9-2009

The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems

Evgeni V. Nikolaev

Thomas Jefferson University, evgeni.nikolaev@jefferson.edu

Let us know how access to this document benefits you

Follow this and additional works at: https://jdc.jefferson.edu/pacbfp

Part of the Cell and Developmental Biology Commons, Medical Cell Biology Commons, and the Pathology Commons

Recommended Citation

https://jdc.jefferson.edu/pacbfp/33

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University’s Center for Teaching and Learning (CTL). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Department of Pathology, Anatomy, and Cell Biology Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: JeffersonDigitalCommons@jefferson.edu.
As submitted to:

*Metabolic Engineering*

And later published as:

“The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems”

*Volume 12, Issue 1, January 2010, Pages 26-38*

DOI: 10.1016/j.ymben.2009.08.010

Evgeni V. Nikolaev

Department of Pathology, Anatomy and Cell Biology, Daniel Baugh Institute for Functional Genomics and Computational Biology, Thomas Jefferson University, Philadelphia, PA 19107, USA
December 22, 2009
Abstract

Metabolic engineering of cellular systems to maximize reaction fluxes or metabolite concentrations still presents a significant challenge by encountering unpredictable instabilities that can be caused by simultaneous or consecutive enhancements of many reaction steps. It can therefore be important to select carefully small subsets of key enzymes for their subsequent stable modification compatible with cell physiology. To address this important problem, we introduce a general Mixed Integer Non-Linear Problem (MINLP) formulation to compute automatically which enzyme levels should be modulated and which enzyme regulatory structures should be altered to achieve the given optimization goal using nonlinear kinetic models of relevant cellular systems. The developed MINLP formulation directly employs a stability analysis constraint and also includes nonlinear biophysical constraints to describe homeostasis conditions for metabolite concentrations and protein machinery without any preliminary model simplification (e.g. linlog kinetics approximation). The framework is demonstrated on a well-established large-scale kinetic model of the *Escherichia coli* central metabolism used for the optimization of the glucose uptake through the phosphotransferase transport system (PTS) and serine biosynthesis. Computational results show that substantial stable improvements can be predicted by manipulating only small subsets of enzyme levels and regulatory structures. This means that while more efforts can be required to elucidate larger stable optimal enzyme level/regulation choices, no further significant increase in the optimized fluxes can be obtained and, therefore, such choices may not be worth the effort due to the potential loss of stability properties. The source for instability through saddle-node and Hopf bifurcations is identified, and all results are contrasted with predictions from Metabolic Control Analysis.

**Keywords:** Mixed Integer Non-Linear Problem (MINLP), simulated annealing, sequential quadratic programming, biochemical engineering of cellular systems, kinetic models
1. Introduction

A systematic development of optimal bioprocesses and application of metabolic engineering in biotechnology and biomedical studies requires a deep understanding of microbial organization and function (Bailey 1991; Stephanopoulos et al. 1998). While bioinformatics tools and related technology will continue to dominate in the field within the next decade (Overbeek et al. 2005), these efforts are, by themselves, insufficient and should be complemented by the development of alternative tools to relate static genomes to dynamic nonlinear cellular physiology and population response. One way to approach this goal is to construct plausible mathematical models incorporating relevant molecular details (Bailey 1998; Palsson 2006; Shuler 2005; Tomita 2001). Intrinsic complexity of cellular systems and, as a result, of the corresponding mechanistic models necessitates further development of modeling concepts and computational tools to rapidly extract valuable information from such complex models.

Mathematical models have been extensively used in microbiology since the Monod’s discovery of the relationship between the specific growth rate and the concentrations of limiting substrates (Monod 1949). Microbial population studies can now be complemented by nuclear magnetic resonance (NMR) experiments to measure intracellular fluxes and metabolite levels, which can be economically designed using multi-level computational optimization-based frameworks employing stoichiometric reaction networks (Ghosh et al., 2006). Credible overproduction strategies have recently been suggested based on cellular stoichiometry alone (Alper et al. 2005; Burgard et al. 2003, 2004; Ibarra et al. 2002; Lee et al. 2005, 2006; Pharkya et al. 2004; Pharkya and Maranas 2006). Because stoichiometry does correctly define overall barriers and limits for steady state reaction fluxes under fixed ‘defined medium’ constraints, genome-scale stoichiometric models have been very successful in many instances in fundamental
and applied research (Palsson 2006). However, predictive capability of such stoichiometric models is limited to calculations of ‘instant phenotype snapshots’ and, therefore, such models cannot capture non-stoichiometric dynamic effects (Mahadevan et al. 2002), enzyme regulation (Pharkya et al. 2003), and dynamic responses in protein machinery and genetic control (Laffend and Shuler 1993; Schmid et al. 2004). Alternative advances toward the rational analysis of cellular function are known as Metabolic Control Analysis (MCA) (Heinrich and Rapoport 1974; Kacser and Burns 1973) and Biochemical Systems Theory (BST) (Savageau 1976).

The MCA, linlog and BTS approaches have been successfully used to improve control architectures in metabolic reaction networks (Hatzimanikatis et al. 1996ab). Their method has been recently generalized by Vital-Lopez et al. (2006) to include knockouts by proposing a general computational procedure determining which genes/enzymes should be eliminated (i.e. knocked out), repressed or overexpressed to maximize the metabolic flux of interest. In that work, automatic linearizations according to basic function approximations for arbitrary kinetic representations are combined with binary variables introduced to remove reactions from the linearized kinetic model. Insightful ‘universal’ perturbation methods were developed to increase desired concentrations and fluxes within complex networks (Kacser and Acerenza 1993; Small and Kacser 1994). Based on this universal approach, a conception of group flux and concentration control coefficients was introduced and then used for the optimal selection of small subsets of key enzymatic reactions with the maximum impact on the targeted flux (Stephanopoulos and Simpson 1997). The metabolic design analysis based on the moiety conservation information has been recently discussed in the context of genome-scale conserved moiety pools spanning many metabolic subsystems (Nikolaev et al. 2005).
Although the discussed analyses are based on important approximations (Heijnen 2005) of inherently nonlinear metabolic pathways, genetic manipulations can however cause metabolic networks to deviate significantly from the original stationary state. Besides, both stoichiometric flux balance (FBA) and MCA do not provide any means to check the stability of predicted optimal states. Simultaneous or consecutive modifications of many reaction steps can often lead to unpredictable instabilities and, therefore, can be experimentally infeasible (Stephanopoulos and Simpson 1997). In response to these limitations and challenges, a number of research groups have undertaken the development of plausible large-scale kinetic models. Prominent modeling projects include large-scale kinetic models of *Saccharomyces cerevisiae* (Hynne et al. 2001; Rizzi et al. 1997) and *Escherichia coli* (Chassagnole et al. 2002; Visser et al. 2004), the kinetic model of central metabolism coupled with tryptophan gene expression in *E.coli* (Schmid et al. 2004), and computer models of single cells capturing dynamic effects of chromosome replication and changes in the cell geometry (Atlas et al., 2008; Castellanos et al. 2007; Domach et al. 1984; Nikolaev et al. 2006). The discussed models are highly nonlinear, stiff and include many kinetic parameters which are hard to identify from the measurements. To reduce the model stiffness and to minimize the number of kinetic parameters, an important advantage in the current problem of parameter identifiability, rigorous mathematical and computational frameworks have been recently developed (Gerdtzen et al., 2004; Nikerel et al., 2009). Different approximation frameworks require the development of powerful algorithms allowing for the evaluation of information contents and predictive capabilities of such approximations by translating biochemical models from one kinetic format to another (Hadlich et al., 2009).

While the discussed optimization-based frameworks essentially utilize similar concepts based on the traditional linearization and linlog approximation, combined with a Mixed Integer
Linear Programming (MILP) approach, the goal of this work is substantially different. For the first time, we introduce a large-scale nonlinear optimization computational approach which 1) can be applied to the original kinetic model without any preliminary model simplification such as linearization and size reduction which could result in potential loss of important (e.g. nonlinear dynamics) information; 2) utilizes a more general nonlinear optimization formulation; 3) includes explicitly a stability constraint to detect unstable optimal solutions and the type of bifurcation; and 4) allows for the straightforward incorporation of additional important context-dependent biophysical nonlinear constraints. The developed approach is based on a general Mixed Integer Non-Linear Problem (MINLP) formulation to compute automatically which enzyme levels should be modulated and which enzyme regulatory structures should be altered to achieve the given optimization goal. To solve a complex MINLP formulation, a hybrid deterministic-stochastic computational method is used. Specifically, a stochastic simulated annealing is employed to navigate through the discrete space of enzyme levels and regulatory structures, while a sequential quadratic programming method is utilized to identify optimal enzyme levels and regulatory kinetic parameters. The framework is demonstrated on a well-established kinetic model of the E. coli central metabolism (Chassagnole et al. 2002), used for the optimization of the glucose uptake through the phosphotransferase transport system (PTS) and serine biosynthesis.

2. Mathematical Model

A mathematical model of relevant processes in cellular metabolism and protein machinery can be represented in the form (Laffend and Shuler 1993; Mauch et al. 2001; Schmid et al. 2004).
Stable optimization of kinetic models

\[
\frac{dC_i}{dt} = \sum_{j=1}^{M} S_{ij} r_{j_{\text{max}}} (C, R, K), \quad i \in \mathcal{N}.
\]  

(1)

Here \( C_i \) is the concentration of metabolite \( i \), \( S_{ij} \) is the stoichiometric coefficient of metabolite \( i \) in reaction \( j \) with rate \( r_{j_{\text{max}}} (C, R, K) \), \( r_{j_{\text{max}}} \) is the maximal specific rate of reaction \( j \); \( C \) is the vector of metabolite concentrations, \( R \) is the vector of all regulatory parameters (e.g. allosteric parameters), and \( K \) is the vector of all other kinetic parameters (e.g. Michaelis-Menten constants etc.). The index sets \( \mathcal{N} = \{1, \ldots, N\} \), \( \mathcal{M} = \{1, \ldots, M\} \), and \( \mathcal{R} = \{1, \ldots, R\} \) correspond to all metabolites, reactions, and regulatory parameters, respectively.

Let \( e_j \) be the level of the enzyme catalyzing reaction \( j \). We introduce enzyme level/regulation subset \( S_D \) of size \( D \), \( S_D = E_L \cup R_Q \), which is constructed of the two subsets, \( E_L \) of modulated enzyme levels and \( R_Q \) of altered regulatory parameters. \( E_L \) corresponds to a subset of enzyme levels \( (e_{j_1}, \ldots, e_{j_L}) \) or, equivalently, maximal specific rates, \( (r_{j_1_{\text{max}}}, \ldots, r_{j_L_{\text{max}}}) \), \( E_L = \{j_1, \ldots, j_L\} \), \( E_L \subseteq \mathcal{M}, \quad L \leq M \). Set \( R_Q \) corresponds to a subset of regulatory parameters \( (k_{\theta_1}, \ldots, k_{\theta_Q}) \), \( R_Q = \{\theta_1, \ldots, \theta_Q\}, \quad R_Q \subseteq \mathcal{R}, \quad Q \leq R, \quad D = L + Q \). Following these definitions, we use model (1) for the optimal selection of set \( S_D \) such that the best possible reaction rate \( r_{j_0_{\text{max}}} (C, R, K) \) can be achieved for targeted enzyme \( j_0 \) of interest.

Since detailed mechanistic equations describing dynamic changes in all enzyme levels \( e_j \) are not available, reasonable context-dependent modeling assumptions and ‘coarse-grained’ approximations are necessary. We follow a general approach (Chassagnole et al. 2002; Mauch et al. 2001), accounting for homeostasis and limited protein biosynthesis machinery in the cell. To capture the cell’s limited protein biosynthesis efforts, constraint (2) is used,
Stable optimization of kinetic models

\[ \frac{1}{M} \sum_{j=1}^{M} r_j^{\max} = 1. \]  

(2)

Constraint (2) can be interpreted as a stress-related restriction for allowable genetic manipulations that could otherwise potentially lead to crowding of overexpressed enzymes. Due to the homeostasis condition, allowable concentration changes relative to the original stationary concentrations \( C_0 \) should also be restricted,

\[ \frac{1}{N} \sum_{i=1}^{N} \left| \frac{C_i - C_0^i}{C_0^i} \right| \leq \delta. \]  

(3)

Equation (1), constraints (2) and (3) are still, by themselves, insufficient to describe coordinated changes in both modulated and non-modulated enzyme levels. In the previous optimization-modeling studies (Schmid et al. 2004; Vital-Lopez et al. 2006), this issue is approached by assuming that the levels of all non-modulated enzymes can be kept unchanged. However, experimental observations show that overexpression of even one enzyme can already lead to a significant decrease in all other enzyme levels as reported, for example, for \textit{Zymomonas mobilis} (Bakker et al. 1995). Based on this observation, we introduce constraint (4) to generalize a complex systems response of the protein-synthesizing system (PSS) as described below in detail,

\[ \frac{r_j^{\max}}{r_j^{\max,0}} = \ldots = \frac{r_{j_k}^{\max}}{r_{j_k}^{\max,0}} = \gamma. \]  

(4)

Here \( j_1, \ldots, j_K \) are the indices of non-modulated enzymes, \( K = M - L \), \( L \) is the number of modulated enzymes, and \( \gamma \) is a proportionality coefficient chosen for simplicity uniformly for all non-modulated enzymes as discussed below. Coefficient \( \gamma \) can be calculated from (2) and (4) as
Stable optimization of kinetic models

\[
M - \sum_{s=1}^{L} f_{j_s}^{\text{max},0} = \gamma \cdot \frac{M - L}{M - L}.
\]

(5)

The control of PSS, consisting of ribosomes, mRNA, tRNA, and the enzymes that make and modify these molecules is very complex. Therefore, additional modeling assumptions are required to formalize the complex processes. Following (Segre et al. 2002), we derive the approximation of the adjustments for the non-modulated enzyme levels (4) by assuming that the perturbed cell attempts to preserve the ratios between non-modulated enzyme activities equal or close to those established at the reference non-perturbed ‘wild-strain’ cell,

\[
\frac{r_{j_p}^{\text{max}}}{r_{j_p}^{\text{max},0}} \approx \frac{r_{j_q}^{\text{max}}}{r_{j_q}^{\text{max},0}}.
\]

(6)

Here, \( p \neq q \), and \( p \) and \( q \) correspond to non-modulated enzyme levels. Eq. (6) may be interpreted as corresponding to important cellular function when it is the ratio of specific rates and not their absolute magnitudes may be important and preserved (Browning and Shuler 2001). Using simple algebraic manipulations, (6) can be transformed to the equivalent form

\[
\frac{r_{j_p}^{\text{max}}}{r_{j_p}^{\text{max},0}} \approx \frac{r_{j_q}^{\text{max}}}{r_{j_q}^{\text{max},0}}.
\]

(7)

Using \( \frac{r_{j_p}^{\text{max}}}{r_{j_q}^{\text{max},0}} = \gamma \) in (7) and then approximating (7) by the exact equality, we obtain

\[
\frac{r_{j_p}^{\text{max}}}{r_{j_p}^{\text{max},0}} = \frac{r_{j_q}^{\text{max}}}{r_{j_q}^{\text{max},0}} = \gamma.
\]

(8)

We note that (4) is equivalent to (8). When more molecular details on the complex protein machinery become available, (4) can be replaced by more accurate mechanistic equations without affecting the presented modeling framework.
Model (1) incorporates 30 enzymes and 13 enzyme regulatory structures shown in Table 1 and Figure 1. Because enzyme regulation can be extremely complex by encompassing many kinetic parameters (Chassagnole et al. 2002), we have implemented 13 additional dimensionless ‘generalized regulatory’ parameters \( \alpha_1, \ldots, \alpha_{13} \) to activate or disable the entire regulatory structures. The modified model includes 43 optimization parameters \( (p_1, \ldots, p_{43}) \), \( (p_1, \ldots, p_{43}) = (r_{i_{\max}}^{\max}, \ldots, r_{30}^{\max}, \alpha_1, \ldots, \alpha_{13}). \) The simplest example is phosphoglucoisomerase (PGI) inhibition by 6-phosphogluconate (6pg),

\[
\begin{align*}
    r_{\text{PGI}} &= \frac{r_{\text{PGI}}^{\max} \left( C_{\text{g6p}} - \frac{C_{\text{f6p}}}{K_{\text{PGI, eq}}} \right)}{K_{\text{PGI, g6p}} \cdot (1 + \frac{C_{\text{f6p}}}{K_{\text{PGI, f6p}}} + \alpha_2 \cdot \frac{C_{\text{6pg}}}{K_{\text{PGI, 6pg, 6pginh}}}) + C_{\text{g6p}}}.
\end{align*}
\]

Here parameter \( \alpha_2 \) is introduced to alter the regulatory properties of PGI mediated by the concentration of 6-phosphogluconate (i.e. \( C_{\text{6pg}} \)). Specifically, the unit value of \( \alpha_2 \) (i.e. \( \alpha_2 = 1 \)) corresponds to the original enzyme regulatory properties, while altered values, \( \alpha_2 < 1 \) or \( \alpha_2 > 1 \), correspond to decreased or increased regulation of PGI by \( C_{\text{6pg}} \), respectively.

### 3. Computational Method

#### 3.1. Mixed Integer-Non-Linear Problem (MINLP) formulation

To select alternative optimal targets for practically feasible modulations of enzyme levels and genetic mutations of regulatory properties (e.g., inhibitory affinities), optimal solutions of the following Mixed Integer Non-Linear Problem (MINLP) are calculated,
\[
\begin{align*}
\text{maximize} & \quad r_{j_0} \left( r_{j_0}^{\text{max}}, C, R, K \right) \\
\text{subject to} & \quad \sum_{j=1}^{M} S_{ij} \cdot r_j \left( r_j^{\text{max}}, C, R, K \right) = 0, \quad i \in \mathcal{N} \\
& \quad \frac{r_{j_{\text{max},0}}}{r_{j_{\text{max},0}}} + \cdots + \frac{r_{j_{K}}}{r_{j_{K}}} + K \cdot \gamma = M \\
& \quad r_{j_i}^{\text{max}} = \gamma \cdot r_{j_i}^{\text{max,0}}, \quad s = 1, \ldots, K \\
& \quad \frac{1}{N} \sum_{i=1}^{N} \left| C_i - C_i^0 \right| \leq \delta \\
& \quad \max_i \text{Re} \lambda_i \leq -\lambda_0 < 0
\end{align*}
\]

The first constraint in (10) corresponds to the steady state in (1). The second and third constraints result from the protein limited biosynthesis constraint (2) and the conditions for non-modulated enzymes (4) and (5), respectively. The forth constraint is the homeostasis condition (3). The fifth constraint enforces the optimal solution stability property. Here Re \( \lambda_i \) is the real part of eigenvalue \( \lambda_i \) calculated from the linearization of the right-hand side of equation (1) at the steady state concentrations, and \( \lambda_0 \) is an appropriate small positive number. Formulation (10) allows for the simultaneous elucidation of two optimal subsets \( E_L \) (i.e. a subset of modulated enzyme levels) and \( R_Q \) (i.e., a subset of altered regulation parameters), \( S_D = E_L \cup R_Q \), such that the best possible targeted rate \( r_{j_0} \left( r_{j_0}^{\text{max}}, C, R, K \right) \) can be achieved. In (10), the indices of enzyme levels and regulatory parameter correspond to integer variables, while the magnitudes of the enzyme levels and regulatory parameters correspond to continuous variables.

### 3.2. Computational implementation

The optimization framework and MINLP formulation (10) have been demonstrated on a large-scale nonlinear model of central carbon metabolism for a glucose-limited culture of \( E. coli \) (Chassagnole et al. 2002). The model is comprised of 30 enzymes and 17 metabolites with the
objective of maximizing fluxes through the phosphotransferase system (PTS) and serine production reaction (see Figure 1). Initial reference values of non-perturbed specific maximal rates (i.e., \( r_{\text{max},0} \)), kinetic parameters, and steady state metabolite concentrations (i.e. \( C_0 \)) have been used as suggested in (Chassagnole et al. 2002).

To solve (10) numerically, we have implemented a simulated annealing (SA) algorithm (Kirkpatrick et al. 1983) to navigate through the discrete space of enzyme levels \( M \) and regulatory parameters \( R \). In our approach, subsets \( S_D \) of increasing fixed sizes \( D, D = 1, \ldots, 10 \), are separately investigated, where \( D = L + Q \). \( L \) and \( Q \) are the numbers of modulated enzyme levels and altered regulatory parameters, respectively. For each randomly selected \( S_D \), the optimal values of specific reaction rates \( (r_{\text{max},1}, \ldots, r_{\text{max},L}) \) and regulatory parameters \( (\alpha_{\theta_1}, \ldots, \alpha_{\theta_Q}) \) are calculated by utilizing gradient-based algorithms (e.g. an SQP-algorithm). To evaluate rate \( r_{j_0}(r_{j_0}^\text{max}, C_{\text{ss}}, R, K) \), steady state concentrations \( C_{\text{ss}} \) are calculated using Newton-based solvers. Because Newton-based solvers can converge to both stable and unstable solutions, the stability of \( C_{\text{ss}} \) is investigated by computing the eigenvalues of the Jacobian matrix readily available from such solvers.

Random multistarts have been used to check the robustness of the SQP search and no alternative global optima was found as earlier reported by Visser et al. (2004). Tight absolute (i.e. abs = 10^{-11} - 10^{-9}) and relative (i.e. rel = 10^{-9}-10^{-7}) tolerances have been enforced to keep integration errors low due to the enormous ‘stiffness’ of the model. Also, \( \delta = 0.1 \) is used in (3) for an allowable 10% change in the metabolite concentrations to preserve stability properties. The optimization modeling framework is implemented in Matlab© on a Linux cluster with Intel CPU 3.06 GHz computers. Typical computational requirements are in order of minutes for small
Stable optimization of kinetic models

$S_{L+Q}$ (i.e. with $L + Q \leq 3$), up to 30-40 hours for large enzyme subset $S_L$ (i.e. with $Q = 0$), and up to 60-90 hours for large enzyme level/regulation subsets $S_{L+Q}$ (i.e. with $Q > 0$).

4. Results and Discussion

4.1. Comparative analysis of optimally selected subsets

The best found flux ratios for PTS and serine (i.e. SerSynth) rates in the optimized and original models, respectively, are presented in Table 2 for the following three distinct cases: (i) all 13 generalized regulatory parameters are altered, while all enzyme activities are kept unchanged, (ii) all 30 enzyme levels are modulated, while regulatory parameters are kept fixed, and (iii) the maximum possible number of enzyme levels and regulatory parameters is manipulated until the pathway stability is lost. We find from Table 2 that the alteration of enzyme regulation alone does not lead to any significant improvement in the targeted fluxes. This can be explained by a limited capacity of an enzymatic reaction to ‘channel’ a large flux without a substantial increase in the enzyme specific activity. At the same time, no further impressive increase in the serine flux has been observed compared to the modulation of enzyme levels alone (see Table 2). These observations are biologically meaningful since PTS is a tightly regulated enzyme, while the SerSynth reaction lacks any kind of pathway regulation (see Figure 1).

The analysis of small optimal enzyme level/regulation subsets leads to the following facts and conclusions. First of all, substantial improvements in the desired fluxes can be predicted by manipulating only small enzyme level subsets (see Figure 2a and 3a). We find that no substantial increase in the desired PTS and serine flux can be obtained for the best mixed enzyme/regulation subsets of small sizes (i.e., $D = 1, 2, \text{ and } 3$) (solid circles in Figure 2a and 3a, respectively), compared to enzyme level modulations alone (white circles in Figure 2a and 3a, respectively).
For larger mixed enzyme level/regulation choices, a significant additional flux increase can be predicted, though at the expense of the loss of stability properties due to the attenuation of negative feedbacks, evolved to ensure the robustness of cellular systems, and the amplification of positive feedbacks always destabilizing the cellular system. Specifically, we could not find any stable optimal steady state solution for the serine overproduction by navigating enzyme level/regulation subsets of size 8, 9 and 10, where the saddle-node bifurcation (Kuznetsov 2004) leading to the disappearance of steady state solutions was detected. This conclusion is supported by plotting the values of the leading real eigenvalue that sharply tends to the zero level around the subset size 7 as depicted in Figure 2b. Beyond the bifurcation point, solutions to (1) became unbounded violating homeostasis constraint (3). The disappearance of the pathway stable steady state can be interpreted by the loss of the balance between the increased large demand for the serine biosynthesis and the limited activity of the cellular system which cannot support the non-physiologic product demands. This observation is in line with the bifurcation analysis of the model (Vital-Lopez et al., 2006). Despite the importance of the situation for the dynamic optimization, we have not studied the dynamics of the system in detail after the loss of stability. Indeed, the model and its parameters are carefully selected to study stationary regimes only. Specifically, the concentrations of all co-factors, ATP, ADP, AMP, NAD, NADH, NADP and NADPH are kept fixed under stationary physiological conditions (Chassagnole et al. 2002). Therefore, freeing the co-factor concentration is needed to study the dynamic regimes beyond the model applicability condition limited to stable steady states.

The Hopf bifurcation giving rise to small amplitude stable periodic oscillations has been encountered in the case of the PTS optimization for mixed enzyme level/regulation subsets of size 8, 9, and 10. Plotting the real parts of the leading eigenvalue has revealed multiple solution
branches and their bifurcations as depicted in Figure 3. We observe that for the mixed enzyme-regulation subsets, the corresponding optimal steady-state solution loses its stability via a saddle-node bifurcation when the subset size is 5. However, another branch of optimal steady-state solution does exist when the subset size is 6. The latter steady-state solution branch then loses its stability via a Hopf bifurcation when a couple of complex-conjugated pairs of eigenvalues crosses the imaginary axis in the complex plane (Kuznetsov 2004). The steady-state solution does not disappear at the Hopf bifurcation point and, instead, becomes unstable for the cases of the mixed subsets of size 8, 9, and 10. Here, a branch of stable two-dimensional small amplitude sinusoidal limit cycles emanates from the critical steady-state at the bifurcation moment.

Although the stability of limit cycles can be practically checked by direct integration, we note that the stability of the periodic solutions near the Hopf point can also be systematically determined by calculating appropriate nonlinear terms in the Taylor series at the critical steady state (Khazin and Shnol, 1991; Kuznetsov, 2004). The appearance of oscillations can be attributed to the well-known autocatalytic properties of glycolysis (Heinrich and Schuster 1996), enhanced by the increased glucose uptake through the optimized PTS. Similarly to the bifurcation analysis done in (Vital-Lopez et al., 2006), only small-amplitude periodic regimes existing within small parameter regions have been identified and no large-amplitude regime has been found from the stochastic search. This model’s inability to generate large-amplitude oscillations can again be attributed to the model intrinsic properties limited to the steady-state conditions by freezing the co-factor concentrations.

Although the model utilized in this study cannot be used to suggest dynamic optimal engineering strategies far from the reference steady-state, the observed small-amplitude periodic regimes can provide important insights into the potential of oscillatory optimal states compared
Stable optimization of kinetic models

to the optimal steady-state conditions. To compare optimal small-amplitude periodic and steady-state predictions, we have averaged the solutions along the observed periodic solutions and plotted the data on Figure 2a. We observe a noticeable increase in the optimal averaged PTS rate for the small-amplitude periodic processes compared with the stead-state case. Given the fact that the annual production of amino-acids worldwide is approximately two million metric tons (Pharkua et al., 2003), the observed increase in the PTS rate can lead to a significant improvement in certain amino-acid production, when recalculated using absolute values. We emphasize that these preliminary studies require further analysis that goes beyond the scope of this work as summarized in the conclusion and outlook section below.

In all cases considered, the calculations demonstrate a saturation type of the optimal behavior for the entire reaction network (see Figure 2 and 3) due to the fixed external conditions and the imposed protein limiting machinery constrains (2) - (4). This means that while more efforts can be required to elucidate larger stable optimal enzyme level/regulation choices, no further significant achievement in the corresponding targeted flux can be obtained and therefore such choices may not be worth the effort.

4.2. Calculation of control coefficients

To obtain quantitative insights into how successive small enzyme level/regulation subsets can be selected to meet overproduction requirements, MCA can be used to calculate sensitivity and control coefficients (Stephanopoulos et al. 1998). Taking into account that both enzyme levels and kinetic regulatory parameters can significantly contribute to the targeted pathway flux \( J \), the following coefficients have been evaluated, (i) the flux control coefficient (FCC),
\[ C_r^J = \frac{\partial \ln J}{\partial \ln r^\text{max}}, \]  

\[(11)\]

(ii) the flux response coefficient (FRC) with respect to a regulatory parameter (i.e. \( \alpha \)),

\[ R_a^J = \frac{\partial \ln J}{\partial \ln \alpha}, \]  

\[(12)\]

(iii) the enzyme elasticity coefficient (EEC) quantifying the potential of parameter \( \alpha \) to affect the individual reaction rate \( r \) under isolated conditions,

\[ \pi_a^r = \frac{\partial \ln r}{\partial \ln \alpha}. \]  

\[(13)\]

Coefficients \( R_a^J \), \( C_r^J \), and \( \pi_a^r \) satisfy the simple identity (Heinrich and Schuster 1996),

\[ R_a^J \equiv C_r^J \cdot \pi_a^r. \]  

\[(14)\]

Using a log-log finite-difference approximation of (11) - (13), identity (14) has been numerically checked to ensure the robustness of all numeric calculations in the stiff model (see Table 3). The calculated control coefficients reveal a distributed control on the targeted fluxes, allocated within a group of several rate-limiting steps exerting the highest control as seen from Figure 3. Namely, the same group of rate limiting steps (i.e., PTS, PFK, GAPDH, PDH, PEPxylase, and G6PDH) is identified for potential practical enzyme level modulations and regulatory structures genetic mutations for both cases of the PTS and serine optimizations. Since the control coefficients are readily available from the measurements (Stephanopoulos et al. 1998), we will compare local MCA-based predictions with those obtained from the nonlinear optimization framework.

4.3. The PTS rate optimization results

We begin with the discussion of the best enzyme level choices presented in the left side of Table 4. First, the best enzyme level choices are in most cases found to be in a good agreement with the
MCA-based predictions. However, the detailed analysis of Table 4 reveals that both local FCC-based and nonlinear optimization predictions lack the additivity property in a sense that the best enzyme choices alone cannot be combined with one another to significantly improve the PTS rate. For example, the triplet of enzymes PTS, phosphofructokinase (PFK), and pyruvate dehydrogenase (PDH) exerting the highest total control (see Figure 3) is absent from Table 4. These enzymes are present in all larger subsets (i.e., with \( D \geq 5 \)). Importantly, while MCA suggests decreasing the level of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) exerting high negative control (see Figure 3), the level of this enzyme is increased in all nonlinear optimization studies (see Table 4). This observation is in line with the previous optimization study (Visser et al., 2004). Therefore, even for a rate limiting step with a high control coefficient, the direction of the corresponding enzyme level modulation cannot be solely predicted by MCA alone. We also find from Table 4 that almost half of all modulated enzymes present in larger subsets are near-equilibrium enzymes exerting negligibly small control coefficients. Note that which of such near-equilibrium enzymes should be modulated cannot be predicted from MCA.

The optimization results also show how the best enzyme level choices emerge. Although, the best choices lack the additivity property, the best smaller choices repeatedly enter the best larger subsets (see Table 4). This means that control of flux does not shift between different groups of enzymes due to the compensating effects of global regulation and homeostasis. The absence of the shift in distributed control additionally emphasizes the importance of reaction steps with high values of control coefficients for enzyme subsets of different size.

Enzymes with large flux control coefficients are not always the ones to be modified, especially if they are involved in feedback control loops. It may be the removal or attenuation of certain negative-feedback loops that should be considered and not the amplification of the
activity of the corresponding enzymes (Bailey 1991; Hatzimanikatis et al. 1996b; Stephanopoulos and Vallino 1991). These and similar important theoretic predictions are exemplified by the optimal selection of the best mixed enzyme level/regulation choices, presented in the right side of Table 4. Importantly, the regulatory properties (see Table 1) and not enzyme levels of all three tightly regulated enzymes, PTS (i.e. \texttt{PTS}\textsubscript{g6p(-)}), PFK (i.e. \texttt{PFK}\textsuperscript{amp(+)}, and PEPxylase (i.e. \texttt{PEPxyrase}\textsuperscript{fdp(+)}, have been chosen to alter, while their levels were automatically adjusted due to the limited protein machinery constraints (4) and (5). These nonlinear optimization observations are also found in a good agreement with calculations of response coefficients presented in Table 3. For example, the amplification of the positive regulation of PFK by amp (i.e. \texttt{PFK}\textsuperscript{amp}) with FRC = 0.065 is more preferable than the amplification of the positive feedback by adp (i.e. \texttt{PFK}\textsuperscript{adp}) with FRC = 0.00605. Although many enzymes are present in both left and right sides of Table 4, different manipulations of these enzymes are automatically suggested by the computational procedure, based on the presence or absence of regulatory structures in the selected enzymes.

Figure 5 further compares flux control coefficients and relative optimal enzyme levels, calculated for the six best enzyme choices selected to increase the PTS flux (i.e. \texttt{PTS}, \texttt{PFK}, TIS, \texttt{GAPDH}, \texttt{PDH}, and \texttt{PEPxyrase} in Table 4). The corresponding distributions of steady state fluxes are shown in Figure 6. The levels of the modulated enzymes (see Figure 5(b)) are for the most part proportionally changed accordingly to the changes in their flux control coefficients. Specifically, the change in the sign of the FCCs for GAPDH (i.e., negative values are changed to positive values) correctly predicts a substantially increased level of the enzyme when both the enzyme levels and the regulation of all selected enzymes are allowed to vary.
4.4. The serine optimization results

Similar results have been obtained for the optimization of the serine production flux as shown in Table 5. We note that small enzyme/regulation choices are intuitive as the PTS (i.e. PTS(+) in Table 5) supplies metabolism with the initial substrate, while SerSynth (i.e. SerSynth(+) in Table 5) leads to the final serine production. Similarly to the case of the PTS flux optimization, larger enzyme level/regulation choices encompass enzymes with both high and negligibly low values of FCCs.

Because of the importance of the robustness and stability issues for application, we have chosen two intermediate cases of (i) the best six enzyme levels PTS(+), PFK(+), GAPDH(+), TIS(-), PDH(+), SerSynth (+), and (ii) the best mixed enzyme level/regulation choices, PTS↓g6p(-), PFK↑amp(+), GAPDH(+), TIS(+), PDH(+), SerSynth (+), to discuss the optimal enzyme activities and regulatory properties in more details (see Table 5). Recall that these cases already provide an increase in the serine production of about 80% of the best predictions shown in Figure 2(a) (white and solid circles). The optimal distributions of the steady state fluxes allocated within the pathway towards the serine overproductions are shown in Figure 8. We find that an increase in the serine demand (see Figure 7(b)) reallocates the strength of metabolic control from the serine synthesis (SerSynth in Figure 7(a)) towards the supply block (i.e. PTS) and the pyruvate removal block (i.e. PDH) (see Figure 7(a)).

Although the nonlinear optimization observations are in many cases found in a good agreement with calculations of response coefficients (RFC) presented in Table 3, there are several discrepancies. For example, the amplification of the positive regulation of PFK by amp (i.e. PFK↑amp) with small FRC = 0.0376 calculated in Table 3 has proved to be more preferable
than the attenuation of the negative feedback by pep (i.e. PFK↓pep as in Table 1) with FRC = -0.287 of much larger magnitude.

The optimization results do confirm the importance of high flux control coefficients, estimated at the reference (‘wild-type’) strain which in many cases correctly delineate the most important blocks of central metabolism from less important subordinate pathways. Specifically, flux is increased through phosphotransferase system (PTS), phosphofructokinase (PFK), a committed enzyme in the network, and pyruvate dehydrogenase (PDH) required to removing an excess of pyruvate accumulated through the enhanced PTS. Comparing the best enzyme choices for the optimal PTS rate and serine flux, we come to a very important observation that in both cases the best choices emerge in a similar fashion signifying a common trend in the selection of candidate enzyme/regulation subsets, where the best smaller choices repeatedly enter the best larger subsets. This property can be apparently attributed to the homeostasis condition and negative feedback which attempt to stabilize the system. However, more research is needed to support or disprove this intuitive explanation.

5. Conclusions and outlook

A general stable hybrid deterministic-stochastic nonlinear optimization and modeling framework for the optimal selection of enzyme levels and regulatory structures using dynamic kinetic models of cellular systems has been introduced, and the corresponding computational method has been for the first time demonstrated on a large-scale nonlinear kinetic model of central carbon metabolism of *E. coli* without any preliminary model simplification. Computational results show that the modification of all enzyme levels and regulatory properties leads to a stable 8-fold increase in the PTS uptake rate and a stable 22-fold increase in the serine
biosynthesis rate. Substantial improvements can be predicted by manipulating only small subsets of enzyme levels and regulatory structures due to the saturation property of metabolism limited by the protein synthesis machinery. For example, the modulation of six enzyme levels already leads to a flux increase of 80% of the best predicted serine flux. This means that while more efforts can be required to elucidate larger stable optimal enzyme level/regulation choices, no further significant increase in the corresponding targeted fluxes can be obtained and therefore such choices may not be worth the effort due to the computationally predicted loss of stability properties via the Hopf or saddle-node bifurcation. We have also observed a strong synergism between the redesign of tightly regulated enzymes (e.g. phosphofructokinase) and the overexpression of those enzymes that lack regulation (e.g. glyceraldehyde-3-phosphate dehydrogenase). The obtained nonlinear optimization results are contrasted with respect to local-linear predictions obtained from the well-established metabolic control analysis (MCA).

Although the implemented stochastic simulated annealing approach does not guarantee the convergence to the global optimal solutions and better choices can still be found, the solutions discussed, which have been found from a very extensive search, already provide valuable candidate enzyme/regulation choices. Such choices can be used in prioritizing theoretic and practical studies of important properties of enzymatic reactions, kinetic regulatory structures, and providing a systematic framework for designing experiments to better understand regulation of cellular function. The framework can also be used as a powerful modeling tool for the direct computational validation of context-dependent theoretic assumptions allowing the modeler to better understand how the biological mechanisms give rise to biological function. Based on this, the modeler can then devise experiments to test the model’s predictions of relevant modes of cellular function. The framework also allows for the bifurcation analysis of the critical cases
where the stability of the optimized regimes can be lost. Such cases additionally emphasize the complexity of large-scale optimization studies and the importance of the careful selection of stable small subsets of enzyme levels and regulatory structures for their subsequent practically feasible alterations.

Originally highly nonlinear and (linlog) approximated models have been carefully compared in a number of previous studies (Visser et al., 2004; Vital-Lopez et al., 2006). Under steady-state conditions, full models and their linlog approximations lead to similar conclusions. However, contrasting to the previous steady-state optimization studies, we have made an important observation that even small amplitude optimal periodic regimes computed using full models can lead to a better prediction compared to those obtained for steady-states cases. It is important to note that the idea to use periodic regimes in bioengineering and biotechnology is not radically novel and was introduced and discussed long time ago in the context of frequency responses (Douglas, 1967; Bailey, 1973; Hatzimanikatis et al., 1993; Hatzimanikatis and Bailey, 1996). To this end, simulation of sustained periodic DNA replication in the engineered bacterial strain can also be used as an important biological criterion of the feasibility of the corresponding metabolic engineering strategies and interventions (Atlas et al., 2008).

We emphasize that our study of optimal dynamics regimes requires further theoretical and practical analysis that goes beyond the scope of this work. Specifically, as it has been earlier found by Vital-Lopez et.al. (2006), the observed periodic processes exist in the model within small parametric ranges and are very close to the boundary of the total stability loss by the system. Therefore, a more detailed study of practically-feasible optimal periodic states necessitates, at least, (i) an adequate modification of the model to include dynamic cofactors, and (ii) the utilization of the global stability constrains based on Lyapunov functions, which could be
used to estimate the geometric size of the attraction basins of the corresponding stationary and periodic states. In this respect we note that the idea of including stability constraints into an optimization formulation is now standard in control theory and, in particularly, in the model predictive control, and it goes well beyond the eigenvalue-type constraint used in formulation (10), which can only guarantee local stability. To this end, modern control approaches includes stability constraints based on Lyapunov (energy)-type functions (Mhaskar et al., 2005, 2006). Therefore, these and similar well-established and new ideas from the modern control theory need to be further employed in optimization modeling studies to suggest a tighter control of the genetically altered cellular system operation and functioning near the loss of stability properties under more realistic conditions in the bioreactor including spatiotemporal uncertainties such temperature fluctuation.

Another important aspect of optimization-based modeling studies employing large-scale kinetic models is in the typical stiffness of the corresponding differential equations during the solution process. An important rigorous approach to reduce the model stiffness and to improve the computational performance while maintaining the integrity of the final results has been recently suggested (Gerdtzen et al., 2004). Because in many practical situations, the model’s stiffness can significantly complicate the model integration without any noticeable effect on the final results, an alternative approach to reduce the stiffness in the original (i.e. non-reduced) model could be the introduction of a small bias into parameter fitting procedures that would penalize the selection of parameter values leading to stiffness in the model nonlinear behavior (Brown and Sethna, 2003).

Finally, we hope that as soon as more information on enzyme regulation becomes available due to the emergence of powerful inference methods like the double regulation method
allowing for the practical estimation of enzyme elasticities from experimental measurements (Link and Weuster-Boltz 2007), the discussed optimization framework and modeling strategy, when coupled with emerging synthetic biology methods (Elowitz and Leibler, 2000; Fung et al., 2005; Wong and Liao, 2006), will have a broader implication in directing practically feasible metabolic manipulations of small numbers of key cellular functions to achieve a targeted metabolic engineering objective.

**Acknowledgments.** The author is very grateful to Prof. Reuss and Dr. Schmid for providing me with the model SBML representation. I appreciate very much the insightful suggestions of three anonymous reviewers. I would like to thank Dr. Burgard and Dr. Pharkya for stimulating discussions of the modeling aspect of the work. I am very thankful to Prof. Armaou for the discussions of dynamical systems approaches. I am especially grateful to Prof. Maranas for helping me better understand various optimization techniques and Prof. Shuler for many discussions of synthetic and *E. coli* biology.

**Nomenclature**

*Enzymes*

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALDO</td>
<td>aldolase</td>
</tr>
<tr>
<td>DAHPS</td>
<td>DAHP synthases</td>
</tr>
<tr>
<td>ENO</td>
<td>enolase</td>
</tr>
<tr>
<td>G1PAT</td>
<td>glucose-1-phosphate adenylytransferase</td>
</tr>
<tr>
<td>G3PDH</td>
<td>glycerol-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>G6PDH</td>
<td>glucose-6-phosphate dehydrogenase</td>
</tr>
</tbody>
</table>
GAPDH  glyceraldehyde-3-phosphate dehydrogenase
MetSynth  methionine synthesis
MurSynth  mureine synthesis
PFK  phosphofructokinase
PGDH  6-phosphogluconate dehydrogenase
PGI  glucose-6-phosphate isomerase
PGK  phosphoglycerate kinase
PGM  phosphoglycerate mutase
PDH  pyruvate dehydrogenase
PEPxylase  PEP carboxylase
PGlucMoM  phosphoglucomutase
PK  pyruvate kinase
PTS  phosphotransferase system
R5PI  ribose-phosphate isomerase
RPPK  ribose-phosphate pyrophosphokinase
Ru5P  ribulose-phosphate epimerase
Synth1  synthesis1
Synth2  synthesis2
TA  transaldolase
TIS  triosephosphate isomerase
TKa  transketolase A
TKb  transketolase B
TrpSynth  tryptophan synthesis
Metabolites

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2pg</td>
<td>2-phosphoglycerate</td>
</tr>
<tr>
<td>3pg</td>
<td>3-phosphoglycerate</td>
</tr>
<tr>
<td>6pg</td>
<td>6-phosphogluconate</td>
</tr>
<tr>
<td>accoa</td>
<td>acetyl-coenzyme A</td>
</tr>
<tr>
<td>dhap</td>
<td>dihydroxyacetonephosphate</td>
</tr>
<tr>
<td>e4p</td>
<td>erythrose-4-phosphate</td>
</tr>
<tr>
<td>f6p</td>
<td>fructose-6-phosphate</td>
</tr>
<tr>
<td>fdp</td>
<td>fructose-1,6-bisphosphate</td>
</tr>
<tr>
<td>g1p</td>
<td>glucose-1-phosphate</td>
</tr>
<tr>
<td>g6p</td>
<td>glucose-6-phosphate</td>
</tr>
<tr>
<td>gap</td>
<td>glyceraldehyde-3-phosphate</td>
</tr>
<tr>
<td>glc</td>
<td>glucose</td>
</tr>
<tr>
<td>oaa</td>
<td>oxaloacetate</td>
</tr>
<tr>
<td>pep</td>
<td>phosphoenolpyruvate</td>
</tr>
<tr>
<td>ppg</td>
<td>1,3-diphosphoglycerate</td>
</tr>
<tr>
<td>pyr</td>
<td>pyruvate</td>
</tr>
<tr>
<td>rib5P</td>
<td>ribose-5-phosphate</td>
</tr>
<tr>
<td>ribu5p</td>
<td>ribulose-5-phosphate</td>
</tr>
<tr>
<td>sed7p</td>
<td>sedoheptulose-7-phosphate</td>
</tr>
<tr>
<td>xyl5p</td>
<td>xylulose-5-phosphate</td>
</tr>
</tbody>
</table>
References


TABLES

Table 1. Enzyme regulation.

<table>
<thead>
<tr>
<th>№</th>
<th>Enzyme</th>
<th>Regulation</th>
<th>Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PTS</td>
<td>inhibition by g6p</td>
<td>PTS↓g6p</td>
</tr>
<tr>
<td>2</td>
<td>PGI</td>
<td>inhibition by 6pg</td>
<td>PGI↓6pg</td>
</tr>
<tr>
<td>3</td>
<td>PFK</td>
<td>inhibition by pep</td>
<td>PFK↓pep</td>
</tr>
<tr>
<td>4</td>
<td>PFK</td>
<td>activation by adp</td>
<td>PFK↑adp</td>
</tr>
<tr>
<td>5</td>
<td>PFK</td>
<td>activation by amp</td>
<td>PFK↑amp</td>
</tr>
<tr>
<td>6</td>
<td>PK</td>
<td>activation by amp</td>
<td>PK↑amp</td>
</tr>
<tr>
<td>7</td>
<td>PK</td>
<td>activation by fdp</td>
<td>PK↑fdp</td>
</tr>
<tr>
<td>8</td>
<td>PK</td>
<td>inhibition by atp</td>
<td>PK↓atp</td>
</tr>
<tr>
<td>9</td>
<td>G1PAT</td>
<td>activation by nadph</td>
<td>G1PAT↑nadph</td>
</tr>
<tr>
<td>10</td>
<td>G6PDH</td>
<td>inhibition by nadph</td>
<td>G6PDH↓nadph</td>
</tr>
<tr>
<td>11</td>
<td>PGDH</td>
<td>inhibition by atp</td>
<td>PGDH↓atp</td>
</tr>
<tr>
<td>12</td>
<td>PGDH</td>
<td>inhibition by nadph</td>
<td>PGDH↓nadph</td>
</tr>
<tr>
<td>13</td>
<td>PEPxylase</td>
<td>activation by fdp</td>
<td>PEPxylase↑fdp</td>
</tr>
</tbody>
</table>
### Table 2. Best flux ratios of steady state fluxes in the optimized and original models.

<table>
<thead>
<tr>
<th>Flux</th>
<th>Regulation</th>
<th>Enzyme Level</th>
<th>Enzyme Level &amp; Regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTS</td>
<td>1.43</td>
<td>3.16</td>
<td>7.31 (8.26 H)</td>
</tr>
<tr>
<td>Serine</td>
<td>1.06</td>
<td>20.59</td>
<td>22.01</td>
</tr>
</tbody>
</table>

Here flux ratio is the ratio of the objective function with the respect to the reference value. H corresponds to the Hopf bifurcation.
Table 3. Control coefficients for regulated enzymes at the original steady state.

<table>
<thead>
<tr>
<th>№</th>
<th>Enzyme</th>
<th>Modifier</th>
<th>Regulation</th>
<th>EEC</th>
<th>FCC (PTS)</th>
<th>FRC (PTS)</th>
<th>FCC (Serine)</th>
<th>FRC (Serine)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PTS</td>
<td>g6p</td>
<td>negative</td>
<td>-0.978</td>
<td>0.416</td>
<td>-0.407</td>
<td>0.192</td>
<td>-0.187</td>
</tr>
<tr>
<td>2</td>
<td>PGI</td>
<td>6pg</td>
<td>negative</td>
<td>-0.551</td>
<td>0.000692</td>
<td>-0.000381</td>
<td>0.000374</td>
<td>-0.000206</td>
</tr>
<tr>
<td>3</td>
<td>PFK</td>
<td>pep</td>
<td>negative</td>
<td>-2.047</td>
<td>0.242</td>
<td>-0.496</td>
<td>0.14</td>
<td>-0.287</td>
</tr>
<tr>
<td>4</td>
<td>PFK</td>
<td>adp</td>
<td>positive</td>
<td>0.025</td>
<td>0.242</td>
<td>0.00605</td>
<td>0.14</td>
<td>0.0035</td>
</tr>
<tr>
<td>5</td>
<td>PFK</td>
<td>amp</td>
<td>positive</td>
<td>0.268</td>
<td>0.242</td>
<td>0.065</td>
<td>0.14</td>
<td>0.0376</td>
</tr>
<tr>
<td>6</td>
<td>PK</td>
<td>amp</td>
<td>positive</td>
<td>0.000226</td>
<td>0.0109</td>
<td>0.00000246</td>
<td>-0.122</td>
<td>-0.0000275</td>
</tr>
<tr>
<td>7</td>
<td>PK</td>
<td>fdp</td>
<td>positive</td>
<td>0.0000682</td>
<td>0.0109</td>
<td>0.00000074</td>
<td>-0.122</td>
<td>-0.00000831</td>
</tr>
<tr>
<td>8</td>
<td>PK</td>
<td>atp</td>
<td>negative</td>
<td>-0.0000544</td>
<td>0.0109</td>
<td>-0.0000006</td>
<td>-0.122</td>
<td>0.00000664</td>
</tr>
<tr>
<td>9</td>
<td>G1PAT</td>
<td>fdp</td>
<td>positive</td>
<td>0.731</td>
<td>0.00721</td>
<td>0.00527</td>
<td>-0.00934</td>
<td>-0.00683</td>
</tr>
<tr>
<td>10</td>
<td>G6PDH</td>
<td>nadph</td>
<td>negative</td>
<td>-0.419</td>
<td>0.115</td>
<td>-0.0483</td>
<td>-0.0721</td>
<td>0.0302</td>
</tr>
<tr>
<td>11</td>
<td>PGDH</td>
<td>atp</td>
<td>negative</td>
<td>-0.012</td>
<td>0.000389</td>
<td>-0.00000464</td>
<td>0.000211</td>
<td>-0.0000025</td>
</tr>
<tr>
<td>12</td>
<td>PGDH</td>
<td>nadph</td>
<td>negative</td>
<td>-0.485</td>
<td>0.000389</td>
<td>-0.000189</td>
<td>0.000211</td>
<td>-0.000102</td>
</tr>
<tr>
<td>13</td>
<td>PEPxylase</td>
<td>fdp</td>
<td>positive</td>
<td>0.019</td>
<td>0.0387</td>
<td>0.00073</td>
<td>-0.126</td>
<td>-0.00238</td>
</tr>
</tbody>
</table>

Here EEC is Enzyme Elasticity Coefficient, FCC is Flux Control Coefficient, and FRC is Flux Response Coefficient, FRC = EEC·FCC (see (10) in the text).
Table 4. Alternative 10 best optimal enzyme/regulation subsets for the increased PTS rate.

<table>
<thead>
<tr>
<th>Size</th>
<th>Enzyme Subset</th>
<th>Flux Ratio</th>
<th>Enzyme/Regulation Subset</th>
<th>Flux Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PTS(+)</td>
<td>1.073</td>
<td>PTS↓g6p(-)</td>
<td>1.080</td>
</tr>
<tr>
<td>2</td>
<td>PTS(+), RPPK/PEPxylose (+)</td>
<td>1.233</td>
<td>PTS↓g6p(-), G1PAT↑nadph(+)</td>
<td>1.462</td>
</tr>
<tr>
<td>3</td>
<td>PTS(+) GAPDH(+) PEPxylose(+)</td>
<td>1.628</td>
<td>PTS↓g6p(-), G1PAT↑nadph(+) GAPDH(+)</td>
<td>2.173</td>
</tr>
<tr>
<td>4</td>
<td>PTS(+), PFK(+) GAPDH(+) PEPxylose(+)</td>
<td>2.246</td>
<td>PFK↑adp(+), ALDO(+) GAPDH(+), Ru5p(-), PEPxylose↑fdp(+)</td>
<td>3.880</td>
</tr>
<tr>
<td>5</td>
<td>PTS(+), PFK(+) GAPDH(+), PDH(+) PEPxylose(+)</td>
<td>2.541</td>
<td>PFK↑adp(+), ALDO(+) GAPDH(+), PEPxylose↑fdp(+)</td>
<td>5.863</td>
</tr>
<tr>
<td>6</td>
<td>PTS(+), PFK(+) GAPDH(+), TIS(+) PDH(+) PEPxylose(+)</td>
<td>2.843</td>
<td>PTS↓g6p(-), PFK↑amp(+) GAPDH(+), TIS(+) PDH(+) PEPxylose↑fdp(+)</td>
<td>5.550 (new branch)</td>
</tr>
<tr>
<td>7</td>
<td>PTS(+), PFK(+), ALDO(-) GAPDH(+), TIS(+), PGK(-) PDH(+) PEPxylose(+)</td>
<td>2.892</td>
<td>PTS↓g6p(-), PFK↑amp(+), ALDO(+) GAPDH(+), TIS(+) PDH(+) PEPxylose↑fdp(+)</td>
<td>7.314</td>
</tr>
<tr>
<td>8</td>
<td>PTS(+), PFK(+), ALDO(-) GAPDH(+), TIS(+), PGK(-) PDH(+) PEPxylose(+)</td>
<td>2.964</td>
<td>PTS↓g6p(-), PFK↑amp(+) ALDO(+) GAPDH(+), TIS(+) PGK(+) PDH(+) PEPxylose↑fdp(+)</td>
<td>7.852 (H)</td>
</tr>
<tr>
<td>9</td>
<td>PTS(+), PFK(+), ALDO(-) GAPDH(+), TIS(+), PGK(-) PDH(+), PGM(-) PEPxylose(+)</td>
<td>3.048</td>
<td>PTS↓g6p(-), PFK↑amp(+) ALDO(+) GAPDH(+), TIS(+) PGK(+) PDH(+), PGM(-) PEPxylose↑fdp(+)</td>
<td>8.059 (H)</td>
</tr>
<tr>
<td>10</td>
<td>PTS(+), PFK(+), ALDO(-) GAPDH(+), TIS(+), PGK(-) PDH(+), PGM(-), ENO(-) PEPxylose(+)</td>
<td>3.155</td>
<td>PTS↓g6p(-), PFK↑amp(+) ALDO(+) GAPDH(+), TIS(+) PGK(+) PDH(+), PGM(-), ENO(-) PEPxylose↑fdp(+)</td>
<td>8.263 (H)</td>
</tr>
</tbody>
</table>

Enzymes highlighted in bold exert high control on the PTS rate. Signature (+)/(-) corresponds to the increase/decrease in the corresponding enzyme property (i.e. the enzyme level or regulation), respectively. Flux ratio is the ratio of the objective function with the respect to the reference value. H corresponds to the Hopf bifurcation.
Table 5. Alternative 10 best optimal enzyme/regulation subsets for the serine flux.

<table>
<thead>
<tr>
<th>Size</th>
<th>Enzyme Subset</th>
<th>Flux Ratio</th>
<th>Enzyme/Regulation Subset</th>
<th>Flux Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SerSynth(+)</td>
<td>1.880</td>
<td>G6PDH↓ nadph (+)</td>
<td>1.034</td>
</tr>
<tr>
<td>2</td>
<td>PTS(+) SerSynth (+)</td>
<td>4.652</td>
<td>PTS↓ g6p(-)</td>
<td>4.754</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SerSynth (+)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>PTS(+) GAPDH(+) SerSynth (+)</td>
<td>9.086</td>
<td>PTS↓ g6p(-)</td>
<td>9.963</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAPDH(+)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SerSynth (+)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>PTS(+), PFK(+) GAPDH(+) SerSynth (+)</td>
<td>14.451</td>
<td>PTS↓ g6p(-), PFK↑ adp/amp(+)</td>
<td>16.650</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GAPDH(+)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SerSynth (+)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PTS(+), PFK(+) GAPDH(+) PDH(+)</td>
<td>15.933</td>
<td>PTS↓ g6p(-), PFK↑ adp(+)</td>
<td>18.518</td>
</tr>
<tr>
<td></td>
<td>PDH(+) SerSynth (+)</td>
<td></td>
<td>GAPDH(+), TIS(+)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SerSynth (+)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>PTS(+), PFK(+) GAPDH(+), TIS(-)</td>
<td>17.418</td>
<td>PTS↓ g6p(-), PFK↑ amp(+)</td>
<td>20.039</td>
</tr>
<tr>
<td></td>
<td>PDH(+) SerSynth (+)</td>
<td></td>
<td>GAPDH(+), TIS(+)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SerSynth (+)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>PTS(+), PFK(+) GAPDH(+), TIS(-)</td>
<td>19.085</td>
<td>PTS↓ g6p(-), PFK↑ amp(+)</td>
<td>22.013</td>
</tr>
<tr>
<td></td>
<td>ALDO(-) PDH(+) SerSynth (+)</td>
<td></td>
<td>GAPDH(+), TIS(+), ALDO(-)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PDH(+)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SerSynth (+)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>PTS(+), PFK(+) GAPDH(+), TIS(-)</td>
<td>19.838</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALDO(-) PDH(+), PGK(-) SerSynth (+)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>PTS(+), PFK(+) GAPDH(+), TIS(-)</td>
<td>20.538</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALDO(-) PDH(+), PGK(-) SerSynth (+)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>PTS(+), PFK(+) GAPDH(+), TIS(-)</td>
<td>20.591</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALDO(-) PDH(+), PGK(-), ENO(-)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SerSynth (+)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Enzymes highlighted in bold exert high control on the PTS rate. Signature (+)/(-) corresponds to the increase/decrease in the corresponding enzyme property (i.e. the enzyme level or regulation), respectively. Flux ratio is the ratio of the objective function with the respect to the reference value.
FIGURE CAPTIONS

**Figure 1.** *Escherichia coli* central carbon metabolism.

**Figure 2.** (a) Shown are the best optimal reaction rate ratios for the Serine biosynthesis rate plotted as a function of size $D$ of modulated enzyme/regulation subsystems (i.e. $D = 1, 2, \ldots, 10$). (b) Shown are the values of the leading real eigenvalues of the model’s linearization computed at optimal steady-state solutions. White circles correspond to the case where enzyme levels are modulated while enzyme regulation is kept unchanged. Solid circle correspond to the case where both enzyme levels and enzyme regulations are manipulated.

**Figure 3.** (a) Shown are the best optimal reaction rate ratios for the PTS rate plotted as a function of size $D$ of modulated enzyme/regulation subsystems (i.e. $D = 1, 2, \ldots, 10$). (b) Shown are the values of the real parts of the leading eigenvalues of the model’s linearization computed at optimal steady-state solutions. White circles correspond to the case where enzyme levels are modulated while enzyme regulation is kept unchanged. Solid circle and triangles correspond to the case where both enzyme levels and enzyme regulations are manipulated.

**Figure 4.** Flux Control Coefficients (FCCs) for the PTS reaction (white bars) and serine production (solid bars), respectively.

**Figure 5.** (a) Flux Control Coefficients (FCCs) for the PTS reaction. (b) Optimal enzyme levels relative to the reference enzyme levels. The white bars in (a) correspond to the original non-perturbed case as shown in Figure 4. The gray bars in (a) and (b) correspond to the case where
the enzyme levels are optimally chosen in the absence of any regulation alteration. The black bars in (a) and (b) correspond to the case where the regulatory structures for the enzymes PTS, PFK, and PEPxylase are altered and the levels of the enzymes TIS, GAPDH, and PDH are modulated.

**Figure 6.** The distributions of stable steady state fluxes relative to the PTS flux at the reference non-perturbed steady state. The italic numbers correspond to the values of fluxes at the reference steady state. The upright numbers correspond to the values of fluxes in the case where the levels of all the six enzymes are modulated in the absence of any regulatory changes. The values of fluxes highlighted in bold correspond to the case where the regulatory structures for the enzymes PTS and PFK, and the levels of the enzymes TIS, GAPDH, PDH, and PEPxylase are simultaneously manipulated.

**Figure 7.** (a) Flux Control Coefficients (FCCs) for the SerSynth reaction. (b) Relative enzyme levels at the optimal stable steady states. The white bars in (a) correspond to the original non-perturbed case as shown in Figure 4. The gray bars correspond to the case where the enzyme levels are optimally chosen in the absence of any regulation alteration. The black bars correspond to the case where the regulatory structures for the enzymes PTS and PFK are altered and the levels of enzymes TIS, GAPDH, PDH, and SerSynth are simultaneously modified.

**Figure 8.** The distributions of stable steady state fluxes relative to the PTS flux at the reference state. The italic numbers correspond to the values of fluxes at the reference steady state. The upright numbers correspond to the values of fluxes in the case where the levels of all the six
enzymes are modulated in the absence of any regulatory changes. The values of fluxes highlighted in bold correspond to the case where the regulatory structures for the enzymes PTS and PFK, and the levels of the enzymes TIS, GAPDH, PDH, and SerSynth are simultaneously modified.
Figure 1
Figure 2

(a) Graph showing Serine Flux Ratios against Enzyme/Regulation Subsets (D).

(b) Graph showing Real part of λ against λ.

Stable optimization of kinetic models

42
Figure 3

(a)

Enzyme/Regulation Subsets (D)

PTS Flux Ratios

(b) (H)

Re λ

Figure 3
Stable optimization of kinetic models

Figure 4

Flux Control Coefficients (FCCs)

Enzymes

PTS
PGI
PFK
ALDO
TIS
GAPDH
PGK
PGM
ENO
PK
PDH
PEPxylase
PGlucoM
G1PAT
RPPK
G3PDH
SerSynth
Synth1
Synth2
DAHPS
G6PDH
PGDH
Ru5P
R5P
Tka
TKa
TA
MurSynth
TrpSynth
MetSynth
(a) Graph showing flux control coefficients for different enzymes.

(b) Graph showing enzyme level ratio for different enzymes.

Figure 5
Figure 6
Figure 7
Figure 8