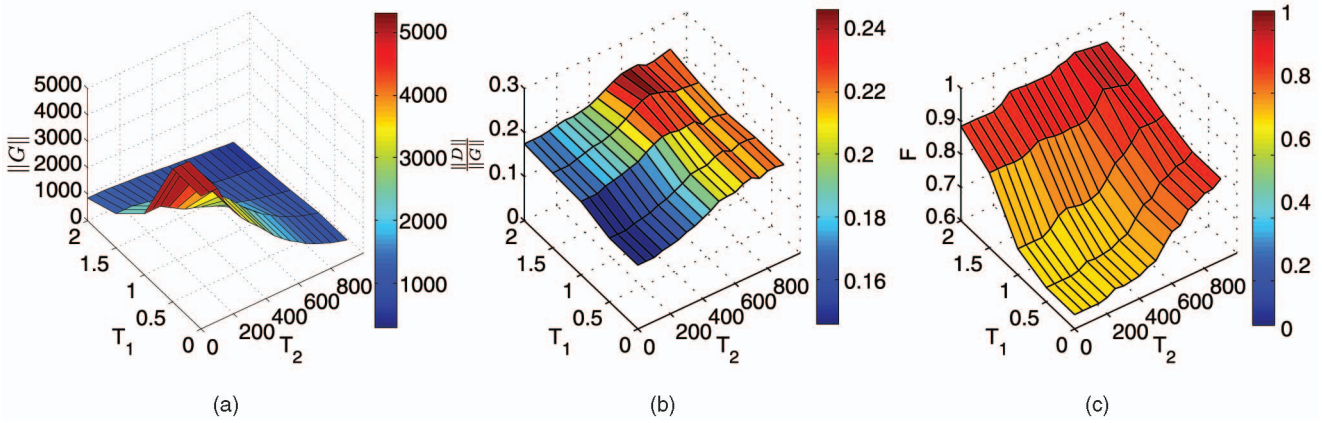


Fig. 1. (a) Number of regulatory *up/down* constraints, $\|G\|$, that can be derived from the gene expression data for the nine different cultivation conditions as function of T_1 and T_2 ($\|G\| = \sum_{k \in C} \sum_{l \in C} \|G^{k \to l}\|$). T_1 was incremented from 0 to 2 in steps of 0.5, whereas T_2 was incremented from 0 to 900 in steps of 50. (b) Fraction of regulatory constraints, $\frac{\|D\|}{\|G\|}$, for which the direction of the flux can be inferred using flux variability analysis ($\|D\| = \sum_{k \in C} \sum_{l \in C} \|D^{k \to l}\|$). (c) Fraction, F, of cases for which the flux indeed changes according to the expression data.

Verduyn et al. [27]. We removed all irrelevant extracellular transport reactions (transport reactions through which the flux is zero in all steady state conditions) and subsequently applied flux variability analysis to remove all dead ends. Dead ends are reactions having reaction rates equal to zero in all possible steady states. The resulting metabolic model contains 485 metabolites and 614 reactions. Seventy nine percent of these reactions have at least one known catalyst (or gene) assigned to it.

3.3 Transcript Levels

Gene expression data were derived from Knijnenburg et al. [10] (GEO Accession number GSE8895 and GSE1723). Transcript levels were measured using Affymetrix GeneChip microarrays. The results for each growth condition were derived from three independently cultured replicates. Before comparison, all arrays were globally scaled to a target value of 150 using the average signal from all gene features using Microarray Suite Version 5.0 [22]. To eliminate insignificant variations, genes with values below 12 were set to 12 as described by Piper et al. [15]. From the 9,335 transcript features on the YG-S98 arrays, a filter was applied to extract 6,383 yeast open reading frames, 6,084 of which were different genes [2]. This discrepancy was due to several genes being represented more than once when suboptimal probe sets were used in the array design.

Some reactions are catalyzed by multiple enzymes. In those cases, we averaged the expression of the different isoenzymes to obtain a unique expression value for each reaction.

4 Results and Discussion

4.1 Changes in Gene Expression are Predictive for Changes in Fluxes

We carried out an in silico experiment to infer whether fluxes change according to expression data. To this end, we first performed all possible pairwise transcriptome comparisons ($\|C\| * (\|C\| - 1)/2 = 36$) to identify the set of reactions for which the controlling genes change in expression from one condition to the other (see Fig. 1a) according to (13). Note that each pairwise comparison yielded two sets of reactions, $G^{I \to II}$ and $G^{II \to I}$, and that the sizes of these sets depended on the values of T_1 and T_2 in (13). Second, the flux range $[v_j^{c,min}, v_j^{c,max}]$, for each reaction j in each cultivation condition c, was calculated using FVA by considering only the governing stoichiometric, thermodynamic, and extracellular flux constraints (see Section 2).

This means that in the calculation of the fluxes, no assumptions about regulatory mechanisms and cellular objectives are made. Third, we determined whether the fluxes through the reactions in $G^{k \to l}$ increased or decreased when shifting from condition k to l. This is not evident, since we can only be certain that a flux increases from k to l if $v_j^{l,min}$ exceeds $v_j^{k,max}$, and only be certain that a flux decreases if $v_j^{k,min}$ exceeds $v_j^{l,max}$. Therefore, we could only derive the direction of the fluxes for a subset of the reactions in $G^{k \to l}$. Let $D^{k \to l}$ denote this subset (see Fig. 1b). Finally, we calculated the fraction of reactions in $D^{k \to l}$ through which the flux indeed increases as the expression increases. This fraction is comparable to the performance measure P used later to validate FBA, rFBA, and tFBA.

Fig. 1a shows that a large number of regulatory *up/down* constraints can be defined. Even for high values of T_1 and T_2 ($T_1 = 2, T_2 = 900$), we still obtain approximately 1,000 regulatory constraints. For approximately 20 percent of these regulatory constraints, the direction of the fluxes can be elucidated by employing FVA (see Fig. 1b). Most of these fluxes indeed change according to gene expression data (~ 80 percent), especially when the change in expression is large (see Fig. 1c).

From these observations, we conclude that it makes sense to add regulatory *up/down* constraints to FBA. After all, changes in expression are predictive for changes in fluxes and the direction of most fluxes (~ 80 percent, see Fig. 1b) cannot be derived by imposing only stoichiometric, thermodynamic, and extracellular flux constraints.

4.2 tFBA Yields a Better Performance P Than rFBA and FBA

We determined the performance measure P (23) for FBA, rFBA, and tFBA in terms of gene expression fit. To this end, we first estimated the flux distributions with each of the methods. All FBA problems were solved using the

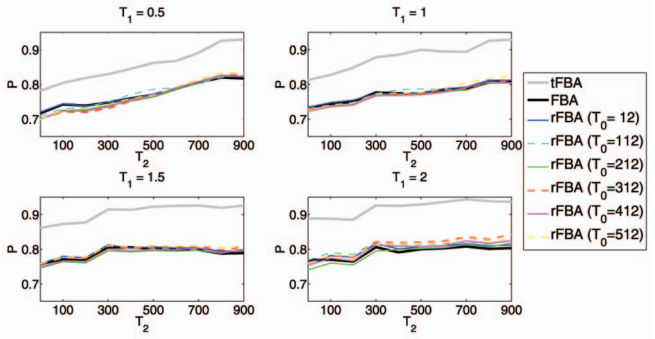


Fig. 2. The performance P for FBA, rFBA, and tFBA as function of the parameters, T_0, T_1, and T_2.

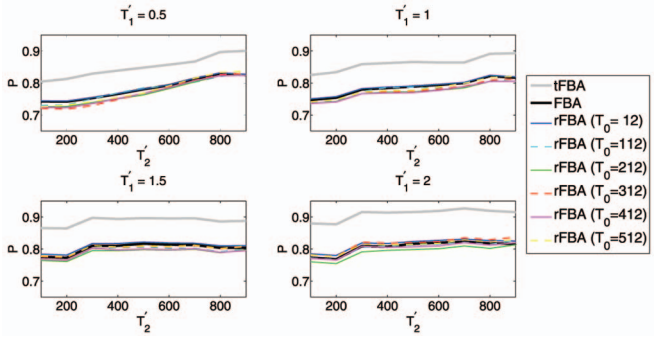


Fig. 3. The performance P for FBA, rFBA, and tFBA as function of the parameters, T_1' and T_2'. T_1 is equal to 0.5 and T_2 is equal to 100. Note that T_1' and T_2' are not used in the optimization itself, but are only used afterwards in calculating the performance.

commercially available package MOSEK[3] (MOSEK [13]). To limit the number of binary variables in the tFBA optimization, we first performed flux variability analysis using only the stoichiometric, thermodynamic, and extracellular flux constraints. In this way, we could infer which reversible reactions become irreversible as a result of the cultivation condition. Irreversible reactions can be represented by only a forward reaction. As a result, no binary variables are needed to impose either the forward or backward reaction rate to be zero in those cases. Subsequently, we solved the MILP problem as defined in (18), however, using minimal flux instead of minimal violation of regulatory constraints as objective function. Hence, the regulatory constraints are actually not active, such that the MILP can be easily solved. This enables us to find an initial solution for the MILP problem. We employ this as a starting point for solving the tFBA problem. This problem contains 26,205 variables (10,411 Boolean variables) and 25,111 equations, and was solved in approximately a week using the MOSEK solver.

Fig. 2 illustrates that tFBA predictions indeed better fulfill the regulatory *up/down* constraints than rFBA and FBA predictions. This is independent of the values of T_1 and T_2. Apparently, expression data strongly point to the use of some longer (or/and additional) pathways, as otherwise the performance of FBA and tFBA would have been the same.

Interestingly, the performance P of rFBA is almost equal to the performance of FBA. Moreover, the performance of rFBA only slightly depends on the value of T_0 and is not really consistent over all values of T_1 and T_2. This indicates that the regulatory *on/off* constraints have only a minor impact.

The performance of tFBA increases as T_1 and T_2 increase. One explanation could be that less regulatory constraints have to be fulfilled if both T_1 and T_2 are high. Another explanation could be that gene expression data are more predictive, when the change in gene expression is large (see Fig. 1c). To reveal whether the second explanation has an impact, we examined the tFBA prediction as found by setting T_1 equal to 0.5 and T_2 equal to 100. For this prediction, we calculated the performance as function of the thresholds T_1' and T_2', where $T_1' \geq T_1$ and $T_2' \geq T_2$. Note that T_1' and T_2' are not used in the optimization, such that all regulatory constraints (based on T_1 and T_2) are equally weighted in the optimization. Still the constraints related to high gene expression differences (high T_1' and T_2'), are more often fulfilled than constraints corresponding to low gene expression differences (low T_1' and T_2') (see Fig. 3). This indicates that gene expression data are more predictive, when the change in expression is large.

Another observation is that the performance of (r)FBA also improves if both T_1' and T_2' increase (see Fig. 3). From this observation, we conclude that expression data sometimes confirm the assumption made by FBA, namely that the shortest path is utilized.

4.3 Cells Do Not Only Optimize for Enzymatic Efficiency

A comparison of the flux distributions obtained with tFBA and (r)FBA ($T_1 = 0.5$ and $T_2 = 100$) reveals that the largest differences between (r)FBA and tFBA predictions are obtained for the aerobic carbon-excess conditions (N-lim Ae and P-lim Ae, see Fig. 4a). Fig. 4b shows that the solution spaces for those conditions are also the largest. Of course, (r)FBA and tFBA predictions will be very similar if the solution space is small. Therefore, most gene-expression-based adjustments can be made in conditions where

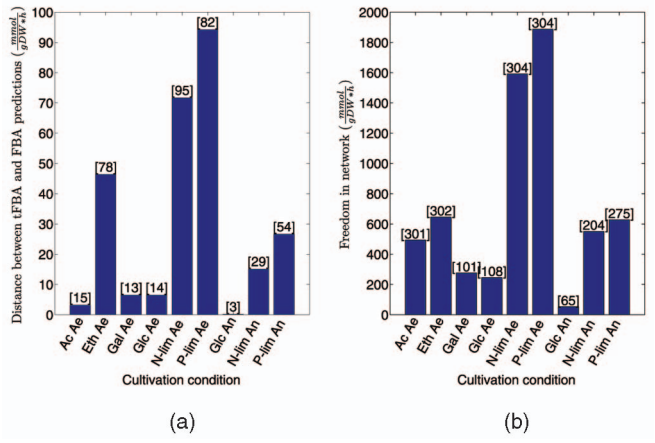


Fig. 4. (a) The difference between the flux distributions of FBA and tFBA for each of the different cultivation conditions, c. The difference between the two flux distributions was calculated as $\sum_{j \in R} |v_j^{c,FBA} - v_j^{c,tFBA}|$. The numbers of reactions for which the flux predictions differed are listed at the top of each bar. (b) The amount of freedom in the metabolic network. The freedom was calculated by adding up the flux ranges of all reactions as obtained with FVA, $\sum_{j \in R} v_j^{c,max} - v_j^{c,min}$. We use this measure as an estimate of the size of the solution space.

3. All optimization problems were solved using MOSEK (version 5.0) in MATLAB (version 7.5) on an Intel Xeon 2.33 GHz (EM64T Quad Core) processor and 16 Gb of RAM under Linux (OpenSuse 10.3 x86-64, kernel 2.6.22).

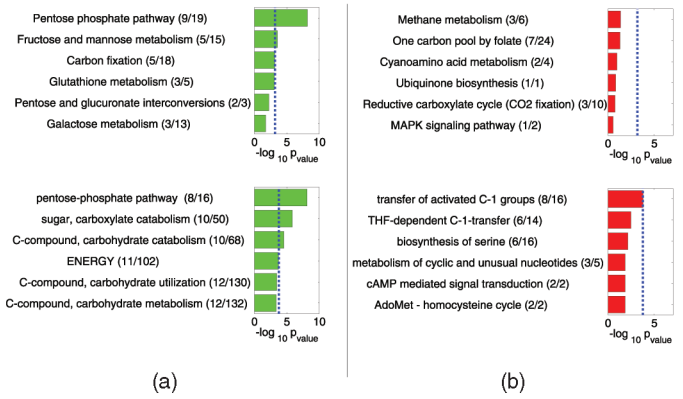


Fig. 5. (a) The most significantly overrepresented pathways and functional categories in the set of reactions through which the flux *decreases* if we shift from (r)FBA to tFBA for aerobic carbon-excess conditions. (b) The same as (a), but for the set of reactions through which the flux *increases* if we shift from FBA to tFBA. The bars represent the p-values ($-\log_{10}$) calculated using the hypergeometric test. The blue dotted line represents the Bonferroni-corrected p-value cutoff if a significance level of 0.05 is employed. Top: KEGG pathways, bottom: MIPS functional categories.

the solution space is large. Since we observe drastic changes in fluxes if we shift from (r)FBA to tFBA, we expect cells not to optimize for minimal flux (only), which might indicate that cells do not only strive for enzymatic efficiency.

4.4 tFBA Yields More Rational Results for the Response of Yeast to Carbon-Excess Conditions

Since we noticed that especially in aerobic carbon-excess conditions the (r)FBA and tFBA predictions are different, we examined these conditions in more detail. We assessed for which pathways the predictions of (r)FBA and tFBA differed according to (25) (see Fig. 5).

Figs. 5a and 6 show that mainly fluxes through the pentose phosphate pathway (PP) are lower if tFBA instead of (r)FBA is applied. The PP pathway is the main route for NADPH formation. Apparently, our model predicts less NADPH formation through the PP pathway in respirofermentative conditions (glucose excess condition). This agrees with results of Gombert et al. [8]. They performed ^{13}C labeling experiments to show that the PP pathway flux (relative to glucose uptake) is much lower in respirofermentative metabolism than in respiratory metabolism (glucose-limited condition). Since the same amount of NADPH for growth is needed, an alternative route for NADPH formation has to be utilized. Our model predicts a higher flux through one-carbon (C1) metabolism (see Fig. 5b). C1 metabolism not only lies at the center of a large number of essential cellular processes, including methyl group biogenesis and the synthesis of nucleotides, vitamins, and some amino acids [15], but it can also be used to generate NADPH, namely through the reaction in which 5,10-methylenetetrahydrofolate is converted into 5,10-methenyltetrahydrofolate (see Fig. 7). Although C1 metabolism has not previously been related to NADPH formation, several studies have reported the activation of the C1 pathway in respirofermentative metabolism. For example, Kresnowati et al. [11] already showed that C1 metabolism plays a crucial role in the *dynamic* response of yeast to a sudden relief from carbon limitation. Transition from a carbon-limited to a carbon-excess condition is accompanied by a sudden depletion of ATP [28]. To compensate for this drop, a fast transcriptional increase in purine, C1, and sulfur metabolism was observed. They concluded that C1 activation could totally be ascribed to the drop in ATP. Our results suggest that part of this

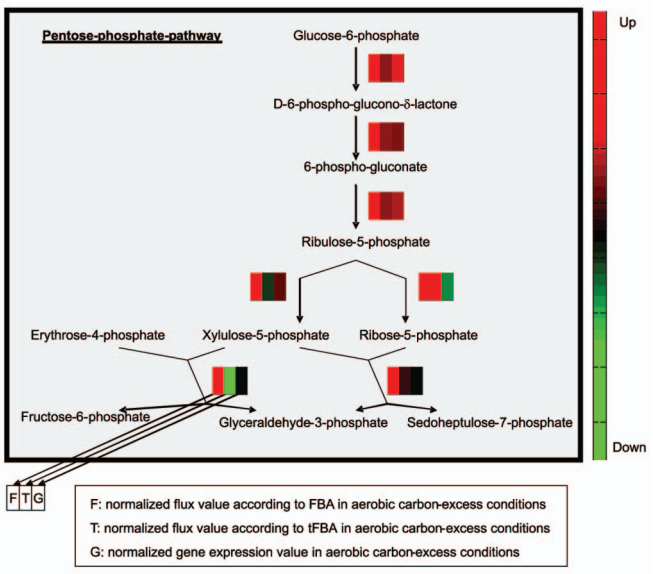


Fig. 6. The normalized flux (expression) values for each of the reactions in the PP pathway. tFBA predicts lower PP pathway fluxes than (r)FBA. Normalization was performed for each reaction separately.

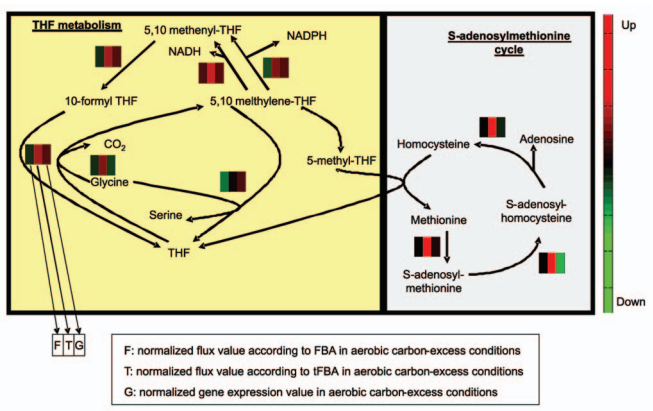


Fig. 7. The normalized flux (expression) values for each of the reactions in C1 metabolism. tFBA predicts higher fluxes through C1-metabolism based on coordinated up-regulation of this pathway. Normalization was performed for each reaction separately.

response is related to NADPH formation and that this dynamic response remains present at high glucose concentrations in steady state. This agrees with results of van den Brink et al. [25], who applied a glucose pulse of larger amplitude ($40\ g.l^{-1}$). Their results indicated that the genes encoding for C1 metabolism remained up-regulated long after the transition phase. Two hours after the pulse (glucose was still in excess) these genes were still exhibiting an increased expression. All in all, these results suggest that C1 metabolism participates actively in adaptation of *S. cerevisiae* to growth under excess glucose condition, and that there could be other reasons than ATP formation for the activation of C1 metabolism.

5 Conclusions

In this study, we examined the benefit of using additional experimental data in flux-balance analysis. We showed that most fluxes change according to gene expression data. Furthermore, we provided a model that directly incorporates regulatory *up/down* constraints in FBA optimization. Our model predictions agreed with results of previous publications related to glucose excess conditions, suggesting that tFBA predictions are more reliable than FBA predictions. Apparently, the assumptions underlying standard FBA may be too simplistic to obtain reliable flux estimates. Gene expression data enable us to locate unlikely FBA flux estimates.

Of course, the performance of the data-driven FBA models (rFBA, tFBA) can be improved by utilizing additional data, such as protein or metabolite data. However, we believe that as long as the kinetics of the metabolite reactions are not fully understood, experimental data can only be used to add soft constraints to the FBA optimization.

References

[1] M. Akesson, J. Förster, and J. Nielsen, "Integration of Gene Expression Data into Genome-Scale Metabolic Models," *Metabolic Eng.*, vol. 6, no. 4, pp. 285-293, 2004.

[2] V.M. Boer, J.H. de Winde, J.T. Pronk, and M.D.W. Piper, "The Genome-Wide Transcriptional Responses of *Saccharomyces cerevisiae* Grown on Glucose in Aerobic Chemostat Cultures Limited for Carbon, Nitrogen, Phosphorus, or Sulfur," *J. Biological Chemistry*, vol. 278, no. 5, pp. 3265-3274, 2003.

[3] M.W. Covert, E.M. Knight, J.L. Reed, M.J. Herrgard, and B.O. Palsson, "Integrating High-Throughput and Computational Data Elucidates Bacterial Networks," *Nature*, vol. 429, no. 6987, pp. 92-96, 2004.

[4] M.W. Covert and B.O. Palsson, "Constraints-Based Models: Regulation of Gene Expression Reduces the Steady-State Solution Space," *J. Theoretical Biology*, vol. 221, no. 3, pp. 309-325, 2003.

[5] M.W. Covert, C.H. Schilling, and B. Palsson, "Regulation of Gene Expression in Flux Balance Models of Metabolism," *J. Theoretical Biology*, vol. 213, no. 1, pp. 73-88, 2001.

[6] N.C. Duarte, M.J. Herrgård, and B. Palsson, "Reconstruction and Validation of *Saccharomyces cerevisiae* ind750, a Fully Compartmentalized Genome-Scale Metabolic Model," *Genome Research*, vol. 14, no. 7, pp. 1298-1309, 2004.

[7] I. Famili, J. Forster, J. Nielsen, and B.O. Palsson, "*Saccharomyces cerevisiae* Phenotypes Can Be Predicted by Using Constraint-Based Analysis of a Genome-Scale Reconstructed Metabolic Network," *Proc. Nat'l Academy of Sciences USA*, vol. 100, no. 23, pp. 13134-13139, 2003.

[8] A.K. Gombert, M.M. dos Santos, B. Christensen, and J. Nielsen, "Network Identification and Flux Quantification in the Central Metabolism of *Saccharomyces cerevisiae* under Different Conditions of Glucose Repression," *J. Bacteriology*, vol. 183, no. 4, pp. 1441-1451, 2001.

[9] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamanishi, "KEGG for Linking Genomes to Life and the Environment," *Nucleic Acids Research*, vol. 36, Database Issue, pp. D480-D484, 2008.

[10] T.A. Knijnenburg, J.M.G. Daran, M.A. van den Broek, P.A. Daran-Lapujade, J.H. de Winde, J.T. Pronk, M.J.T. Reinders, and L.F.A. Wessels, "Combinatorial Effects of Environmental Parameters on Transcriptional Regulation in *Saccharomyces cerevisiae*: A Quantitative Analysis of a Compendium of Chemostat-Based Transcriptome Data," *BMC Genomics*, vol. 10, no. 53, 2009.

[11] M.T.A.P. Kresnowati, W.A. van Winden, M.J.H. Almering, A. ten Pierick, C. Ras, T.A. Knijnenburg, P. Daran-Lapujade, J.T. Pronk, J.J. Heijnen, and J.M. Daran, "When Transcriptome Meets Metabolome: Fast Cellular Responses of Yeast to Sudden Relief of Glucose Limitation," *Molecular Systems Biology*, vol. 2, no. 49, 2006.

[12] H.W. Mewes, S. Dietmann, D. Frishman, R. Gregory, G. Mannhaupt, K.F.X. Mayer, M. Münsterkötter, A. Ruepp, M. Spannagl, V. Stümpflen, and T. Rattei, "Mips: Analysis and Annotation of Genome Information in 2007," *Nucleic Acids Research*, vol. 36, Database Issue, pp. D196-D201, 2008.

[13] MOSEK, http://www.mosek.com/, 2010.

[14] M.D.W. Piper, P. Daran-Lapujade, C. Bro, B. Regenberg, S. Knudsen, J. Nielsen, and J.T. Pronk, "Reproducibility of Oligonucleotide Microarray Transcriptome Analyses. An Interlaboratory Comparison Using Chemostat Cultures of *Saccharomyces cerevisiae*," *J. Biological Chemistry*, vol. 277, no. 40, pp. 37001-37008, 2002.

[15] M.D.W. Piper, S.P. Hong, T. Eissing, P. Sealey, and I.W. Dawes, "Regulation of the Yeast Glycine Cleavage Genes is Responsive to the Availability of Multiple Nutrients," *FEMS Yeast Research*, vol. 2, no. 1, pp. 59-71, 2002.

[16] N.D. Price, J.L. Reed, and B. Palsson, "Genome-Scale Models of Microbial Cells: Evaluating the Consequences of Constraints," *Nature Rev. Microbiology*, vol. 2, no. 11, pp. 886-897, 2004.

[17] O. Rokhlenko, T. Shlomi, R. Sharan, E. Ruppin, and R.Y. Pinter, "Constraint-Based Functional Similarity of Metabolic Genes: Going Beyond Network Topology," *Bioinformatics*, vol. 23, no. 16, pp. 2139-2146, 2007.

[18] R. Schuetz, L. Kuepfer, and U. Sauer, "Systematic Evaluation of Objective Functions for Predicting Intracellular Fluxes in *Escherichia coli*," *Molecular Systems Biology*, vol. 3, no. 119, 2007.

[19] T. Shlomi, M.N. Cabili, M.J. Herrgård, B. Palsson, and E. Ruppin, "Network-Based Prediction of Human Tissue-Specific Metabolism," *Nature Biotechnology*, vol. 26, no. 9, pp. 1003-1010, 2008.

[20] T. Shlomi, Y. Eisenberg, R. Sharan, and E. Ruppin, "A Genome-Scale Computational Study of the Interplay between Transcriptional Regulation and Metabolism," *Molecular Systems Biology*, vol. 3, no. 101, 2007.

[21] A. Stouthamer, "The Search for Correlation between Theoretical and Experimental Growth Yields," *Int'l Rev. Biochemistry and Microbial Biochemistry*, vol. 21, pp. 1-47, 1979.

[22] S.L. Tai, V.M. Boer, P. Daran-Lapujade, M.C. Walsh, J.H. de Winde, J.M. Daran, and J.T. Pronk, "Two-Dimensional Transcriptome Analysis in Chemostat Cultures. Combinatorial Effects of Oxygen Availability and Macronutrient Limitation in *Saccharomyces cerevisiae*," *J. Biological Chemistry*, vol. 280, no. 1, pp. 437-447, 2005.

[23] B. Teusink and E.J. Smid, "Modelling Strategies for the Industrial Exploitation of Lactic Acid Bacteria," *Nature Rev. Microbiology*, vol. 4, no. 1, pp. 46-56, 2006.

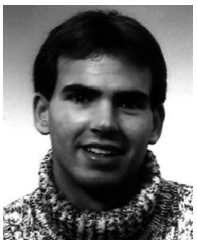
[24] R. Urbanczik and C. Wagner, "An Improved Algorithm for Stoichiometric Network Analysis: Theory and Applications," *Bioinformatics*, vol. 21, no. 7, pp. 1203-1210, 2005.

[25] J. van den Brink, P. Daran-Lapujade, J. Pronk, and J. de Winde, "New Insights into the *Saccharomyces cerevisiae* Fermentation Switch: Dynamic Transcriptional Response to Anaerobicity and Glucose-Excess," *BMC Genomics*, vol. 9, no. 1, p. 100, 2008.

[26] A. Varma and B.O. Palsson, "Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type *Escherichia coli* W3110," *Applied and Environmental Microbiology*, vol. 60, no. 10, pp. 3724-3731, 1994.

[27] C. Verduyn, E. Postma, W.A. Scheffers, and J.P. van Dijken, "Physiology of *Saccharomyces cerevisiae* in Anaerobic Glucose-Limited Chemostat Cultures," *J. General Microbiology*, vol. 136, no. 3, pp. 395-403, 1990.

[28] D. Visser, G.A. van Zuylen, J.C. van Dam, M.R. Eman, A. Pröll, C. Ras, L. Wu, W.M. van Gulik, and J.J. Heijnen, "Analysis of In Vivo Kinetics of Glycolysis in Aerobic *Saccharomyces cerevisiae* by Application of Glucose and Ethanol Pulses," *Biotechnology and Bioeng.*, vol. 88, no. 2, pp. 157-167, 2004.
[29] W. Wiechert, "An Introduction to 13c Metabolic Flux Analysis," *Genetic Eng. (NY)*, vol. 24, pp. 215-238, 2002.
[30] W. Wiechert, "Modeling and Simulation: Tools for Metabolic Engineering," *J. Biotechnology*, vol. 94, no. 1, pp. 37-63, 2002.



Rogier J.P. van Berlo received the MSc degree in electrical engineering from Delft University of Technology, the Netherlands, in 2003. In 2004, he started the PhD research in the Department of Electrical Engineering, Mathematics and Computer Science at the same university, on knowledge extraction from high-throughput genomic data through integrative analyses.

Dick de Ridder received the MSc degree in computer science in 1996 and the PhD degree in applied physics in 2001, from Delft University of Technology, The Netherlands. In 2005, he became an assistant professor in bioinformatics at the Faculty of Electrical Engineering, Mathematics and Computer Science of Delft University of Technology. In the past, he has worked on pattern recognition, neural networks, and image processing. Currently, his interest is in integrating various high-throughput measurements and prior knowledge to model the living cell. Specifically, he studies the problem how to integrate data obtained from various measurement devices and databases with wildly varying levels of coverage, reliability, noise, etc.

Jean-Marc Daran received the MD and PhD degree from the National Institute of Applied Science in Toulouse (France) in 1992 and 1996, respectively. After an 18-month postdoc in the UK, he joined the research and development of DSM Food Specialties. He worked on various research sites as screening scientist and molecular geneticist. After nearly seven years, he moved back to the academic world by joining the Industrial Microbiology Section (headed by Professor J.T. Pronk) of the Department of Biotechnology at Delft University of Technology as an assistant professor in 2004. His main scientific interests have been genomics of industrial organisms (*Saccharomyces cerevisiae* and *Penicillium chrysogenum*), molecular engineering of industrial microorganisms, and metabolism of flavor compounds in *S. cerevisiae*. Since he joined Delft University of Technology in 2004, he has been the author of more than 20 publications in peer-reviewed journals.

Pascale A.S. Daran-Lapujade received the MD degree in agricultural engineering from the Ecole Superieure dAgriculture de Purpan (ESAP, France) in 1995, the Diplome detudes approfondies in molecular and cellular biology—biotechnology from the Institut National des Sciences Appliquees (INSA, France) in 1996, and the PhD degree in microbial biotechnology from the Institut National Polytechnique de Lorraine (INPL, France) in 2000. She was granted a European personal Marie Curie fellowship to join the Industrial Microbiology Section (headed by Prof. J.T. Pronk) of the Department of Biotechnology at Delft University of Technology. After 5 years as postdoctoral fellow, she became an assistant professor in the same group where her main scientific interest is experimental systems biology, with a special focus on bakers yeast *S. cerevisiae* and its central carbon metabolism.

Bas Teusink received the PhD degree from the University of Amsterdam, The Netherlands, in 1999. From 1999 to 2002, he performed a postdoc in the Gaubius Laboratorium TNO Leiden. In 2002, he moved to NIZO food research in Ede, The Netherlands, as a (senior) scientist and worked on several TI Food and Nutrition projects. In 2008, he became a full professor in systems biology, in particular integrative bionformatics, at the VU University Amsterdam, The Netherlands. His research interests encompasses microbial systems biology, in particular modeling of metabolic networks and design principles.

Marcel J.T. Reinders received the MSc degree in applied physics and the PhD degree in electrical engineering from Delft University of Technology, The Netherlands, in 1990 and 1995, respectively. In 2004, he became a professor in bioinformatics within the Mediamatics Department of the Faculty of Electrical Engineering, Mathematics and Computer Science at Delft University of Technology. His background is within pattern recognition. Besides studying fundamental issues, he applies pattern recognition techniques to the areas of bioinformatics, computer vision, and context-aware recommender systems. His special interest goes toward understanding complex systems (such as biological systems) that are severely undersampled.