Thomas Jefferson University
Jefferson Digital Commons

Department of Pathology, Anatomy and Cell Biology Faculty Papers

Department of Pathology, Anatomy and Cell Biology

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: http://jdc.jefferson.edu/pacbfp

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

Recommended Citation
Nikolaev, Evgeni V., "The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems" (2009). *Department of Pathology, Anatomy and Cell Biology Faculty Papers.* Paper 33.
http://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

Telephone: (215) 503-2969
Fax: (215) 923-3808
Email: evgeni.nikolaev@jefferson.edu

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.

Kew words: 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 stoichiomertic 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 substantionally 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).

$$\frac{dC_i}{dt} = \sum_{j=1}^{M} S_{ij} \cdot r_j(r_j^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K}), \quad i \in \mathcal{N} . \tag{1}$$

Here C_i is the concentration of metabolite i, S_{ij} is the stoichiometric coefficient of metabolite i in reaction j with rate $r_j(r_j^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K})$, $r_j^{\max}$ 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, $S_D = E_L \bigcup R_Q$, 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 $(e_{j_1},...,e_{j_L})$ or, equivalently, maximal specific rates, $(r_{j_1}^{\max},...,r_{j_L}^{\max})$, $E_L = \{j_1,\ldots,j_L\}$, $E_L \subseteq \mathcal{M}$, $L \leq M$. Set R_Q corresponds to a subset of regulatory parameters $(k_{\theta_1},...,k_{\theta_Q})$, $R_Q = \{\theta_1,\ldots, \theta_Q\}$, $R_Q \subseteq \mathcal{R}$, $Q \leq R$, $D = L + Q$. Following these definitions, we use model (1) for the optimal selection of set S_D such that the best possible reaction rate $r_{j_0}(r_{j_0}^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K})$ 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,

$$\frac{1}{M}\sum_{j=1}^{M}\frac{r_j^{\max}}{r_j^{\max,0}}=1\,. \tag{2}$$

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 $\mathbf{C}_0$ should also be restricted,

$$\frac{1}{N}\sum_{i=1}^{N}\frac{|\,C_i-C_i^0\,|}{C_i^0}\le\delta. \tag{3}$$

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 *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,

$$\frac{r_{j_1'}^{\max}}{r_{j_1'}^{\max,0}}=\ldots=\frac{r_{j_K'}^{\max}}{r_{j_K'}^{\max,0}}=\gamma\,. \tag{4}$$

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

$$\gamma = \frac{M - \sum_{s=1}^{L} \frac{r_{j_s}^{\max}}{r_{j_s}^{\max,0}}}{M - L}. \tag{5}$$

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,

$$\frac{r_{j_p'}^{\max}}{r_{j_q'}^{\max}} \approx \frac{r_{j_p'}^{\max,0}}{r_{j_q'}^{\max,0}}. \tag{6}$$

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

$$\frac{r_{j_p'}^{\max}}{r_{j_p'}^{\max,0}} \approx \frac{r_{j_q'}^{\max}}{r_{j_q'}^{\max,0}}. \tag{7}$$

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

$$\frac{r_{j_p'}^{\max}}{r_{j_p'}^{\max,0}} = \frac{r_{j_q'}^{\max}}{r_{j_q'}^{\max,0}} = \gamma. \tag{8}$$

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 ($\alpha_1,...,\alpha_{13}$) to activate or disable the entire regulatory structures. The modified model includes 43 optimization parameters ($p_1,..., p_{43}$), $(p_1,..., p_{43}) = (r_1^{\max},..., r_{30}^{\max}, \alpha_1,..., \alpha_{13})$. The simplest example is phosphoglucoisomerase (PGI) inhibition by 6-phosphoglucoconate (6pg),

$$r_{\mathrm{PGI}} = \frac{r_{\mathrm{PGI}}^{\max} \cdot (C_{\mathrm{g6p}} - \frac{C_{\mathrm{f6p}}}{K_{\mathrm{PGI,\,eq}}})}{K_{\mathrm{PGI,\,g6p}} \cdot (1 + \frac{C_{\mathrm{f6p}}}{K_{\mathrm{PGI,\,f6p}} \cdot (1 + \alpha_2 \cdot \frac{C_{\mathrm{6pg}}}{1 + K_{\mathrm{PGI,\,f6p,\,6pginh}}})} + \alpha_2 \cdot \frac{C_{\mathrm{6pg}}}{K_{\mathrm{PGI,\,g6p,\,6pginh}}}) + C_{\mathrm{g6p}}} \quad . \tag{9}$$

Here parameter α_2 is introduced to alter the regulatory properties of PGI mediated by the concentration of 6-phosphogluconate (i.e. C_{6pg}). Specifically, the unit value of α_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_{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,

$$\left(\begin{array}{ll} \underset{S_D}{\text{maximize}} & r_{j_0}(r_{j_0}^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K}) \\ \text{subject to} & \sum_{j=1}^{M} S_{ij} \cdot r_j(r_j^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K}) = 0,\ i \in \mathcal{N} \\ & \frac{r_{j_1}^{\max}}{r_{j_1}^{\max,0}} + \ldots + \frac{r_{j_L}^{\max}}{r_{j_L}^{\max,0}} + K \cdot \gamma = M \\ & r_{j_s'}^{\max} = \gamma \cdot r_{j_s'}^{\max,0},\ s = 1,\ldots,K \\ & \frac{1}{N}\sum_{i=1}^{N} \frac{|C_i - C_i^0|}{C_i^0} \le \delta \\ & \max_i \operatorname{Re} \lambda_i \le -\lambda_0 < 0 \end{array}\right) \tag{10}$$

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 $\operatorname{Re}\lambda_i$ is the real part of eigenvalue λ_i calculated from the linearization of the right-hand side of equation (1) at the steady state concentrations, and λ_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 $r_{j_0}(r_{j_0}^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K})$ 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., $r^{\max,0}$), kinetic parameters, and steady state metabolite concentrations (i.e. $\mathbf{C_0}$) 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 $\mathcal{M}$ and regulatory parameters $\mathcal{R}$. 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 $(r_{j_1}^{\max},...,r_{j_L}^{\max})$ and regulatory parameters $(\alpha_{\theta_1},...,\alpha_{\theta_Q})$ are calculated by utilizing gradient-based algorithms (e.g. an SQP-algorithm). To evaluate rate $r_{j_0}(r_{j_0}^{\max}, \mathbf{C_{SS}}, \mathbf{R}, \mathbf{K})$, steady state concentrations $\mathbf{C_{ss}}$ are calculated using Newton-based solvers. Because Newton-based solvers can converge to both stable and unstable solutions, the stability of $\mathbf{C_{ss}}$ 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

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, 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 it 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 stead-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

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),

$$C_r^J = \frac{\partial \ln J}{\partial \ln r^{\max}}, \tag{11}$$

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

$$R_\alpha^J = \frac{\partial \ln J}{\partial \ln \alpha}, \tag{12}$$

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

$$\pi_\alpha^r = \frac{\partial \ln r}{\partial \ln \alpha}. \tag{13}$$

Coefficients R_α^J, C_r^J, and π_α^r satisfy the simple identity (Heinrich and Schuster 1996),

$$R_\alpha^J \equiv C_r^J \cdot \pi_\alpha^r. \tag{14}$$

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 at 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. **PTS**↓g6p(-)), PFK (i.e. **PFK**↑amp(+)), and PEPxylase (i.e. **PEPxylase**↑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. PFK↑amp) with FRC = 0.065 is more preferable than the amplification of the positive feedback by adp (i.e. PFK↑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. **PTS**, **PFK**, TIS, **GAPDH**, **PDH**, and **PEPxylase** 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

ALDO	aldolase
DAHPS	DAHP synthases
ENO	enolase
G1PAT	glucose-1-phosphate adenyltransferase
G3PDH	glycerol-3-phosphate dehydrogenase
G6PDH	glucose-6-phosphate dehydrogenase

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
PGlucoM	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

2pg	2-phosphoglycerate
3pg	3-phosphoglycerate
6pg	6-phosphogluconate
accoa	acetyl-coenzyme A
dhap	dihydroxyacetonephosphate
e4p	erythrose-4-phosphate
f6p	fructose-6-phosphate
fdp	fructose-1,6-bisphosphate
g1p	glucose-1-phosphate
g6p	glucose-6-phosphate
gap	glyceraldehyde-3-phosphate
glc	glucose
oaa	oxaloacetate
pep	phosphoenolpyruvate
pgp	1,3-diphosphoglycerate
pyr	pyruvate
rib5P	ribose-5-phosphate
ribu5p	ribulose-5-phosphate
sed7p	sedoheptulose-7-phosphate
xyl5p	xylulose-5-phosphate

References

Alper, H., Jin, Y.S., Moxley, J.F., Stephanopoulos, G., 2005. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. Metab Eng 7(3):155-64.

Atlas, J.C., Nikolaev, E.V., Browning, S.T., and Shuler M.L. [2008], "Incorporating Genome-Wide DNA Sequence Information into a Dynamic Whole-Cell Model of *Escherichia coli*: Application to DNA Replication, " *IET Systems Biology* **2**(5), 369 - 382

Bailey, J.E., 1973. Periodic operation of chemical reactors: a review. Chem. Eng. Commun. 1: 111-124.

Bailey, J.E., 1991. Toward a science of metabolic engineering. Science 252(5013):1668-75.

Bailey, J.E., 1998. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. Biotechnol Prog 14(1):8-20.

Bakker, B.M., Westerhoff, H.V., Michels, P.A., 1995. Regulation and control of compartmentalized glycolysis in bloodstream form Trypanosoma brucei. J Bioenerg Biomembr 27(5):513-25.

Brown, K., Sethna, J.P. (2003). Statistical mechanical approaches to models with many poorly known parameters. Physical Review E 68: 021904-1-021904-9.

Browning, S.T., Shuler, M.L., 2001. Towards the development of a minimal cell model by generalization of a model of Escherichia coli: use of dimensionless rate parameters. Biotechnol Bioeng 76(3):187-92.

Burgard, A.P., Nikolaev, E.V., Schilling, C.H., Maranas, C.D., 2004. Flux Coupling Analysis of Genome-scale Metabolic Reconstructions. Genome Research 14:301-312.

Burgard A.P., Pharkya, P., Maranas, C.D., 2003. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. Biotechnology and Bioengineering 84:647-657.

Castellanos, M., Kushiro, K., Lai, S.K., Shuler, M.L., 2007. A genomically/chemically complete module for synthesis of lipid membrane in a minimal cell. Biotechnol Bioeng 97(2):397-409.

Chassagnole. C., Noisommit-Rizzi. N., Schmid. J.W., Mauch, K., Reuss, M., 2002. Dynamic modeling of the central carbon metabolism of Escherichia coli. Biotechnology and Bioengineering 79(1):53-73.

Domach, M.M., Leung, S.K., Cahn, R.E., Cocks, G.G., Shuler, M.L., 1984. Computer Model for Glucose-Limited Growth of a Single Cell of Escherichia Coli B/r-A. Biotechnology and Bioengineering 26:203-216.

Douglas, J.M., 1967. Periodic reactor operation. I & EC Process Design and Development 6(1): 43-48.

Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transciptional regulators. Nature 403(20):335-338.

Fung, E., Wong, W.W., Suen, J.K., Bulter, T., Lee, S., Liao, J.C., 2005. A synthetic gene-metabolic oscillator. Nature 435(5): 118-122.

Gerdtzen, Z.P., Daoutidis, P., Hu, W.-S., 2004. Non-linear reduction for kinetic models of metabolic reaction networks. Metabolic Engineering 16: 140-154.

Ghosh S., Grossmann I.E., Ataai M.M., Domach M.M., 2006. A three-level problem-centric strategy for selecting NMK precursor labeling and analytes. Metabolic Engineering 8: 491-507.

Hadlich, F., Noack, S., Wiechert, W., 2009. Translating biochemical network models between different kinetic formats. Metabolic Engineering 11: 87-100.

Hatzimanikatis, V., Bailey, J.e., 1997. Effects of Spatiotemporal Variations on Metabolic Control: Approximate Analysis Using (Log)Linear Kinetics Models 54(2): 92-104.

Hatzimanikatis, V., Floudas, C.A., Bailey, J., 1996a. Analysis and Design of Metabolic Reaction Networks via Mixed-Integer Linear Optimization. AIChE Journal 42(5):1277-1292.

Hatzimanikatis, V., Floudas, C.A., Bailey, J., 1996b. Optimization of regulatory Architectures in metabolic reaction networks. Biotechnology and Bioengineering 52:485-500.

Hatzimanikatis, V., Lyberatos, G., Pavlou, S., Svoronos, S.A. 1993. A method for periodic optimization of chemical reaction systems 48(4): 789-797

Heijnen, J.J., 2005. Approximative kinetic formats used in metabolic network modeling. Biotechnol Bioeng 91(5):534-45.

Heinrich, R., Rapoport, T.A., 1974. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. Eur J Biochem 42(1):89-95.

Heinrich, R., Schuster, S., 1996. The Regulation of Cellular Systems. New York: Chapman & Hall.

Hynne, F., Dano, S., Sorensen, P.G., 2001. Full-scale model of glycolysis in Saccharomyces cerevisiae. Biophys Chem 94(1-2):121-63.

Ibarra, R.U., Edwards, J.S., Palsson, B.Ø., 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420:186-189.

Kacser, H., Acerenza, L., 1993. A universal method for achieving increases in metabolite production. Eur J Biochem 216(2):361-7.

Kacser, H., Burns, J.A., 1973. The Control of flux. Symp Soc Exp Biol 27:65-104.

Kirkpatrick, S., Gelatt, C.D.J., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220(4598):671-680.

Kuznetsov, Y.A., 2004. Elements of Applied Bifurcation Theory. New York: Springer.

Laffend, L., Shuler, M.L., 1993. Ribosomal Protein Limitations in *Escherichia coli* under Conditions of High Translational Activity. Biotechnology and Bioengineering 43:388-398.

Lee, S.J., Lee, D.Y., Kim, T.Y., Kim, B.H., Lee, J., Lee, S.Y., 2005. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation. Appl. Envirn. Microbiol. 71(12):7880-7.

Lee, S.Y., Song, H., Lee, S.Y., 2006. Genome-based metabolic engineering of *Mannheimia succiniciproducens* for succinic acid production. Appl. Envirn. Microbiol. 72(3):1939-48.

Link, H., Weuster-Botz, D., 2007. Steady-state analysis of metabolic pathways: Comparing the double modulation method and the lin-log approach. Metabolic Engineering 9: 433-441.

Mahadevan, R., Edwards, J.S., Doyle, F.J., 3rd. 2002. Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J 83(3):1331-40.

Mauch, K., Buziol, S., Schmid, J., Reuss, M., Computer-Aided Design of Metabolic Networks. AIChE Symposium Series; 2001; Tucson, Arizona. p 82-91.

Mhaskar, P., El-Farra, N.H., Christofides, P.D., 2005. Predictive Control of Switched Nonlinear Systems With Scheduled Mode Transitions. IEEE Transactions and automatic control 50(11): 1670-1680

Mhaskar, P., El-Farra, N.H., Christofides, P.D., 2006. Stabilization of nonlinear systems with state and control constraints using Lyapunov-based predictive control. Systems & Control Letters 55: 650-659.

Monod, J., 1949. The growth of bacerial cultures. Ann. Rev. of Microbiol. 111:371-394.

Nikerel, I.E., van Winden, W.A., Verheijen, P.J.T., Hejnen, J.J., 2009. Model reduction and a priori kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics. Metabolic Engineering 11: 20-30.

Nikolaev, E.V., Atlas, J.C., Shuler, M.L., 2006. Computer models of bacterial cells: from generalized coarse-grained to genome-specific modular models. J Physics: Conference Series 46:322-326.

Nikolaev, E.V., Atlas, J.C., Shuler, M.L., 2007. Sensitivity and control analysis of periodically forced reaction networks using the Green's function method. J Theor Biol 274(3):442-461.

Nikolaev, E.V., Burgard, A.P., Maranas, C.D., 2005. Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions. Biophys J 88(1):37-49.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R., and others. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res 33(17):5691-702.

Palsson, B.Ø., 2006. Systems Biology: Properties of Reconstructed Networks. New York: Cambridge University Press.

Pharkya, P., Burgard, A.P., Maranas, C.D., 2004. OptStrain: a computational framework for redesign of microbial production systems. Genome Res 14(11):2367-76.

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

Pharkya, P., Nikolaev, E.V., Maranas, C.D., 2003. Review of the BRENDA Database. Metab Eng 5(2):71-3.

Rizzi, M., Baltes, M., Theobald, U., Reuss, M., 1997. In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae: II. Mathematical Model. Biotechnology and Bioengineering 55(4):592-608.

Savageau, M.A., 1976. Biochemical systems analysis. A study of function and design in molecular biology. New York: Addison-Wesley, Reading, MA.

Schmid, J.W., Mauch, K., Reuss, M., Gilles, E.D., Kremling, A., 2004. Metabolic design based on a coupled gene expression-metabolic network model of tryptophan production in Escherichia coli. Metab Eng 6(4):364-77.

Segre, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A 99(23):15112-7.

Shuler, M.L., 2005. Computer Models of Bacterial Cells To Integrate Genomic Detail with Cell Physiology; Seoul, Korea. p 54-62.

Small, R., Kacser, H., 1994. A method for increasing the concentration of a specific internal metabolite in steady-state systems. Eur J Biochem 226:649-656.

Khazin, L.G., Shnol, E.E., 1991. Stability of Critical Equilibrium States. Nonlinear Science.

Stephanopoulos, G., Simpson, T.W., 1997. Flux amplification in comlex metabolic networks. Chemical Engineering Science 52(15):2607-2627.

Stephanopoulos, G., Vallino, J.J., 1991. Network rigidity and metabolic engineering in metabolite overproduction. Science 252(5013):1675-81.

Stephanopoulos, G.N., Aristidou, A.A., Nielsen, J., 1998. Metabolic Engineering. Principles and Methods. New York: Academic Press.

Tomita, M., 2001. Whole-cell simulation: a grand challenge of the 21st century. Trends Biotechnol. 19(6):205-210.

Visser, D., Schmid, J.W., Mauch, K., Reuss, M., Heijnen, J.J., 2004. Optimal re-design of primary metabolism in *Escherichia coli* using linglog kinetics.Metabolic Engineering 6: 378-390.

Vital-Lopez, F.G., Armaou, A., Nikolaev, E.V., Maranas, C.D., 2006. A computational procedure for optimal engineering interventions using kinetic models of metabolism. Biotechnol Prog 22(6):1507-17.

Vital-Lopez, F.G., Maranas, C.D., Armaou, A. 2006. Bifurcation analysis of the metabolism of *E.coli* at optimal enzyme levels. Proceedings of the 2006 American Control Conference, Minneapolis, Minnesota, USA, June 14-16, 3439-3444.

Wong, W.W., Liao, J.C. (2006). The design of intracellular oscillators that interact with metabolism. Visions and Reflections. Cell. Mol. Life Sci. 63:1215-1220.

TABLES

Table 1. Enzyme regulation.

№	Enzyme	Regulation	Notation
1	PTS	inhibition by g6p	PTS↓g6p
2	PGI	inhibition by 6pg	PGI↓6pg
3	PFK	inhibition by pep	PFK↓pep
4	PFK	activation by adp	PFK↑adp
5	PFK	activation by amp	PFK↑amp
6	PK	activation by amp	PK↑amp
7	PK	activation by fdp	PK↑fdp
8	PK	inhibition by atp	PK↓atp
9	G1PAT	activation by nadph	G1PAT↑nadph
10	G6PDH	inhibition by nadph	G6PDH↓nadph
11	PGDH	inhibition by atp	PGDH↓atp
12	PGDH	inhibition by nadph	PGDH↓nadph
13	PEPxylase	activation by fdp	PEPxylase↑fdp

Table 2. Best flux ratios of steady state fluxes in the optimized and original models.

Flux	Regulation	Enzyme Level	Enzyme Level & Regulation
PTS	1.43	3.16	7.31 (8.26 H)
Serine	1.06	20.59	22.01

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.

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

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.

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

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.

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

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, …, 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, …, 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.

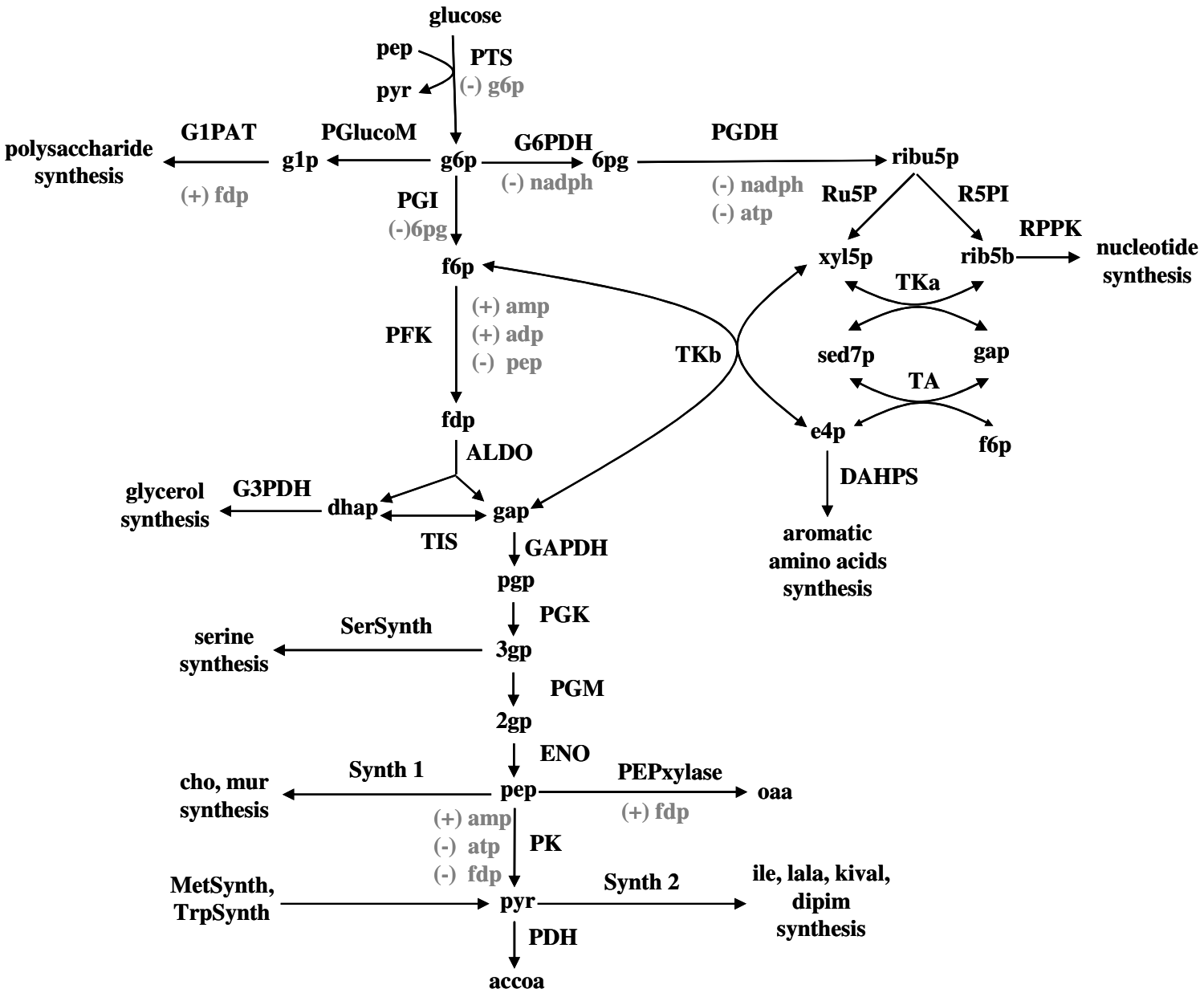
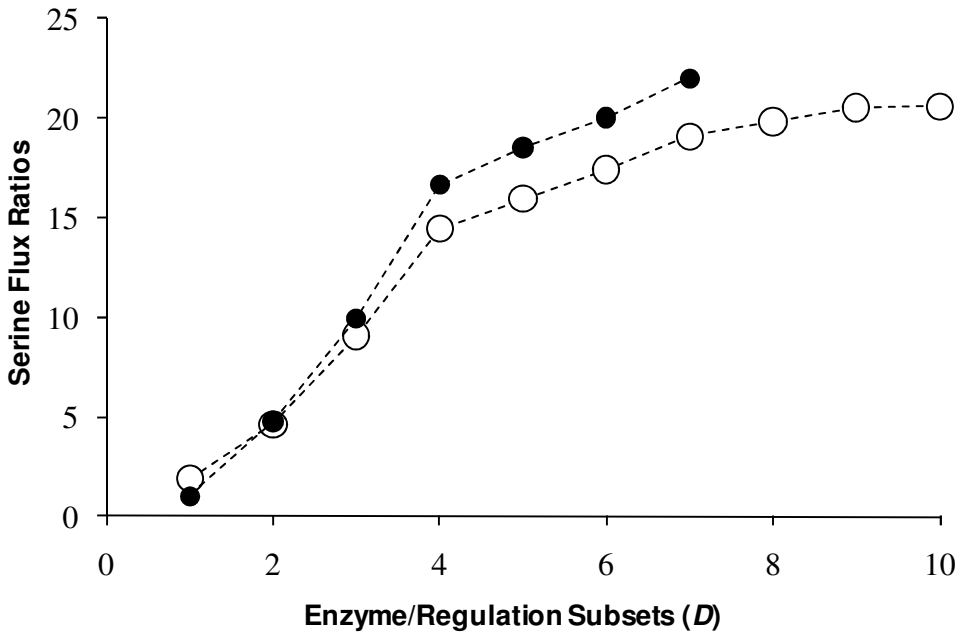

glucose
pep
PTS
pyr
(-) g6p
G1PAT
PGlucoM
polysaccharide synthesis
g1p
g6p
G6PDH
6pg
PGDH
ribu5p
(+) fdp
(-) nadph
(-) nadph
(-) atp
Ru5P
R5PI
PGI
(-)6pg
RPPK
xyl5p
rib5b
nucleotide synthesis
f6p
TKa
(+) amp
(+) adp
(-) pep
PFK
TKb
sed7p
gap
TA
fdp
e4p
f6p
ALDO
DAHPS
glycerol synthesis
G3PDH
dhap
gap
TIS
GAPDH
aromatic amino acids synthesis
pgp
PGK
serine synthesis
SerSynth
3gp
PGM
2gp
ENO
cho, mur synthesis
Synth 1
pep
PEPxylase
oaa
(+) fdp
(+) amp
(-) atp
(-) fdp
PK
MetSynth, TrpSynth
pyr
Synth 2
ile, lala, kival, dipim synthesis
PDH
accoa

Figure 1

(a)


25
20
15
10
5
0
Serine Flux Ratios
0
2
4
6
8
10
Enzyme/Regulation Subsets (*D*)

(*b*)

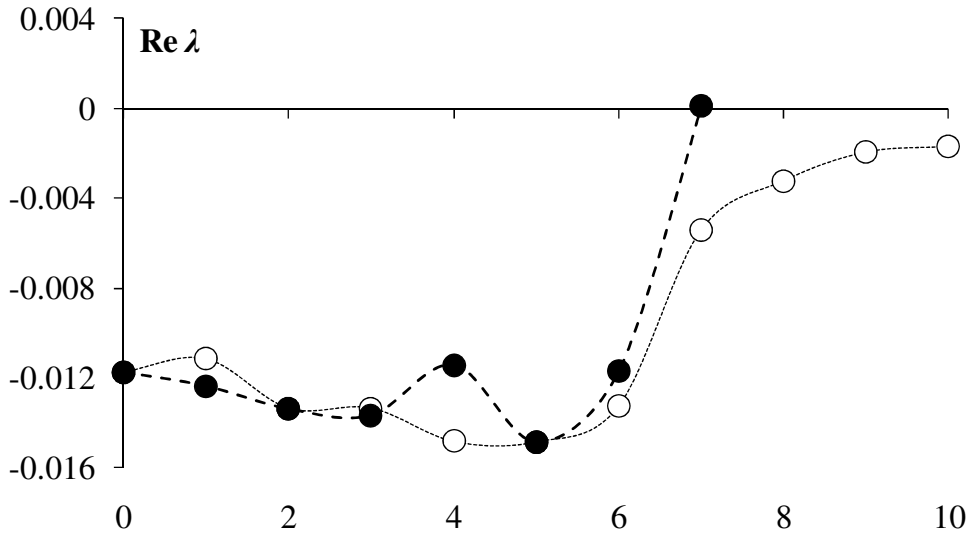
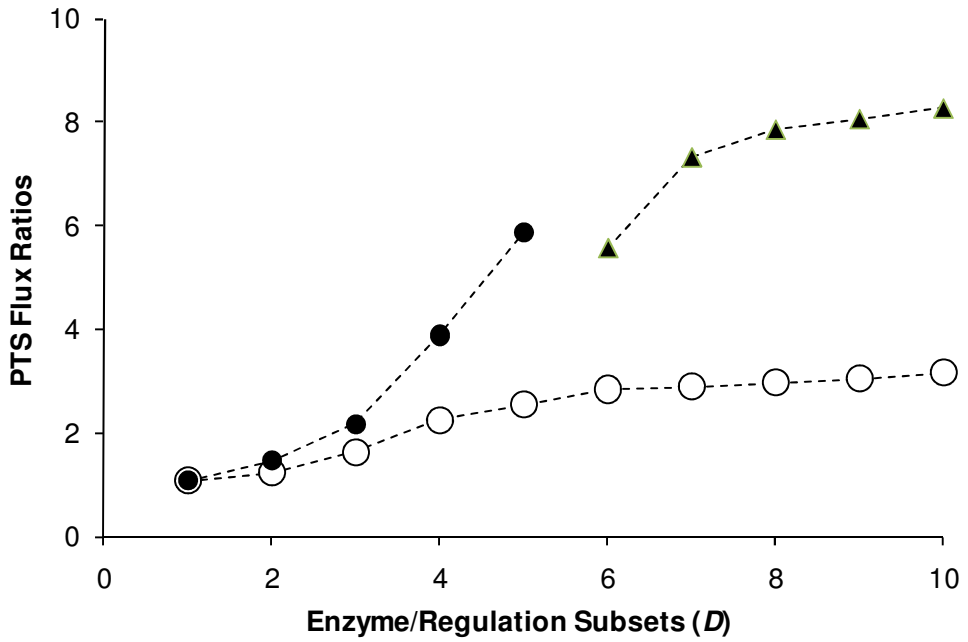

0.004
Re λ
0
-0.004
-0.008
-0.012
-0.016
0
2
4
6
8
10

Figure 2

(a)


10
8
6
4
2
0
PTS Flux Ratios
0
2
4
6
8
10
Enzyme/Regulation Subsets (*D*)

(b) (H)

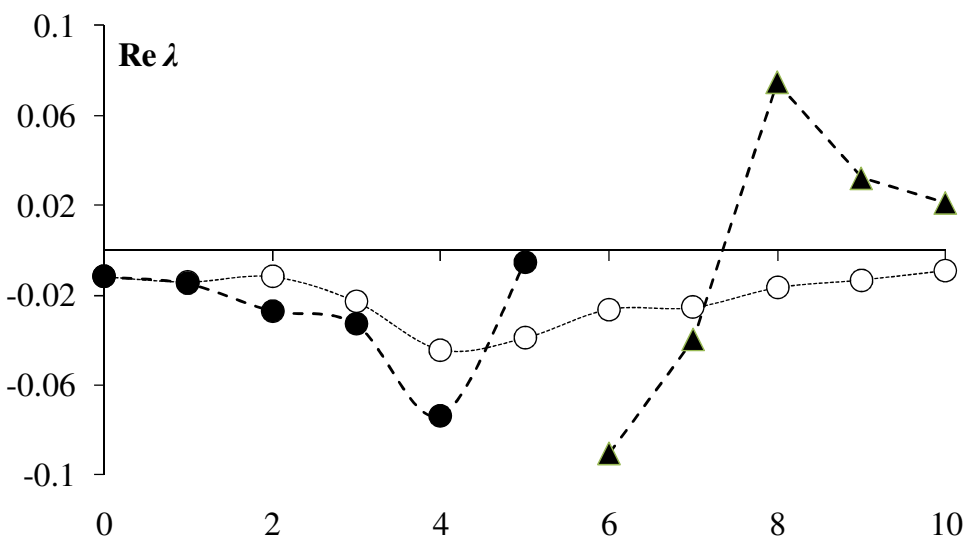

0.1
Re λ
0.06
0.02
-0.02
-0.06
-0.1
0
2
4
6
8
10

Figure 3

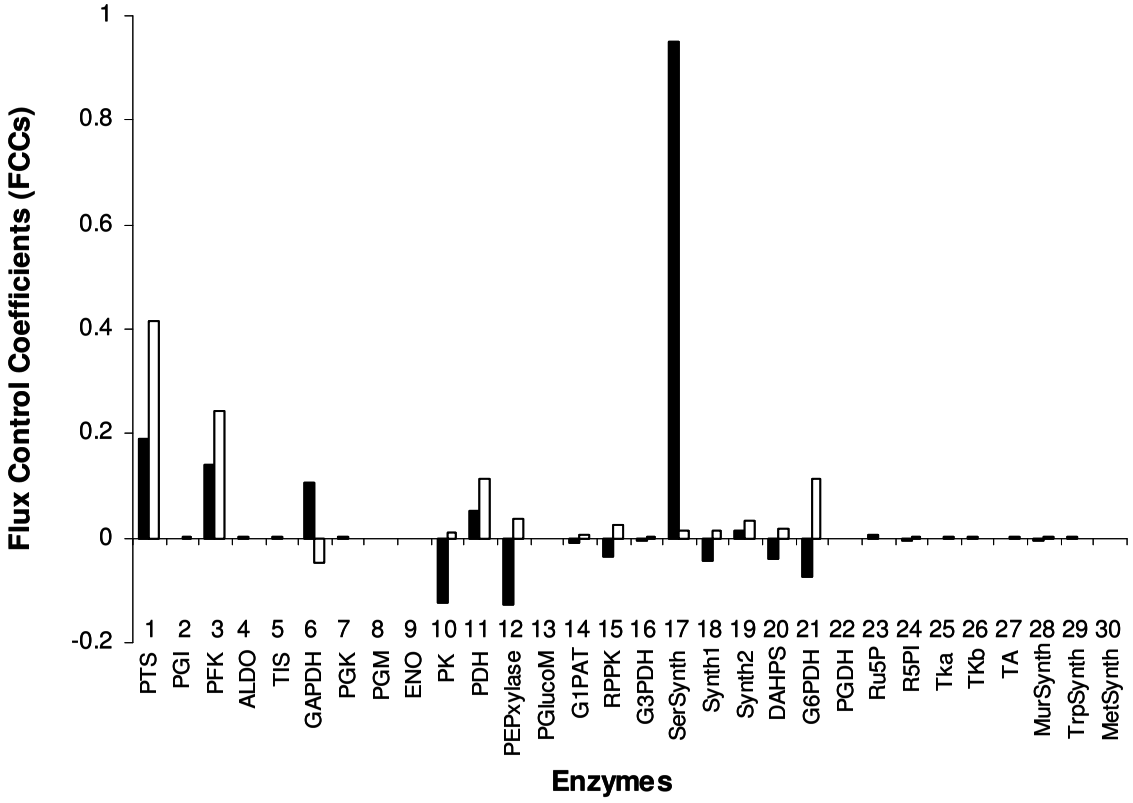

Flux Control Coefficients (FCCs)
1
0.8
0.6
0.4
0.2
0
-0.2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
PTS
PGI
PFK
ALDO
TIS
GAPDH
PGK
PGM
ENO
PK
PDH
PEPxylase
PGlucoM
G1PAT
RPPK
G3PDH
SerSynth
Synth1
Synth2
DAHPS
G6PDH
PGDH
Ru5P
R5PI
Tka
TKb
TA
MurSynth
TrpSynth
MetSynth
Enzymes

Figure 4

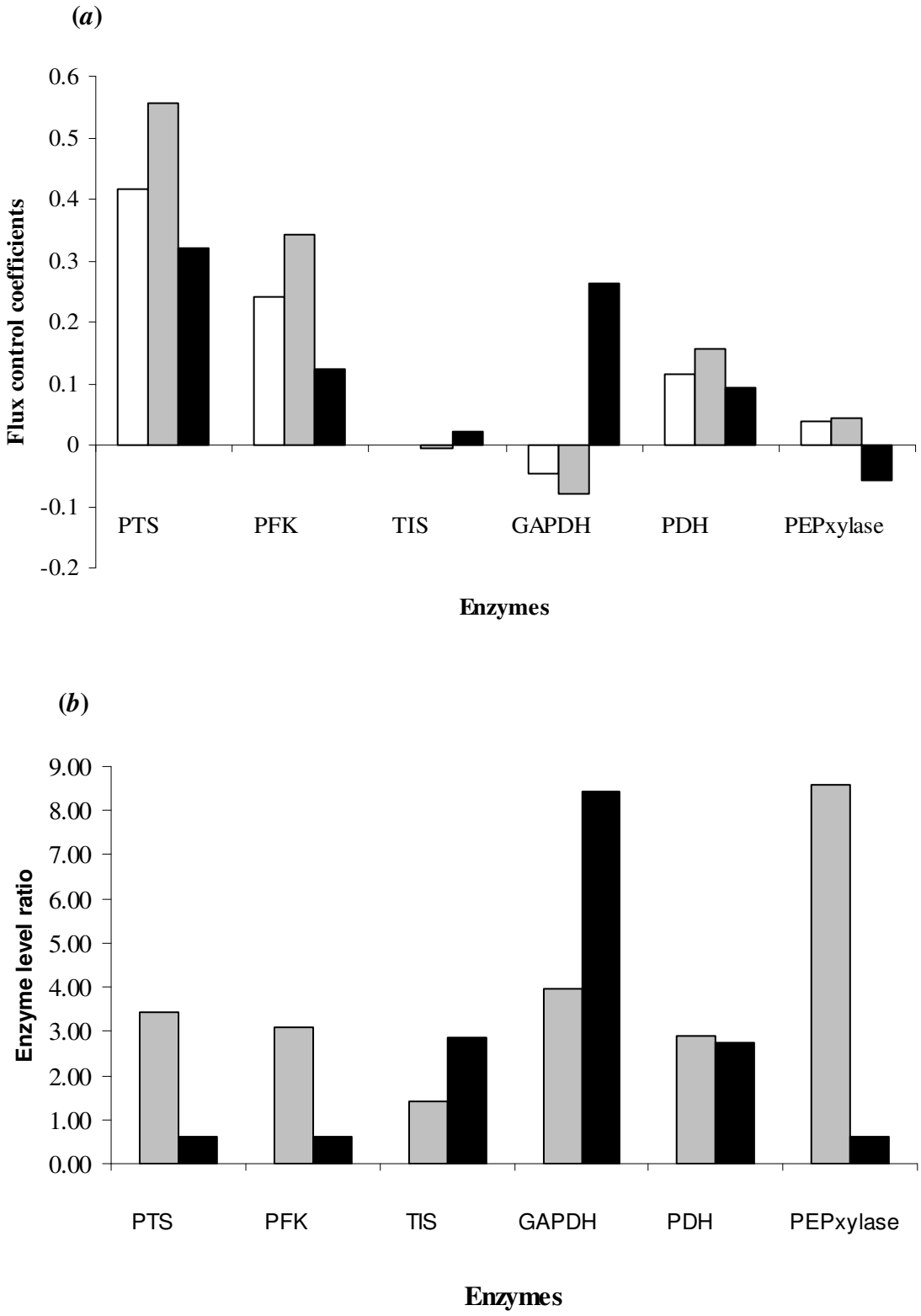

(a)
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
Flux control coefficients
PTS
PFK
TIS
GAPDH
PDH
PEPxylase
Enzymes
(b)
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Enzyme level ratio
PTS
PFK
TIS
GAPDH
PDH
PEPxylase
Enzymes

Figure 5

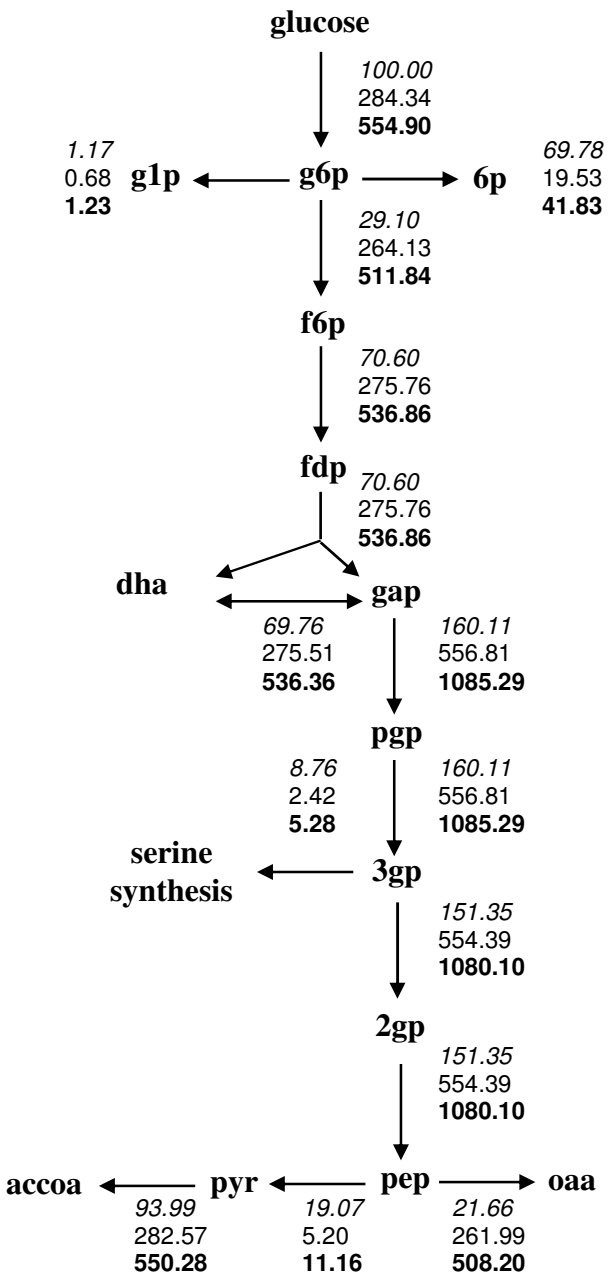

glucose
100.00
284.34
554.90
1.17
0.68
1.23
g1p
g6p
6p
69.78
19.53
41.83
29.10
264.13
511.84
f6p
70.60
275.76
536.86
fdp
70.60
275.76
536.86
dha
gap
69.76
275.51
536.36
160.11
556.81
1085.29
pgp
8.76
2.42
5.28
160.11
556.81
1085.29
serine
synthesis
3gp
151.35
554.39
1080.10
2gp
151.35
554.39
1080.10
accoa
pyr
pep
oaa
93.99
282.57
550.28
19.07
5.20
11.16
21.66
261.99
508.20

Figure 6

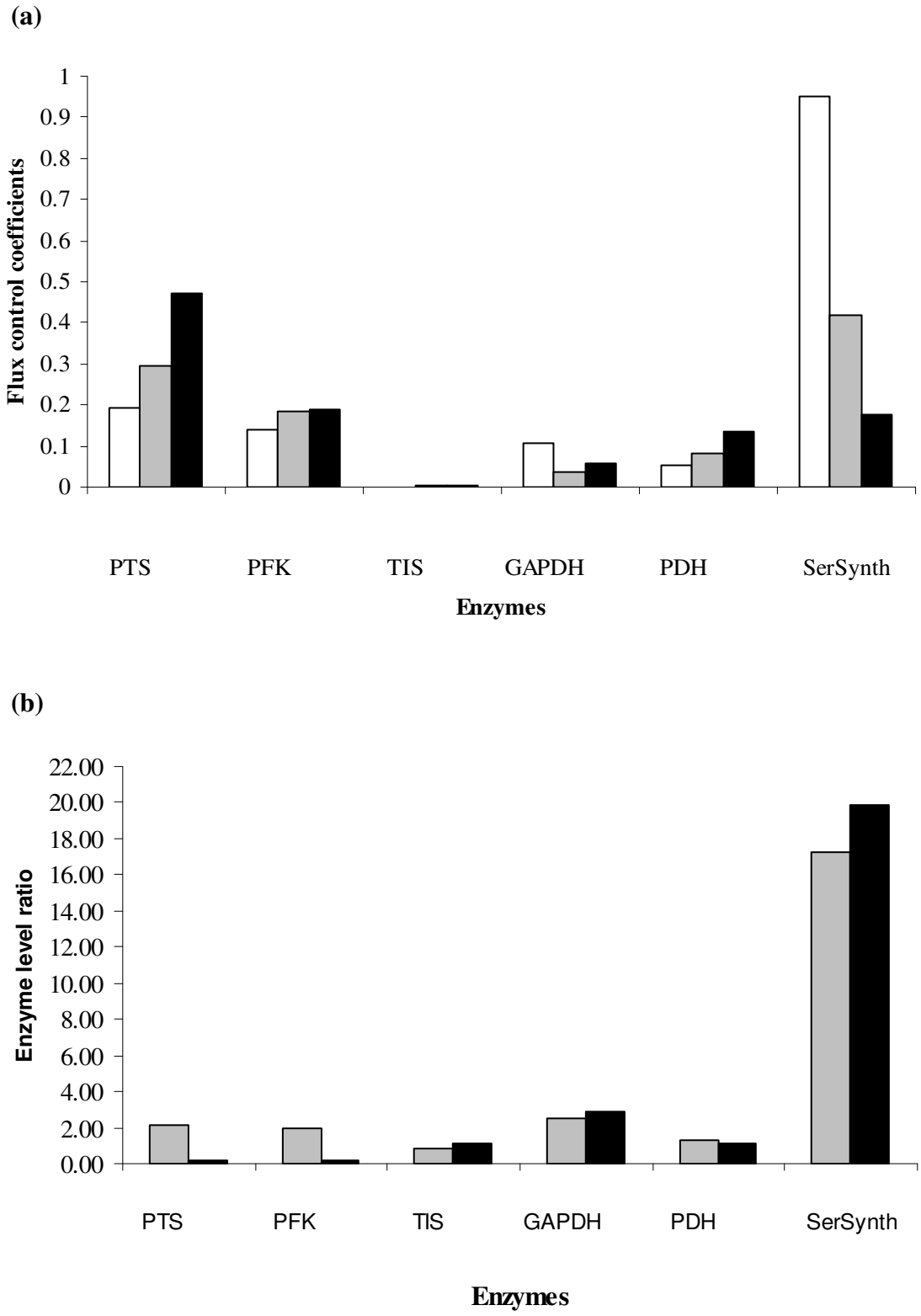

(a)
Flux control coefficients
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
PTS
PFK
TIS
GAPDH
PDH
SerSynth
Enzymes
(b)
Enzyme level ratio
22.00
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
PTS
PFK
TIS
GAPDH
PDH
SerSynth
Enzymes

Figure 7

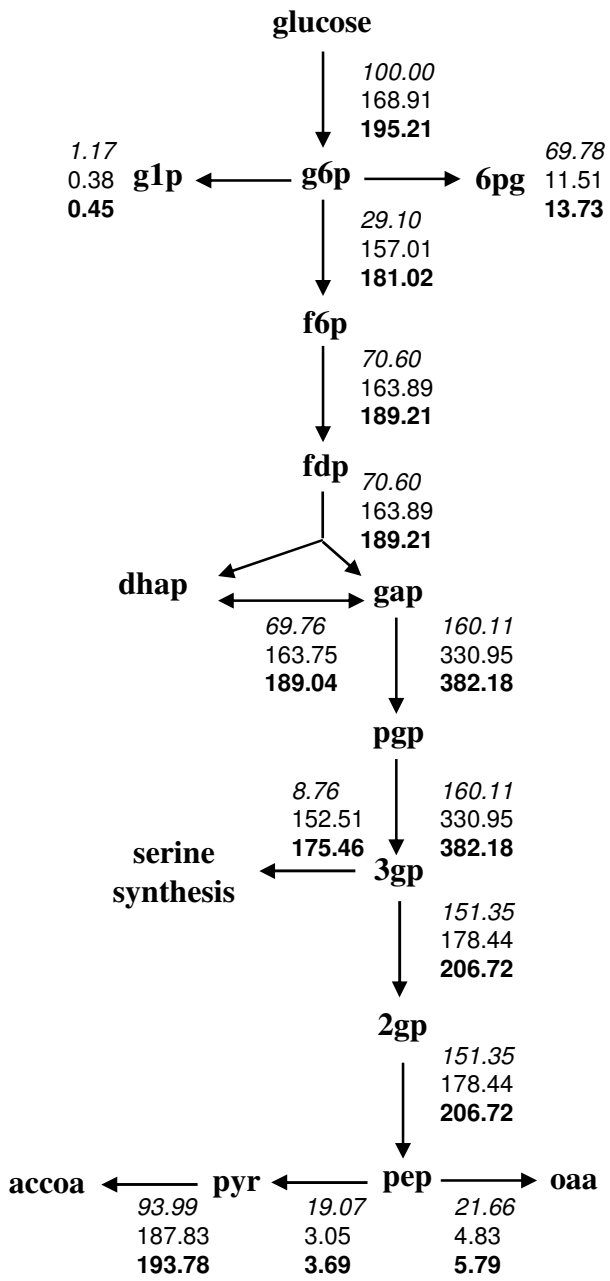

glucose
100.00
168.91
195.21
1.17
0.38
0.45
g1p
g6p
6pg
69.78
11.51
13.73
29.10
157.01
181.02
f6p
70.60
163.89
189.21
fdp
70.60
163.89
189.21
dhap
gap
69.76
163.75
189.04
160.11
330.95
382.18
pgp
8.76
152.51
175.46
160.11
330.95
382.18
serine
synthesis
3gp
151.35
178.44
206.72
2gp
151.35
178.44
206.72
accoa
pyr
pep
oaa
93.99
187.83
193.78
19.07
3.05
3.69
21.66
4.83
5.79

Figure 8