OPTIMAL SELECTION OF ENZYME LEVELS USING LARGE-SCALE KINETIC MODELS

Evgeni V. Nikolaev, Priti Pharkya, Costas D. Maranas, and Antonios Armaou*

Department of Chemical Engineering, The Pennsylvania State University, University Park, U.S.A. 16802

Abstract: A hybrid optimization framework is introduced to identify enzyme sets and levels to meet overproduction requirements using kinetic models of metabolism. A simulated annealing algorithm is employed to navigate through the discrete space of enzyme sets while a sequential quadratic programming method is utilized to identify optimal enzyme levels. The framework is demonstrated on a model of *E.coli* central metabolism for serine biosynthesis. Computational results show that by optimally manipulating relatively small enzyme sets, a substantial increase in serine production can be achieved. The proposed approach thus provides a versatile tool for the elucidation of controlling enzymes with implications in biotechnology.

Keywords: biotechnology, optimization, and mathematical models.

1. INTRODUCTION

The systematic development of optimal microbial strains in biotechnology and efficient therapeutic interventions in medicine still presents a fundamental challenge for metabolic engineering in the postgenomic era (Cornish-Bowden and Cardenas 2000; Kholodenko and Westerhoff 2004). In this endeavor, mathematical modeling technologies have gradually become indispensable tools in formulating plausible hypotheses and improving decision-making abilities by providing a *systematic* quantitative description of how system's properties (i.e., metabolic fluxes, concentrations, or cell growth) respond to changes in system's components and environments (i.e., gene knockouts, enzyme activities, or gene expression). To this end, a variety of modeling frameworks are already available. Mathematical modeling has been widely used in microbiology and biotechnology since the Monod's discovery of the relationship between the specific growth rate and limiting substrate concentration (Monod 1949). These and similar population and cellular culture biomass growth studies can now be complemented with genome-scale stoichiometric modeling to devise strategies of genetic modifications for targeted overproductions of useful biochemicals, based on cellular stoichiometry alone (Stephanopoulos, *et al.*, 1998; Palsson 2004). While successful in many instances, stoichiometric models cannot capture dynamic effects mediated by metabolite concentrations, enzyme activities, and changes in the environment and genetic control. Alternative important advances toward the rational analysis of cellular systems are known as Metabolic Control Analysis (MCA) (Kacser and Burns 1973; Heinrich and Rapoport 1974) and Biochemical Systems Analysis (BSA) (Savageau 1976). These analyses are based on local linear and log-linear approximations of inherently nonlinear metabolic phenomena, respectively.

The linearity of MCA and *S*-systems representation has been exploited before using a rational design analysis for improvements in bioprocess performance by genetic modifications of metabolic control structures (Hatzimanikatis, *et al.*, 1996). Genetic manipulations, however, typically cause metabolic networks to deviate significantly from the original steady states. In response to these limitations, a number of research groups have undertaken the development of large-scale kinetic mechanistic models (Chassagnole, *et al.*, 2002). Prominent modeling projects include the ECell International Project (Tomita 2001), the minimal cell

*To whom correspondence should be addressed:
email: armaou@engr.psu.edu, tel. (814) 865-5316,
fax: (814) 865-7846

(Castellanos, *et al.*, 2004), and virtual cell models (Slepchenko, *et al.*, 2003). Similar to molecular biologists and genetic engineers, modelers can now simulate the perturbation of cellular systems with the intention of introducing desired fluxes and metabolite concentrations within complex pathways.

Motivated by the advent of large-scale mechanistic models of cellular systems, the objective of this study is to introduce a general optimization framework to automatically identify minimal enzyme sets leading to a significant overproduction potential. Because metabolism plays an important role in cellular systems by supplying them with energy and biosynthetic precursors, the large-scale mechanistic kinetic model of central metabolism of *Escherichia coli* (Chassagnole, *et al.*, 2002; Visser, *et al.*, 2004) is chosen as the basis for benchmarking and presenting the developed optimization framework.

2. MODELING CELLULAR KINETICS AND GENETIC MACHINERY

A mathematical model (1) of relevant processes in metabolism and genetic control can be postulated as a set of kinetic mass balances coupled with equations describing the genetic machinery (i.e., ribosomes and RNA polymerases contents) (Schmid, *et al.*, 2004).

$$\begin{aligned} \frac{dC_i}{dt} &= \sum_{j=1}^{M} S_{ij} \cdot r_j(r_j^{\max}, \mathbf{C}, \mathbf{K}), \quad \forall i \in \mathrm{N}, \\ \frac{de_j}{dt} &= r_j^{\mathrm{syn}} - r_j^{\deg}, \qquad \forall j \in \mathbf{M}. \end{aligned} \tag{1}$$

Here C_i is the concentration of metabolite i, $i \in$ N, S_{ij} is the stoichiometric coefficient of metabolite i in reaction j, $j \in$ **M**, $r_j(r_j^{\max}, \mathbf{C}, \mathbf{K})$ is the rate of reaction j, **C** is the vector of metabolite concentrations, **K** is the vector of kinetic parameters, and $r_j^{\max}$ is the maximal reaction rate determined by enzyme level e_j. Changes in enzyme levels e_j are defined by the rates of enzyme synthesis r_j^{syn} and degradation $r_j^{\deg}$. N = {1,…,*N*} and **M** = {1,…,*M*} are sets of metabolites and reactions, respectively.

Model (1) can be used to select optimal enzyme sets E_L (i.e., $E_L = \{j_1,\ldots,j_L\}$) and the corresponding enzyme levels (i.e., $e_{j_1},\ldots,e_{j_L}$) such that the best possible reaction rate $r_{j_0}(r_{j_0}^{\max}, \mathbf{C}, \mathbf{K})$ can be achieved for the overproduction of a specific biochemical of interest (i.e., a product of reaction j_0). Since detailed models (1) are rarely available, reasonable approximations are necessary (Young, *et al.*, 2004). In this study, we follow an approximation approach (Mauch, *et al.*, 2001) to account for important processes such as the redistribution of limited mRNA contents and homeostasis. Specifically, the following constraints are introduced to anticipate the effects of mutation. Constraint (2) assures that an increase in certain enzyme levels is compensated by the decrease among the remaining ones,

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

Since maximal reaction rates are proportional to enzyme levels (Stephanopoulos, *et al.*, 1998), each $r_j^{\max} / r_j^{\max,0}$ can be interpreted as the ratio of enzyme levels e_j and e_j^0 for engineered and reference organisms, respectively (i.e., $r_j^{\max} / r_j^{\max,0} = e_j / e_j^0$).

Cellular systems maintain homeostasis (Reich and Selkov 1981; Heinrich and Schuster 1996), meaning that any large increase in metabolite concentrations will trigger the expression of specific genes that are responsible for the synthesis of enzymes counteracting the undesired changes. This can be mathematically captured by constraint (3), which enforces allowable concentration changes (i.e., within $d \cdot 100\%$) relative to the reference steady state concentrations $\mathbf{C_0}$,

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

Constraints (2) and (3) alone are still not sufficient to describe coordinated changes in *all* enzymes when only L enzymes are modulated. To account for such coordinated changes, the following constraint is introduced

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

Here $j_1',\ldots,j_K'$ are the indices of non-modulated enzymes and $K = M - L$. Condition (4) can be interpreted as maintaining gene expressions rates of non-modulated enzymes at ratios equal to the ones at the reference steady state.

3. SOLUTION METHOD

3.1 Mixed integer nonlinear problem (MINLP)

To select alternative optimal targets for practically feasible enzyme modulations and genetic mutations, a mixed integer nonlinear problem (MINLP) (5) is solved to find small sets of modulated enzymes and the corresponding specific maximal rates ($r_{j_i}^{\max},\ldots,r_{j_L}^{\max}$) such that the best possible reaction rate $r_{j_0}(r_{j_0}^{\max}, \mathbf{C}, \mathbf{K})$ for the production of a biochemical of interest can be achieved. In formulation (5), both enzyme choices E_L and specific maximal reaction rates $r_j^{\max}$ are design variables. Here, the first constraint describes the steady-state condition in model (1), the second and third constraints result from combining constraints (2) and (4). In this paper, we devise and utilize a hybrid '*stochastic/deterministic*' strategy to efficiently solve formulation (5).

$$\left(\begin{array}{ll} \underset{r_{j_1}^{\max},\ldots,r_{j_L}^{\max}}{\text{maximize}} & r_{j_0}(r_{j_0}^{\max}, \mathbf{C}, \mathbf{K}) \\ \text{subject to} & \sum_{j=1}^{M} S_{ij} \cdot r_j(r_j^{\max}, \mathbf{C}, \mathbf{K}) = 0,\ i \in \mathrm{N} \\ & \frac{r_{j_1}^{\max}}{r_{j_1}^{\max,0}} + \ldots + \frac{r_{j_L}^{\max}}{r_{j_L}^{\max,0}} + K \cdot g = M \\ & r_{j_s}^{\max} = g \cdot r_{j_s}^{\max,0},\ s = 1,\ldots,K \\ & \frac{1}{N} \sum_{i=1}^{N} \frac{|C_i - C_i^0|}{C_i^0} \le d \end{array} \right) \text{? (5)}$$

3.2 Search for optimal enzyme sets and levels

A simulated annealing algorithm (Kirkpatrick, *et al.*, 1983) is implemented to navigate through the discrete space of enzyme sets (see Fig. 1). Here, E_L is a randomly chosen initial enzyme set of size L, E_b is the set with the best rate r_b found so far, E_c is the currently investigated set with r_c, and E_t is the trial set with rate r_t. Parameter T is the 'annealing temperature,' reduced by factor a after each J random moves performed, and *MaxIter* is the maximum number of all allowable iterations. To generate E_t, the move class 'Select or Terminate' is implemented, where a random swap between two enzymes, one from E_c and another one from M\ E_c is repetitively performed until a new trial set E_t (i.e., not encountered before) is found. The search is terminated when all neighbours of E_c are evaluated or the maximum number of iterations (i.e., *MaxIter*) is performed.

1. Generate an initial enzyme set E_L
2. Set $E_b = E_c = E_t = E_L$
3. $r_b = r_c = r_t$ = Optimize(E_t)
4. **for** i = 1:*MaxIter*
5. E_t = Select or Terminate (E_c)
6. r_t = Optimize (E_t)
7. **if** $r_t > r_b$
8. $E_b = E_c = E_t$
9. $r_b = r_c = r_t$
10. **else**
11. $anneal = e^{(r_t - r_c)/T}$
12. Generate a random $d \in (0,1)$
13. **if** $d < anneal$
14. $E_c = E_t$
15. $r_c = r_t$
16. **end if**
17. **end if**
18. **if** [i/J] = 0
19. $T = \mathrm{a} \cdot T$
20. **end if**
21. **end loop**

Fig.1. A simulated annealing pseudo-code.

Optimal enzyme levels for every trial set E_t are computed by using standard gradient-based algorithms (i.e., SQP). The evaluation of the objective function (i.e., reaction rate $r_{j_0}(r_{j_0}^{\max}, \mathbf{C}, \mathbf{K})$) relies on the calculation of steady state concentrations **C** by a two-step *prediction-correction* procedure. At the prediction step, the kinetic equation in (1) is integrated over time span [0, T_{end}]. The integration can be automatically terminated at an intermediate t, $t \in [0, T_{\text{end}}]$, if $\max_i |dC_i(t)/dt| \le e$. Subsequently, at the correction step, the final time t integration condition (i.e., **C**(t)) is used as an initial guess for a Newton-based solver to find a solution of the nonlinear equation in (5). Finally, the stability of the corrected steady state **C** in (1) is investigated by computing the eigenvalues of the Jacobian matrix available from the Newton-based solver.

3.3 Computational implementation

The optimization framework is demonstrated on a model of central metabolism for *E. coli* (Chassagnole, *et al.*, 2002), comprised of 30 enzymes and 17 metabolites, with the objectives of maximizing the serine overproduction and flux through the PTS transport system (see Fig. 2). The fixed values of the following parameters have been chosen, d = 0.1, J = 25, *MaxIter* = 10^3, a = 0.9, e = 10^{-3}, and $T_{\text{end}} = 10^3$. To ensure both the robustness and the fast convergence of the algorithm, different values for the initial 'simulated annealing temperature' T were used, $T_0 = 10^{-4}$ - 10^{-2} for the serine production and $T_0 = 10^{-5}$ - 10^{-3} for the PTS flux. These values account for 1% - 100% of the corresponding rates (mM/sec) in the reference state. The complete enumeration of all one- and two-enzymes sets was performed to test the ability of the algorithm to locate the global optima. Also, random multistarts were performed to check the robustness of the SQP search. The entire framework was implemented in Matlab© on a Linux cluster with Intel CPU 3.06 GHz computers. Computational requirements were in order of minutes for small and 10-30 hours for larger enzyme sets E_L.

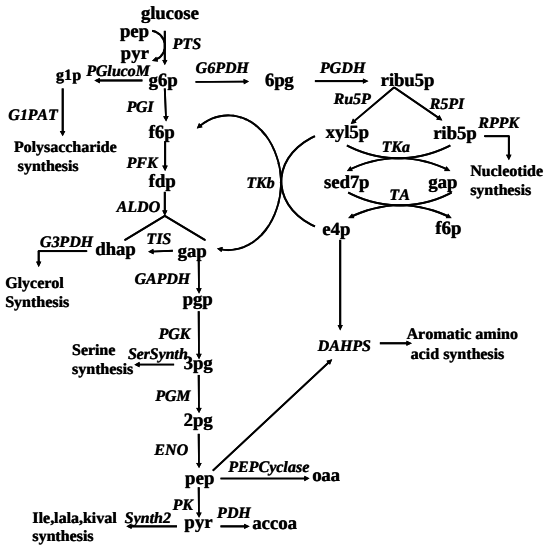

Fig. 2. *Escherichia coli* central metabolism.

4. RESULTS AND DISCUSSION

If all 30 enzymes in the model (Chassagnole, *et al.*, 2002; Visser, *et al.*, 2004) are allowed to vary their levels using formulation (5), a 20-fold increase in the serine production and a 3-fold increase in the PTS flux can be achieved. Substantial improvements though are predicted by manipulating only *small* enzyme sets (see Fig. 3). For example, the modulation of only three enzymes leads to a flux increase, which is almost 50% of the best predictions. The manipulation of six enzymes already leads to a flux increase of about 80% of the best predictions. Importantly, by manipulating 10 enzymes the organism's maximum overproduction capability is reached.

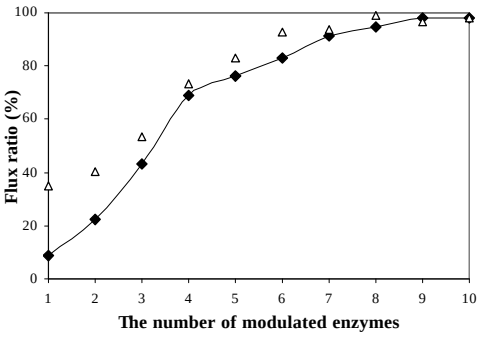

Fig. 3. Ratios r_L / r_{30} (%) are plotted as a function of the modulated enzymes set size *L*. Solid rhombi and white triangles correspond to serine production and PTS rates, respectively.

To get quantitative insights into how successive small sets E_L can be chosen to meet overproduction requirements, flux control coefficients (FCCs) calculated by formulae (6) can be used,

$$C_e^J = \frac{d\ln J}{d\ln e} \approx \frac{e}{J}\cdot\frac{\Delta J}{\Delta e}. \qquad (6)$$

Here, *J* is a pathway flux or the rate of a particular reaction, and *e* is the enzyme's level. Originally, Metabolic Control Analysis (MCA) was developed and Flux Control Coefficients (FCCs) (6) were introduced to quantify bottlenecks or rate limiting steps in complex pathways (Kacser and Burns 1973; Heinrich and Rapoport 1974). It was shown that flux control is distributed among several enzymes in the pathway. Specifically, calculations of FCCs for the model under investigation (Chassagnole, *et al.*, 2002; Visser, *et al.*, 2004) reveal several rate limiting enzymes with high control on the serine and PTS fluxes (see Fig. 4). Importantly, in both cases the same group of enzymes (i.e., PTS, PFK, GAPDH, PDH, and G6PDH) exerts high control and, hence, these enzymes can be viewed as potential candidates for practical modulation and genetic mutation implementations. Since FCCs are also readily available from measurements, it is therefore important to compare local FCC-based predictions with those provided by the proposed optimization framework. To facilitate the comparison, the best enzyme sets leading to a substantial increase in the serine production have been organized in Table 1, where the indices highlighted in bold correspond to enzymes exerting high control on the serine bioprocess performance (see Fig. 4).

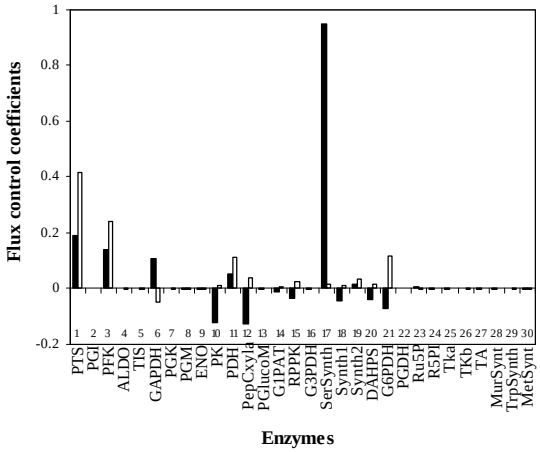

Fig. 4. Flux control coefficients (FCCs) for the serine (solid bars) and PTS (white bars) fluxes.

Table 1. Alternative best enzyme sets leading to increased serine production

Size	Enzyme Set	Flux Ratio
1	**17**	**1.89**
1	8	1.083
1	9	1.082
2	**1**, **17**	**4.651**
2	**10**, **17**	4.03
2	**17**, 19	3.672
3	**1**, **6**, **17**	**9.086**
3	**3**, **6**, **17**	6.599
3	**1**, **10**, **17**	6.258
4	**1**, **3**, **6**, **17**	**14.451**
4	**1**, **6**, 9, **17**	12.221
4	**1**, **6**, 8, **17**	10.719
5	**1**, **3**, **6**, **11**, **17**	**15.933**
5	**1**, **3**, 5, **6**, **17**	15.661
5	**1**, **3**, 4, **6**, **17**	15.146
6	**1**, **3**, 5, **6**, **11**, **17**	**17.418**
6	**1**, **3**, 4, **6**, **11**, **17**	16.905
6	**1**, **3**, 4, 5, **6**, **17**	16.689
7	**1**, **3**, 4, 5, **6**, **11**, **17**	**19.085**
7	**1**, **3**, 5, **6**, **11**, **12**, **17**	17.744
7	**1**, **3**, 5, **6**, **10**, **11**, **17**	17.671
8	**1**, **3**, 4, 5, **6**, 7, **11**, **17**	**19.838**
8	**1**, 2, **3**, 4, 5, **6**, **11**, **17**	19.244
8	**1**, **3**, 4, 5, **6**, **11**, **12**, **17**	19.188
9	**1**, 2, **3**, 4, 5, **6**, 7, **11**, **17**	**20.539**
9	**1**, **3**, 4, 5, **6**, 7, **11**, **17**, **21**	19.906
9	**1**, **3**, 4, 5, **6**, 7, **11**, **12**, **17**	19.898
10	**1**, 2, **3**, 4, 5, **6**, 7, 9, **11**, **17**	**20.591**
10	**1**, 2, **3**, 4, 5, **6**, 7, 8, **11**, **17**	20.574
10	**1**, 2, **3**, 4, 5, **6**, 7, **11**, **12**, **17**	20.57

We find that the best enzyme choices are in complete agreement with the MCA predictions only for the sets comprised of one and two enzymes (i.e., PTS and SerSynth). These enzyme choices are intuitive as the PTS transport system supplies metabolism with the initial substrate, while SerSynth leads to the final serine production. Further analysis reveals that both FCC-based predictions and enzyme choices provided by the optimization framework lack the *additivity* property in a sense that the best choices cannot be just combined one with another to improve

the serine overproduction. For example, the triplet of the most important enzymes for metabolism and serine production (i.e., PTS, PFK, and SerSynth) exerts the highest total control and, yet, is absent from Table 1. These enzymes are, however, present in all larger enzyme sets.

Another important observation resulting from the analysis of Table 1 is that the best choices for large enzyme sets encompass enzymes with both high and low values of FCCs. Therefore, it cannot be deduced from MCA alone why some near-equilibrium enzymes (e.g., TIS, PGK, and ENO) can enter the best enzyme choices while others (i.e., PGM) do not. This can be explained by additional important factors such as complex interactions between demand and supply, regulation, and homeostasis. Specifically, due to homeostasis conditions (2) – (4), the concentration of the near equilibrium PGM, located just below the branching point toward the serine production (see Fig. 2), is lowered to allocate transcriptional rates in favor of the biosynthesis of the modulated enzymes (e.g., the enzymes above the branching point including near-equilibrium TIS, PGK, and ENO). In contrast, the higher level of ENO is enforced to increase the phosphoenolpyruvate concentration for supply to the enhanced PTS system (see Fig. 2). Thus, certain enzymes with low flux control should be also considered for potential modulation to maintain metabolism at homeostasis and, hence, to prevent cellular systems from undesirable or even catastrophic changes due to targeted perturbations.

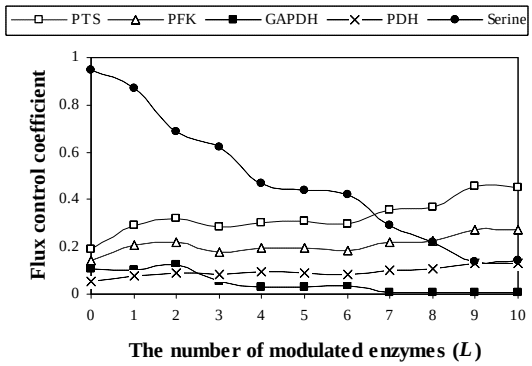

Fig. 5. FCCs are plotted as a function of the enzyme set size L.

The results presented in Table 1 also show how the best enzyme choices emerge. Specifically, while the best choices lack the additivity property, the best smaller sets repeatedly enter the best larger sets. This means that control of flux in the pathway does not shift between different groups of enzymes due to the compensating effects of global regulation and homeostasis. Indeed, stability of distributed control is facilitated by many negative feedback loops, which significantly contribute toward the stabilization of cellular systems at homeostasis (Reich and Selkov 1981; Heinrich and Schuster 1996; Stephanopoulos, *et al.*, 1998). As a result, the FCCs are preserved around their original unperturbed values. The observed effect of stability of distributed control becomes even more pronounced after plotting high FCCs as a function of the enzyme set size *L* (see Fig. 5). The absence of the shift in distributed control emphasizes the importance of the rate limiting enzymes (or pathway steps) with high values of FCCs, computed for the unperturbed reference organism (see Fig. 4 and Table 1). Indeed, an increase in the serine demand reallocates the strength of control from the serine synthesis (i.e., SerSynth) to the supply block (i.e., PTS and PFK), and the pyruvate removal block (i.e., PDH). The FCC for GAPDH is gradually decreased meaning that its activity is saturated.

Interestingly, the best enzyme set predicted encompasses eight glycolytic enzymes (i.e., PTS, PGI, PFK, ALDO, TIS, GAPDH, PGK, and ENO) and two enzymes outside of glycolysis (i.e., PDH and SerSynth – 17). This implies that the high values of FCCs correctly delineate the most important blocks of central metabolism for the serine overproduction from less important subordinate pathways (i.e., PPP and other biosynthetic routes). Specifically, flux is increased through the PTS transport system, PFK, which is a committed enzyme in glycolysis, and PDH to remove an excess of pyruvate accumulated through the enhanced PTS transport system.

5. CONCLUSIONS

A general hybrid stochastic/deterministic optimization framework for optimal selection of enzyme levels using large-scale mathematical models of cellular systems has been introduced and demonstrated on the model of central metabolism in *Escherichia coli* (Chassagnole, *et al.*, 2002; Visser, *et al.*, 2004). A simulated annealing algorithm is employed to navigate through the discrete space of enzyme sets, while general gradient-based search methods are used to estimate optimal enzyme levels. The proposed framework allows for the optimization of the entire cellular system where by systematically selecting *small* enzyme sets, feasible for experimental implementations, significant many-fold production improvements are predicted. The framework can also be used as a powerful tool for the direct validation of modeling predictions and theoretic assumptions, the interconnections between distributed control, cellular economy of supply and demand, and negative feedback stabilization. Alternatively, this framework can be utilized in biomedical studies to identify enzymes controlling undesired large metabolite concentrations and fluxes. Such enzymes can then be ranked as candidates for potential biomarkers of the underlying diseases or drug targets (Bandara, *et al.*, 2003).

5. NOMENCLATURE

Enzymes: aldolase (ALDO), DAHP synthases (DAHPS), enolase (ENO), glucose-1-phosphate adenyltransferase (G1PAT), glycerol-3-phosphate dehydrogenase (G3PDH), glucose-6-phosphate de-hydrogenase (G6PDH), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), isloleucine synthesis

(IleSynth), methionine synthesis (MetSynth), mureine synthesis (MurSynth), phosphofructokinase (PFK), 6-phosphogluconate dehydrogenase (PGDH), glucose-6-phosphate isomerase (PGI), phosphoglycerate kinase (PGK), phosphoglycerate mutase (PGM), pyruvate dehydrogenase (PDH), PEP carboxylase (PEPCyclase), phosphoglucomutase (PGlucoM), pyruvate kinase (PK), phopshotransferase system (PTS), ribose-phosphate isomerase (R5PI), ribose-phosphate pyrophosphokinase (RPPK), ribulose-phosphate epimerase (Ru5P), synthesis1 (Synth1), synthesis2 (Synth2), transaldolase (TA), triosephosphate isomerase (TIS), transketolase A (TKa), transketolase B (TKb), tryptophan synthesis (TrpSynth). *Metabolites*: 1,3-diphosphoglycerate (pgp), 2-phosphoglycerate (2PG), 3-phosphoglycerate (3PG), 6-phosphogluconate (6PG), acetyl-coenzyme A (accoa), dihydroxyacetonephosphate (dhap), erythrose-4-phosphate (e4p), fructose-6-phosphate (f6p), fructose-1,6-bisphosphate (fdp), glucose-1-phosphate (g1p), glucose-6-phosphate (g6p), glyceraldehyde-3-phosphate (gap), glucose (glc), oxaloacetate (oaa), phosphoenolpyruvate (pep), pyruvate (pyr), ribose-5-phosphate (rib5P), ribulose-5-phosphate (ribu5p), sedoheptulose-7-phosphate (sed7p), xylulose-5-phosphate (xyl5p).

ACKNOWELEDGMENTS

Financial support from the Pennsylvania State University, Dean's Fund, the NSF Award BES0120277, and the U.S. DOE is gratefully acknowledged. The authors are very grateful to Profs. Reuss and Schmid for providing us with the detailed kinetic model, and also would like to thank Dr. Burgard for helpful discussions and suggestions.

REFERENCES

Bandara, L. R., M. D. Kelly, E. A. Lock and S. Kennedy (2003). "A potential biomarker of kidney damage identified by proteomics: preliminary findings." Biomarkers **8**(3-4): 272-86.

Castellanos, M., D. B. Wilson and M. L. Shuler (2004). "A modular minimal cell model: purine and pyrimidine transport and metabolism." Proc Natl Acad Sci U S A **101**(17): 6681-6.

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

Cornish-Bowden, A. and M. L. Cardenas (2000). Technological and Medical Implications of Metabolic Control Analysis. Boston, Kluwer Academic Publishers.

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

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

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

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

Kholodenko, B. N. and H. V. Westerhoff (2004). Metabolic engineering in the post-genomics era. Wymondham, UK, Horizon Bioscience.

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

Mauch, K., S. Buziol, J. Schmid and M. Reuss (2001). Computer-Aided Design of Metabolic Networks. Chemical Process Control-6 Conference, Tucson, Arizona.

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

Palsson, B. Ø. (2004). "In silico biotechnology: Era of reconstruction and interrogation." Current Opinion in Biotechnology **15**(1): 50-51.

Reich, J. G. and E. E. Selkov (1981). Energy Metabolism of the Cell. A Theoretical Treatise. New York, Academic Press.

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

Slepchenko, B. M., J. Schaff, I. G. Macara and L. M. Loew (2003). "Quantitative Cell Biology with the Virtual Cell." Trends in Cell Biology **13**: 570-576.

Stephanopoulos, G. N., A. A. Aristidou and J. Nielsen (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., J. W. Schmid, K. Mauch, M. Reuss and J. J. Heijnen (2004). "Optimal re-design of primary metabolism in *Escherichia coli* using linglog kinetics." Biotechnology and Bioengineering.

Young, J., K. Henne, J. Morgan, A. Konopka and D. Ramkrishna (2004). "Cybernetic modeling of metabolism: towards a framework for rational design of recombinant organisms." Chemical Engineering Science **59**(22-23): 5041-5049.