Available online at www.sciencedirect.com
SCIENCE DIRECT®
Metabolic Engineering 8 (2006) 1–13
METABOLIC ENGINEERING
www.elsevier.com/locate/ymben

An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems

Priti Pharkya, Costas D. Maranas*

Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Received 20 April 2005; received in revised form 5 August 2005; accepted 9 August 2005
Available online 30 September 2005

Abstract

We introduce a computational framework termed OptReg that determines the optimal reaction activations/inhibitions and eliminations for targeted biochemical production. A reaction is deemed up- or downregulated if it is constrained to assume flux values significantly above or below its steady-state before the genetic manipulations. The developed framework is demonstrated by studying the overproduction of ethanol in *Escherichia coli*. Computational results reveal the existence of synergism between reaction deletions and modulations implying that the simultaneous application of both types of genetic manipulations yields the most promising results. For example, the downregulation of phosphoglucomutase in conjunction with the deletion of oxygen uptake and pyruvate formate lyase yields 99.8% of the maximum theoretical ethanol yield. Conceptually, the proposed strategies redirect both the carbon flux as well as the cofactors to enhance ethanol production in the network. The OptReg framework is a versatile tool for strain design which allows for a broad array of genetic manipulations.

Keywords: Metabolic engineering; Strain optimization; Metabolic pathway modeling

1. Introduction

Metabolic engineering in microbial hosts for the production of renewable fuels and chemicals has received considerable attention in recent years (Stafford and Stephanopoulos, 2001; Gross et al., 2003; Vera et al., 2003; Wyman, 2003; Alper et al., 2005a). This is because biotechnology offers an opportunity for unparalleled product diversity and is integral to achieving the goal of sustainable development. Furthermore, the potential of biocatalysts to produce very complex products of desired stereospecificity (Breuer et al., 2004) with possibly more favorable economics has motivated many large-scale efforts in engineering microbial production systems. Already, a number of compounds are being produced industrially using microbial production systems (Chotani et al., 2000; Nakamura and Whited, 2003), and many efforts are ongoing for synthesizing several others using biological routes (Lee and Schmidt–Dannert, 2002; Martin et al., 2003; Baez–Viveros et al., 2004).

In recent years, our group has introduced optimization-based frameworks to predict genetic modifications (i.e., deletions and additions), aimed at maximizing the secretion of biochemicals from metabolic networks. The objective was to guide experimental metabolic engineering strategies by adopting a systems approach for anticipating the effect of genetic modifications on metabolism. Metabolism is described by adopting genome-scale metabolic models widely available for many organisms (Schilling et al., 2002; Van Dien and Lidstrom, 2002; Forster et al., 2003; Reed et al., 2003; Duarte et al., 2004). Specifically, the bilevel computational framework called OptKnock (Burgard et al., 2003; Pharkya et al., 2003) was developed to suggest reaction deletion strategies that maximize biochemical production. This is accomplished by using maximization of biomass yield (Varma and Palsson, 1994), minimization of metabolic adjustment (MOMA)

*Corresponding author. Fax: +1 814 865 7846.
E-mail address: costas@psu.edu (C.D. Maranas).

1096-7176/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymben.2005.08.003

(Segre et al., 2002; Alper et al., 2005b) or any other plausible cellular objective to estimate the redirected fluxes in the face of the imposed knock-outs. In the case of biomass maximization, this leads to, under certain situations, a highly desirable coupling of biomass formation to targeted biochemical production enabling the use of adaptive evolution for strain design (Ibarra et al., 2002; Fong and Palsson, 2004). Notably, the knock-out strategies predicted by OptKnock were recently demonstrated experimentally by constructing strains for overproducing lactic acid (Fong et al., 2005). The experimental observations for lactic acid yields and biomass growth rates were quite close to the model predictions, thus demonstrating the applicability of the proposed framework. Recently, we proposed an integrated framework called OptStrain which extends OptKnock by pinpointing minimal reaction set recombination tasks to confer a desired non-native biochemical production capability on a microbial host (Pharkya et al., 2004).

However, gene additions and complete eliminations do not describe the entire range of genetic manipulation strategies available. The importance of tuning upward or downward gene expression and consequently enzyme levels and corresponding flux rates is widely being recognized in metabolic engineering community (Jensen and Hammer, 1998a; Koffas et al., 2003). A recent successful effort for producing 1,3 propanediol from *Escherichia coli* employed the downregulation of glyceraldehyde-3-phosphate dehydrogenase as a key genetic modification (Nakamura and Whited, 2003). Also, in many cases, a gene deletion is lethal whereas its downregulation is not. For example, in an earlier paper (Pharkya et al., 2003) we predicted a strategy involving elimination of enolase to enhance the theoretical production of serine in the metabolic network of *E. coli*. However, the deletion of enolase is known to be lethal in *E. coli* due to regulatory interactions not accounted for in the considered metabolic reconstruction. Consequently, the repression of enolase rather than its deletion appears to be a more appropriate strategy. Gene up- or downregulation can be tuned by using widely available promoter libraries (de Ruyter et al., 1996; Jensen and Hammer, 1998b) providing experimental strategies to implement predictions on desired upward or downward flux changes.

In this paper, we describe the modeling and algorithmic changes required to extend OptKnock (Burgard et al., 2003) to allow for up- and/or downregulation in addition to gene knock-outs to meet a bioproduction goal. Specifically, the objective here is to computationally identify which reactions should be modulated, (i.e., repressed or activated) or knocked-out such that the biochemical of interest is overproduced. This extended computational framework termed OptReg uses the OptKnock formulation as a starting point. However, the breadth and complexity of the newly considered genetic manipulations introduce many new variables and nonlinearities requiring a new and non-trivial theoretical treatment for the generation of the single-level optimization problem. This treatment is described in detail in Section 2. The computational difficulties arising due to hundreds of binary variables and bilinear products of binary and continuous variables make this a challenging problem to solve.

Conceptually, in the OptReg framework, reaction fluxes are referred to as repressed or activated when their fluxes are forced to be sufficiently higher or lower with respect to their corresponding steady-state fluxes. Parameter C, termed the regulation strength parameter, quantifies the threshold that needs to be overcome before a reaction is considered up- or downregulated (see Fig. 1). It can be assigned values between zero and one. At C equal to zero, even if the reaction flux is equal to its steady-state value, the reaction is considered to be modulated. On the other hand, for C equal to one, a reaction flux must be equal to its upper or lower stoichiometric bound $v_j^{\max}$ or $v_j^{\min}$ before it is deemed as up- or downregulated, respectively. It follows that the higher the value of C, stronger is the requirement imposed on a reaction when it is regulated (see Section 2 for further details).

Fig. 1 graphically illustrates the imposed bounds as a consequence of activation, inhibition or elimination of a reaction. The figure also shows that not a single value but rather a range of values between $v_{j,L}^0$ and $v_{j,U}^0$ sometimes needs to be used to describe the original steady-state. This is because as our group (Burgard et al., 2001) and others (Papin et al., 2002) have found, the maximization of biomass or any other cellular objective does not yield a unique solution (i.e., value) for a majority of reactions (especially for internal ones) at steady-state. Instead, due to the high redundancy in the network, a range of flux values is typically identified corresponding to alternate but equivalent optima for biomass. Therefore, in OptReg we have to use a range of flux values rather than a single value

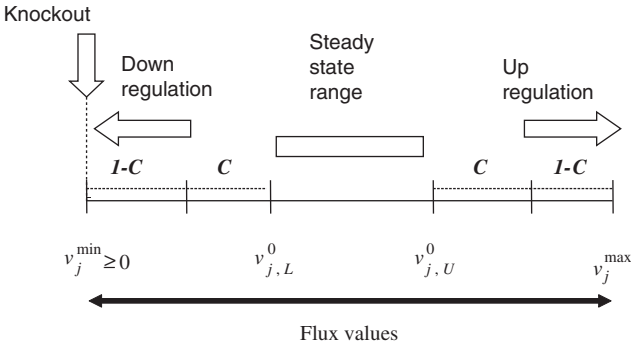

Fig. 1. A pictorial overview of the definitions of up/downregulations and deletions. A reaction is called "upregulated" if it assumes flux values in the range $[(v_{j,U}^0)(1-C)+(v_j^{\max})(C), v_j^{\max}]$; "downregulated" if it assumes flux values between $v_j^{\min}$ and $(v_j,_L^0)(1-C)+(v_j^{\min})(C)$ where $0 \leqslant C \leqslant 1$. If a reaction is knocked out, it is forced to assume a flux of zero. $v_j^{\min}$ and $v_j^{\max}$ represent the stoichiometric bounds on the flux through reaction j whereas $[v_{j,L}{}^0, v_{j,U}{}^0]$ is the range of allowable steady-state values. Note that the stoichiometric lower bound $v_j^{\min}$ for a flux j may be greater than zero if the specific reaction is required for biomass formation. In such a scenario, the reaction cannot be knocked out.

to describe the base state of the network before any genetic manipulations are implemented. The following section highlights these modeling and algorithmic developments.

2. Methods

2.1. Steady-state flux determination

In OptReg, up or down flux modulations are modeled as upward or downward departures, respectively, from the wild-type steady-state values. Therefore, before employing the model it is necessary to establish estimates for the steady-state flux values (or range of values) for all reactions in the wild-type network of *E. coli* (Reed et al., 2003). To this end, we use flux measurements (Fischer et al., 2004) to fix some of the reaction fluxes in central metabolism at values determined from comprehensive isotopomer balancing experiments performed on exponentially growing *E. coli* cells in a bioreactor culture. These experimental results (shown in Fig. 2) provide us with carbon flux partitioning values at key branch points in the central metabolism and an estimate for the biomass formation rate at steady-state ($0.81\,h^{-1}$ as predicted in experimental studies).

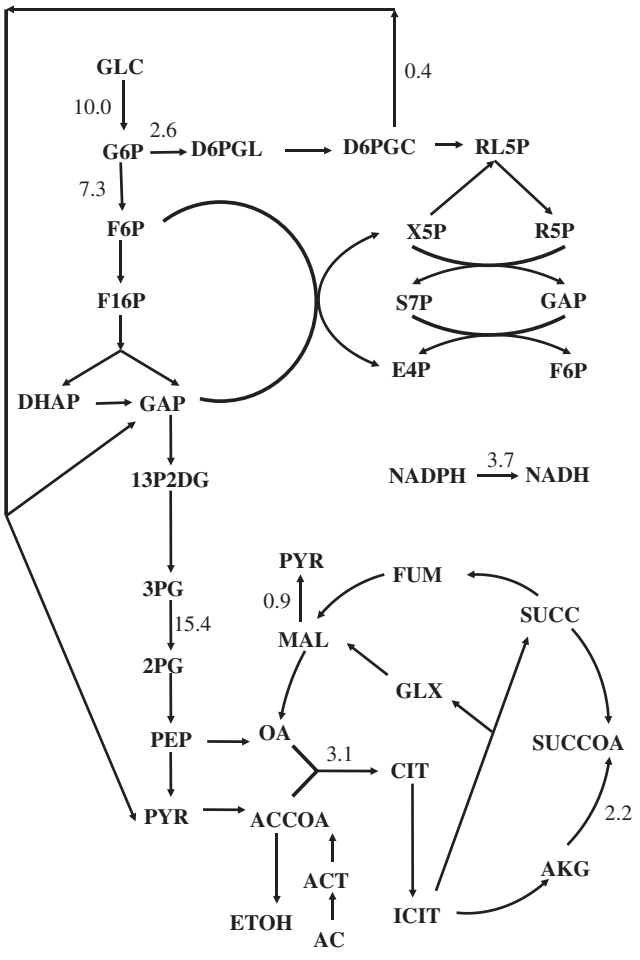

Fig. 2. The flux values (in mmol/gDW · h) at steady-state fixed at experimental values extracted from Fischer et al. (2004).

The steady-state flux values or ranges for the remaining fluxes are estimated computationally. A linear programming formulation, referred as the min/max problem (Burgard et al., 2001; Mahadevan and Schilling, 2003) is solved for base flux estimation for reactions in the most recent genome-scale model of *E. coli* (Reed et al., 2003). Specifically, each flux is successively maximized and then minimized subject to predetermined experimental values of a few fluxes.

$$\text{minimize/maximize} \quad v_j \quad \forall j \in \mathtt{M}$$

$$\text{subject to} \quad \sum_j S_{ij} v_j = 0, \quad \forall i \in \mathtt{N}$$

$$v_{\text{atp}} \geqslant v_{\text{atp\_maint}},\ v_{\text{glc}} = 10\,\text{mmol/gDW} \cdot \text{h},$$

$$v_j = v_j^{\text{exp}}, \quad \forall j \in \mathtt{M}_{\text{exp}},$$

$$v_j \geqslant 0, \quad \forall j \in \mathtt{M}$$

Here, $\mathtt{M} = (1, \ldots, M)$ denotes the set of reactions and $\mathtt{N} = (1, \ldots, N)$ is the set of metabolites. S_{ij} is the stoichiometric coefficient of metabolite i in reaction j. The first constraint imposes a stoichiometric balance on the network. The glucose uptake rate v_{glc} is fixed at $10\,\text{mmol/gDW} \cdot \text{h}$ and a minimum amount of ATP formation ($v_{\text{atp\_maint}} = 7.6\,\text{mmol/gDW} \cdot \text{h}$) is imposed for maintenance. The subset $\mathtt{M}_{\text{exp}}$ is comprised of the reactions whose fluxes are fixed at experimental values. In addition, all reversible reactions are split into their forward and backward counterparts to facilitate the modeling of regulation as described in the next section, increasing the size of the metabolic network to 1,470 one-directional positive-valued reactions. The identified minimum and maximum values for each flux through reaction j are denoted as $v_{j,L}^0$ and $v_{j,U}^0$, respectively. If $v_{j,L}^0$, is equal to $v_{j,U}^0$, then a unique value for the base steady-state value of reaction j is obtained. Otherwise, the range of values between $v_{j,L}^0$, and $v_{j,U}^0$ quantifies the ambiguity in assigning a flux value to reaction j.

2.2. Modeling of genetic manipulations

Three sets of binary variables for each reaction $j \in \mathtt{M}$ are introduced to model all possible combinations of knockouts, up- and downregulations.

$$y_j^k = \begin{cases} 0 & \text{if reaction } j \text{ is knocked out,} \\ 1 & \text{if reaction } j \text{ is not knocked out,} \end{cases}$$

$$y_j^d = \begin{cases} 0 & \text{if reaction } j \text{ is downregulated,} \\ 1 & \text{if reaction } j \text{ is not downregulated,} \end{cases}$$

$$y_j^u = \begin{cases} 0 & \text{if reaction } j \text{ is upregulated,} \\ 1 & \text{if reaction } j \text{ is not upregulated.} \end{cases}$$

These binary variables then act as switches to ensure that fluxes are appropriately restricted in response to a deletion or an up/downregulation based on the following

constraints:

$$v_j^{\min} \leqslant v_j \leqslant [(v_{j,L}^0)\cdot(1-C)+(v_j^{\min})\cdot(C)]\cdot(1-y_j^d)+v_j^{\max}\cdot y_j^d,$$
$\forall j \in \mathtt{M}$ (Downregulations),

$$[(v_{j,U}^0)\cdot(1-C)+(v_j^{\max})\cdot(C)]\cdot(1-y_j^u)+v_j^{\min}\cdot y_j^u \leqslant v_j \leqslant v_j^{\max},$$
$\forall j \in \mathtt{M}$ (Up regulations),

$$v_j^{\min}\cdot y_j^k \leqslant v_j \leqslant v_j^{\max}\cdot y_j^k,$$
$\forall j \in \mathtt{M}$ (Knockouts).

Here, $v_j^{\min}$ and $v_j^{\max}$ are the stoichiometric bounds on the fluxes determined by minimizing and maximizing each flux subject to (i) the stoichiometric network balances, (ii) a fixed glucose uptake rate, (iii) the fulfilment of the ATP maintenance requirement and (iv) the formation of at least 1% of the maximum theoretical biomass in the network. Note that the stoichiometric lower bound $v_j^{\min}$ for a flux j may be greater than zero if the specific reaction is required for biomass formation.

As described earlier, the regulation strength parameter C is assigned values between zero and one. This parameter determines the fraction of the range of flux values between the stoichiometric bounds (lower or upper) and the corresponding lower or upper steady-state flux values that are available to a regulated reaction (see Fig. 1). We require that when a reaction is inhibited, the flux should vary between the stoichiometric lower bound $v_j^{\min}$ and the point denoted by $(v_{j,L}^0)(1-C)+(v_j^{\min})(C)$ (see Fig. 1). Similarly, the reaction flux should be greater than $(v_{j,U}^0)(1-C)+(v_j^{\max})(C)$ when it is upregulated. The use of parameter C for the regulated fluxes is employed to ensure a significant deviation of the fluxes from their steady-state values. Obviously, the higher the value of C, the higher will be the departure from the steady-state values and consequently, the "stronger" the regulation. Note that an appropriate estimate of the value of C can be made beforehand depending upon the strength of the promoter and the inhibitor (Jensen and Hammer, 1998b). Finally, if a reaction is knocked out, then the corresponding flux is forced to zero. If a reaction is required for biomass formation it has a non-zero lower bound $v_j^{\min}$ and as a consequence, it cannot be knocked out.

A reaction flux can be the target of, at the most, a single type of genetic manipulation, thus:

$$(1-y_j^k)+(1-y_j^d)+(1-y_j^u)\leqslant 1, \quad \forall j \in \mathtt{M}.$$

Additionally, constraints $\sum_j(1-y_j^k)\leqslant K$, $\sum_j(1-y_j^d)\leqslant D$, and $\sum_j(1-y_j^u)\leqslant U$ specify that the total number of reactions that can be deleted, downregulated and upregulated are K, D and U respectively. Alternatively, a limit on the total number of regulated and deleted reactions is imposed as follows:

$$\sum\nolimits_j(1-y_j^k)+(1-y_j^d)+\sum\nolimits_j(1-y_j^u)\leqslant L,$$

where L is the total number of reactions that can be modulated or knocked out. Recall that all reactions are one-directional and their fluxes are constrained to be greater than or equal to zero. A constraint $y_j^k = y_{j+1}^k \quad \forall j \in \mathtt{M}_{\mathrm{rev}}$ is imposed in the outer problem so that the forward and backward reactions mapping to a reversible reaction (listed in the set $\mathtt{M}_{\mathrm{rev}}$) can only be knocked out simultaneously. For the regulated reactions, either the forward or the backward flux can be modulated and this is imposed by the following set of constraints:

$$y_j^d + y_{j+1}^d \geqslant 1 \ \ \forall j \in \mathtt{M}_{\mathrm{rev}} \quad \text{and} \quad y_j^u + y_{j+1}^u \geqslant 1 \ \ \forall j \in \mathtt{M}_{\mathrm{rev}}.$$

2.3. OptReg framework

In analogy to OptKnock and based on the above variable definitions and constraints, the bilevel optimization formulation for OptReg is as follows:

$$\max_{y_j^K, y_j^U, y_j^D} v_{\mathrm{biochemical}} \qquad \text{(OptReg)}$$

s.t.

$$\left(\begin{array}{ll}
\max\limits_{v_j} & v_{\mathrm{biomass}} - \varepsilon\cdot\sum\limits_j v_j \quad \text{(Primal)} \\
\text{s.t.} & \sum\limits_{j=1}^{M} S_{ij}v_j = 0, \quad \forall i \in \mathtt{N} \\
& v_{\mathrm{atp}} \geqslant v_{\mathrm{atp_maint}}, \\
& v_{\mathrm{glc}} = 10\,\mathrm{mmol/gDW\cdot h}, \\
& v_{\mathrm{biomass}} \geqslant (0.01)\cdot v_{\mathrm{biomass}}^{\max}, \\
& v_j \leqslant v_j^{\max}\cdot y_j^k, \ v_j \geqslant v_j^{\min}\cdot y_j^k, \quad \forall j \in \mathtt{M}, \\
& v_j \leqslant [(v_{j,L}^0)\cdot(1-C)+(v_j^{\min})\cdot(C)]\cdot(1-y_j^d) \\
& \qquad +v_j^{\max}\cdot y_j^d, \quad \forall j \in \mathtt{M}, \\
& v_j \geqslant v_j^{\min}, v_j \leqslant v_j^{\max}, \quad \forall j \in \mathtt{M}, \\
& v_j \geqslant [(v_{j,U}^0)\cdot(1-C)+(v_j^{\max})\cdot(C)]\cdot(1-y_j^u) \\
& \qquad +v_j^{\min}\cdot y_j^u, \quad \forall j \in \mathtt{M},
\end{array}\right)$$

$$(1-y_j^k)+(1-y_j^d)+(1-y_j^u)\leqslant 1, \quad \forall j \in \mathtt{M},$$
$$y_j^k \in \{0,1\};\ y_j^d \in \{0,1\};\ y_j^u \in \{0,1\}, \quad \forall j \in \mathtt{M},$$
$$\sum_j[(1-y_j^k)+(1-y_j^u)+(1-y_j^d)]\leqslant L$$
$$y_j^k = y_{j+1}^k, \quad y_j^d + y_{j+1}^d \geqslant 1, \quad y_j^u + y_{j+1}^u \geqslant 1, \quad \forall j \in \mathtt{M}_{\mathrm{rev}},$$

where $v_{\mathrm{biochemical}}$ refers to the flux towards the synthesis of the desired biochemical. The second term in the objective function of the inner problem ensures that the maximum biomass flux distribution with the minimum network "trafficking" (i.e., minimum sum of all fluxes) is chosen out of the set of alternative optima. Recall that reversible reactions have been divided into forward and backward fluxes in the network. Therefore, it is possible to have assigned arbitrarily large values to the forward and backward fluxes for a reversible reaction while the net flux is finite. This in turn can lead to an erroneous prediction of

upregulation for a reaction. To safeguard against this occurrence, we introduce the second term in the objective function of the formulation (OptReg). By minimizing the total flux circulation in the network, only one of the two unidirectional reactions forming a reversible reaction can carry a non-zero flux. Through a trial-and-error process, we have determined that ε has to be assigned values between 0.001 and 0.0001. If ε is greater than 0.001, then the second term starts affecting the solution for the maximization of biomass. If, on the other hand, ε is less than 0.0001, it is ineffective at preventing the presence of flux for both forward and backward directions of a reversible reaction.

The first three constraints in the inner problem (Primal) have already been described previously. The next constraint imposes a minimum requirement on biomass formation and is set at 1% of the maximum theoretical biomass ($v^{\max}_{\text{biomass}}$) feasible in the network. This bilevel problem is solved by first transforming it into a single level problem. To this end, we generate the dual of the Primal problem (Ignizio and Cavalier, 1994) as follows:

minimize

$$\begin{aligned}&(v_{\text{atp\_maint}}\cdot\lambda\_atp)+(0.01\cdot v^{\max}_{\text{biomass}}\cdot\lambda\_bio)\\&\quad+\sum_j(q^k_{U,j}\cdot y^k_j\cdot v^{\max}_j+q^k_{L,j}\cdot y^k_j\cdot v^{\min}_j)\\&\quad+\sum_j\Big[(v^{\max}_j\cdot y^d_j\cdot q^d_{U,j})+[(v^0_{j,L}\cdot(1-C)\\&\quad+v^{\min}_j\cdot(C)]\cdot(1-y^d_j)\cdot q^d_{U,j}+(q^d_{L,j})\cdot(v^{\min}_j)\Big]\\&\quad+\sum_j\Big[(q^U_{U,j}.v^{\max}_j)+(v^{\min}_j\cdot y^u_j\cdot q^u_{L,j})\\&\quad+[(v^u_{j,0}\cdot(1-C)+v^{\max}_j\cdot(C)]\cdot(1-y^u_j)\cdot q^u_{L,j}\Big]\quad\text{(Dual)}\end{aligned}$$

subject to

$$\sum_{i=1}^{N}\lambda_i S_{i,j}+q^k_{U,j}+q^k_{L,j}+q^d_{U,j}+q^d_{L,j}+q^u_{U,j}+q^u_{L,j}\geqslant-\varepsilon,\quad\forall j\in\mathtt{M},\ j\neq atp,\ biomass,$$

$$\sum_{i=1}^{N}\lambda_i S_{i,\text{biomass}}+q^k_{U,\text{biomass}}+q^k_{L,\text{biomass}}+q^d_{U,\text{biomass}}+q^d_{L,\text{biomass}}+q^u_{U,\text{biomass}}+q^u_{L,\text{biomass}}+\lambda\_bio\geqslant1-\varepsilon,$$

$$\sum_{i=1}^{N}\lambda_i S_{i,\text{atp}}+q^k_{U,\text{atp}}+q^k_{L,\text{atp}}+q^d_{U,\text{atp}}+q^d_{L,\text{atp}}+q^u_{U,\text{atp}}+q^u_{L,\text{atp}}+\lambda\_atp\geqslant-\varepsilon,$$

$$q^k_{U,j},q^d_{U,j},q^u_{U,j}\geqslant0;\quad q^k_{L,j},q^d_{L,j},q^u_{L,j}\leqslant0;\quad\forall j\in\mathtt{M},$$

$$\lambda_i\in\mathtt{R},\ \forall\, i\in\mathtt{N};\quad\lambda\_atp\leqslant0;\quad\lambda\_bio\leqslant0.$$

In the dual formulation, λ_i are the dual variables corresponding to the stoichiometric constraints and are unrestricted in sign. The dual variables for the ATP maintenance and the biomass formation constraints are denoted as λ_atp and λ_bio, respectively. Furthermore, $q^k_{L,j}$, $q^u_{L,j}$ and $q^d_{L,j}$ are the dual variables assigned to the constraints imposing lower bounds on the fluxes that become active if the binary variables y^k_j, y^u_j and y^d_j assume a value of zero in this order. Similarly, $q^k_{U,j}$, $q^u_{U,j}$ and $q^d_{U,j}$ are the dual variables corresponding to the constraints imposing upper bounds on the fluxes when a reaction is knocked out, upregulated or downregulated, respectively. An additional complication, absent in OptKnock, is that the objective function involves terms where a binary variable is multiplied by a continuous variable. To recast these nonlinear constraints into an equivalent linear form, we introduce three sets of additional variables (Glover, 1975) as follows:

$$z^k_{U,j}=(q^k_{U,j})\cdot(y^k_j),\quad z^k_{L,j}=(q^k_{L,j})\cdot(y^k_j),\quad\forall j\in\mathtt{M},$$
$$z^u_{L,j}=(q^u_{L,j})\cdot(y^u_j),\quad z^d_{U,j}=(q^d_{U,j})\cdot(y^d_j)\quad\forall j\in\mathtt{M}.$$

By imposing the following constraints the equivalent linear transformation of the nonlinearities is accomplished.

$$(q^k_{U,j})_{LB}\cdot y^k_j\leqslant z^k_{U,j}\leqslant(q^k_{U,j})_{UB}\cdot y^k_j,\quad\forall j\in\mathtt{M},$$

and

$$q^k_{U,j}-(q^k_{U,j})_{UB}\cdot(1-y^k_j)\leqslant z^k_{U,j}\leqslant q^k_{U,j}-(q^k_{U,j})_{LB}\cdot(1-y^k_j),\quad\forall j\in\mathtt{M}.$$

Similar constraints are imposed for the sets of variables $z^k_{L,j}$, $z^u_{L,j}$ and $z^d_{U,j}$. These variables assume non-negative or non-positive values according to the negativity restrictions on the corresponding dual continuous variables. All other bounds (upper/lower) on the dual variables are calculated by maximizing or minimizing them subject to the constraints of the dual problem.

From the strong duality theory in Linear Programming, if the primal and the dual optimal solutions are bounded, then at optimality both objective function values must be equal (Ignizio and Cavalier, 1994). This implies that the unique optimum solution value to the inner primal can be obtained by solving a system of equations encompassing an equality relation for the objective functions of both the primal and the dual problems and the accumulation of their respective constraints. The inner problem is thus transformed from an optimization problem to an equivalent set of equations (equalities and inequalities). This provides a single-level mixed-integer linear MILP formulation for OptReg that is solved using CPLEX (Brooke et al., 2005) accessed through GAMS (Brooke et al., 2003). The complete formulation of OptReg is provided in Appendix A.

This utility of the framework is next demonstrated by applying it to elucidate the optimal down/upregulation and knock-out strategies for overproducing ethanol, an industrially important chemical. In the next section, we discuss and contrast strategies that involve (i) only reaction eliminations, (ii) only reaction modulations i.e., activations/inhibitions and (iii) both deletions and modulations for overproducing ethanol by manipulating the central

metabolism in the *E. coli* metabolic network as abstracted in Reed et al. (2003). The obtained results based on biomass maximization are also contrasted against the ones obtained based on the MOMA (Segre et al., 2002) criterion.

3. Results

The OptReg formulation was employed to determine efficient reaction modification strategies for overproducing ethanol in the *E. coli* metabolic network comprised of more than 1,470 one-directional reactions. The glucose uptake rate was fixed at 10 mmol/gDW · h and the network was allowed to uptake a maximum of 20 mmol/gDW · h of oxygen. All other nutrients such as potassium, ammonia and iron available from minimal media in experimental studies were allowed in unlimited quantities and the network could secrete any metabolite. As expected, ethanol formation in the network takes place through acetaldehyde/alcohol dehydrogenase (*adhE*). It is worth mentioning that the predicted yields of ethanol in the network are strongly affected not only by the carbon availability but also by the abundance of NADH which is a cofactor for *adhE*. Table 1 shows the two and three-reaction modification strategies predicted by the framework for overproducing ethanol. We first discuss the strategies predicted by the framework for a universal value of $C = 0.5$. The impact of C on these strategies will be briefly discussed in a separate subsection.

3.1. Two-reaction modification strategies

A two-reaction deletion mutant (mutant M_A) involving the removal of oxygen transport and phosphotransacetylase (*pta*) (see Fig. 3) was predicted by the framework. This double mutant network has a theoretical yield of 16.30 mmol/gDW · h of ethanol. The predicted anaerobic (fermentative) environment for producing ethanol and the deletion of the competing pathway producing acetate are consistent with experimental evidence. The design for the two-reaction modulation (i.e., repression and activation only) mutant involves the upregulation of pyruvate

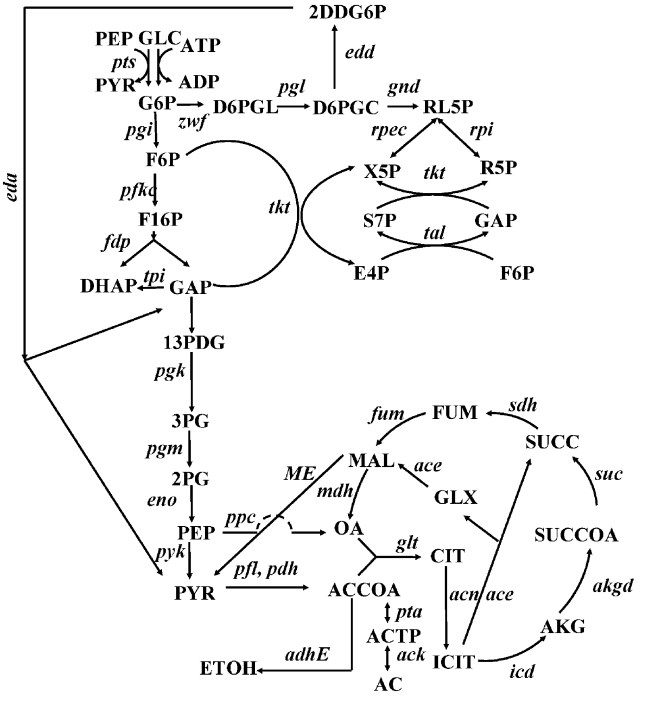

Fig. 3. A pictorial representation of the central metabolic network of *E. coli* and the associated genes to explain the proposed metabolic engineering strategies.

Table 1
Strategies for overproducing ethanol (for $C = 0.5$) listed along with the corresponding ethanol production and biomass formation rates

Deletions	Regulations	Regulations/Deletions
Mutant M_A	Mutant M_B	Mutant M_C
Oxygen transport (X) Phosphotransacetylase (X) Ethanol:16.30, Biomass = 0.19	Pyruvate dehydrogenase (↑) (1.75) Succinate dehydrogenase (↓) (0.33) Ethanol: 16.72; Biomass = 0.17	Pyruvate dehydrogenase (↑) (1.75) Oxygen transport (X) Ethanol: 18.64; Biomass = 0.09
Mutant M_D	Mutant M_E	Mutant M_F
Glucose-6-phosphate isomerase (X) Pyruvate formate lyase (X) Oxygen transport (X) Ethanol: 18.74; Biomass = 0.08	Pyruvate dehydrogenase (↑) (1.75) Succinate dehydrogenase(↓) (0.33) Oxygen transport(↓) (0.14) Ethanol: 16.72; Biomass = 0.17	Phosphoglucomutase (↓) (0.37) Pyruvate formate lyase (X) Oxygen transport(X) Ethanol: 19.83, Biomass = 0.011

The first column indicates strategies that show only eliminations of functionalities. The second column lists only up and down regulations and the last column depicts strategies which involve both reaction deletions and activations/inhibitions. Also note that the first row in the table refers to mutants with two modifications and the second row describes mutant networks with three modifications. The upregulated reactions are denoted with the (↑) symbol. Conversely, downregulated reactions are denoted with the (↓) symbol and the deleted reactions with a (X) symbol. The numbers within parentheses show the relative values of the modified fluxes when compared to the steady state-fluxes. In all the cases except that of succinate dehydrogenase, there was a single value of the flux at steady state. In the case of succinate dehydrogenase, the number within parentheses is the relative value of the modified flux corresponding to $v_{j,L}^0$. Biomass formation rate is expressed on a per hour basis and all the other fluxes are expressed in units of mmol/gDW · h.

dehydrogenase (*pdh*) and the simultaneous downregulation of succinate dehydrogenase (*sdh*) leading to a theoretical yield of 16.72 mmol/gDW · h (mutant M_B). The upregulation of the *pdh* complex enables the conversion of most of pyruvate into acetyl CoA with a concurrent generation of an extra NADH molecule per molecule of acetyl CoA formed. The decrease in flux through *sdh* reduces the activity in the citric acid cycle preventing acetyl CoA from being channeled. The oxygen uptake in the network is very low suggesting microaerobic conditions of growth. Note that the model description does not include regulatory conditions which would signal *adhE* to be inactivated in the presence of oxygen. However, a study suggests that alcohol dehydrogenase is indeed active even in aerobic conditions concomitant to mutations in the *adhE* structural gene and in the promoter region (Holland–Staley et al., 2000). A recent publication reports successful overexpression of *pdh* to increase the carbon flux from pyruvate to acetyl CoA (Vadali et al., 2004).

Column 3 (row1) of Table 1 shows a predicted design in which both activations/inhibitions and knock-outs are allowed (mutant M_C). Not surprisingly, oxygen uptake in the network is eliminated and *pdh* is upregulated. These two-reaction manipulations, in tandem, lead to an ethanol production rate that is approximately 94% of the maximum theoretical yield of 19.87 mmol/gDW · h. Interestingly, Dien et al. (2003) show that pyruvate formate lyase (*pfl*) is induced in anaerobic conditions and a majority of the pyruvate flux is directed through it. However, this fermentative pathway is unbalanced because one NADH and proton is generated for each pyruvate made from sugars, and two NADH and protons are required for converting pyruvate into ethanol (See Appendix B for exact stoichiometry of the reactions). Therefore, *E. coli* balances fermentation by also producing acetate, lactate and succinate which compromises the yield of ethanol (Dien et al., 2003). The proposed strategy (mutant M_C) circumvents this problem by upregulating *pdh* leading to a concurrent generation of NADH alleviating cofactor imbalances in the fermentative pathway. This provides higher yields of ethanol by preventing carbon loss to other fermentative products. Note that experimental observations indicate considerable *pdh* activity in anaerobic conditions which can be further induced by the presence of pyruvate in the medium (Carlsson et al., 1985). Similar results with high flux through *pdh* were reported (Underwood et al., 2002) in experiments with pyruvate and acetaldehyde-supplemented media to optimize carbon partitioning at the acetyl CoA node to promote both ethanol and biomass yields in ethanologenic *E. coli*. The flux distribution through the mutant network M_C shows a very small flux through the glycolytic pathway; instead the flux is diverted to the oxidative branch of the pentose phosphate pathway through which it enters the Entner–Doudoroff pathway. Notably, the ethanol-producing bacterium *Zymomonas mobilis* employs this pathway for glucose metabolism instead of the Embden–Meyerhoff–Parnas glycolytic pathway. The net yield of ATP through this pathway is only one mole per mole of glucose consumed which results in low cell mass allowing for higher ethanol yields (Jeffries, 2005; Seo et al., 2005).

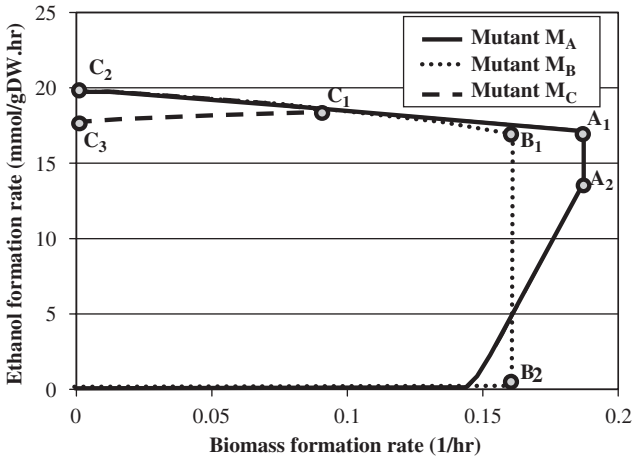

Fig. 4. The biochemical production abilities of the mutant networks with two reactions knocked out and/or modulated. Mutant M_A entails the deletions of phosphotransacetylase in an anaerobic environment chosen by the framework. Mutant M_B requires the downregulation of *sdh* and upregulation of *pdh* and mutant M_C involves the upregulation of *pdh* in an anaerobic environment. Points A_1, B_1 and C_1 denote the maximum ethanol production abilities of the mutant networks M_A, M_B and M_C, respectively. The triangle $C_1C_2C_3$ represents the solution space available to mutant M_C.

We next investigated the maximum and minimum biochemical production abilities of the mutant networks. To this end, we solved the **Primal** problem (see the subsection on OptReg formulation) by first maximizing and subsequently, minimizing ethanol production at all values of biomass feasible to the network. The biochemical production envelopes that we derived for the three mutant networks are shown in Fig. 4. The maximum theoretical yield of the deletion mutant M_A is denoted by point A_1. However, the presence of the vertical line A_1A_2 at the maximum biomass formation rate implies the existence of alternative optimal solutions which means that the theoretical yield of ethanol varies between points A_1 and A_2. Point B_1 denotes the maximum theoretical yield attainable in the mutant network M_B. Line B_1B_2 denotes the rightmost boundary of the allowable phenotypes. Note that point B_2 corresponds to zero ethanol yield implying no coupling between growth and ethanol production for this mutant. In contrast, the network for mutant M_C exhibits ethanol production limits that are superior not only to those for mutant M_B but also to those for the mutant network M_A. Notably, even at small biomass formation rates, the network for mutant M_C produces ethanol (point C_3). Thus, the combination of an upregulation and a knock-out in this case eliminates undesirable phenotypes to a large extent as demarcated by the lower boundary C_1C_3, ultimately providing a single point optimal solution of 18.64 mmol/gDW · h at the maximum growth rate (point

C_1). Note that the cofactor balance in the network forces the increased acetyl CoA flux to be directed towards ethanol production and prevents the formation of other fermentative products such as acetate. The results allude to the synergistic effect of regulations and deletions on enhancing biochemical production.

3.2. Three-reaction modification strategies

The identified strategies for overproducing ethanol which entail modification of three reactions are described here. The three-reaction deletion mutant (mutant M_D) involves the removal of *pfl and* phosphoglucoisomerase (*pgi*) in an anaerobic growth environment. As a consequence of the elimination of *pgi*, 99.8% of the glucose-6-phosphate flux is directed into the pentose phosphate pathway, which eventually enters the Entner–Doudoroff pathway. The expected ethanol yield for this mutant network is 18.74 mmol/gDW · h. The consequence of the removal of *pfl* has already been discussed in the context of the upregulation of *pdh* for mutant M_C. We next discuss the effect of allowing only activations and inhibitions of up to three reactions (mutant M_E). OptReg suggested that oxygen uptake be downregulated along with the inhibition of *sdh* and upregulation of flux through *pdh*. The maximum theoretical yield of ethanol in this case is approximately 16.72 mmol/gDW · h. Note that no improvement over the two-reaction modulation strategy is shown by any of the more than 20 alternative optima that we determined by applying integer cuts (Pharkya et al., 2004).

Once again, allowing for both regulations and knockouts (mutant M_F) leads to the best scenario for ethanol production. In addition to the elimination of *pfl* in an anaerobic environment whose effects have already been discussed earlier, the downregulation of phosphoglucomutase (*pgm*) is predicted to provide an ethanol formation rate of 19.83 mmol/gDW · h. This is approximately 99.84% of the maximum theoretical yield. The computationally generated flux distribution diagram for the network of mutant M_F is shown in Fig. 5. A non-intuitive strategy for enhancing the availability of NADH to promote ethanol production through *adhE* is observed. Approximately 5.1 mmoles/gDW · h of 3-phosphoglycerate, an intermediate in the second half of glycolysis, is converted into serine, thus producing NADH (see Appendix B). A majority of serine is subsequently converted into pyruvate through L-serine deaminase (*sda* in Fig. 3) and is utilized for ethanol production. Note that L-serine deaminase can be produced and activated in *E. coli* grown in glucose minimal medium when appropriate mutations are introduced (Lan and Newman, 2003). The enzyme can also be induced in a variety of ways including growth in anaerobic environment (Su et al., 1989).

An interesting observation that can be made from the flux distribution in Fig. 5 is that even though the serine deaminase reaction is not upregulated, the flux through it is substantially increased. This is because the modeling

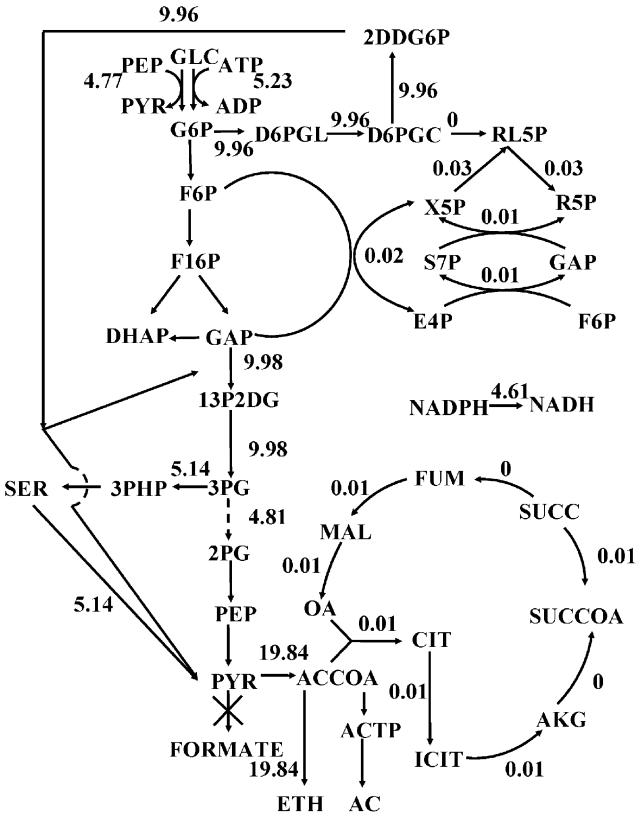

Fig. 5. The flux distribution for the three-reaction mutant network M_F with two reactions deleted (oxygen transport and *pfl*) and one reaction downregulated (*pgm*). The theoretical yield of ethanol in this network is 19.83 mmol/gDW · h which is more than 99% of the maximum theoretical yield. The down regulated reaction is shown with a dashed arrow and the cross represents the reaction deletion. The numbers on the figure indicate the reaction fluxes in mmol/gDW · h based on a glucose uptake rate of 10 mmol/gDW · h.

impact of the downregulation of phosphoglucomutase reaction is to add a constraint on the network that forces the flux through the phosphoglucomutase reaction at a value lower than the reference steady-state value. Consequently, the solution of the inner problem, with this added constraint, causes the flux through a number of reactions, specifically serine deaminase, to depart from their corresponding reference values. Thus, the direct modulation of specific reactions may indirectly affect the flux through other reactions in the network even though they are not classified as up- or downregulated.

The ethanol production envelopes of the networks are shown in Fig. 6. The maximum theoretical yield of mutant M_D is denoted by point D_1. Although this network also exhibits alternative optimal solutions (line D_1D_2), the range of alternative optima at the maximum biomass formation rate is considerably smaller as compared to the one found for mutant M_A. The network for mutant M_E shows significant improvement in terms of the coupling between growth and ethanol production as compared to the network for mutant M_B even though the maximum theoretical yields for both networks are the same. The range of ethanol production rates for mutant M_E at the

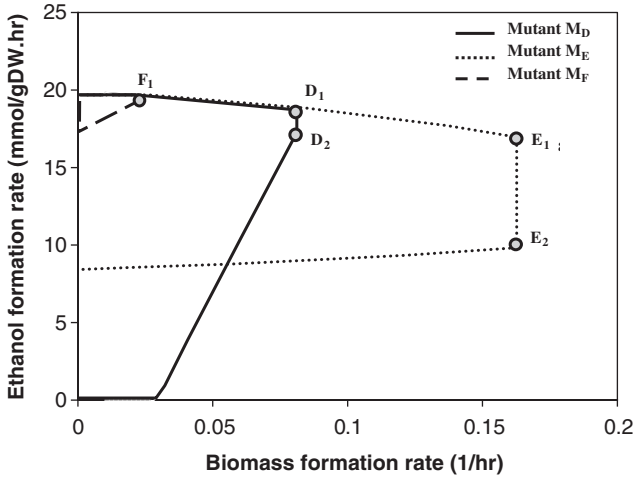

Fig. 6. The biochemical production abilities of the mutant networks with three reactions deleted and/or regulated. The maximum ethanol production yield of each of the networks M_D, M_E and M_F is represented by points D_1, E_1 and F_1, respectively. The vertical lines D_1D_2 and E_1E_2 at the maximum biomass formation rates for the mutant networks M_D and M_E, respectively represent the alternative optimal solutions and consequently, variability in ethanol production for these networks.

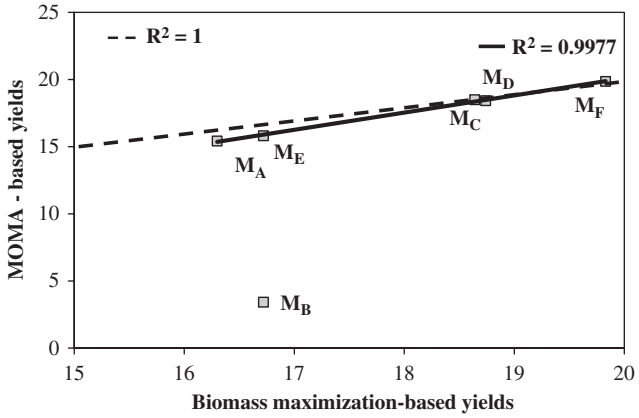

Fig. 7. The ethanol yields (in mmol/gDW · h) of the six mutant networks obtained by using the OptReg framework (max biomass) contrasted against the yields when the minimal flux adjustment criterion (MOMA) is imposed on the networks. The yields from five of the six mutant networks are very close to the prefect correlation line (shown dotted).

maximum growth rate is depicted by line E_1E_2. It can be seen that the network has to produce more than 50% of the maximum theoretical yield of ethanol (point E_2) at the maximum rate of growth. The downregulation of *pgm* in conjunction with the elimination of *pfl* and oxygen availability to the network of mutant M_F forces it to form high amount of ethanol (point F_1). The feasible solution space for this network is significantly reduced and the mutant is forced to secrete ethanol in the presence of growth.

Given that more than 99% of the theoretical yield of ethanol is achieved with three modulations and knockouts, further manipulations in the network do not produce discernible improvements.

3.3. Evaluating the mutant networks using an alternate objective: MOMA

MOMA (Segre et al., 2002) is an alternative criterion introduced recently to anticipate the behavior of microbial systems immediately after imposing genetic modifications. This criterion hypothesizes that microorganisms adjust to a perturbation by minimally adjusting their flux distribution to be as close as possible to that of the wild-type organism which is no longer accessible. Here, we calculate the MOMA criterion for all the six mutant networks identified by OptReg. The objective is to deduce whether the maximization of biomass criterion employed in OptReg yields mutants with predicted yields that are in agreement or disagreement with the predicted yields under the MOMA criterion. The base case flux distribution for the MOMA study is derived by obtaining a feasible set of fluxes that meets the steady-state biomass formation rate ($0.81\,h^{-1}$). Fig. 7 shows that for five out of six of the predicted redesigns (mutants M_A, M_C, M_D, M_E, M_F), both criteria predict very similar yields. The only outlier is the mutant network M_B for which MOMA predicts a much lower yield compared to the max biomass criterion. This close agreement in the results obtained using biomass maximization and MOMA is not surprising in light of the fact that biomass maximization is used to estimate the original steady-state flux distribution needed for applying MOMA. These results are indicative, but not necessarily conclusive, that the strategies obtained from the OptReg framework are quite robust with respect to the choice for the cellular objective (i.e., max biomass or MOMA).

3.4. Effect of the value of the regulation strength parameter, C, on the predicted strategies

In this section we investigate the effect of different values of C on (i) ethanol formation rates and (ii) prediction of design strategies for overproducing ethanol.

With the first objective in mind, we applied four different values of C to the mutant networks identified with C equal to 0.5 and recalculated the ethanol yields. Table 2 shows the obtained results. Notably for values of C below 0.5 and for mutants that involve both modulations and deletions, the theoretical ethanol yields are similar, though not identical, to the ones predicted for the mutant networks for C equal to 0.5. Not surprisingly, the influence of the 'regulation strength parameter' is more pronounced for strategies that involve only reaction modulations. However, it was somewhat surprising to find that the predicted flux strategies led to infeasible solutions for parameter values higher than 0.5. Upon further analysis, we found that higher values of C imply that the fluxes are constrained more tightly. As a result, the strategies suggested for higher C values cannot meet the cofactor requirements in the network and become infeasible.

Table 2
Comparisons of the theoretical yields of ethanol obtained for a value of C equal to 0.5 when implemented on mutant networks after changing C

C	Two reaction M	Two reaction M/D	Three reaction M	Three reaction M/D
0.25	4.19	16.4	4.19	18.64
0.4	13.99	17.59	13.99	19.1
0.5	16.72 (M_B)	18.64 (M_C)	16.72 (M_E)	19.83 (M_F)
0.6	Infeasible	Infeasible	Infeasible	16.95
0.75	Infeasible	Infeasible	Infeasible	Infeasible

The second column shows the yields for the two reaction modulation strategy (mutant M_B), the third column for the two reaction deletion and modulation strategy (mutant M_C), the fourth column for the three-reaction modulation strategy (mutant M_E) and the fifth column for the three reaction deletion/modulation strategy (mutant M_F). M stands for modulation and M/D stands for both modulation and deletion. All yields are in mmol/gDW · h.

Next, we revisited OptReg for values of C equal to 0.25 and 0.75 and identified how the predicted strategies change. When the parameter C is set at 0.75, the framework chooses to upregulate *pfl* and inhibits oxygen uptake. Contrary to the two-reaction modulation strategy for C equal to 0.5, *pdh* is not upregulated here. This is because a higher value of C implies a higher range of fluxes that is allowed to *pdh*. The enhanced *pdh* flux corresponds to an increased requirement of the oxidized cofactor NAD^+. Due to the inhibition of oxygen uptake, the reduced cofactor NADH cannot be oxidized at the rate that is required for the *pdh* flux to be upregulated at C equal to 0.75. As expected, this strain has a predicted lower yield of ethanol than mutant M_B because of the unbalanced fermentative pathway through *pfl*. Thus, its appears that the cofactor balance in the network is a critical factor for deciding whether a particular reaction can be up- or downregulated beyond a certain threshold.

For C equal to 0.25, all regulated fluxes are less constrained as compared to the case when C is equal to 0.5. Therefore, all the strategies suggested for C equal to 0.5 are feasible (see Table 2) though not necessarily optimal. For example, we found a three-reaction deletion/modulation strategy that can lead to the production of 19.68 mmol/gDW · h of ethanol in the network. The framework predicted the upregulation of malic enzyme (ME) in addition to the elimination of *pfl* in an anaerobic environment. To increase the flux through ME, the framework first redirects carbon flux from phosphoenolpyruvate (PEP) to oxaloacetate (OAA) through the phosphoenolpyruvate carboxylase (*ppc*) reaction. OAA is then transformed into malate through the reverse malate dehydrogenase (*mdh*) reaction, which finally gets converted into pyruvate through ME (see Fig. 3). These results indicate that the value of C must be carefully chosen by taking into account the strength of the available promoters.

4. Discussion

This paper describes an integrated framework to identify optimal modulation and deletion strategies for biochemical overproduction. The strategies predicted for the specific case study (ethanol overproduction) undertaken in this work demonstrate that OptReg can allocate and manipulate the fluxes in the network to meet the carbon and the redox requirements for accomplishing the desired biotechnological goal. The critical role that cofactor availability plays in the accomplishment of the desired biotechnological goal has been reported extensively in the literature (San et al., 2002; Berrios–Rivera et al., 2004). Also, quite interestingly, results obtained from the OptReg framework are found to be quite robust with respect to the choice for the cellular objective (i.e., max biomass or MOMA).

It should be emphasized that although reactions deletions or modulations alone can be successful at enhancing the secretion of the desired biochemical, *it is the synergistic effect of both kinds of manipulations that bears the maximum effect on the targeted overproduction*. Specifically, for fermentative products such as ethanol where oxygen uptake needs to be eliminated, just downregulation of oxygen uptake may not generate the desired results. Similarly, reaction eliminations restrict the range of reactions that can be manipulated in a network because of lethality considerations. For example, in the network for mutant M_F, the downregulation of *pgm* and the elimination of oxygen availability to the network in conjunction with the removal of *pfl* leads to high ethanol yields. If only reaction deletions are allowed in the framework, a strategy with simultaneous deletion of *pgm* and oxygen uptake is not feasible because the network cannot meet the non-growth associated ATP maintenance requirements under such a scenario.

It is worthwhile to note that the modulation strategies predicted by OptReg should be interpreted carefully because we separate the reversible reactions into their forward and backward counter-parts. For example, in one case we found that OptReg suggested the upregulation of the backward flux of the aconitase reaction. However, when both the forward and backward fluxes of this reaction were analyzed, the net flux was found to be downregulated in the forward direction.

It is important to note that OptReg predicts gene modulation/deletion strategies by using stoichiometric models of metabolism. These models offer a genome-scale, though approximate, description of cellular metabolism and biochemistry (Schilling et al., 2002; Forster et al., 2003; Reed et al., 2003) allowing for the global assessment of the effect of energetics and cofactor balancing on

overproduction. These models used within the context of the Flux Balance Analysis (FBA) framework (Varma and Palsson, 1994; Price et al., 2004), have been quite successful at elucidating the growth characteristics of the *E. coli* cells in disparate environments (Edwards et al., 2001; Ibarra et al., 2002; Fong et al., 2003) and their responses to gene mutations (Fong and Palsson, 2004). However, they do suffer from insensitivity to potential kinetic and regulatory barriers. Therefore, it is important to carefully interpret the results obtained from OptReg and contrast them with results obtained from complementary efforts using kinetic models and those employing the knowledge of qualitative regulatory interactions.

Acknowledgements

The authors would like to thank Professor Patrick Cirino, Dr. Anshuman Gupta and Dr. Anthony Burgard for helpful discussions. We also acknowledge funding from the DOE (PRA# 02-05ER25684.000).

Appendix A

The single level OptReg formulation obtained by converting the inner level optimization problem into a set of equations by utilizing the strong duality theory is as follows:

$$\text{maximize} \quad v_{\text{biochemical}} \qquad \textbf{(OptReg)}$$

subject to

$$\begin{aligned}
&v_{\text{biomass}} - \varepsilon \cdot \sum_j v_j = (v_{\text{atp_maint}} \cdot \lambda_atp) + (0.01.v_{\text{biomass}}^{\max}\lambda_bio) \\
&\quad + \sum_j (z_{U,j}^k \cdot v_j^{\max} + z_{L,j}^k \cdot v_j^{\min}) \\
&\quad + \sum_j \Big[(v_j^{\max} \cdot z_{U,j}^d) + [(v_{j,L}^0 \cdot (1-C) \\
&\quad + v_j^{\min} \cdot (C)] \cdot (q_{U,j}^d - z_U^d) \\
&\quad + (q_{L,j}^d) \cdot (v_j^{\min}) \Big] + \sum_j \Big[(q_{U,j}^U . v_j^{\max}) \\
&\quad + (v_j^{\min} \cdot z_{L,j}^u) + [(v_{j,0}^u \cdot (1-C) \\
&\quad + v_j^{\max} \cdot (C)] \cdot (q_{L,j}^u - z_{L,j}^u) \Big],
\end{aligned}$$

$$\sum_{j=1}^{M} S_{ij} v_j = 0, \quad \forall i \in \mathtt{N},$$

$$v_{\text{atp}} \geqslant v_{\text{atp\_maint}},$$

$$v_{\text{biomass}} \geqslant (0.01) \cdot v_{\text{biomass}}^{\max},$$

$$v_{\text{glc}} = 10\,\text{mmol/gDW} \cdot \text{h},$$

$$v_j \leqslant v_j^{\max} \cdot y_j^k, \quad \forall j \in \mathtt{M},$$

$$v_j \geqslant v_j^{\min} \cdot y_j^k, \quad \forall j \in \mathtt{M},$$

$$v_j^{\min} \leqslant v_j \leqslant [(v_{j,L}^0) \cdot (1-C) + (v_j^{\min}) \cdot (C)] \cdot (1 - y_j^d) + v_j^{\max} \cdot y_j^d, \quad \forall j \in \mathtt{M},$$

$$[(v_{j,U}^0) \cdot (1-C) + (v_j^{\max}) \cdot (C)] \cdot (1 - y_j^u) + v_j^{\min} \cdot y_j^u \leqslant v_j \leqslant v_j^{\max}, \quad \forall j \in \mathtt{M},$$

$$(1 - y_j^k) + (1 - y_j^d) + (1 - y_j^u) \leqslant 1, \quad \forall j \in \mathtt{M},$$

$$y_j^k \in \{0,1\}; \quad y_j^d \in \{0,1\}; \quad y_j^u \in \{0,1\}, \quad \forall j \in \mathtt{M},$$

$$\sum_j [(1 - y_j^k) + (1 - y_j^u) + (1 - y_j^d)] \leqslant L$$

$$y_j^k = y_{j+1}^k, \quad y_j^d + y_{j+1}^d \geqslant 1, \quad y_j^u + y_{j+1}^u \geqslant 1, \quad \forall j \in \mathtt{M}_{\text{rev}},$$

$$\sum_{i=1}^{N} \lambda_i S_{i,j} + q_{U,j}^k + q_{L,j}^k + q_{U,j}^d + q_{L,j}^d + q_{U,j}^u + q_{L,j}^u \geqslant -\varepsilon, \quad \forall j \in \mathtt{M},\ j \neq atp,\ biomass,$$

$$\sum_{i=1}^{N} \lambda_i S_{i,\text{biomass}} + q_{U,\text{biomass}}^k + q_{L,\text{biomass}}^k + q_{U,\text{biomass}}^d + q_{L,\text{biomass}}^d + q_{U,\text{biomass}}^u + q_{L,\text{biomass}}^u + \lambda\_bio \geqslant 1 - \varepsilon,$$

$$\sum_{i=1}^{N} \lambda_i S_{i,\text{atp}} + q_{U,\text{atp}}^k + q_{L,\text{atp}}^k + q_{U,\text{atp}}^d + q_{L,\text{atp}}^d + q_{U,\text{atp}}^u + q_{L,\text{atp}}^u + \lambda\_atp \geqslant -\varepsilon,$$

$$\begin{aligned}
&0 \leqslant z_{U,j}^k \leqslant (q_{U,j}^k)_{UB} \cdot y_j^k, && \forall j \in \mathtt{M}, \\
&q_{U,j}^k - (q_{U,j}^k)_{UB} \cdot (1 - y_j^k) \leqslant z_{U,j}^k \leqslant q_{U,j}^k, && \forall j \in \mathtt{M}, \\
&(q_{L,j}^k)_{LB} \cdot y_j^k \leqslant z_{L,j}^k \leqslant 0, && \forall j \in \mathtt{M}, \\
&q_{L,j}^k \leqslant z_{L,j}^k \leqslant q_{L,j}^k - (q_{L,j}^k)_{LB} \cdot (1 - y_j^k), && \forall j \in \mathtt{M}, \\
&0 \leqslant z_{U,j}^d \leqslant (q_{U,j}^d)_{UB} \cdot y_j^d, && \forall j \in \mathtt{M}, \\
&q_{U,j}^d - (q_{U,j}^d)_{UB} \cdot (1 - y_j^d) \leqslant z_{U,j}^d \leqslant q_{U,j}^d, && \forall j \in \mathtt{M}, \\
&(q_{L,j}^u)_{LB} \cdot y_{\mathtt{j}}^u \leqslant z_{L,j}^{\mathtt{u}} \leqslant 0, && \forall j \in \mathtt{M}, \\
&q_{L,j}^u \leqslant z_{L,j}^u \leqslant q_{L,j}^u - (q_{L,j}^u)_{LB} \cdot (1 - y_j^u), && \forall j \in \mathtt{M}, \\
&q_{U,j}^k, q_{U,j}^d, q_{U,j}^u \geqslant 0; \quad q_{L,j}^k, q_{L,j}^d, q_{L,j}^u \leqslant 0; && \forall j \in \mathtt{M}, \\
&\lambda_i \in \mathtt{R}, \quad \forall i \in \mathtt{N}; \quad \lambda_atp \leqslant 0; \quad \lambda_bio \leqslant 0.
\end{aligned}$$

Appendix B. Stoichiometry of selected reactions

Pyruvate formate lyase: CoA + pyruvate → acCoA + formate

Pyruvate dehydrogenase: $CoA + NAD^+ + pyruvate \rightarrow acCoA + CO_2 + NADH$

Acetaldehyde dehydrogenase: $acCoA + 2\,H^+ + 2\,NADH \rightarrow CoA + ethanol + 2\,NAD^+$

Phosphoglycerate dehydrogenase (*serA*): 3-phosphoglycerate + $NAD^+ \rightarrow NADH + H^+$ + 3-phosphohydroxypyruvate

L-Serine deaminase (*sda*): L-serine → NH_4 + pyruvate

References

Alper, H., Miyaoku, K., Stephanopoulos, G., 2005a. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. Nat. Biotechnol. 23 (5), 612–616.

Alper, H., Jin, Y.S., Moxley, J.F., Stephanopoulos, G., 2005b. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *E. coli*. Metab. Eng. 7 (3), 155–164.

Baez-Viveros, J.L., Osuna, J., Hernandez-Chavez, G., Soberon, X., Bolivar, F., Gosset, G., 2004. Metabolic engineering and protein directed evolution increase the yield of L-phenylalanine synthesized from glucose in *E. coli*. Biotechnol. Bioeng. 87 (4), 516–524.

Berrios-Rivera, S.J., Sanchez, A.M., Bennett, G.N., San, K.Y., 2004. Effect of different levels of NADH availability on metabolite distribution in *E. coli* fermentation in minimal and complex media. Appl. Microbiol. Biotechnol. 65 (4), 426–432.

Breuer, M., Ditrich, K., Habicher, T., Hauer, B., Kesseler, M., Sturmer, R., Zelinski, T., 2004. Industrial methods for the production of optically active intermediates. Angew Chem. Int. Ed. Engl. 43 (7), 788–824.

Brooke, A., Kendrick, D., Meeraus, A., Raman, R., 2003. GAMS: A User's Guide. GAMS Development Corporation, Washington D.C.

Brooke, A., Kendrick, D., Meeraus, A., Raman, R., 2005. GAMS: The Solver Manuals. GAMS Development Corporation, Washington, DC.

Burgard, A.P., Vaidyaraman, S., Maranas, C.D., 2001. Minimal Reaction Sets for *E. coli* Metabolism under Different Growth Requirements and Uptake Environments. Biotechnol. Prog. 17, 791–797.

Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84 (6), 647–657.

Carlsson, J., Kujala, U., Edlund, M.B., 1985. Pyruvate dehydrogenase activity in *Streptococcus mutans*. Infect. Immun. 49 (3), 674–678.

Chotani, G., Dodge, T., Hsu, A., Kumar, M., LaDuca, R., Trimbur, D., Weyler, W., Sanford, K., 2000. The commercial production of chemicals using pathway engineering. Biochim. Biophys. Acta 1543 (2), 434–455.

de Ruyter, P.G., Kuipers, O.P., de Vos, W.M., 1996. Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. Appl. Environ. Microbiol. 62 (10), 3662–3667.

Dien, B.S., Cotta, M.A., Jeffries, T.W., 2003. Bacteria engineered for fuel ethanol production: current status. Appl. Microbiol. Biotechnol. 63 (3), 258–266.

Duarte, N.C., Herrgard, M.J., Palsson, B.O., 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res. 14, 1298–1309.

Edwards, J.S., Ibarra, R.U., Palsson, B.O., 2001. In silico predictions of *E. coli* metabolic capabilities are consistent with experimental data. Nat. Biotechnol. 19 (2), 125–130.

Fischer, E., Zamboni, N., Sauer, U., 2004. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. Anal. Biochem. 325 (2), 308–316.

Fong, S.S., Palsson, B.O., 2004. Metabolic gene-deletion strains of *E. coli* evolve to computationally predicted growth phenotypes. Nat. Genet. 36 (10), 1056–1058.

Fong, S.S., Marciniak, J.Y., Palsson, B.O., 2003. Description and interpretation of adaptive evolution of *E. coli* K-12 MG1655 by using a genome-scale in silico metabolic model. J. Bacteriol. 185 (21), 6400–6408.

Fong, S.S., Burgard, A.P., Herring, C.D., Knight, E.M., Blattner, F.R., Maranas, C.D. and Palsson, B.O., 2005. In silico design and adaptive evolution of *E. coli* for production of lactic acid. Biotechnol. Bioeng., accepted for publication.

Forster, J., Famili, I., Fu, P.C., Palsson, B., Nielsen, J., 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. 13 (2), 244–253.

Glover, F., 1975. Improved linear integer programming formulations of nonlinear integer problems. Manage. Sci. 22 (4), 445.

Gross, R., Leach, M., Bauen, A., 2003. Progress in renewable energy. Environ. Int. 29 (1), 105–122.

Holland-Staley, C.A., Lee, K., Clark, D.P., Cunningham, P.R., 2000. Aerobic activity of *E. coli* alcohol dehydrogenase is determined by a single amino acid. J. Bacteriol. 182 (21), 6049–6054.

Ibarra, R.U., Edwards, J.S., Palsson, B.O., 2002. *E. coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420 (6912), 186–189.

Ignizio, J.P., Cavalier, T.M., 1994. Linear programming. Prentice Hall, Englewood Cliffs, N.J.

Jeffries, T.W., 2005. Ethanol fermentation on the move. Nat. Biotechnol. 23 (1), 40–41.

Jensen, P.R., Hammer, K., 1998a. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl. Environ. Microbiol. 64 (1), 82–87.

Jensen, P.R., Hammer, K., 1998b. Artificial promoters for metabolic optimization. Biotechnol. Bioeng. 58 (2–3), 191–195.

Koffas, M.A., Jung, G.Y., Stephanopoulos, G., 2003. Engineering metabolism and product formation in *Corynebacterium glutamicum* by coordinated gene overexpression. Metab. Eng. 5 (1), 32–41.

Lan, J., Newman, E.B., 2003. A requirement for anaerobically induced redox functions during aerobic growth of *E. coli* with serine, glycine and leucine as carbon source. Res. Microbiol. 154 (3), 191–197.

Lee, P.C., Schmidt-Dannert, C., 2002. Metabolic engineering towards biotechnological production of carotenoids in microorganisms. Appl. Microbiol. Biotechnol. 60 (1–2), 1–11.

Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. 5 (4), 264–276.

Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D., Keasling, J.D., 2003. Engineering the mevalonate pathway in *E. coli* for production of terpenoids. Nat. Biotechnol. 21, 796–802.

Nakamura, C.E., Whited, G.M., 2003. Metabolic engineering for the microbial production of 1,3-propanediol. Curr. Opin. Biotechnol. 14 (5), 454–459.

Papin, J.A., Price, N.D., Edwards, J.S., Palsson, B.B., 2002. The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. J. Theor. Biol. 215 (1), 67–82.

Pharkya, P., Burgard, A.P., Maranas, C.D., 2003. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. Biotechnol. Bioeng. 84 (7), 887–899.

Pharkya, P., Burgard, A.P., Maranas, C.D., 2004. OptSrain: a computational framework for redesign of microbial production networks. Genome Res. 14 (11), 2367–2376.

Price, N.D., Reed, J.L., Palsson, B.O., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat. Rev. Microbiol. 2 (11), 886–897.

Reed, J.L., Vo, T.D., Schilling, C.H., Palsson, B.O., 2003. An expanded genome-scale model of *E. coli* K-12 (iJR904 GSM/GPR). Genome Biol. 4 (9), R54.

San, K.Y., Bennett, G.N., Berrios-Rivera, S.J., Vadali, R.V., Yang, Y.T., Horton, E., Rudolph, F.B., Sariyar, B., Blackwood, K., 2002. Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in *E. coli*. Metab. Eng. 4 (2), 182–192.

Schilling, C.H., Covert, M.W., Famili, I., Church, G.M., Edwards, J.S., Palsson, B.O., 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. J. Bacteriol. 184 (16), 4582–4593.

Segre, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA 99 (23), 15112–15117.

Seo, J.S., Chong, H., Park, H.S., Yoon, K.O., Jung, C., Kim, J.J., Hong, J.H., Kim, H., Kim, J.H., Kil, J.I., Park, C.J., Oh, H.M., Lee, J.S., Jin, S.J., Um, H.W., Lee, H.J., Oh, S.J., Kim, J.Y., Kang, H.L., Lee, S.Y., Lee, K.J., Kang, H.S., 2005. The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. Nat. Biotechnol. 23 (1), 63–68.

Stafford, D.E., Stephanopoulos, G., 2001. Metabolic engineering as an integrating platform for strain development. Curr. Opin. Microbiol. 4 (3), 336–340.

Su, H.S., Lang, B.F., Newman, E.B., 1989. L-serine degradation in *E. coli* K-12: cloning and sequencing of the sdaA gene. J. Bacteriol. 171 (9), 5095–5102.

Underwood, S.A., Zhou, S., Causey, T.B., Yomano, L.P., Shanmugam, K.T., Ingram, L.O., 2002. Genetic changes to optimize carbon

partitioning between ethanol and biosynthesis in ethanologenic *E. coli*. Appl. Environ. Microbiol. 68 (12), 6263–6272.

Vadali, R.V., Bennett, G.N., San, K.Y., 2004. Enhanced isoamyl acetate production upon manipulation of the acetyl-CoA node in *E. coli*. Biotechnol. Prog. 20 (3), 692–697.

Van Dien, S.J., Lidstrom, M.E., 2002. Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph *Methylobacterium extorquens* AM1, with application to reconstruction of C(3) and C(4) metabolism. Biotechnol. Bioeng. 78 (3), 296–312.

Varma, A., Palsson, B.O., 1994. Metabolic flux balancing-basic concepts, scientific and practical use. Bio-Technology 12 (10), 994–998.

Vera, J., de Atauri, P., Cascante, M., Torres, N.V., 2003. Multicriteria optimization of biochemical systems by linear programming: application to production of ethanol by *Saccharomyces cerevisiae*. Biotechnol. Bioeng. 83 (3), 335–343.

Wyman, C.E., 2003. Potential synergies and challenges in refining cellulosic biomass to fuels, chemicals, and power. Biotechnol. Prog. 19 (2), 254–262.