

# ARTICLES

## A Computational Procedure for Optimal Engineering Interventions Using Kinetic Models of Metabolism

Francisco G. Vital-Lopez, Antonios Armaou, Evgeni V. Nikolaev, and Costas D. Maranas\*

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

The identification of optimal intervention strategies is a key step in designing microbial strains with enhanced capabilities. In this paper, we propose a general computational procedure to determine which genes/enzymes should be eliminated, repressed or overexpressed to maximize the flux through a product of interest for general kinetic models. The procedure relies on the generalized linearization of a kinetic description of the investigated metabolic system and the iterative application of mixed-integer linear programming (MILP) optimization to hierarchically identify all engineering interventions allowing for reaction eliminations and/or enzyme level modulations. The effect of the magnitude of the allowed changes in concentrations and enzyme levels is investigated, and a variant of the method to explore high-fold changes in enzyme levels is also analyzed. The proposed framework is demonstrated using a kinetic model modeling part of the central carbon metabolism of *E. coli* for serine overproduction. The proposed computational procedure is a general approach that can be applied to any metabolic system for which a kinetic description is provided.

### Introduction

Recent advances in molecular biology enable the routine deletion or modulation of the expression level of genes in microbial production strains. This new capability has ushered a new era in microbial strain design where genome content can be manipulated at will. A challenge, however, is that the impact of deletions and/or expression modulations is hard to quantitatively predict. Stoichiometric models of metabolism (1–6) attempt to predict the impact of these modifications by optimizing an objective function that serves as a surrogate of true cellular behavior. These hypotheses include maximization of biomass formation (7), minimization of metabolic adjustment (MOMA) (8), regulatory on/off minimization (ROOM) (9) and ratio adjustment minimization (RAM) (10). These hypotheses-driven predictions have, in certain cases, been quite successful at recapitulating new fluxes after an environmental and/or genetic perturbation (11–14). However, in the absence of any regulatory or mechanistic information the stoichiometry-based predictions are bound to be incomplete.

In contrast, kinetic models of metabolism directly capture engineering interventions as modifications in the parameters of the model. For example, gene knock-outs, as is the case with stoichiometric models, can be described by setting the maximum flux of the affected reaction to zero. On the other hand, repressions or overexpressions can be modeled by increasing or decreasing the corresponding maximum reaction rate parameters. A limitation of kinetic models is that they are typically constructed for a small subset of the totality of metabolism present and it is questionable whether parameter values are invariant in response to environmental or genetic changes (1).

Several researchers have employed a hybrid modeling approach that combines stoichiometric information with high quality (kinetic) data, whenever available. For example, Beard and Qian have combined the stoichiometry with thermodynamic constraints to improve the prediction capabilities of stoichiometric models (15, 16). Covert and Palsson (17, 18) have incorporated the regulation of gene expression to flux balance analysis (FBA), whereas Mahadevan et al. (19) have extended FBA (i.e., dFBA) to describe the dynamic behavior of metabolic system. Gadkar et al. (20) included kinetic expressions in dynamic FBA to optimize the concentration of a targeted product molecule.

For purely stoichiometric models we introduced the Opt-Knock (21) framework for optimal deletion strategies and extended it afterward in OptReg (22) to allow for up/down reaction modulations. The ultimate goal is to combine stoichiometric with kinetic descriptions, whenever available, in the context of hybrid kinetic-stoichiometric descriptions. Specifically, given an underlying kinetic description we would like to identify a prespecified number of knock-outs and/or over/under gene expression regulations that maximize the synthesis rate of targeted metabolite (e.g., maximum ethanol production). It is of particular importance to ensure that any proposed algorithms will be able to handle relatively large kinetic models irrespective of the form of the underlying functions used to describe kinetic rates. In this work, we are beginning to unravel some of these computational challenges by exploring how to optimally perform strain engineering modifications for a purely kinetic description. The approach proposed here relies on the iterative use of basis function linearization and mixed-integer linear programming (MILP) optimization.

A number of different kinetic modeling descriptions are available, which are either mechanistic descriptions or general-

\* To whom correspondence should be addressed. Tel: 1+814 863 9958. Fax: 1+814 865 7846. E-mail: costas@psu.edu.

**Table 1. Examples of Generalized Reaction Rates**

generalized reaction rate	mathematical form <sup>a</sup>
linear	$r = r_0 + k(e - e_0) + \sum_i k_i(c_i - c_{0,i})$
linear in logarithms	$r = r_0 + k \ln\left(\frac{e}{e_0}\right) + \sum_i k_i \ln\left(\frac{c_i}{c_{0,i}}\right)$
generalized mass action, S-systems	$r = r_0 k \frac{e}{e_0} \prod_i \left(\frac{c_i}{c_{0,i}}\right)^{k_i}$
thermokinetic, lin-log models	$r = \frac{e}{e_0} \left( k + \sum_i k_i \ln\left(\frac{c_i}{c_{0,i}}\right) \right)$
Michaelis–Menten type	$r = e \left( \prod_i \frac{c_i}{k_i + c_i} - K \prod_j \frac{c_j}{k_j + c_j} \right)$

<sup>a</sup>  $r$ : reaction rate,  $e$ : enzyme level,  $c_i$ : concentration of metabolite  $i$  affecting the rate;  $c_j$ : concentration of metabolite  $j$  consider as a product in reversible reactions in the last equation;  $K$ : ratio of the reverse to forward rate constant;  $k, k_i$ : kinetic parameters;  $r_0, e_0, c_{0,i}$  stand for a reference state.

ized rate equations. In general, mechanistic kinetic models are mathematically more complicated and require more experimental information to parametrize; however, they offer a more comprehensive view of metabolism (23–26). On the other hand, kinetic models that draw from generalized rate equations conform all reaction rates to general forms (27). This results in more compact models with standardized nonlinearities that may lead to some loss of accuracy. Table 1 illustrates some examples of generalized rate equations such as linear approximations (28, 29), linear with respect to logarithmic functions (30), generalized mass action (GMA) and S-systems (31–34), thermokinetic descriptions (35), lin-log kinetics (36), and general Michaelis–Menten type (37). As expected, the simplest linear approximations are valid only within a narrow region around the steady state at which they were parametrized. This region is substantially expanded in linear in logarithms models by substituting the linear variables by logarithmic expressions. Hatzimanikatis and Bailey (38) have shown that these models can accurately describe the dynamic response of a metabolic system that exhibits strong nonlinearities. Moreover, their kinetic parameters can be obtained, using the well-established metabolic control analysis (MCA) framework, either from a mechanistic model or from experimentally elucidated elasticities and control coefficients. Alternatively, reaction rates have been expressed as the product of enzyme levels and various functions of the concentrations.

Given the aforementioned modeling frameworks, a number of optimization approaches have already been proposed that make use of these frameworks to drive the re-engineering of metabolism. Nonlinear programming (NLP) has been widely used for this purpose. For instance, Mendes and Kell developed a platform to determine the value of the enzyme levels and kinetic parameters that optimize a given metabolic output (39). They formulated the problem as a NLP and applied it on a branched metabolic pathway with feedback composed of eight metabolites and reactions along with 16 kinetic parameters. They solved the resulting NLP with a variety of methods including genetic algorithms and evolutionary programming. Visser et al. have used a large-scale lin-log kinetic model to determine the optimal enzyme modulations by solving a NLP (40). Similarly, Schmid et al. used a large-scale mechanistic model to address the same problem (41). In their approach they partitioned the manipulated variables (i.e., enzyme levels) into subsets and

solved a NLP for each subset while keeping the enzyme levels of all other subsets constant. In the pioneering works of Hatzimanikatis et al. (30) and Dean and Dervakos (42, 43) a number of novel questions regarding the redesign of the metabolic systems were raised, such as which enzymes and at what levels should be modulated to optimize a desired flux if only a given number of manipulations are allowed? To answer this question, Dean and Dervakos formulated the problem as a mixed-integer nonlinear programming (MINLP) problem based on a nonlinear kinetic model. Alternatively, Hatzimanikatis et al. used linear in logarithms functions to approximate the kinetic description of a relatively small metabolic model (i.e., six reactions and metabolites) and formulated the problem as a MILP. In both cases the employed models were relatively small with at most 32 reactions and 13 metabolites.

In this paper, we determine the genes/enzymes that should be manipulated (i.e., deleted, repressed or overexpressed) to optimize the flow toward a target metabolite given a constrained number of total genetic manipulations for a general nonlinear kinetic model of a relatively large size. As in previous efforts, we assume that the enzyme levels can be manipulated independently of each other and that the levels of the enzymes not manipulated remain constant. Existing NLP-based approaches do not allow for only a prespecified number of manipulations. Hatzimanikatis et al. (30) and Dean and Dervakos (42, 43) laid the foundation toward answering some of these challenges. However, Hatzimanikatis et al. did not consider knockouts of genes, as the adopted linear in logarithms model (similar to second row in Table 1) becomes ill-defined when the enzyme levels approach zero. In the other case, the MINLP formulation proposed by Dean and Dervakos becomes hard to solve for problems involving a large number of enzymatic reactions (i.e., hundreds of reactions). The approach introduced here overcomes these limitations by first automatically deriving linearizations according to basis function approximations for arbitrary kinetic representations. Then, the modeling of knockouts of genes/enzymes is implemented through the introduction of binary variables that “remove” the corresponding reactions from the system. By combining these approaches, the problem can be formulated as a MILP that can be solved for large metabolic systems. It should be mentioned that although linearization of the kinetic models confers mathematical and computational tractability to the problem, generally only moderate enzyme levels changes can be imparted and still maintain agreement with the nonlinear model. For large enzyme level changes, a MINLP-based modification of the procedure is described. In the next section we briefly describe the adopted mathematical description of metabolic networks followed by the proposed optimization procedure. Subsequently, we provide computational results, and the work is summarized.

## Mathematical Description

**Metabolic Kinetic Models.** Briefly, a metabolic system consists of a set of processes (enzymatic reactions and transport processes) that transform organic substrates into metabolic precursors and energy metabolites. The rate at which these processes take place depends on the concentration of the reactants/products, the amount of inhibitors or activators, and the activity and level of the participating enzymes. Mathematically, the reaction rate of an enzymatic reaction can be expressed as

$$r_j = r_j^{\max}(e_j)f_j(\mathbf{c}_j, \mathbf{p}_j) \quad (1)$$

where  $r_j$  represents the rate of reaction catalyzed by enzyme  $j$ ,

$r_j^{\max}(e_j)$  is the maximum reaction rate, and  $e_j$  is the corresponding enzyme level.  $f_j(\mathbf{c}_j; \mathbf{p}_j)$  is a nonlinear function of the concentration of metabolites  $\mathbf{c}_j$  and kinetic parameters  $\mathbf{p}_j$  that affect this reaction. Typical examples for  $f_j(\mathbf{c}_j; \mathbf{p}_j)$  are Michaelis–Menten, Hill equation, GMA kinetics, and S-system (31–34), lin-log (30, 35, 36, 38, 40) and allosteric enzymes type kinetics as discussed in the previous section.

For a metabolic system of  $m$  reactions,  $n$  metabolites, and  $k$  kinetic parameters, the reaction rates are related through the mass balances of the metabolites. These mass balances can be described by a set of ordinary differential equations:

$$\frac{d\mathbf{c}}{dt} = \mathbf{S} \cdot \mathbf{r}(\mathbf{e}, \mathbf{c}, \mathbf{p}) \quad (2)$$

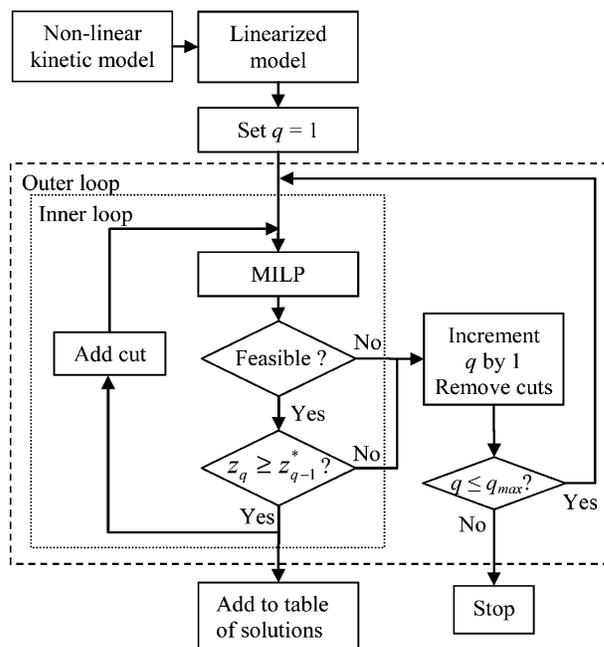
Here  $\mathbf{c}$  is the  $n$ -dimensional vector of metabolite concentrations,  $\mathbf{r}$  is the  $m$ -dimensional vector of reaction rates,  $\mathbf{e}$  is the  $m$ -dimensional vector of enzyme levels,  $\mathbf{p}$  is the  $k$ -dimensional vector of parameters, and  $\mathbf{S}$  is the  $n \times m$  stoichiometric matrix. It should be noted that a dilution accounting term due to cellular growth can be incorporated in eq 2 as in ref 24. Equations 1 and 2 capture the general structure of general kinetic models addressed in this work.

**Optimization.** Given the model description outlined above, we use mathematical optimization tools to answer the following questions: (1) Which genes should be deleted? (2) Which genes should be repressed or overexpressed or combinations thereof in order to maximize the flow through a prespecified flux? This optimization task involves both discrete 0-1 variables (e.g., whether a gene is targeted for deletion or modulation of expression) as well as continuous variables (e.g., metabolic fluxes and concentrations). The mathematical formulation of this task gives rise to mixed-integer nonlinear programming (MINLP) problems with a nonconvex continuous part (42, 43). Although there exist algorithms (44) for globally solving this class of problems, the solution procedure is highly dependent on the type of nonlinearities and typically cannot handle problems of the targeted size and complexity. Local MINLP solvers (45, 46) may provide satisfactory solutions for medium-scale metabolic systems (i.e., tens of enzymatic reactions); however, computational burden hinders the application of these solvers as the size of the system increases. This is incompatible with our objective to introduce a procedure that would be applicable for kinetic models involving arbitrary nonlinear expressions with as many as hundreds of reactions. By allowing for perturbations of limited size around the steady-state values, the same questions can be addressed using a MILP representation using customized linearizations that can be obtained adopting different basis functions. These linearizations according to different basis functions are next described in detail.

**Generalized Linearization.** Instead of using the standard Taylor expansion and retaining only up to linear terms, we employ a Lagrange expansion according to an arbitrary basis function to enable the best possible approximation of the nonlinear kinetics (47). This generalized linearization is accomplished for any arbitrary kinetic model, represented by eqs 1 and 2, as a function of the enzyme levels and metabolite concentrations, while keeping the kinetic parameters  $\mathbf{p}$  constant:

Let  $x_j = g_j(e_j)$  for  $j = 1, \dots, m$  and  $y_i = h_i(c_i)$  for  $i = 1, \dots, n$  such that  $g_j(e_{0,j}) \neq 0$  and  $h_i(c_{0,i}) \neq 0$ , then the linearization of eq 1 based on the variables  $x_j$  and  $y_i$  is given by

$$r_j = r_{0,j} + \left. \frac{\partial r_j}{\partial e_j} \right|_{e_{0,j}} (x_j - x_{0,j}) + \sum_{i=1}^n \left( \left. \frac{\partial r_j}{\partial c_i} \right|_{c_{0,i}} \right) (y_i - y_{0,i}) \quad (3)$$



**Figure 1.** Hierarchical procedure for identifying optimal gene manipulation strategies. At each iteration of the outer loop  $q$  manipulations are enforced and  $q_{\max}$  is the maximum number of manipulations. At each iteration of the inner loop a cut is added to exclude the current solution in the next iteration.  $z_q$  is the objective function of interventions of  $q$  manipulations and  $z_{q-1}^*$  is the best objective function value of the best intervention strategy of  $q-1$  manipulations.

where  $r_{0,j}$ ,  $e_{0,j}$ ,  $x_{0,j}$ ,  $c_{0,i}$  and  $y_{0,i}$  are the values of  $r_j$ ,  $e_j$ ,  $x_j$ ,  $c_i$  and  $y_i$  at the reference value, respectively. It should be noted that the region where the linearization satisfactorily approximates the nonlinear model depends on the form of the reaction rates expressions. For instance, this region may be relatively large for Michaelis–Menten kinetics, whereas it could be considerably smaller for sigmoidal kinetics.

Many standardized models conform to the formalism described above such as the approach introduced in ref 30 and the S-systems evaluated at steady state (31, 34). For example, the log-linear description introduced in ref 30 is a special case of the generalized linearization described above by setting  $g_j(e_j) = \ln e_j$  and  $h_i(c_i) = \ln c_i$ . Other basis functions can be used depending on the nonlinearities in the original kinetic model with the objective to improve on the expected approximation error.

Incorporating the linearized reaction rates in the mass balances, eq 2 yields

$$\frac{d\mathbf{y}}{dt} = \mathbf{Y}_c \cdot \mathbf{S} \cdot (\mathbf{r}_0 + \mathbf{G}_0 \cdot \Delta \mathbf{x} + \mathbf{H}_0 \cdot \Delta \mathbf{y}) \quad (4)$$

where  $\Delta \mathbf{x} = \mathbf{x} - \mathbf{x}_0$  and  $\Delta \mathbf{y} = \mathbf{y} - \mathbf{y}_0$ ,  $\mathbf{Y}_c$  is a  $n \times n$  diagonal matrix with elements  $Y_{c,ii} = dy_i/dc_i$ ,  $\mathbf{G}_0$  is a  $m \times m$  diagonal matrix with elements  $G_{jj} = (\partial r_j / \partial e_j) (1/g_j')$ , and  $\mathbf{H}_0$  is a  $m \times n$  matrix with elements  $H_{ji} = (\partial r_j / \partial c_i) (1/h_i')$ , both evaluated at the reference steady state. Equation 4 approximates the nonlinear model as linear function of  $\Delta \mathbf{x}$  and  $\Delta \mathbf{y}$  in the neighborhood of the reference state. At a steady state of eq 4, the time derivative of  $\mathbf{c}$  vanishes, as long as  $\partial y_i / \partial c_i \neq 0$  and  $\partial y_i / \partial c_i$  are finite for all  $i$ , since  $dc_i/dt = (dc_i/dy_i)(dy_i/dt)$ . The enzyme levels and the metabolite concentrations can be obtained from this linearized model by taking the inverse of the corresponding functions, i.e.,  $e_j = g_j^{-1}(x_j)$  and  $c_i = h_i^{-1}(y_i)$ . This generalized linearization enables the expression of the continuous parts of the problem formulation using linear expressions. Different basis functions

can be used depending on the nonlinearities in the original kinetic model. These linear expressions allow us to formulate the optimization problem using a MILP description.

**Hierarchical Complete Enumeration of Genetic Interventions.** Figure 1 illustrates the sequence of hierarchical calculations needed to elucidate the complete set of genetic manipulations leading to an improved phenotype. Specifically, first we identify all single genetic interventions that improve upon the current phenotype by starting from the optimal and subsequently generating the second, third, etc. best by excluding the previous solution and rerunning the problem. Next we proceed to identify double interventions (either knock-outs or modulations) that are at least as effective as the best single intervention strategy. This additional requirement ensures that only genetic modifications that contribute directly to an improvement are generated. This iterative procedure is continued up to a prespecified number  $q_{\max}$  of genetic interventions, which determine the number of iterations of the outer loop. The core of this scheme is the optimization problem that has to be solved at each inner iteration. This optimization problem is formulated as a MILP, which is described next.

**MILP Formulation.** Figure 2 pictorially illustrates the geometric interpretation of the two types of genetic interventions considered. The presence of knock-outs partitions the feasible region into a union of disjoint subregions. These subregions are hyper-rectangles of dimensions that range from  $m - q_w$  to  $m$ , where  $q_w$  is the number of allowed knock-outs. In the absence of knock-outs a single hyper-rectangle of dimension  $m$  is present. Genetic interventions are modeled using the following two sets of binary variables:

$$v_j = \begin{cases} 1 & \text{if enzyme } j \text{ is modulated} \\ 0 & \text{otherwise} \end{cases}$$

$$w_j = \begin{cases} 1 & \text{if enzyme } j \text{ is knocked-out} \\ 0 & \text{otherwise} \end{cases}$$

If enzyme  $j$  is knocked-out, then its corresponding reaction rate must be set equal to zero. A straightforward way to ensure this is to use the following constraints:

$$r_j^l(1 - w_j) \leq r_{0,j} + G_{0,jj}\Delta x_j + \sum_{i=1}^n (H_{0,ji}\Delta y_i) \leq r_j^u(1 - w_j)$$

where  $r_j^l$  and  $r_j^u$  are lower and upper bounds for  $r_j$ . However, these constraints would impose invalid restrictions on  $\Delta x_j$  and  $\Delta y_i$  when  $w_j = 1$ . This can be avoided by introducing two extra terms into eq 3:

$$r_j = r_{0,j}(1 - w_j) + G_{0,jj}\Delta x_j + \sum_{i=1}^n (H_{0,ji}\Delta y_i) + u_j \quad (5)$$

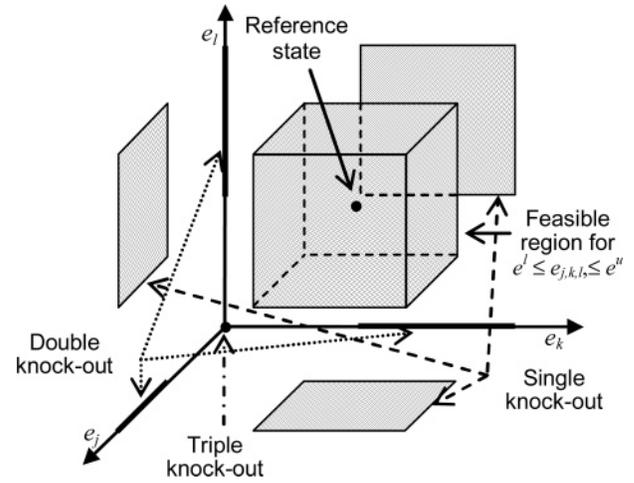
and the following constraints:

$$r_j^l(1 - w_j) \leq r_{0,j}(1 - w_j) + G_{0,jj}\Delta x_j + \sum_{i=1}^n (H_{0,ji}\Delta y_i) + u_j \leq r_j^u(1 - w_j)$$

$$u_j^l w_j \leq u_j \leq u_j^u w_j$$

where  $u_j$  is a continuous variable and  $u_j^l$  and  $u_j^u$  are lower and upper bounds for  $u_j$ .

The lower and upper bounds in  $\Delta x_j$  and  $\Delta y_i$  are computed directly from the bounds on the allowed changes on enzyme



**Figure 2.** Disjoint feasible region corresponding to the MILP formulation. The diagram illustrates the disjoint feasible region of the enzyme levels for  $j$ ,  $k$ , and  $l$ . If no knock-outs are allowed, the optimal solution lies on the cube enclosing the reference state. If one knock-out is allowed, the feasible region will also contain the projections of the cube on the planes formed by the two axes. Increasing the number of allowed knock-outs will expand the feasible region with the addition of more disjoint sets as shown in the figure by the lines on the axis (double knock-out) and the point at their intersection (triple knock-out).

expression levels and metabolite concentrations as  $\Delta x_j^{l,u} = g_j(e_j^{l,u}) - g_j(e_{0,j})$  and  $\Delta y_i^{l,u} = h_i(c_i^{l,u}) - h_i(c_{0,i})$ , respectively. The bounds on  $r_j$  and  $u_j$  can be chosen to be conservative (e.g., very large values for the upper bounds). However, tighter bounds will speed up the solution of the MILP. The maximum lower bounds and minimum upper bounds for  $r_j$  and  $u_j$  can be computed by using  $\Delta x_j^{l,u}$ ,  $\Delta y_i^{l,u}$  and eq 5.

Based on the analyses described above the optimization problem is posed as the following MILP formulation (I):

$$\max z = \sum_{j=1}^m d_j(r_{0,j}(1 - w_j) + G_{0,jj}\Delta x_j + \sum_{i=1}^n (H_{0,ji}\Delta y_i) + u_j)$$

subject to

$$\sum_{j=1}^m S_{ij}(r_{0,j}(1 - w_j) + G_{0,jj}\Delta x_j + \sum_{i=1}^n (H_{0,ji}\Delta y_i) + u_j) = 0 \quad \forall i \quad (c.1)$$

$$\Delta x_j^l v_j \leq \Delta x_j \leq \Delta x_j^u v_j \quad \forall j \quad (c.2)$$

$$\Delta y_i^l \leq \Delta y_i \leq \Delta y_i^u \quad \forall i \quad (c.3)$$

$$r_j^l(1 - w_j) \leq r_{0,j}(1 - w_j) + G_{0,jj}\Delta x_j + \sum_{i=1}^n (H_{0,ji}\Delta y_i) + u_j \leq r_j^u(1 - w_j) \quad \forall j \quad (c.4)$$

$$u_j^l w_j \leq u_j \leq u_j^u w_j \quad \forall j \quad (c.5)$$

$$\sum_{j=1}^m (v_j + w_j) = q \quad (c.6)$$

$$\sum_{j=1}^m v_j = q_v \quad (c.7)$$

$$\sum_{j=1}^m w_j = q_w \quad (c.8)$$

$$v_j + w_j \leq 1 \quad \forall j \quad (c.9)$$

$$\sum_{j|w_j=1}^m w_j + \sum_{j|v_j=1}^m v_j \leq q - 1 \quad \forall t = 1, 2, \dots, T - 1 \quad (c.10)$$

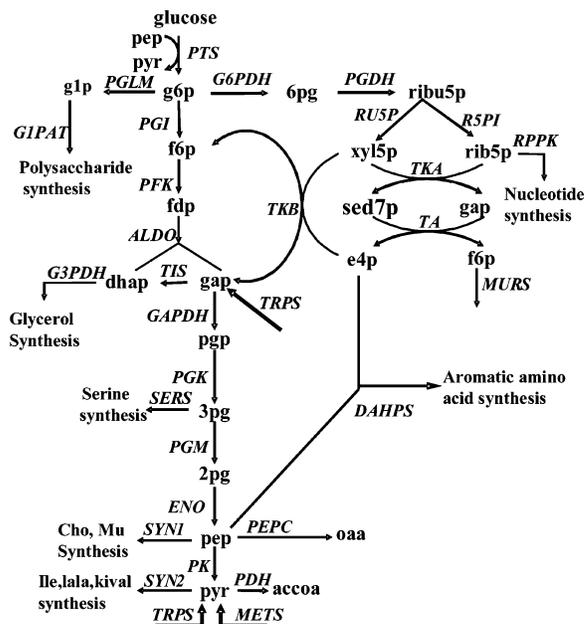
In formulation (I) a weighted combination of fluxes is, in general, being maximized where  $d_j$  is the weight given to reaction  $r_j$  in the objective function. The constraints (c.1) describe the mass balance at steady state for each metabolite concentration. Constraints (c.2) and (c.3) delimit the region where the linearized kinetic model reasonably describes the metabolic system. Constraints (c.2) allow a change in  $\Delta x_j$  only if the corresponding enzyme is to be modulated; otherwise  $\Delta x_j$  is set equal to zero. Constraints (c.3) provide lower and upper bounds on metabolite concentrations. Constraints (c.4) and (c.5) are included to model enzyme knock-outs. If enzyme  $j$  is knocked-out (i.e.,  $w_j = 1$ ), then constraints (c.4) set to zero its corresponding reaction rate, whereas constraints (c.5) allow  $u_j$  to vary between its lower and upper limits. On the other hand, if enzyme  $j$  is not eliminated, constraints (c.5) set  $u_j = 0$  while constraints (c.4) allow  $r_j$  to vary within its lower and upper limits. The total number of manipulated enzymes (modulations plus eliminations) is set equal to  $q$  through constraint (c.6), while constraints (c.7) and (c.8) control the number of modulations and knock-outs, respectively. Simultaneous knock-out and modulation of the same enzyme is avoided through constraints (c.9). Finally, in order to determine the  $T$ th optimal solution considering  $q$  manipulations, constraints (c.10) are used to exclude the solutions derived in previous iterations. These constraints are referred to as cuts in Figure 1 and one of such cuts is added at each iteration of the inner cycle. The accumulation of these cuts is restarted at the beginning of the outer cycle.

By solving formulation (I), the desired values of enzyme expression levels and the concentrations are obtained from  $e_j = g_j^{-1}(x_{0j} + \Delta x_j)$  and  $c_i = h_i^{-1}(y_{0,i} + \Delta y_i)$ . Next we demonstrate the application of the computational framework by determining optimal genetic manipulation strategies through an example of a kinetic model of the central carbon metabolism of *E. coli* (24).

### Computational Results and Discussion

The proposed scheme was employed to determine optimal genetic manipulation strategies that maximize the flux of a prespecified reaction in a relatively large-scale metabolic kinetic model. Specifically, the objective is to maximize serine synthesis in *E. coli*. Current serine production methods include enzymatic and whole cell conversions from glycine and methanol, whereas processes of direct fermentation from glucose are not common (48).

The model used, introduced by Chassagnole et al. (24), describes part of the central carbon metabolism of *E. coli* and is graphically presented in Figure 3. It includes the glycolysis and the pentose phosphate pathway (PPP) as well as drain fluxes

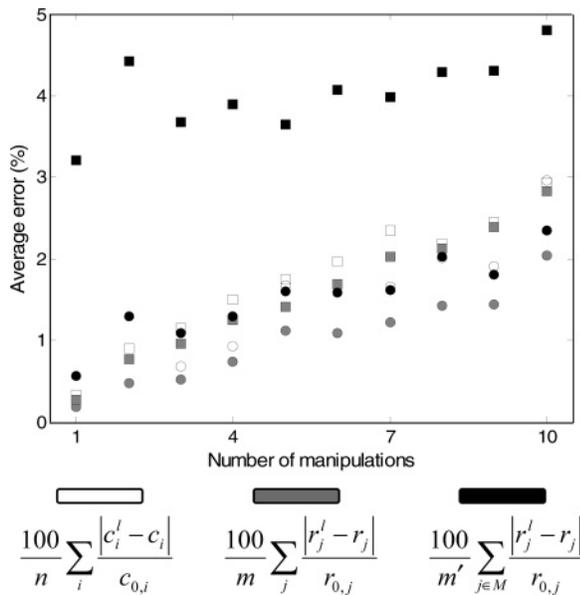


**Figure 3.** Overview of central carbon metabolism of *E. coli*. See Notation section for notation.

of metabolic precursors (e.g., pyruvate, phosphoenol-pyruvate, etc.) to the synthesis of cellular components. The model comprises 17 metabolites and 30 nonlinear reaction rates (e.g., reversible Michaelis–Menten, Hill equation, allosteric regulation, etc). An important limitation of this model is that it does not include mass balances for cofactors such as ATP, NADH and NADPH. Therefore in this work we assumed that the concentrations of the cofactors remain constant. This assumption may significantly affect the quality of the result obtained, but it does not affect the fundamentals of the proposed procedure. We investigate the maximization of serine synthesis by (i) enforcing only knock-outs, (ii) enforcing only modulations of enzyme levels and (iii) allowing for both modulations and knock-outs simultaneously.

In order to select the basis functions for the generalized linearization, the log-linear description used by Hatzimanikatis et al. (30) and the option  $g_j(e_j) = e_j/e_{0,j}$  and  $h_j(c_j) = \ln c_j$  were tested. The test consists of comparing the steady-state reaction rates and concentrations predicted by the linearized models with those corresponding to the nonlinear model after perturbation of a number of enzyme levels. The perturbed enzymes and the perturbation levels were determined randomly using a uniform distribution. Figure 4 shows the average errors of the linearized models for different numbers of perturbed enzymes (100 data points were considered for each number of perturbed enzymes). As can be appreciated, both approximate models (squares represent log-linear description; circles represent  $g_j(e_j) = e_j/e_{0,j}$ ,  $h_j(c_j) = \ln c_j$ ) cause similar average errors in the predicted concentrations (white points) and fluxes (gray points). However, the average errors considering only the fluxes associated with perturbed enzyme levels (black points) are significantly smaller for the case  $g_j(e_j) = e_j/e_{0,j}$  and  $h_j(c_j) = \ln c_j$ . Therefore, we adopt this basis function set to perform the linearization.

We allow enzyme levels to vary only by a factor of 2 (i.e.,  $e_j^l = 0.5e_{0,j}$  and  $e_j^u = 2e_{0,j}$ ) unless they are eliminated, whereas metabolite concentrations are allowed to deviate from the reference state by no more than 50% (i.e.,  $c_i^l = 0.5c_{0,i}$  and  $c_i^u = 1.5c_{0,i}$ ). These bounds correspond to  $\Delta x_j^l = -0.5$ ,  $\Delta x_j^u = 1$ ,  $\Delta y_i^l = \ln 0.5$ , and  $\Delta y_i^u = \ln 1.5$ . The linearization was performed



**Figure 4.** Deviations of the linearized models with respect to a nonlinear kinetic model of *E. coli* (24). Squares correspond to  $g_i(e_j) = \ln e_j$ ,  $h_j(c_j) = \ln c_j$  and circles correspond to  $g_j(e_j) = e_j/e_{0,j}$ ,  $h_j(c_j) = \ln c_j$ .  $c_i$  and  $r_j$  are the predictions of the nonlinear model; the superscript  $l$  denotes the predictions of the linearized models using the indicated functions and the subscript 0 stands for the reference value.  $M$  and  $m'$  define the set and number of perturbed enzymes respectively. The deviations of the linear models from the nonlinear predictions are the average for 100 data points. Each data point was obtained by determining the deviations of the steady-state reaction rates and metabolite concentrations after random perturbations of the enzyme levels within  $0.5e_{0,j} \leq e_j \leq 2e_{0,j}$ . The uniform distribution was used to choose the perturbed enzymes and the perturbations levels. Only points with  $0.5c_{0,i} \leq c_i \leq 1.5c_{0,i}$  were considered.

with the symbolic toolbox of MATLAB, and the MILPs were solved with CPLEX using the interface CPLEXINT (49).

In the following subsections, we present the results for each policy. Subsequently, the effect of the upper limits on the metabolite concentrations is investigated, and finally we consider a variation of the proposed scheme to explore larger changes on the enzyme levels.

**I. Allowable Manipulations: Knock-Outs.** Table 2 lists the identified optimal gene knock-out strategies for serine overproduction (only the top five results are presented for each case). Not surprisingly, most of the single and double knock-outs strategies involve only end-point reactions in the kinetic model. Knock-outs of non-terminal steps are not chosen because they tend to cause large concentration changes. Enzymes of the PPP ( $e_{G6PDH}$  and  $e_{PGDH}$ , see Notation) only appear in the triple knock-outs, whereas glycolytic enzymes ( $e_{PGI}$  and  $e_{PK}$ ) are selected in the best set with four knock-outs and in all the sets with five and six knock-outs, respectively. The best single manipulation is the knock-out of  $e_{DAHPS}$ . This strategy is not intuitive since this reaction does not compete directly with the serine synthesis. It improves the serine production because it increases the concentration of  $c_{PEP}$ , which in turn results in an increase of the concentration of the serine precursor (3-phospho-glycerate,  $c_{3pg}$ ). This knock-out also occurs in all of the best multiple knock-outs. In the best double strategy, the knock-out of  $e_{G1PAT}$  increases the concentration of  $c_{G6p}$ , resulting in a reduction of  $r_{PTS}$ , which is the main consuming reaction of  $c_{PEP}$ . The best strategies with four, five and six knock-outs involve  $e_{DAHPS}$ ,  $e_{G1PAT}$ ,  $e_{PGI}$  and  $e_{PK}$ . In these sets, the knock-out of  $e_{PGI}$  is interesting since one would expect the overexpression of the enzymes upstream of the flux to be maximized in order to

**Table 2. Optimal Knock-Out Strategies to Maximize Serine Synthesis<sup>a</sup>**

$r^*/r_0^*$	knocked-out enzymes <sup>b</sup>				
1.004	G3PDH				
1.010	G1PAT				
1.037	RPPK				
1.043	SYN1				
1.055	DAHPS				
1.055	DAHPS	METS			
1.056	G1PAT	SYN1			
1.059	DAHPS	MURS			
1.061	G3PDH	DAHPS			
1.073	G1PAT	DAHPS			
1.076	G6PDH	PGDH	TRPS		
1.077	G6PDH	PGDH	TA		
1.077	G6PDH	PGDH	TKA		
1.083	RPPK	SYN2	DAHPS		
1.087	SYN1	SYN2	DAHPS		
1.088	SYN1	SYN2	DAHPS	METS	
1.092	SYN1	SYN2	DAHPS	MURS	
1.095	G3PDH	SYN1	SYN2	DAHPS	
1.097	PGI	PK	G1PAT	DAHPS	
1.097	PGI	PK	PGLM	G1PAT	DAHPS
1.098	PGI	PK	G1PAT	DAHPS	METS
1.100	PGI	PK	G3PDH	DAHPS	MURS
1.116	PGI	PK	G1PAT	DAHPS	MURS
1.117	PGI	PK	G1PAT	G3PDH	DAHPS
1.118	PGI	PK	G1PAT	G3PDH	DAHPS METS

<sup>a</sup>  $r^*/r_0^*$  is the ratio of the objective function respect to the reference value. The strategies are sorted in ascending order. Only the top five strategies are presented for each number of allowed knock-outs. <sup>b</sup> See Notation section for notation.

**Table 3. Optimal Modulations Strategies to Maximize Serine Synthesis<sup>a</sup>**

$r^*/r_0^*$	modulated enzymes ( $e_j/e_{0,j}$ )			
1.043	PFK (1.309)			
1.058	GAPDH (1.539)			
1.061	PK (0.5)			
1.063	PEPC (0.5)			
1.949	SERS (2)			
2.001	PDH (2)	SERS (2)		
2.010	PK (0.5)	SERS (2)		
2.012	PTS (1.328)	SERS (2)		
2.012	PEPC (0.5)	SERS (2)		
2.016	PFK (1.479)	SERS (2)		
2.073	GAPDH (1.595)	PK (0.5)	SERS (2)	
2.074	GAPDH (1.579)	PEPC (0.5)	SERS (2)	
2.084	GAPDH (1.774)	PDH (2)	SERS (2)	
2.129	PTS (1.403)	GAPDH (1.964)	SERS (2)	
2.129	PFK (1.567)	GAPDH (1.944)	SERS (2)	

<sup>a</sup>  $r^*/r_0^*$  is the ratio of the objective function with respect to the reference value. The strategies are sorted in ascending order. Only the top five strategies are presented for each number of allowed manipulations. See Notation section for notation.

increase the carbon supply. Moreover, this knock-out redirects the carbon flux through the PPP, which is stoichiometrically less efficient because of the loss of one carbon as  $CO_2$  per each molecule of glucose processed through this pathway. This knock-out is selected however, because other knock-outs lead to larger changes in concentrations than allowed by the imposed bounds.

An important observation of the results listed in Table 2 is that the best single knock-out produces only a 5.5% increment in the rate of serine production, whereas when allowing for as many as six knock-outs only up to an 11.8% increase is achieved. This suggests that serine production cannot be improved considerably by implementing only knock-outs without allowing for drastic concentration changes. This contrasts with the results obtained based on a purely stoichiometric model

**Table 4. Optimal Modulations/Knock-Out Strategies to Maximize Serine Synthesis<sup>a</sup>**

$r^*/r_0^*$	modulated + KO enzymes ( $e_j/e_{0,j}$ )		
1.055	DAHPS (0)		
1.058	GAPDH (1.539)		
1.061	PK (0.5)		
1.063	PEPC (0.5)		
1.949	SERS (2)		
2.010	PK (0.5)	SERS (2)	
2.012	PTS (1.328)	SERS (2)	
2.012	PEPC (0.5)	SERS (2)	
2.016	PFK (1.479)	SERS (2)	
2.074	PK (0)	SERS (2)	
2.106	TIS (2)	PEPC (0)	SERS (2)
2.129	GAPDH (1.191)	PEPC (0)	SERS (2)
2.129	PTS (1.403)	GAPDH (1.964)	SERS (2)
2.129	GAPDH (1.497)	PK (0)	SERS (2)
2.129	PFK (1.567)	GAPDH (1.944)	SERS (2)

<sup>a</sup>  $r^*/r_0^*$  is the ratio of the objective function with respect to the reference value. The strategies are sorted in ascending order. Only the top five strategies are presented for each number of allowed manipulations. See Notation section for notation.

by Pharkya et al. (50). In the cited work a triple knock-out of ENO, ATPase and serine deaminase (the last two enzymes are not considered in the kinetic model) yields a theoretical molar yield of 0.987 with respect to glucose, whereas the kinetic based prediction even after six knock-outs is only 0.122. This indicates, at least for this case, that the kinetics of the enzymatic reactions imposes significantly stronger restrictions than stoichiometry alone. In the following subsection we present the strategies found when only modulations are allowed.

**II. Allowable Manipulations: Modulations.** Table 3 shows the optimal sets and enzyme levels leading to the maximization of serine synthesis when only modulations of the enzyme level are allowed. Not surprisingly, the increase of  $e_{\text{SERS}}$ , corresponding to the flux that is being maximized, represents the best single modulation. Doubling the level of  $e_{\text{SERS}}$  increases the serine synthesis rate by 94.9%. The remaining single modulations improve the serine synthesis indirectly by increasing the concentration of its precursor. Two of them increase the flux upstream of this metabolite by increasing the level of  $e_{\text{PFK}}$  and

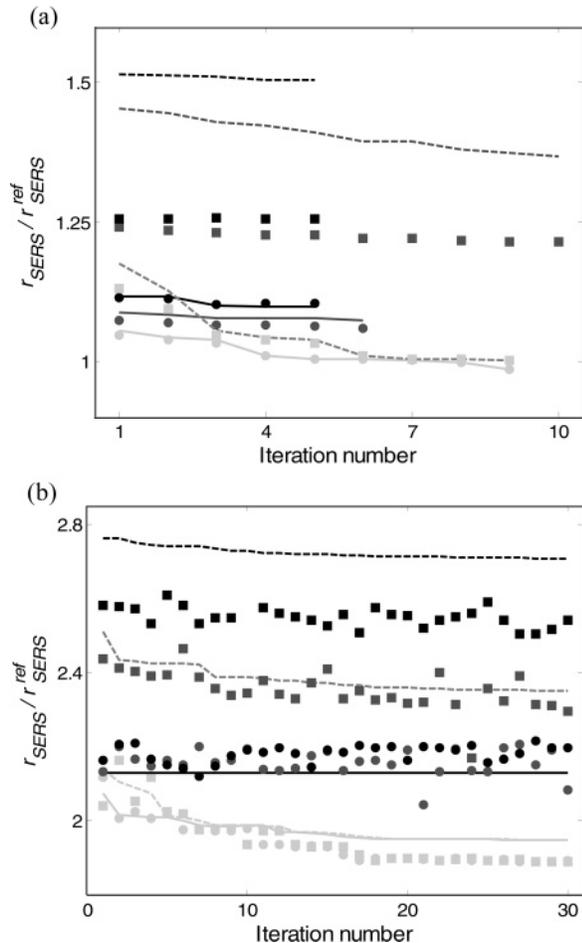
$e_{\text{GAPDH}}$  while the other two reduce fluxes downstream by diminishing the levels of  $e_{\text{PK}}$  and  $e_{\text{PEPC}}$ . In all cases, only modest improvements of less than 6.5% are achieved. This is also the case when two or three enzymes are modulated. All double and triple manipulation strategies include  $e_{\text{SERS}}$ . Modulation of  $e_{\text{PFK}}$  in the best double strategy improves  $r_{\text{SERS}}$  by only an additional 6.7%, while modulation of  $e_{\text{PFK}}$  and  $e_{\text{GAPDH}}$  in the best triple strategy accounts for only 18% of the total improvement. The maximum value of the serine synthesis rate, an improvement of 112.9%, is achieved by two different sets of triple modulations. In both triplets  $e_{\text{SERS}}$  is overexpressed at the maximum allowed level, while  $e_{\text{GAPDH}}$  and  $e_{\text{PTS}}$  or  $e_{\text{PFK}}$  are up-modulated such that  $c_{3\text{pg}}$  reaches its upper limit. Due to the constraints on the upper limits of  $e_{\text{SERS}}$  and  $c_{3\text{pg}}$ , further modulations do not improve the serine synthesis. This occurs in this case because the linear objective function depends only on  $e_{\text{SERS}}$  and  $c_{3\text{pg}}$  and three modulations are enough to make them reach their corresponding upper limits. If  $e_{\text{GAPDH}}$  is not overexpressed together with  $e_{\text{SERS}}$  the maximum value of the serine synthesis is not reached even after six modulations. This suggests that the overexpression of  $e_{\text{GAPDH}}$  is important to maximize the serine synthesis according to the kinetic model. Another interesting result is that the enzymes that are modulated to a lower level,  $e_{\text{PK}}$  and  $e_{\text{PEPC}}$ , are not among the best single, double and triple knock-outs shown in Table 2. This is because the elimination of these enzymes induces changes in concentration that lie outside the allowable concentration ranges.

**III. Allowable Manipulations: Knock-Outs and Modulations.** The strategies for simultaneous knock-outs and modulations produce improvements similar to those obtained when only modulations are allowed. The results are shown in Table 4. The single manipulations are the same as those for modulations, but in the first row overexpression of  $e_{\text{PFK}}$  is replaced with the knock-out of  $e_{\text{DAHPS}}$ . The most important single manipulation is again the overexpression of  $e_{\text{SERS}}$ , while the other manipulations have only a minor effect on the serine synthesis. In the case of two manipulations, the best option is the overexpression of  $e_{\text{SERS}}$  and knock-out of  $e_{\text{PK}}$ , while the remaining manipulations are the best double modulations. The maximum value of this flux is still achieved with only three manipulations, but the

**Table 5. Optimal Strategies To Maximize Serine Synthesis Using Large Upper Bounds for Concentrations<sup>a</sup>**

$r^*/r_0^*$	$\Delta x_j^u = \ln(2); \Delta y_i^u = \ln(11)$						
Knock-Outs							
1.176	PEPC						
1.368	PK	PEPC					
1.453	PK	PEPC	SYN1				
1.505	PK	PEPC	G1PAT	SYN1			
1.516	PK	PEPC	RPPK	SYN1	TRPS		
1.517	PK	PEPC	RPPK	SYN1	TRPS		METS
Modulations ( $e_j/e_{0,j}$ )							
1.949	SERS (2)						
2.140	PTS (2)	SERS (2)					
2.247	PTS (2)	GAPDH (2)	SERS (2)				
2.387	PTS (2)	PFK (2)	GAPDH (2)	SERS (2)			
2.450	PTS (2)	PFK (2)	GAPDH (2)	PEPC (0.5)	SERS (2)		
2.511	PTS (2)	PFK (2)	GAPDH (2)	PK (0.5)	PEPC (0.5)	SERS (2)	
Knock-Outs + Modulations, ( $e_j/e_{0,j}$ )							
1.949	SERS (2)						
2.140	PTS (2)	SERS (2)					
2.342	PTS (1.841)	PEPC (0)	SERS (2)				
2.511	PTS (2)	GAPDH (2)	PEPC (0)	SERS (2)			
2.703	PTS (1.937)	GAPDH (2)	PK (0)	PEPC (0)	SERS (2)		
2.763	PTS (1.382)	GAPDH (1.896)	PK (0)	PEPC (0)	SERS (2)	DAHPS (0)	

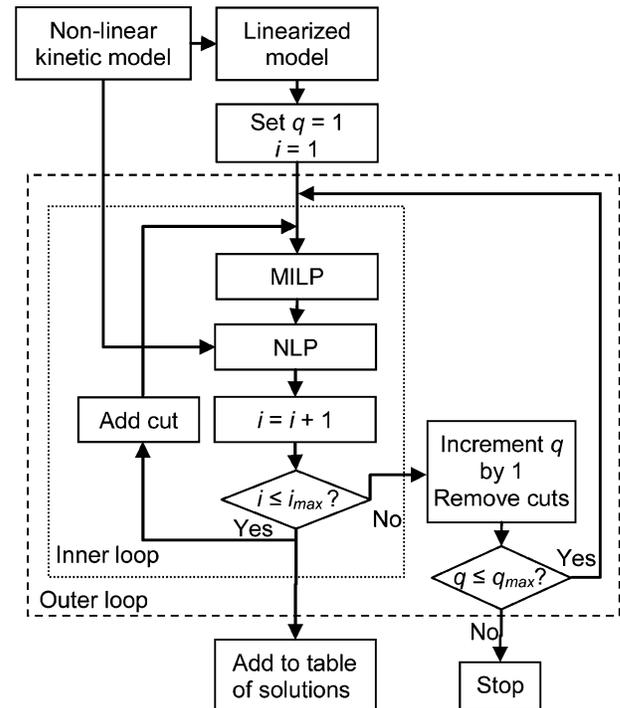
<sup>a</sup>  $r^*/r_0^*$  is the ratio of the objective function respect to the reference value. Only the best strategy is presented for each number of allowed manipulations. See Notation section for notation.



**Figure 5.** Comparison of predictions from MILP formulation and the nonlinear model. Plots illustrate the value of objective function predicted by the MILP formulation (lines) and the corresponding value calculated with the nonlinear model (points) for the optimal strategies. Circles and continuous lines correspond to narrow bounds ( $\Delta x_j^u = 2, \Delta y_i^u = \ln 1.5$ ). Squares and dashed lines correspond to the wide upper bounds ( $\Delta x_j^u = 2, \Delta y_i^u = \ln 11$ ). (a) Strategies involving only knock-outs. Light gray, gray, and black indicate one, three and five knock-outs, respectively. (b) Strategies involving both knock-outs and modulations. Light gray, gray and black indicate two, four and six manipulations respectively.

combination of knock-outs and modulations produce two additional strategies, offering more options. As in the modulation cases, the serine synthesis flux can be increased by a maximum of 112.9%. The yield of the triple manipulations involving either the knock-out of  $e_{PK}$  or  $e_{PEPC}$  (0.19 mol serine/mol glucose) is somewhat larger than the yield of the strategies involving the overexpression of  $e_{PTS}$  or  $e_{PFK}$  (0.16 mol serine/mol glucose). These yields, however, are lower compared to the molar yields with respect to glucose of 0.55 and 0.64 achieved by engineered strains of *B. flavum* in an industrial process (51) and *C. glutamicum* (48). In these strains, the enzymes of the serine synthesis pathway are overexpressed while the serine degrading pathways are deleted or repressed. These manipulations are equivalent to the strategy of the single modulation of  $e_{SERS}$  since serine degrading pathways are not considered in the kinetic model used. The low yields predicted by the linearized kinetic model are due to the restriction on the maximum level of  $e_{SERS}$ . The overexpression of  $e_{GAPDH}$  appears again in all of the sets that achieve the maximum value.

**Effect of the Upper Limit on Concentrations.** In the previous subsections, we identified optimal strategies that comply



**Figure 6.** Procedure to determine optimal manipulation strategies for unrestricted changes in concentration and enzyme levels. The nonlinear model is linearized, and the local information is used to determine the enzyme to be manipulated by solving a MILP problem. The optimal enzyme levels are estimated by solving a NLP problem based on the nonlinear kinetic model through the SQP method. Counter  $i$  controls the number of iterations in the inner loop, and  $i_{max}$  is the maximum number of iterations.  $q$  is the number of required manipulations, and  $q_{max}$  is the maximum number of manipulations.

with the constraints on the allowable concentration deviations (i.e.,  $\pm 50\%$ ). Clearly, the obtained results and accuracy depend on the imposed concentration bound ranges. For relatively small allowed concentration changes the accuracy of the linearized kinetic model is high at the expense of perhaps omitting engineering interventions due to the overly conservative concentration bounds. On the other hand, when allowing for larger concentration deviations, more strategies are sampled at the expense of moving outside of physiologically relevant ranges. The effect of the upper limit on concentrations on the calculated strategies is examined by solving the previous case study after allowing the concentrations to change up to 10-fold (i.e.,  $\Delta y_j^u = \ln(1 + 10)$ ).

The estimated optimal strategies are presented in Table 5 (only the best set for each case is shown). In general, relaxing the concentration bounds leads to higher serine overproductions. Six manipulations allow a serine synthesis rate of 99.9% of its maximum value. The second part of Table 5 shows the best strategies when only modulations are allowed. The results suggest that after  $e_{SERS}$ , the next best candidates to be modulated are  $e_{PTS}$  and  $e_{GAPDH}$ . The strategies obtained when both knock-outs and modulations are allowed are shown in the third part of Table 5.

We next estimated the value of the serine synthesis rate ( $r_{SERS}$ ) using the nonlinear model based on the enzyme levels predicted by the linearized model to compare the quality of the predictions using narrow (i.e., 50% fold changes) and wide (i.e., 10-fold changes) bounds in concentrations. Figure 5 shows that when only knock-outs are allowed, the linearized model, in general, offers reasonable accuracy for narrow concentration bounds, but the error is considerable for the wide concentration bounds.

**Table 6. Iteration Number of the Best Ten Identified Sets using the MILP-SQP**

N.M. <sup>a</sup>		iteration number <sup>b</sup>									
1	1	4	5	7	6	15	3	10	2	12	
2	2	4	7	11	3	6	14	10	8	5	
3	4	10	2	34	5	69	8	41	68	7	
4	4	5	26	6	15	95	31	30	13	36	
5	2	4	12	26	88	8	25	95	34	24	
6	85	65	86	50	13	53	1	59	3	9	

<sup>a</sup> N. M.: number of manipulations. For each number of manipulations 100 iterations were performed. Bounds used in MILP:  $0.9e_0 < e < 1.1e_0$ ,  $0.5c_0 < c < 11c_0$ . NLP:  $0 < e < 5e_0$ ,  $0.001c_0 < c < 10c_0$ . <sup>b</sup> Solutions ordered according to decreasing objective function value.

**Modified Algorithmic Procedure for Large Concentration Ranges.** As described above using local information to identify optimal strategies, allowing large changes in concentration and enzyme levels (results not included) can lead to large approximation errors. To remedy this, we propose a variant of the algorithmic procedure illustrated in Figure 6. This modified procedure uses information at the reference state to predict which enzymes have to be manipulated by solving a MILP. Then the enzyme levels that maximize the serine synthesis rate and the corresponding metabolite concentrations are determined by solving a NLP formulated using the original nonlinear kinetic model. The NLP was solved using a SQP method. In Table 6 we present the iteration number at which the best 10 identified sets were obtained from the procedure (only strategies involving up to six manipulations are considered). In the cases of one and two manipulations, the best 10 sets were determined in the first 10 and 15 iterations, respectively. However, for the cases of three or more manipulations the best solutions are found in late iterations.

### Summary

In this work we presented a computational procedure to determine the gene/enzymes manipulations that maximize the production of a metabolite of interest based on a kinetic model. The procedure is based on iteratively solving MILPs resulting from the linearization of arbitrary kinetic models of metabolism and the use of binary variables to describe the knock-out or modulation of enzymes. Its implementation was illustrated by identifying intervention strategies that maximize the serine synthesis rate according to a kinetic model of a part of the central carbon metabolism of *E. coli*. Through the application of the proposed scheme, optimal strategies involving only knock-outs, only modulations, or both were determined. We found that the obtained engineering strategies and robustness depended heavily on the width of bounds on metabolite and enzyme levels. We concluded that, at least for the case study used in this work, relatively narrow bounds on concentrations generated relatively accurate predictions while for wider concentration ranges there was significant divergence between nonlinear and linearized model predictions. This implies that local information alone is not sufficient to identify optimal strategies of three or more manipulations when large changes in enzyme levels are allowed.

The proposed computational procedure can be a useful tool to aid in the identification of various types of engineering interventions in the redesign of metabolic production systems. Moreover, the generality of the procedure ensures its applicability for any metabolic system for which a kinetic description is provided. This work is a step toward the development of a framework that permits the accurate prediction and selection of the effects of genetic manipulations through the integration of the metabolic knowledge available and/or accessible to such stoichiometry, kinetics, regulatory information, etc. All relevant

MATLAB files are available as Supporting Information; however, a license from ILOG for CPLEX and Matlab 7 are required to execute the scripts.

### Notation

Enzymes	
PT	a phospho-transferase system
PGI	glucose-6-phosphate isomerase
PFK	phosphofructo-kinase
ALDO	aldolase
TIS	triosephosphate iso-merase
GAPDH	glyceraldehyde-3-phosphate dehydro-genase
PGK	phosphoglycerate kinase
PGM	phospho-glycerate mutase
ENO	enolase
PK	pyruvate kinase
PDH	pyruvate dehydro-genase
PEPC	PEP carboxylase
PGLM	phosphoglucomutase
G1PAT	glucose-1-phosphate adenylyltransferase
RPPK	ribose-phosphate pyrophosphokinase
G3PDH	glycerol-3-phosphate dehydrogenase
SERS	serine synthesis
SYN1	synthesis1
SYN2	synthesis2
DAHPS	DAHPS synthases
G6PDH	glucose-6-phosphate dehy-drogenase
PGDH	6-phosphogluconate dehydrogenase
RU5P	ribulose-phosphate epi-merase
R5PI	ribose-phosphate isomerase
TKA	transketolase A
TKB	transketolase B
TA	transaldolase
MURS	mureine synthesis
TRPS	tryptophan synthesis
METS	methionine synthesis

### Metabolites

g6p	glucose-6-phosphate
f6p	fructose-6-phosphate
fdp	fructose-1,6-bisphosphate
gap	glyceraldehydes-3-phosphate
dhap	dihydroxyacetonephosphate
pgp	1,3-diphosphoglycerate
3pg	3-phospho-glycerate
2pg	2-phospho-glycerate
pep	phosphoenolpyruvate
pyr	pyruvate
6pg	6-phosphogluconate
ribu5p	ribulose-5-phosphate
xyl5p	xylulose-5-phosphate
sed7p	sedoheptulose-7-phosphate
rib5p	ribose-5-phosphate
e4p	erythrose-4-phosphate
g1p	glucose-1-phosphate
accoa	aaloacetate
glc	glucose

### Acknowledgment

The authors gratefully acknowledge funding source DOE (DE-FG02-05ER25684).

**Supporting Information Available:** All relevant MATLAB files. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- Gombert, A. K.; Nielsen, J. Mathematical modelling of metabolism. *Curr. Opin. Biotechnol.* **2000**, *11* (2), 180–186.
- Schilling, C. H.; Covert, M. W.; Famili, I.; Church, G. M.; Edwards, J. S.; Palsson, B. O. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol.* **2002**, *184* (16), 4582–4593.
- Reed, J. L.; Vo, T. D.; Schilling, C. H.; Palsson, B. O. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* **2003**, *4* (9), R54.
- Duarte, N. C.; Herrgard, M. J.; Palsson, B. O. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* **2004**, *14* (7), 1298–1309.
- Forster, J.; Famili, I.; Fu, P.; Palsson, B. O.; Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **2003**, *13* (2), 244–253.
- Van Dien, S. J.; Lidstrom, M. E. 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.* **2002**, *78* (3), 296–312.
- Schilling, C. H.; Edwards, J. S.; Palsson, B. O. Toward metabolic phenomics: Analysis of genomic data using flux balances. *Biotechnol. Prog.* **1999**, *15* (3), 288–295.
- Segre, D.; Vitkup, D.; Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (23), 15112–15117.
- Shlomi, T.; Berkman, O.; Ruppin, E.; Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (21), 7695–7700.
- Pharkya, P.; Dasika, M.; Kumar, V. S.; Veeraghavan, N.; Suthers, P.; Burgard, A.; Maranas, C. D. Development of Computational tools for Analyzing and Redesigning Biological Networks. In *Proc. of Genomics: GTL Contractor-Grantee Workshop IV 2006*, 125–126.
- Alper, H.; Jin, Y. S.; Moxley, J. F.; Stephanopoulos, G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab. Eng.* **2005**, *7* (3), 155–164.
- Cox, S. J.; Levanon, S. S.; Sanchez, A.; Lin, H.; Peercy, B.; Bennett, G. N.; San, K. Y. Development of a metabolic network design and optimization framework incorporating implementation constraints: A succinate production case study. *Metab. Eng.* **2006**, *8* (1), 46–57.
- Pramanik, J.; Keasling, J. D. Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.* **1997**, *56* (4), 398–421.
- Fong, S. S.; Burgard, A. P.; Herring, C. D.; Knight, E. M.; Blattner, F. R.; Maranas, C. D.; Palsson, B. O. In silico design and adaptive evolution of *E. coli* for production of lactic acid. *Biotechnol. Bioeng.* **2005**, *91* (5), 643–648.
- Beard, D. A.; Liang, S. D.; Qian, H. Energy balance for analysis of complex metabolic networks. *Biophys. J.* **2002**, *83* (1), 79–86.
- Qian, H.; Beard, D. A. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophys. Chem.* **2005**, *114* (2–3), 213–20.
- Covert, M. W.; Palsson, B. O. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* **2002**, *277* (31), 28058–28064.
- Covert, M. W.; Palsson, B. O. Constraints-based models: Regulation of gene expression reduces the steady-state solution space. *J. Theor. Biol.* **2003**, *221* (3), 309–325.
- Mahadevan, R.; Edwards, J. S.; Doyle, F. J., III, Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* **2002**, *83* (3), 1331–1340.
- Gadkar, K. G.; Doyle, F. J., III; Edwards, J. S.; Mahadevan, R. Estimating optimal profiles of genetic alterations using constraint-based models. *Biotechnol. Bioeng.* **2005**, *89* (2), 243–251.
- Burgard, A. P.; Pharkya, P.; Maranas, C. D. OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **2003**, *84* (6), 647–657.
- Pharkya, P.; Maranas, C. D. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab. Eng.* **2006**, *8* (1), 1–13.
- Hynne, F.; Dano, S.; Sorensen, P. G. Full-scale model of glycolysis in *Saccharomyces cerevisiae*. *Biophys. Chem.* **2001**, *94* (1–2), 121–163.
- Chassagnole, C.; Noisommit-Rizzi, N.; Schmid, J. W.; Mauch, K.; Reuss, M. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol. Bioeng.* **2002**, *79* (1), 53–73.
- Hoefnagel, M. H.; Starrenburg, M. J.; Martens, D. E.; Hugenoltz, J.; Kleerebezem, M.; Van Swam, I. I.; Bongers, R.; Westerhoff, H. V.; Snoep, J. L. Metabolic engineering of lactic acid bacteria, the combined approach: Kinetic modelling, metabolic control and experimental analysis. *Microbiol.* **2002**, *148* (Pt 4), 1003–13.
- Bruggeman, F. J.; Boogerd, F. C.; Westerhoff, H. V. The multifarious short-term regulation of ammonium assimilation of *Escherichia coli*: dissection using an in silico replica. *FEBS J.* **2005**, *272* (8), 1965–1985.
- Heijnen, J. J. Approximative kinetic formats used in metabolic network modeling. *Biotechnol. Bioeng.* **2005**, *91* (5), 534–545.
- Kacser, H.; Burns, J. A. The control of flux. *Symp. Soc. Exp. Biol.* **1973**, *27*, 65–104.
- Heinrich, R.; Rapoport, S. M.; Rapoport, T. A. Metabolic-regulation and mathematical-models. *Prog. Biophys. Mol. Biol.* **1977**, *32* (1), 1–82.
- Hatzimanikatis, V.; Floudas, C. A.; Bailey, J. E. Analysis and design of metabolic reaction networks via mixed-integer linear optimization. *AIChE J.* **1996**, *42* (5), 1277–1292.
- Savageau, M. A. *Biochemical System Analysis*; Addison-Wesley: Reading, MA, 1976.
- Hernandez-Bermejo, B.; Fairen, V.; Sorribas, A. Power-law modeling based on least-squares minimization criteria. *Math. Biosci.* **1999**, *161* (1–2), 83–94.
- Voit, E. O. *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*; Cambridge University Press: Cambridge, 2000.
- Voit, E. O.; Savageau, M. A. Accuracy of alternative representations for integrated biochemical systems. *Biochemistry* **1987**, *26* (21), 6869–6880.
- Nielsen, J. Metabolic control analysis of biochemical pathways based on a thermokinetic description of reaction rates. *Biochem. J.* **1997**, *321* (Pt 1), 133–138.
- Visser, D.; Heijnen, J. J. Dynamic simulation and metabolic redesign of a branched pathway using linlog kinetics. *Metab. Eng.* **2003**, *5* (3), 164–176.
- Varner, J. D. Large-scale prediction of phenotype: Concept. *Biotechnol. Bioeng.* **2000**, *69* (6), 664–678.
- Hatzimanikatis, V.; Bailey, J. E. Effects of spatiotemporal variations on metabolic control: Approximate analysis using (log)-linear kinetic models. *Biotechnol. Bioeng.* **1997**, *54* (2), 91–104.
- Mendes, P.; Kell, D. Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation. *Bioinformatics* **1998**, *14* (10), 869–883.
- Visser, D.; Schmid, J. W.; Mauch, K.; Reuss, M.; Heijnen, J. J. Optimal re-design of primary metabolism in *Escherichia coli* using linlog kinetics. *Metab. Eng.* **2004**, *6* (4), 378–390.
- Schmid, J. W.; Mauch, K.; Reuss, M.; Gilles, E. D.; Kremling, A. Metabolic design based on a coupled gene expression-metabolic network model of tryptophan production in *Escherichia coli*. *Metab. Eng.* **2004**, *6* (4), 364–377.
- Dean, J. P.; Dervakos, G. A. Design of process-compatible biological agents. *Comput. Chem. Eng.* **1996**, *20*, S67–S72.
- Dean, J. P.; Dervakos, G. A. Redesigning metabolic networks using mathematical programming. *Biotechnol. Bioeng.* **1998**, *58* (2–3), 267–271.
- Sahinidis, N. V. BARON: A general purpose global optimization software package. *J. Global Optim.* **1996**, *8* (2), 201–205.
- Viswanathan, J.; Grossmann, I. E. A Combined penalty-function and outer-approximation method for Minlp optimization. *Comput. Chem. Eng.* **1990**, *14* (7), 769–782.

- (46) Brooke, A.; Kendrick, D.; Meeraus, A.; Raman, R. *GAMS—A user's guide*; 1998; www.gams.com.
- (47) Weisstein, E. W. Lagrange Expansion. *MathWorld—A Wolfram Web Resource*; <http://mathworld.wolfram.com/LagrangeExpansion.html>.
- (48) Peters-Wendisch, P.; Stolz, M.; Etterich, H.; Kennerknecht, N.; Sahm, H.; Eggeling, L. Metabolic Engineering of *Corynebacterium glutamicum* for L-serine production. *Appl. Environ. Microbiol.* **2005**, *71* (11), 7139–7144.
- (49) The Hybrid Systems Group; <http://control.ee.ethz.ch/~hybrid/cplexint.php>.
- (50) Pharkya, P.; Burgard, A. P.; Maranas, C. D. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol. Bioeng.* **2003**, *84* (7), 887–899.
- (51) Ikeda, M. Amino acid production processes. *Adv Biochem. Eng. Biotechnol.* **2003**, *79*, 1–35.

Received May 23, 2006. Accepted September 18, 2006.

BP060156O