Analysis and Design of Metabolic Reaction Networks via Mixed-Integer Linear Optimization

Vassily Hatzimanikatis
Dept. of Chemical Engineering, California Institute of Technology, Pasadena, CA 92215

Christodoulos A. Floudas
Dept. of Chemical Engineering, Princeton University, Princeton, NJ 08544

James E. Bailey
Institute of Biotechnology, ETH-Zurich, CH-8093 Zurich, Switzerland

Improvements in bioprocess performance can be achieved by genetic modifications of metabolic control structures. A novel optimization problem helps quantitative understanding and rational metabolic engineering of metabolic reaction pathways. Maximizing the performance of a metabolic reaction pathway is treated as a mixed-integer linear programming formulation to identify changes in regulatory structure and strength and in cellular content of pertinent enzymes which should be implemented to optimize a particular metabolic process. A regulatory superstructure proposed contains all alternative regulatory structures that can be considered for a given pathway. This approach is followed to find the optimal regulatory structure for maximization of phenylalanine selectivity in the microbial aromatic amino acid synthesis pathway. The solution suggests that from the eight feedback inhibitory loops in the original regulatory structure of this pathway, inactivation of at least three loops and overexpression of three enzymes will increase phenylalanine selectivity by 42%. Moreover, novel regulatory structures with only two loops, none of which exists in the original pathway, could result in a selectivity up to 95%.

Introduction

Improvements in the product yield, rate of production, and final product concentration are common goals in achieving more efficient and cost-effective bioprocesses. These improvements can be achieved by two main approaches: genetics and process. Process improvements involve the adjustment of the environment of the organisms and the optimization of parallel and downstream processes in order to achieve the best possible performance. Genetic improvements are based on the use of organisms with altered DNA such that their functional characteristics are enhanced. Traditionally, the latter approach has been based on the introduction of random changes in the DNA of a population of organisms and the subsequent selection of an improved organism from the resulting heterogeneous population. However, recent advances in recombinant DNA technology make targeted modifications in the DNA of an industrial microorganism possible. Moreover, recombinant DNA methods enable the introduction into an organism of DNA fragments from other organisms with the possibility of creating hybrid metabolic networks combining features from metabolic networks in different organisms.

All the major cellular pathways are subject to a collection of natural independent control loops with different signals and different loci of action (Sanwal et al., 1971; Savageau, 1976; Stephanopoulos and Vallino, 1991). These mechanisms of metabolic regulation operate at essentially two different levels. Genetic-level controls regulate the expression of genes, thereby determining which enzymes are present and in what quantity. Protein-level controls regulate the activity of particular enzymes and other proteins in the cell. With respect to protein-level controls, each enzyme can be classified as having no such control, or as having activity modulation by one (or more) particular metabolites in the cell.

Correspondence concerning this article should be addressed to J. E. Bailey.

The metabolic control structures in a native, wild-type organism have evolved through natural selection and are therefore configured to maximize the probability of survival of the species, at least insofar as evolution has progressed. While explicit formulation of the objective function for natural metabolism is nontrivial, prior investigators have proposed maximization of growth rates or most efficient utilization of cellular energetic and chemical resources as the objective function for evolution of natural metabolism (Savageau, 1976; Ramkrishna, 1983; Heinrich et al., 1987; Marr, 1991; Schuster and Heinrich, 1991). However, in chemical and pharmaceutical manufacturing that utilize cultivated microorganisms, it is desirable to identify a different configuration of fluxes which directs raw materials to products efficiently at high rates and in the presence of high concentrations of product. A production-oriented evolution is needed to achieve these goals.

Through currently available genetic engineering technology, it is possible to modify both genetic and protein-level regulation. Thus, the amount of a particular enzyme which is expressed under a particular process condition can be altered by changing the genetic information of the organism. Similarly, by changing the gene that codes for a particular enzyme, the response to metabolites that influence its activity can be altered.

Prior research and industrial practice have clearly shown that very large increases in process performance can be realized by genetic modifications of metabolic control systems (Bailey, 1991; Katsumata and Ikeda, 1993). Past improvements in the performance of a process by modification of the control structures were mainly based on trial and error methods and on well-understood, relatively simple pathways. As the complexity of a set of pathways of interest increases, intuitive and trial and error methods are increasingly ineffective. Modifying the regulatory characteristics of an enzyme is presently a much more difficult experimental challenge than changing the amount of enzyme present in the cell. Therefore, guidance as to what changes in regulation might be of greatest benefit to improve the network is important. To this end, a systematic, multilevel, multiparametric methodology for evolving effective control structures is needed.

To achieve a quantitative understanding and rational metabolic engineering of biochemical reaction pathways, mathematical descriptions of metabolic systems have been developed in many cases, and the expected responses of pathways to changes at individual reactions or within certain pathway segments have been calculated. A population of living cells is an extremely complex system, so complicated that some people doubt the possibility of a credible mathematical description, and therefore of quantitative engineering design, of any of its attributes. Two different bodies of experience contradict this view. First, several complex biological phenomena have been well described by mathematical models which are based on the essential molecular mechanism. Examples include regulated gene expression (Lee and Bailey, 1984b,c), replication of DNA (Lee and Bailey, 1984a), growth of bacterial cells (Shuler and Domach, 1983), animal cell cycle regulation (Hatzimanikatis et al., 1996), and receptor trafficking (Starbuck and Lauffenberger, 1992). Second, design and control of most industrial chemical processes, ranging from catalytic cracking to olefin polymerization are based upon mathematical models which are known to be only crude approximations of physical reality. Most chemical engineering applications involve partially understood, approximately described complex physical systems. Useful engineering has been achieved in many facets of chemical engineering endeavor in spite of this. There is no reason to expect a different outcome in the engineering of complicated, imperfectly known cellular processes.

The modeling approaches previously used can be classified into two kinds: linear and nonlinear. Linear models can be accessed through analysis of input-output relations and certain stimulus-response experiments by applying advanced regression analysis (Schlosser et al., 1993) or other experimental methods developed within and around the metabolic control analysis (MCA) framework (Kacser and Burns, 1973; Heinrich and Rapoport, 1974; Cornish-Bowden and Cardenas, 1990; Fell, 1992). Nonlinear models, on the other hand, can be constructed when detailed kinetic expressions for each step in the reaction pathway are known or can be estimated (Joshi and Palsson, 1989; Gallazo and Bailey, 1990). Because of the greater data requirements for nonlinear model formulation and validation linear (or log-linear, see below) models will often be the only practically accessible description.

Using the available mathematical description of a biochemical system, various analytical and computational techniques can be used for analysis and optimization of the system. Optimization techniques have been used in the analysis of biochemical systems using mainly stoichiometric information and metabolic requirements for growth (Majewski and Domach, 1990; Stephanopoulos and Vallino, 1991; Varma and Palsson, 1994). These approaches do not require kinetic data; therefore, they cannot be used in order to quantify the effects of genetic modifications of enzyme levels and of regulatory structures. Optimization of metabolic pathways on the basis of a kinetic model developed using experimental data represented mathematically using the S-system formalism of biochemical systems theory (BST) has been presented using linear programming (Voit, 1992; Regan et al., 1993). These studies provide information only about optimum manipulation of the external inputs to the system (such as independent effectors and external substrates) and do not address the problem of optimizing the regulatory structure of the metabolic network.

The objective of this work is to provide a mathematical framework for determining changes in regulatory structure and strength which should be considered to optimize a particular metabolic process. A mixed-integer linear programming (MILP) formulation is proposed for the general case of linear model optimization. The solution of the MILP formulation provides information on which enzymes should be present at different levels, the extent of such changes needed, and the accompanying modifications in the regulatory structure that will optimize the process.

Any mathematical description of cellular processes is an approximation, and, genetic manipulation of the cell may cause secondary responses which were not considered in the mathematical (or the conceptual) model. Therefore, the output which the metabolic engineer seeks and all that can be expected from such optimization calculations are reasonable suggestions for changes in the metabolic network which might give useful improvements in cellular performance. Strategic

guidance, not fine quantitative rules, is the intended outcome of these type of calculations. A large body of prior experience with engineering of other complex chemical systems clearly indicates the value of such an approach, even when models are crude approximations, relative to completely ad hoc approaches which are, of course, necessarily based on more crude mental optimization of much more crude mental models.

Problem Statement

This article deals with a mathematical description of a metabolic pathway with a postulated number of regulatory loops. These regulatory loops are classified as either activation (increase the activity of the regulatory enzyme) or inhibition (decrease the activity of the regulatory enzyme) loops. The objective is to determine which of the regulatory loops should be retained, and the number, type, and level of manipulation of amounts of enzymes to optimize a certain function of the outputs of the metabolic pathway (such as production of primary or secondary metabolites, growth, and selectivity).

Many metabolic pathways are common to many organisms. However, enzymes that catalyze the same reaction in different organisms are not necessarily the same in their catalytic and regulatory properties. As discussed in the introduction, recombinant DNA methods enable the introduction into an organism of the DNA from other organisms with the possibility then of combining regulatory features from the metabolic pathways present in these different organisms. Moreover, protein engineering methods allow modifications of the properties of natural enzymes and design of enzymes with novel regulatory characteristics. Therefore, the possible number of regulatory loops for a certain pathway in an organism ranges from the number of the existing loops to this number plus the number of additional, different regulatory loops which can be introduced into the same pathway by genetic engineering.

Most generally, we can consider a *regulatory superstructure* in which every metabolite in the system can potentially regulate any enzyme in that system. The mathematical formulation of such a general regulatory superstructure leads to a large combinatorial problem. Its solution will provide the maximum or minimum performance achievable for a given system, and thus provide valuable insight for protein and metabolic engineering. This problem addresses the question of how regulation and catalyst levels should be chosen *de novo* in order to maximize the performance of the metabolic network.

Mathematical Description of Metabolic Reaction Networks

We consider here a linear model description for metabolic systems. This is the most common situation because of limitations in the available information for most systems. Furthermore, when a nonlinear model is available, it can be linearized and studied within the same framework. In what follows we describe the linearization procedure in a way similar to that presented by Reder (1988).

Consider a metabolic system consisting of n metabolites and m enzymatically-catalyzed reactions. We are in interested in studying how modifications of the expression levels and of the properties of the enzymes that catalyze these reactions affect metabolic functions of the system, such as metabolite concentrations, fluxes, and specific growth rate. Mass balances on the metabolites of the system may be written

$$\frac{dx}{dt} = f[v(x;p),x;p] \tag{1}$$

where x is the n-dimensional metabolite concentration vector, f is a function determined by the mass balances, v is the m-dimensional reaction rate vector, and p is the s-dimensional manipulated parameter vector (such as enzyme concentrations). In addition to metabolite reaction rates, the mass balance equations also include terms that account for other processes by which concentrations of metabolites change (such as the dilution brought about by increases in the biomass volume (Fredrickson, 1976) and transport through the cell wall envelope).

In addition, consider the r-dimensional vector of metabolic outputs h for which we have

$$h = h[v(x;p),x;p] \tag{2}$$

In Eq. 2 h is a function of the rates of interest, of the metabolite concentration, and of the parameters. Linearization of Eqs. 1 and 2 around a steady state (x_o, p_o) results in the following linear system for the logarithmic deviations (see Appendix A for explanation of the transformation)

$$\frac{dz}{dt} = N\epsilon z + \mathcal{K}z + N\Pi q + \Lambda q \tag{3}$$

$$w = \Xi\epsilon z + \mathcal{K}z + \Xi\Pi q + \Theta q \tag{4}$$

where z, q, and w are the logarithmic deviations of the metabolite concentrations, the enzyme levels, and the metabolic outputs, respectively

$$z_i = \ln(x_i/x_{i,o})$$
$$q_k = \ln(p_k/p_{k,o})$$
$$w_l = \ln(h_l/h_{l,o})$$

and N, Ξ, $\mathcal{K}$, Λ, $\mathfrak{K}$, Θ, ϵ, and Π are matrices, defined as

$$N = \left\{ n_{i,j} | n_{i,j} = \frac{v_{j,o}}{x_{i,o}} \left(\frac{\partial f_i}{\partial v_j} \right)_{x_o,p_o} \right\},$$

$$\Xi = \left\{ \xi_{l,j} | \xi_{l,j} = \frac{v_{j,o}}{h_{l,o}} \left(\frac{\partial h_l}{\partial v_j} \right)_{x_o,p_o} \right\},$$

$$\mathcal{K} = \left\{ \kappa_{i,k} | \kappa_{i,k} = \frac{x_{k,o}}{x_{i,o}} \left(\frac{\partial f_i}{\partial x_k} \right)_{x_o,p_o} \right\},$$

$$\Lambda = \left\{ \lambda_{i,k} | \lambda_{i,k} = \frac{p_{k,o}}{x_{i,o}} \left(\frac{\partial f_i}{\partial p_k} \right)_{x_o,p_o} \right\},$$

$$\mathfrak{IC} = \left\{ \eta_{l,i} | \eta_{l,i} = \frac{x_{i,o}}{h_{l,o}} \left(\frac{\partial h_l}{\partial x_i} \right)_{x_o, p_o} \right\},$$

$$\Theta = \left\{ \theta_{l,k} | \theta_{l,k} = \frac{p_{k,o}}{h_{l,o}} \left(\frac{\partial h_l}{\partial p_k} \right)_{x_o, p_o} \right\},$$

$$\epsilon = \left\{ \epsilon_{j,i} | \epsilon_{j,i} = \frac{x_{i,o}}{v_{j,o}} \left(\frac{\partial v_j}{\partial x_i} \right)_{x_o, p_o} \right\},$$

$$\Pi = \left\{ \pi_{j,k} | \pi_{j,k} = \frac{p_{k,o}}{v_{j,o}} \left(\frac{\partial v_j}{\partial p_k} \right)_{x_o, p_o} \right\},$$

Here the subscript o indicates the reference steady state about which the approximate model is developed. Thus, the righthand side of Eq. 1 is zero when evaluated at $(\boldsymbol{x}_o, \boldsymbol{p}_o)$, and $\boldsymbol{h}_o$ denotes the value of $\boldsymbol{h}$ at $(\boldsymbol{x}_o, \boldsymbol{p}_o)$.

It should be noted that a mathematical description linear in logarithms of the system variables is in fact a nonlinear (power law) representation, a functionality well suited to approximating closely the nonsingular rational polynomial kinetic expressions typical of metabolic processes. The quality of this form of approximate representation of metabolic kinetics will be considered in detail in a future publication. Here the nature of this approximation will be tested for one of the examples presented by calculating the consequences for the original nonlinear model (Eqs. 1 and 2) of the optimization strategy determined using the log-linear approximate model of Eqs. 3 and 4.

At steady state, solution of Eqs. 3 and 4 yields

$$\boldsymbol{w} = \boldsymbol{C}\boldsymbol{q} \tag{5}$$

where

$$\boldsymbol{C} = -(\Xi\epsilon + \mathfrak{IC})(N\epsilon + \mathfrak{IC})^{-1}(N\Pi + \Lambda) + \Xi\Pi + \Theta \tag{6}$$

with

$$\boldsymbol{C} = \left\{ c_{l,k} | c_{l,k} = \frac{p_{k,o}}{w_{l,o}} \left(\frac{dw_l}{dp_k} \right)_{x_o, p_o} \right\}$$

The mathematical description described above depends on the same information as that employed within the framework of metabolic control analysis (MCA) (Reder, 1988; Schlosser and Bailey, 1990). Matrices ϵ and Π are the elasticity matrices with respect to metabolites and to parameters, respectively. The matrix $\boldsymbol{C}$ is the control coefficient matrix of the metabolic functions $\boldsymbol{h}$ with respect to parameters $\boldsymbol{p}$. Experimental determination of the parameters for this linear system has been the subject of several studies (Fell, 1992; Cornish-Bowden and Cardenas, 1990; Schlosser and Bailey, 1990) and in many cases it is the only available description of a metabolic system.

Analysis and Synthesis Problems

The regulatory structure of a metabolic network is typically deduced from experimental analysis of the integrated system or from the reported kinetic properties of the enzymes involved in the pathway. In this case the matrix ϵ can be written as a sum of two matrices

$$\epsilon = \epsilon^s + \epsilon^r \tag{7}$$

where the elements in matrix ϵ^s correspond to the substrate elasticities of the enzymes, that is, the sensitivities of enzyme activities with respect to their substrates, and the elements of matrix ϵ^r correspond to the regulatory elasticities of the enzymes, that is, the sensitivities of enzyme activities with respect to regulatory metabolites. In this representation, the substrates themselves can also be considered as regulatory metabolites (in cases such as substrate inhibition)

$$\epsilon^s = \left\{ \epsilon_{j,i_s} = \frac{x_{i_s,o}}{v_{j,o}} \left(\frac{\partial v_j}{\partial x_{i_s}} \right)_{x_o, p_o} | x_{i_s} \text{ is a substrate for reaction } j \right\}$$

$$\epsilon^r = \left\{ \epsilon_{j,i_r} = \frac{x_{i_r,o}}{v_{j,o}} \left(\frac{\partial v_j}{\partial x_{i_r}} \right)_{x_o, p_o} | x_{i_s} \text{ is a regulator for reaction } j \right\}$$

Changes of the elements in matrix ϵ^r from nonzero values to zero or vice versa correspond to modifications in the regulatory structure of the system. In this study which emphasizes the role of control structure, we define the *analysis problem* in the context of a given control structure which we reasonably assume can be modified only by deleting certain control interactions: Which of the existent regulatory loops should be inactivated, and what associated changes should be made in the manipulated variables (such as enzymes expression levels, environmental conditions, and effectors external to the system), in order to optimize the performance of the metabolic network?

The *synthesis problem* considers the possibility of postulating a *regulatory superstructure* and addressing the problem of selecting among alternative regulatory structures for each enzyme. The regulatory superstructure embeds a set of alternative regulatory elasticities for each enzyme that correspond to different kinds of regulation by each metabolite. In particular, the *synthesis problem* addresses the following question: What kind of regulation (activation or inhibition, by which metabolite and of what strength) should be assigned to each enzyme in the network, and what associated changes should be made in the manipulated parameters (such as enzyme expression levels, environmental conditions, and effectors external to the system), in order to optimize the performance of the metabolic network?

The analysis and synthesis problems include discrete decisions concerning the regulatory elasticities of the system. Inactivation or activation of a regulatory loop is equivalent to elimination or introduction of nonzero terms in the ϵ^r matrix. Moreover, the synthesis problem will typically be subject to some constraints such as the possible number of regulatory actions on each enzyme and the requirement that an enzyme cannot be activated and inhibited by the same metabolite.

On the other hand, the continuously adjustable manipulated parameters can potentially be subject to discrete constraints, such as the maximum number of these parameters that we can manipulate simultaneously.

The mixed discrete and continuous nature of the problem and the linear description of the system lead to the formulation of the analysis and synthesis problems as MILP problems, solutions of which provide the optimal regulatory structure and the optimal parameter configuration of a metabolic reaction network.

Mathematical Formulation

The mathematical formulation for the synthesis problem as a MILP problem is presented in this section. To derive the mathematical formulation, the following index sets and variables are introduced to characterize the postulated regulatory superstructure. The metabolites will be denoted by the index set $I=\{i\}$, the reaction rates by the index set $J=\{j\}$, the manipulated parameters by the index set $K=\{k\}$, and the metabolic outputs by the index set $L=\{l\}$. The following sets will be defined to establish the connections of the sets of metabolites with the reaction rates in the network

$$I_s^j=\{i_s|i_s \in I \text{ is a substrate for reaction } j,\ j\in J\}$$

$$I_r^j=\{i_r|i_r \in I \text{ is a regulator for reaction } j,\ j\in J\}$$

The sets $M^+=\{m^+\}$ and $M^-=\{m^-\}$ denote the indices for the activation and the inhibition elasticities, respectively, that can be applied to each enzyme by each metabolite. The regulatory elasticities of the postulated regulatory superstructure will be denoted as ϵ_{mji_r}, where m belongs to the index set $M=M^+\cup M^-$. Therefore, ϵ_{mji_r} will denote the regulatory elasticity of reaction j with respect to metabolite i and can be positive (for activation) or negative (for inhibition) with a fixed magnitude for each m.

The continuous variables of the model are the logarithmic deviations of the metabolite concentrations z_i, the logarithmic deviations of the manipulated variables q_k, the logarithmic deviations of the metabolic outputs w_l and the reaction rates v_j.

A binary variable y_{mji_r} is associated with each regulatory elasticity ϵ_{mji_r}. If a regulatory loop with an elasticity ϵ_{mji_r} is active in the network, y_{mji_r} is set to 1; otherwise it is zero. A binary variable d_k is associated with each manipulated variable q_k. The introduction of these variables serves as a control on the number of the manipulated parameters that are allowed to vary. In many cases, practical experimental limitations allow only a limited number of simultaneous manipulations of enzyme activities.

The linearization procedure described earlier transforms the rate expressions v_j and metabolic output functions h_l to the following form

$$v_j=v_{j,o}\left(1+\sum_{i_s\in I_s^j}\epsilon_{ji}^s z_{i_s}+\sum_{m\in M}\sum_{i_r\in I_r^j}\epsilon_{mji_r}^r y_{mji_r} z_{i_r}+\sum_{k\in K}\pi_{jk}d_k q_k\right) \quad (8)$$

and

$$w_l=\ln(h_l/h_{l,o})=\sum_{j\in J}\sum_{i_s\in I_s^j}\xi_{lj}\epsilon_{ji_s}^s z_{i_s}+\sum_{m\in M}\sum_{j\in J}\sum_{i_r\in I_r^j}\xi_{lj}\epsilon_{mji_r}^r y_{mji_r} z_{i_r}+\sum_{i\in I}\eta_{li}z_i+\sum_{k\in K}\sum_{j\in J}\xi_{lj}\pi_{jk}d_k q_k+\sum_{k\in K}\theta_{lk}d_k q_k \quad (9)$$

The second and third terms on the righthand side of Eq. 8 correspond to the dependence of the rates on the metabolites, and the fourth term corresponds to the dependence on the manipulated parameters. The first and the second terms in Eq. 9 correspond to the first term in Eq. 4 and the third, fourth, and fifth terms correspond to the second, third, and fourth terms in Eq. 4, respectively.

The presence of binary variables in the formulation of the problem introduces bilinear products of continuous and binary variables in Eqs. 8 and 9 which make the problem nonlinear. To circumvent these nonlinearities we follow the modeling approach applied in Psarris and Floudas (1990) using the idea proposed by Petersen (1971) and extended by Glover (1975). This is also discussed in Chapter 7 of the book by Floudas (1995).

The basic idea is to introduce new continuous variables for each bilinear product

$$u_{mji_r}=y_{mji_r}z_{i_r}\ \forall(m,j,i_r) \quad (10)$$

and

$$g_k=d_k q_k\ \forall k \quad (11)$$

and to introduce additional constraints for each (m, j, i_r) and each k, which are described in the subsection on constraints.

Objective function

The process of interest which we optimize (maximize or minimize) can be any of the metabolic outputs or combination of them. Note that using Eq. 9 we can treat metabolite concentrations and rates as metabolic outputs. Therefore, in the case of having a single metabolic function w_l as objective we can express the objective function as

$$f_{\text{obj}}=\sum_{j\in J}\sum_{i_s\in I_s^j}\xi_{lj}\epsilon_{ji_s}^s z_{i_s}+\sum_{m\in M}\sum_{j\in J}\sum_{i_r\in I_r^j}\xi_{lj}\epsilon_{mji_r}^r u_{mji_r}+\sum_{i\in I}\eta_{li}z_i+\sum_{k\in K}\sum_{j\in J}\xi_{lj}\pi_{jk}g_k+\sum_{k\in K}\theta_{lk}d_k q_k \quad (12)$$

Coupling between the various processes in the cell is sometimes a limitation in the performance of a pathway, even when classical metabolic engineering techniques are applied. However, we can study within the present framework the following questions: If and how can we decouple the pathway of interest from other cellular processes by manipulating the regulatory structure of the pathway? If possible, how can this decoupling be achieved? The objective function that corre-

sponds to this question will be formulated and a corresponding solution approach will be discussed later.

Constraints

The metabolic optimization problem is typically subject to one or more of the following types of constraints:

(i) Mass Balance for Each Metabolite i. Equation 3 is the set of mass balance equations for the metabolites. Therefore, at steady state, the lefthand side of Eq. 3 is set to zero, and, for each metabolite x_i, we have a constraint of the form

$$\sum_{j \in J} \sum_{i_s \in I_s^j} n_{ij} \epsilon_{ji_s} z_{i_s} + \sum_{m \in M} \sum_{j \in J} \sum_{i_r \in I_r^j} n_{ij} \epsilon_{ji_r} u_{mji_r} + \sum_{i' \in I} \kappa_{ii'} z_{i'} + \sum_{j \in J} \sum_{k \in K} n_{ij} \pi_{jk} g_k + \sum_{k \in K} \lambda_{ik} g_k = 0 \quad (13)$$

where the first two terms correspond to the first term in the righthand side of Eq. 3, and the third, fourth, and fifth terms correspond to the second, third, and fourth terms in the righthand side of Eq. 3, respectively.

(ii) Bounds on Metabolites, Manipulated Variables, Rates, and Metabolic Outputs. In modeling metabolic pathways it is extremely difficult to describe all relevant processes completely. Metabolic engineering of the pathway of interest will result in changes in metabolite concentrations, metabolic outputs of the pathway, and reaction rates. These changes will propagate into the rest of the cellular processes with unpredictable and, in many cases, undesirable effects.

The concentrations of metabolites should neither exceed toxicity levels nor be very low, because it is possible to induce responses such as stringent responses that will alter qualitatively cellular activities that are not included in the model. Changes in the manipulated variables can have similar effects. Overexpression of enzymes can influence growth, and an excess of external substrate can result in toxic byproduct synthesis. Therefore, variables should be constrained within the bounds determined by their physiological ranges for the pathway of interest and by the available biological knowledge.

$$z_i^L \le z_i \le z_i^U \quad i \in I \quad (14)$$

$$q_k^L \le q_k \le q_k^U \quad k \in K \quad (15)$$

Note that variables z_i and q_k are the logarithmic deviations of the concentrations x_i and the logarithmic deviations of the parameters p_k, respectively, from the steady-state value around which the linear model has been constructed. As a result, their lower bounds can take negative values.

The reaction rates of the pathway cannot be increased infinitely, and zero values for the fluxes are not generally acceptable physiologically. The minimum lower bound for a rate will be zero only if the metabolites produced by the corresponding reaction are provided externally. Reaction rate expressions v_j and metabolic outputs of the system w_l will generally be constrained within physiological bounds depending on the system under study. Therefore, the bounds of these variables will be of the form

$$v_j^L \le v_j \le v_j^U \quad j \in J \quad (16)$$

$$w_l^L \le w_l \le w_l^U \quad l \in L \quad (17)$$

The variables w_l represent the logarithmic deviations of the metabolic outputs and, therefore, their lower bounds can be negative.

(iii) Constraints for the u_{mji_r} Variables. The variables represented by u_{mji_r} are connected with the continuous variables z_{i_r} and the binary variables y_{mji_r} via the following conditions

$$z_{i_r} - z_{i_r}^U(1 - y_{mji_r}) \le u_{mji_r} \le z_{i_r} - z_{i_r}^L(1 - y_{mji_r}) \quad (18)$$

$$i_r \in I_r^j,\ j \in J,\ m \in M$$

$$z_{i_r}^L y_{mji_r} \le u_{mji_r} \le z_{i_r}^U y_{mji_r} \quad (19)$$

$$i_r \in I_r^j,\ j \in J,\ m \in M$$

Note that the above constraints are linear in z_{i_r} and y_{mji_r}. It is interesting to examine the effect of these constraints.

If $y_{mji_r} = 1$, then they become

$$z_{i_r} \le u_{mji_r} \le z_{i_r}$$

$$z_{i_r}^L \le u_{mji_r} \le z_{i_r}^U$$

and the first two constraints imply that $u_{mji_r} = z_{i_r}$ while the second two constraints simply provide bounds.

If $y_{mji_r} = 0$, then we have

$$z_{i_r} - z_{i_r}^U \le u_{mji_r} \le z_{i_r} - z_{i_r}^L$$

$$0 \le u_{mji_r} \le 0$$

and the second two constraints imply that $u_{mji_r} = 0$, while the first two constraints are relaxed since $z_{i_r} - z_{i_r}^U \le 0$ and $z_{i_r} - z_{i_r}^L \ge 0$.

Similarly, for each of the g_k variables we have the following four inequality linear constraints

$$q_k - q_k^U(1 - d_k) \le g_k \le q_k - q_k^L(1 - d_k) \quad k \in K \quad (20)$$

$$q_k^L d_k \le g_k \le q_k^U d_k \quad k \in K \quad (21)$$

for which a similar analysis holds.

(iv) Logical Constraints. There are constraints based on the binary variables which are associated with the existence or nonexistence of various regulatory loops and the activation or deactivation of different continuously adjustable manipulated variables. An important logical constraint is one that forbids activation and inhibition of an enzyme by the same metabolite. Moreover, when we consider alternative loops with the same type of action (inhibition or activation) at different levels (low or high inhibition), only one of the values should be considered. These constraints appear in the formulation for each (j, i_r)

$$\sum_{m^+ \in M^+} y_{m+ji_r} + \sum_{m^- \in M^-} y_{m-ji_r} \le 1 \quad \forall (j,i_r) \qquad (22)$$

$$m^+ \in M^+,\ m^- \in M^-,\ j \in J,\ i_r \in I_r^j$$

The maximum number of the regulatory actions for each enzyme $|I^j_{r,\max}|$ and the maximum number of enzymes that each metabolite regulates $|J^{i_r}_{\max}|$ will, in general, impose one additional constraint for each j and one for each i

$$\sum_{m \in M} \sum_{i_r \in I_{i_r}^j} y_{mji_r} \le |I^j_{r,\max}| \quad j \in J \qquad (23)$$

and

$$\sum_{m \in M} \sum_{j \in J} y_{mji_r} \le |J^{i_r}_{\max}| \quad i_r \in I_{i_r}^j \qquad (24)$$

The simultaneous manipulation of the variables q_k will be subject to the following constraint

$$\sum_{k \in K} d_k \le |K_{\max}| \qquad (25)$$

where $|K_{\max}|$ is the maximum number of the manipulated variables that can be modified simultaneously. This constraint arises from practical and physiological limitations, and $|K_{\max}|$ varies from system to system.

Integer cuts are also introduced such that by solving the proposed mathematical model in an iterative way, we can exclude all the previous solutions to calculate the next best solution. In this case, when we solve for the n-th best solution we have to include $n-1$ constraints of the form

$$\sum_{i \in B_i} y_i - \sum_{i \in NB_i} y_i \le |B_i| - 1$$

$$B_i = \{i | y_i = 1\}$$

$$NB_i = \{i | y_i = 0\} \qquad (26)$$

where $|B_i|$ is the cardinality of the set B_i (the number of the elements in the set).

The proposed mathematical formulation involves continuous and binary variables. The problem, in its initial formulation, features bilinearities as products of continuous and binary variables. By introducing a new continuous variable and four linear inequality constraints for each bilinear term, the final formulation involves only linear terms and, therefore, becomes a MILP formulation. The solution of the MILP will provide the desired optimal regulatory structure. Because of its linear nature, the problem is convex and a global solution is guaranteed. The model allows constraint flexibility in relation with the physical system it describes and is designed to encompass feasible manipulations within the set of current methods for metabolic and protein engineering

Computational Studies

The proposed approach will be illustrated using the aromatic amino acid biosynthetic network in bacteria as an example system. Four specific problems will be postulated and solved with the proposed mixed integer linear optimization framework. The pathway and the original regulatory structure are presented in Figure 1. The pathway has eight regulatory loops all of which are feedback inhibitory loops. In order to derive a linear model for the pathway we started from the nonlinear model for this presented by Schlosser and Bailey (1990). The nonlinear model and the parameters for the linear model are presented in Appendix B.

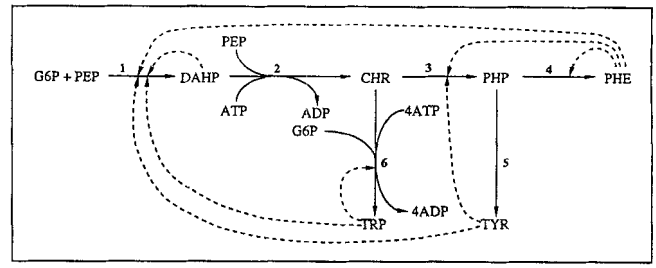

Figure 1. Aromatic amino acid synthesis pathway.

Solid arrows indicate reactions and dashed arrows indicate feedback inhibition loops. Chemical species: G6P = glucose-6-phosphate; PEP = phosphoenolpyruvate; ATP = adenosine triphosphate; ADP = adenosine diphosphate; DAHP = 3-deoxy-D-arabino-heptulosonate-7-phosphate; CHR = chorismate; PHP = prephenate; PHE = phenylalanine; TYR = tyrosine; TRP = tryptophan.

In the following examples, the following bounds on the logarithmic deviations of the metabolite concentrations were imposed

$$-2 \le z_i \le 2 \qquad (27)$$

which implies that we allow the concentrations of the metabolites to vary between 13.5% and 639% of the reference steady state. These wide bounds might allow concentrations that can be toxic to the organism or that induce stress responses that will affect other cellular activities. We use such wide bounds in the interest of exploring the structure of this example and the characteristics of its optimal solution. If needed to accommodate physiological limitations, tighter constraints on metabolite concentrations can certainly be considered within this formalism.

The only physiological constraint which will be introduced constrains the specific growth rate μ to its reference steady value

$$w_\mu = 0 \qquad (28)$$

The only consideration in the model used in this example for the effects of manipulation of the aromatic amino acid biosynthesis network on the rest of the organism's metabolism is the dependence of the specific growth rate on the aromatic amino acid concentration levels (Appendix B). By imposing the above constraint, we essentially constrain the solutions to the ones that are consistent with other requisite coupled metabolic activities of the cell.

The first example deals with the modification of the existing regulatory and activity structure. The second, the third, and the fourth examples assume that the metabolic pathway has no regulation at the outset, and consider what regulatory

connections, with what strength, should be introduced so as to optimize the objective.

The procedure for solving the examples, equally applicable to any other metabolic network optimization problem of the class formulated earlier, was implemented using the high-level modeling language GAMS (general algebraic modeling system), into an algorithmic procedure named METAOPT (METAbolic network OPTimization). The procedure accepts the mathematical model and the postulated regulatory superstructure as a set of matrices and is interfaced with CPLEX, a mixed-integer linear programming solver. At each solution, the optimal regulatory structure is used to form an integer cut constraint, and the problem is solved again for the next best structure. This way a sequence of several solutions is generated. The following computational studies were run on a HP/730 workstation with a Unix-based operating system.

Problem 1

The question addressed in the first problem can be stated as follows: Which of the existing regulatory loops should be inactivated and what should be the associated changes in the enzyme expression levels to maximize the phenylalanine selectivity?

The phenylalanine selectivity is defined here as the rate of phenylalanine production divided by the overall rate of all aromatic amino acids (phenylalanine, tyrosine, and tryptophan) production

$$S_{\text{phe}} = \frac{v_4}{v_4 + v_5 + v_6} \tag{29}$$

The initial number of the regulatory loops is equal to eight, which can be either active or inactive and in any possible combination. Therefore, there are $2^8 = 256$ alternative regulatory structures. The six enzymes of the pathway are the continuously adjustable variables for which we set the bounds

$$0 \le q_k \le \ln(2) \quad k = 1, \ldots, 6 \tag{30}$$

The zero value for the lower bound means that enzyme down-regulation is not feasible. On the other hand, the value ln(2) for the upper bound allows enzyme overexpression up to twice the level of the reference state. In practice, overexpression of an enzyme can result in higher levels; however, since we study a linear model for which the enzyme levels are the inputs, we do not permit large deviations in the input values so as to stay better within the range in which the linear model is a good approximation to the original nonlinear model. Only the six enzymes are considered as the continuously adjustable inputs, and the concentrations of the precursors (metabolites feeding into the reaction network considered here) remain constant at their reference values. Therefore, the following constraints are needed to inactivate changes in the precursors

$$d_k = 0 \quad k = 7,8,9 \tag{31}$$

The objective function to be maximized is the ratio of the phenylalanine production rate divided by the overall rate of the aromatic amino acids production. Equation 29 in linearized form is written as in Eq. 4

$$w_{S_{\text{phe}}} = \ln\left(\frac{S_{\text{phe,opt}}}{S_{\text{phe},o}}\right) = \Xi_{S_{\text{phe}}} \epsilon z + \mathcal{H}_{S_{\text{phe}}} z + \Xi_{S_{\text{phe}}} \Pi q + \Theta_{S_{\text{phe}}} q \tag{32}$$

where

$$\Xi_{S_{\text{phe}}} = [0, 0, 0, 0.572665, -0.439811, -0.132855]$$

and $\mathcal{H}_{S_{\text{phe}}}$ and $\Theta_{S_{\text{phe}}}$ are zero vectors. The selectivity for the reference state is

$$S_{\text{phe},o} = 0.427335.$$

No improvement in this value could be achieved only by enzyme overexpression, without having an effect on the growth rate.

The MILP optimization model discussed in the previous section was solved and four alternative regulatory structures were identified corresponding to the optimal value for the phenylalanine selectivity. The problem consisted of 65 variables (48 continuous, 17 binary) and 94 equations, and the first optimal solution was found within 0.26 CPUs. The structures are presented in Figure 2 (cases a–d) and the value of the objective function for all four solutions is

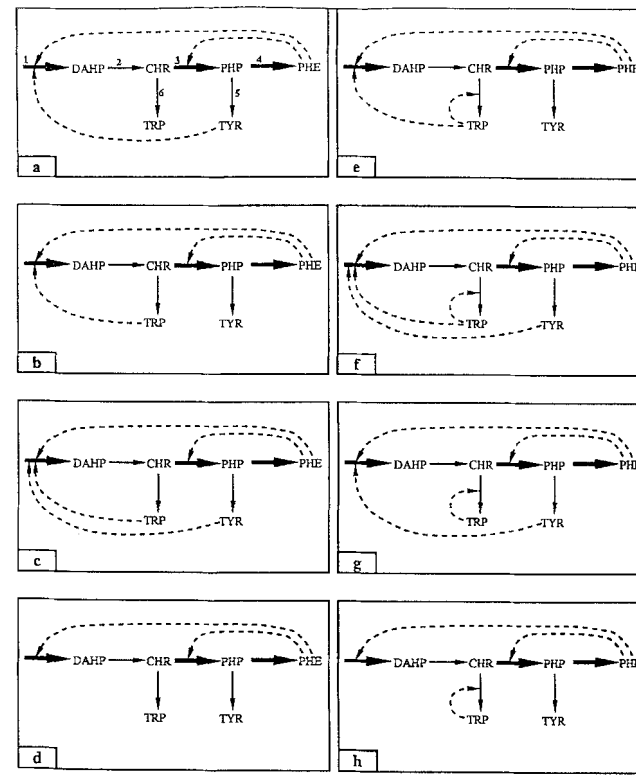

Figure 2. Eight best solutions from problem 1.

Solid arrows indicate reactions, dashed arrows indicate inhibitory loops, and thick solid arrows indicate enzyme overexpression for the respective reaction. In solutions b–h the reaction numbering has been omitted for clarity.

Table 1. Continuous Variables for the First Four Best Solutions of Problem 1

	z_i	q_i			
i	Cases a–d	Case a	Case b	Case c	Case d
1	−0.28302	0.13926	0.21329	0.13837	0.21418
2	−1.0	0.0	0.0	0.0	0.0
3	−0.67889	0.36149	0.36149	0.36149	0.36149
4	0.45238	0.69315	0.69315	0.69315	0.69315
5	−0.59375	0.0	0.0	0.0	0.0
6	−0.10233	0.0	0.0	0.0	0.0

$$S_{\text{phe,opt}} = 0.605883$$

The overexpressed enzyme levels and metabolite concentration levels associated with each regulatory structure are presented in Table 1. The concentrations of the metabolites are the same for all four solutions. The results suggest that the optimal selectivity can be achieved by inactivation of at least four regulatory connections (v_1 by DAHP, v_3 by TYR, v_4 by PHE, and v_6 by TRP) and that three enzymes should be overexpressed, specifically the enzymes that catalyze reactions 1, 3 and 4.

Four more alternative regulatory structures correspond to a value of 0.601315 for the selectivity. These structures are also presented in Figure 2 (cases e–h), and the associated enzyme and metabolite levels are presented in Table 2. Again, the three enzymes that should be overexpressed correspond to reactions 1, 3, and 4. However, the minimum number of the regulatory loops that should be inactivated is three (v_1 by DAHP, v_3 by TYR, and v_4 by PHE). Case f among these results is attractive since implementing it requires less effort than other solutions presented from the genetic and protein engineering point of view.

Examination of the eight regulatory structures indicates that the phenylalanine selectivity can be significantly improved, while maintaining constant specific growth rate, by inactivating at least three regulatory structures and overexpressing three enzymes. While inactivation of the inhibition of v_4 by PHE is quite obvious, the rest of the manipulations, all considered subject to the constraint on the growth rate, cannot be easily anticipated *a priori*.

In this problem which begins with a specified nonlinear model including enzyme regulation, the metabolic design strategy determined by the MILP optimization of an approximate log-linear model can be tested using the original nonlinear metabolic model. Such calculations for the strategies designated a–d give $S_{\text{phe}} = 0.66730$, and $S_{\text{phe}} = 0.86544$ for the strategies designated e–h. These selectivities substantially exceed that for the reference state, indicating that, as desired, the approach described here provides useful guidance towards effective metabolic design. It should be noted that the full nonlinear model, when subjected to the modifications computed using the log-linear model, no longer displays all the properties of the log-linear model. In particular, all of the constraints imposed are no longer exactly satisfied. For example, specific growth rate is decreased somewhat when these strategies are applied to the original nonlinear model. The changes observed are well within an acceptable range from a biological and process point of view. The purpose of this work is to provide reasonable guidance for starting optimization of a metabolic system. Efforts to seek an optimum for a real system will always require further experimental refinement.

Table 2. Continuous Variables for the Second Four Best Solutions of Problem 1

	z_i	q_i			
i	Cases a–d	Case e	Case f	Case g	Case h
1	−0.31840	0.17594	0.09728	0.09766	0.17631
2	−1.0	0.0	0.0	0.0	0.0
3	−0.71270	0.31102	0.31102	0.31102	0.31102
4	0.41735	0.69315	0.69315	0.69315	0.69315
5	−0.62333	0.0	0.0	0.0	0.0
6	−0.04290	0.0	0.0	0.0	0.0

Problem 2

In this problem the aromatic amino acid pathway is considered without any regulatory connections, and we postulate a regulatory superstructure such that any of the first six reactions can be inhibited by any of the six metabolites with an inhibition strength of −0.75 or −0.075. We allow only two regulatory connections for each enzyme and three enzyme level manipulations. The questions addressed in this problem to maximize the phenylalanine selectivity are the following:

(i) Which pair of metabolites should inhibit each reaction?
(ii) What are their inhibition strengths?
(iii) Which three enzymes should be overexpressed?
(iv) What should be their expression levels?

The use of discrete values for the level of the inhibition allows preservation of system linearity and provides qualitative information about the order of magnitude of the regulatory loop strength. This problem can provide us with information about the maximum selectivity which can be achieved for this model.

In the formulation of the problem two regulatory elasticity matrices are introduced

$$\epsilon_1^r = \{\epsilon_{1ji_r} = -0.75 | 1 \le j \le 6 \quad \text{and} \quad 1 \le i_r \le 6\}$$

$$\epsilon_2^r = \{\epsilon_{2ji_r} = -0.075 | 1 \le j \le 6 \quad \text{and} \quad 1 \le i_r \le 6\}$$

The problem was solved with the same bounds on the continuous variables. Moreover, we introduced three additional logical constraints. The first one takes the form

$$\sum_{m^+ \in M^+} y_{m^+ji_r} + \sum_{m^- \in M^-} y_{m^-ji_r} \le 1 \qquad (33)$$

$$m^+ \in M^+,\ m^- \in M^-,\ j \in J,\ i_r \in I_r^j$$

and allows for only one level of inhibition chosen from the two different orders of magnitude allowed. The second one is

$$\sum_{m \in M} \sum_{i_r \in I_{i_r}^j} y_{mji_r} \le 2 \qquad (34)$$

and allows for only two regulatory connections for each enzyme. The third one is

$$\sum_{m \in M} \sum_{j \in J} y_{mji_r} \leq 6 \quad i_r \in I^j_{i_r} \tag{35}$$

and allows any of the six enzymes to be regulated by any of the six metabolites. Under these constraints the number of the alternative regulatory structures we can build around the pathway for this problem are 2^{47}.

The resulting mathematical formulation accounts for 193 variables (112 continuous, 81 binary) and 393 constraints. The problem was solved and the optimal solution $S_{\text{phe,opt}}$ found in 19.56 CPU·s was 1.05938. This value is greater than the actual upper bound for the phenylalanine selectivity (Eq. 29), because we used as the objective function the logarithmic selectivity (Eq. 32) which is a linearized approximation of Eq. 29 and is not subject to any upper bound. On the other hand, the value for the optimal selectivity is 0.818672, if calculated from the solution linearized rate expressions used in Eq. 29. This solution suggests that for the parameters chosen for the system and with an optimized regulatory structure we can increase the selectivity up to 95% by simply manipulating only three enzymes without affecting the specific growth rate.

The problem has multiple regulatory structures that result in the same optimal objective value. However, we are interested in the structures that achieve the optimal performance but have the minimum number of regulatory loops since the creation of these loops is very difficult. Therefore we formulate a new objective function

$$\min \sum_{m \in M} \sum_{j \in J} \sum_{i_r \in I^j_{i_r}} y_{mji_r} \tag{36}$$

with the additional equality constraint

$$w_{\text{phe}} = \ln\left(\frac{S_{\text{phe,opt}}}{S_{\text{phe},o}}\right) \tag{37}$$

where $w_{\text{phe}} = 0.908115$

The solution to this problem resulted in only four alternative regulatory structures presented in Figure 3; the corresponding values for the continuous variables are presented in Table 3. The minimum number of loops is equal to two, the elasticity for each loop is -0.75, and the three enzymes that must be manipulated are the enzymes that catalyze reactions 1, 4 and 6. Moreover, we observe that the regulatory metabolites are only two, CHR and PHE, and the enzymes that should be regulated are the ones that catalyze reactions 5 and 6. The number of different structures is equal to the number of all possible combinations of two reactions regulated by two metabolites, with only one regulatory connection allowed for each reaction. Therefore, the solution suggests that the enzymes that catalyze reactions 5 and 6 in the pathway should be engineered, if possible, so that both will be inhibited by either CHR and PHE. Once this regulatory

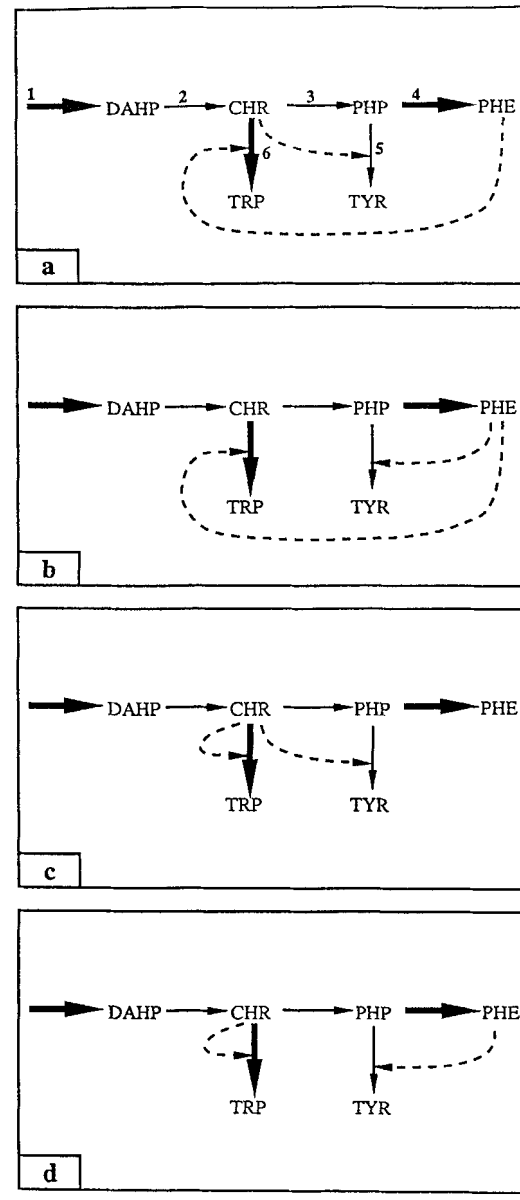

Figure 3. Four best solutions from problem 2.

Solid, dashed, and thick arrows as in Figure 2. In solutions b-d the reaction numbering has been omitted for clarity.

Table 3. Continuous Variables for the Second Four Best Solutions of Problem 2

	Case a		Case b		Case c		Case d	
i	z_i	q_i	z_i	q_i	z_i	q_i	z_i	q_i
1	0.08347	0.03100	0.08302	0.03083	0.08346	0.03100	0.08302	0.03083
2	1.34812	0.0	1.34606	0.0	1.34811	0.0	1.34606	0.0
3	0.66751	0.0	0.48125	0.0	0.66751	0.0	0.48125	0.0
4	1.50038	0.48497	1.50038	0.60071	1.50038	0.48497	1.50038	0.60071
5	−1.0	0.0	−1.0	0.0	−1.0	0.0	−1.0	0.0
6	−1.0	0.43690	−1.0	0.32129	−1.0	0.32270	−1.0	0.43703

structure has been successfully realized, the enzymes that catalyze reactions 1, 4 and 6 should be cloned and overexpressed simultaneously.

Problem 3

This problem is the same as the previous one except that we would like to design a regulatory structure for which only *enzyme activation* is allowed. We consider again that any of the first six reactions can be activated by any of the six metabolites with strength 0.75 or 0.075, and only two regulatory connections for every enzyme and three enzyme manipulations are allowed. To maximize the phenylalanine selectivity the questions to be addressed are:

(i) Which pair of metabolites should activate each reaction?

(ii) What are their activation strengths?

(iii) Which three enzymes should be overexpressed?

(iv) What should their expression levels be?

The mathematical formulation is the same as for problem 2 except that the elasticity matrices are

$$\epsilon_1^r = \{\epsilon_{1ji_r} = 0.75 | 1 \le j \le 6 \quad \text{and} \quad 1 \le i_r \le 6\}$$

$$\epsilon_2^r = \{\epsilon_{2ji_r} = 0.075 | 1 \le j \le 6 \quad \text{and} \quad 1 \le i_r \le 6\}$$

The size of this problem is the same as the size of the previous problem, and it consists of 393 constraints and 193 variables (112 continuous, 81 binary). The optimal solution to this problem was found, within 249 CPU·s, to be 1.05938 which is the same as before. The value for the optimal selectivity, calculated from the ratio of the linearized rate expressions was again 0.818672, and multiple regulatory structures were found to correspond to the optimal value. Therefore, we solved the problem with the objective function (Eq. 36) and the additional equality constraint (Eq. 37).

The minimum number of regulatory activation loops that correspond to the optimal selectivity is equal to 3, and 37 alternative regulatory structures were identified. The regulatory elasticities for every structure were equal to 0.75. Analysis of the alternative structures indicates that the enzymes that catalyze reactions 4, 5 and 6 should be activated (this is true in all 37 cases), and that these steps should be the target of any attempt to engineer the regulatory features of the pathway. Moreover, out of the 60 possible combinations only 12 combinations of enzyme levels should be manipulated. These combinations and the number of the number of the alternative regulatory structures for each combination are presented in Table 4. In Figure 4 the structures that correspond to the manipulation of enzyme levels for reactions 1, 2 and 4, are presented, and the associated changes in the expression levels and the metabolite concentrations are given in Table 5.

Table 4. Alternative Regulatory Structures for Each of the 12 Combinations of Enzyme Overexpressions from Problem 3

Overexpressed Enzymes	1,2,4	3,4,6	1,4,5	1,4,6	1,2,5	1,2,6
Number of Solutions	6	6	5	4	3	3
Overexpressed Enzymes	1,5,6	2,3,4	2,3,6	3,5,6	3,4,5	1,2,3
Number of Solutions	3	3	1	1	1	1

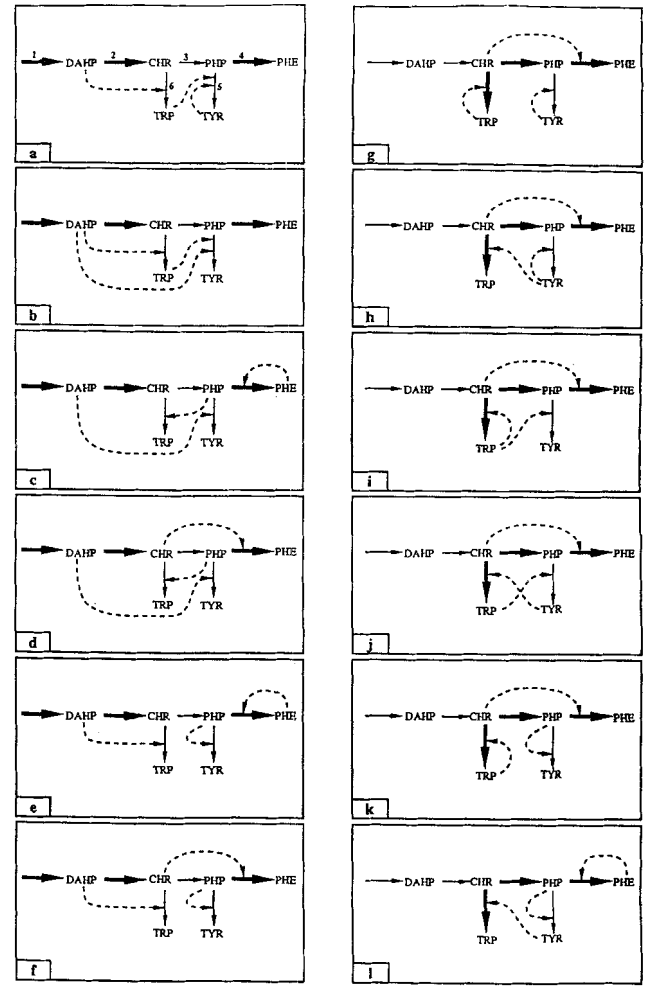

Figure 4. Twelve of the 37 best solutions from problem 3.

Solid and thick arrows as in Figure 2. Dashed arrows indicate activation. In solutions b–l the reaction numbering has been omitted.

Problem 4

To maximize the phenylalanine selectivity, the last problem, a combination of problems 2 and 3, is formulated as:

(i) Which pair of metabolites should regulate each reaction?

(ii) What should be the type of regulation (activation or inhibition)?

(iii) What is the strength of the regulation?

(iv) Which three enzymes should be overexpressed?

(v) What should their expression levels be?

We have four elasticity matrices

Table 5. Continuous Variables for Six Solutions of Problem 3 (Cases a–f from Figure 4)

	Case a		Case b		Case c	
i	z_i	q_i	z_i	q_i	z_i	q_i
1	−0.91840	0.02832	−0.91830	0.02823	−0.18806	0.02872
2	1.35478	0.37064	1.35368	0.37051	1.33065	0.09889
3	1.27053	0.0	1.17193	0.0	−0.91641	0.0
4	1.50038	0.11026	1.50038	0.17153	1.50038	0.34393
5	−1.0	0.0	−1.0	0.0	−1.0	0.0
6	−1.0	0.0	−1.0	0.0	−1.0	0.0
	Case d		Case e		Case f	
i	z_i	q_i	z_i	q_i	z_i	q_i
1	−0.18806	0.02873	−0.91677	0.02674	−0.91677	0.02674
2	1.33066	0.09889	1.33506	0.36844	1.33504	0.36844
3	−0.91641	0.0	−0.51808	0.0	−0.51808	0.0
4	1.50038	0.47122	1.50038	0.09641	1.50038	0.22042
5	−1.0	0.0	−1.0	0.0	−1.0	0.0
6	−1.0	0.0	−1.0	0.0	−1.0	0.0

$$\epsilon_1^r = \{\epsilon_{1ji_r} = -0.75 | 1 \le j \le 6 \quad \text{and} \quad 1 \le i_r \le 6\}$$

$$\epsilon_2^r = \{\epsilon_{2ji_r} = -0.075 | 1 \le j \le 6 \quad \text{and} \quad 1 \le i_r \le 6\}$$

$$\epsilon_1^r = \{\epsilon_{3ji_r} = 0.75 | 1 \le j \le 6 \quad \text{and} \quad 1 \le i_r \le 6\}$$

$$\epsilon_2^r = \{\epsilon_{4ji_r} = 0.075 | 1 \le j \le 6 \quad \text{and} \quad 1 \le i_r \le 6\}$$

and the rest of the constraints and the bounds are the same as in the two previous problems. The number of alternative regulatory structures is 2^{59}. The problem featured 681 constraints and 337 variables (184 continuous, 153 binary), and the first optimal solution was found in 156 CPU·s.

As before, the value for the optimal selectivity was equal to 1.05938, and the ratio of the linearized rates again was 0.818672. This optimal value corresponds to multiple regulatory structures and combinations for enzyme manipulations. Therefore, we solved the problem again in order to find the minimum number of the regulatory loops that correspond to this optimal value.

The minimum number of loops was 2, the regulatory elasticities were 0.75 and −0.75 for the activation and the inhibition loops, respectively, and the number of structures with only two regulatory loops was 10. As expected, the four structures found in problem 2 are also included in this set of structures. In Figure 5 the six new regulatory structures are depicted, and in Table 6 the corresponding values of the expression level of the enzymes are presented.

In all of the optimal regulatory structures involving activation and inhibition we observe three main characteristics:

(i) Only the enzymes that catalyze reactions 5 and 6 are regulated.

(ii) The enzyme that catalyzes reaction 5 is always inhibited, and the enzyme that catalyzes reaction 6 is always activated.

(iii) DAHP, TRP, and TYR act as activators, and CHR and PHE act as inhibitors.

These observations clearly suggest that, if possible, the enzyme that catalyzes reaction 5 should be designed so that it is inhibited by CHR or PHE, and the enzyme for reaction 6 should be designed so that it is activated by DAHP or TRP or TYR. Moreover, if modification in the regulatory structure is accompanied by overexpression of the enzymes for reactions 1, 4 and 6, for any combination of the just listed regulatory patterns will be successful, except for DAPH-activated reaction 6. If, on the other hand, the enzymes for reactions 1, 2 and 4 are overexpressed, the regulatory structure should be designed so that reaction 6 will always be activated by DAHP.

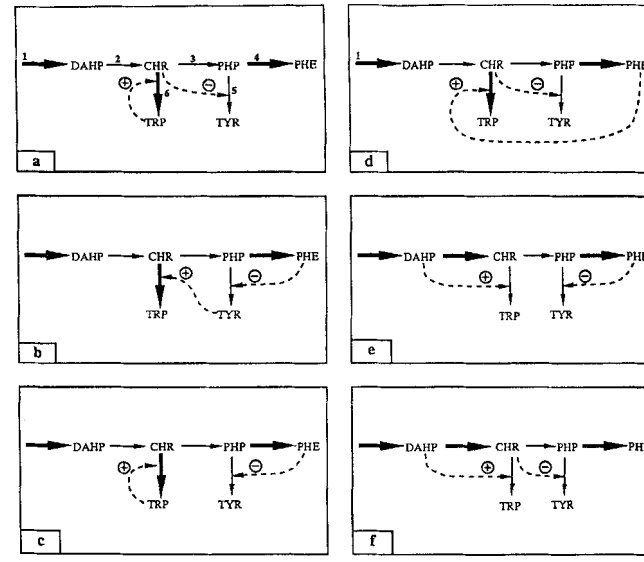

Figure 5. Six best solutions from problem 4 that feature activation and inhibition.

Solid and thick arrows as in Figure 2. Dashed arrows indicate regulation, (+) indicates activation and (−) inhibition.

Discussion

Linear models have been used within MCA in order to characterize and identify the enzymes that limit the performance of metabolic pathways. Such linear models can provide, for each metabolic function, its control coefficients, defined as the fractional changes of the metabolic function expected for a unit fractional change in the amount of each enzyme or external effector participating in a given pathway. Experimental, theoretical, and computational analyses have shown

Table 6. Continuous Variables for Six Solutions of Problem 4 (Cases a–f from Figure 5)

	Case a		Case b		Case c	
i	z_i	q_i	z_i	q_i	z_i	q_i
1	0.08302	0.030832	0.08346	0.03100	0.08346	0.03100
2	1.34606	0.0	1.34811	0.0	1.34811	0.0
3	0.48125	0.0	0.66751	0.0	0.66751	0.0
4	1.50038	0.60071	1.50038	0.48497	1.50038	0.48497
5	−1.0	0.0	−1.0	0.0	−1.0	0.0
6	−1.0	0.06174	−1.0	0.06161	−1.0	0.06161
	Case d		Case e		Case f	
i	z_i	q_i	z_i	q_i	z_i	q_i
1	0.08302	0.03083	−0.91785	0.02779	−0.91768	0.026724
2	1.34606	0.0	1.34811	0.36989	1.34607	0.36967
3	0.48125	0.0	0.66751	0.0	0.48127	0.0
4	1.50038	0.60071	1.50038	0.48497	1.50038	0.60070
5	−1.0	0.0	−1.0	0.0	−1.0	0.0
6	−1.0	0.06174	−1.0	0.0	−1.0	0.0

that the existence of a single enzyme which limits a metabolic process should not be presumed, and overexpression of a limiting enzyme results in a shift of the limitation to other steps in the pathway. Many of these studies have attributed these responses to the coupling between different pathways through regulatory connections and the fact that they share metabolites (Kacser and Burns, 1973; Savageau, 1976; Cornish-Bowden and Cardenas, 1990; Bailey, 1991).

One effective way to manipulate metabolic pathways is to implement pathways that are desensitized and decoupled from other cellular activities, and limited by a minimum number of enzymes. These objectives can be formulated and studied following the approach described above.

If in Eqs. 3 and 4 we set a manipulated variable q_k equal to one and the rest of the manipulated variables equal to zero, the metabolic functions w will be equal to the control coefficients of these functions with respect to this manipulated variable. Then, the problem of adjusting the control coefficients close to a desired value w_f can be written as

$$\min (w - w_f)^T (w - w_f) \tag{38}$$

subject to the same constraints introduced in the above mathematical formulation. In the special case that $w_f = 0$, we study the problem of decoupling functions w from other cellular processes for which q_k is an output. This problem is nonlinear in the objective function and combines both discrete and continuous variables and as a result can be formulated as a mixed-integer nonlinear programming (MINLP) problem. An optimization framework that can address the analysis and synthesis problem of metabolic pathways for nonlinear models is presently being developed.

The results presented above for the analysis and synthesis of the regulatory structure of the aromatic amino acid pathway do not take into account the stability and the dynamic characteristics of the network with alternative regulatory structures. Even if a system is stable, obtaining desirable transient responses associated with changes in the manipulated variables and the dynamic responses to fluctuations in the parameters of the system have been proposed as criteria for optimization of metabolic processes (Torres, 1994). The approach introduced in this work cannot explicitly formulate objectives associated with such dynamic characteristics, but can allow the generation of a sequence of optimal alternative regulatory structures which can be reordered based on their dynamic performance using simulation analysis and process control tools.

Conclusions

In this article we present a novel approach to the analysis and synthesis of metabolic pathways. The problem of designing the regulatory structures built around a given metabolic reaction network was formulated as a MILP optimization problem. A synthesis approach has been proposed which assumes that the metabolic pathway of interest has no regulation, and considers which regulatory structure optimizes the objective. Assuming that a linear model for the pathway of interest is given, integer variables were introduced to denote the existence or nonexistence of the postulated regulatory loops.

The approach, implemented in METAOPT, was applied to the study of the aromatic amino acid pathway in bacteria. The solution allows the identification of the regulatory structures and the associated changes in the enzyme levels that result in an optimal phenylalanine selectivity. Multiple regulatory structures were found to correspond to optimal solutions. The consistent patterns identified within these solutions helped in the postulation of design principles that were effective when applied to the full nonlinear model on which the first example is based. For the other examples, in which new patterns of enzyme regulation were considered as options, there is not a unique transformation from the log-linear model used here to a corresponding nonlinear model. The implications of this will be examined in the future employing the MINLP framework now being developed.

Acknowledgments

This research was supported by the Swiss Priority Program in Biotechnology (SPP Biotech); CAF acknowledges support from the National Science Foundation under grants CBT-8857013, CTS-9221411. VH gratefully acknowledges I. P. Androulakis and C. D. Maranas for their helpful discussions, and the Chemical Engineering Department at Princeton University for its hospitality.

Literature Cited

Bailey, J. E., "Toward a Science of Metabolic Engineering," *Science*, **252**, 1668 (1991).

Cornish-Bowden, A., and M. L. Cardenas, eds., "Control of Metabolic Processes," *NATO ASI Ser. A: Lifesciences*, Vol. 190, Plenum Press, New York (1990).

Fell, D. A. "Metabolic Control Analysis—A Survey of its Theoretical and Experimental Development," *Biochem. J.*, **152**, 313 (1992).

Fredrickson, A. G., "Formulation of Structured Growth Models," *Biotechnol. Bioeng.*, **28**, 1481 (1976).

Floudas, C. A., *Nonlinear and Mixed Integer Optimization: Fundamentals and Applications*, Oxford Univ. Press, New York, in press (1995).

Gallazo, J. L., and J. E. Bailey, "Fermentation Pathway Kinetics and Metabolic Flux Control in Suspended and Immobilized *Saccharomyces cerevisiae*," *Enzyme Microb. Technol.*, **12**, 162 (1990).

Glover, F., "Improved Linear Programming Formulations of Nonlinear Integer Problems," *Mgmt. Sci.*, **22**, 445 (1975).

Hatzimanikatis, V., K. H. Lee, W. A. Renner, and J. E. Bailey, "A Mathematical Model for the G1/S Transition of the Mammalian Cell Cycle," *Biotech. Lett.*, in press (1996).

Heinrich, R., H.-G. Holzhütter, and S. Schuster, "A Theoretical Approach to the Evolution and Structural Design of Enzymatic Networks," *Bull. Math. Biol.*, **49**, 539 (1987).

Heinrich, R., and T. A. Rapoport, "A Linear Steady-State Treatment of Enzymatic Chains," *Eur. J. Biochem.*, **42**, 89 (1974).

Joshi, A., and B. O. Palsson, "Metabolic Dynamics in the Human Red Cell: I. A Comprehensive Kinetic Model," *J. Theor. Biol.*, **141**, 515 (1989).

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

Katsumata, R., and M. Ikeda, "Hyperproduction of Tryptophan in *Corynebacterium glutamicum* by Pathway Engineering," *Bio/Technol.*, **11**, 921 (1993).

Lee, S. B., and J. E. Bailey, "A Mathematical Model for Lambda-dv Plasmid Replication—Analysis of Wild-Type Plasmid," *Plasmid*, **11**, 151 (1984a).

Lee, S. L., and J. E. Bailey, "Genetically Structured Models for lac Promoter-Operation Function in the *Escherichia coli* Chromosome and in Multicopy Plasmids—lac Operator Function," *Biotechnol. Bioeng.*, **26**, 1327 (1984b).

Lee, S. L., and J. E. Bailey, "Genetically Structured Models for lac Promoter-Operation Function in the *Escherichia coli* Chromosome and in Multicopy Plasmids—lac Promoter Function," *Biotechnol. Bioeng.*, **26**, 1383 (1984c).

Majewski, R. A., and M. M. Domach, "Simple Constrained-Optimi-

zation View of Acetate Overflow in *E. coli*," *Biotechnol. Bioeng.*, **35**, 732 (1990).
Marr, A. G., "Growth Rate of *Escherichia coli*," *Microbiol. Rev.*, **55**, 316 (1991).
Petersen, C. C., "A Note on Transforming the Product of Variables to Linear Forms in Linear Programs," Working paper, Purdue Univ. (1971).
Psarris, P., and C. A. Floudas, "Improving Dynamic Operability in MIMO Systems with Time Delays," *Chem. Eng. Sci.*, **45**, 3505 (1990).
Ramkrishna, D., "A Cybernetic Respective of Microbial Growth," *ACS Symp. Ser.*, **207**, 151 (1983).
Reder, C., "Metabolic Control Theory: A Structural Approach," *J. Theor. Biol.*, **135**, 175 (1988).
Regan, L., D. L. Bogle, and P. Dunnill, "Simulation and Optimization of Metabolic Pathways," *Comp. Chem. Eng.*, **17**, 627 (1993).
Sanwal, B. D., M. Kapoor, and H. W. Duckworth, "The Regulation of Branched and Converging Pathways," *Curr. Top. Cell. Regul.*, **3**, 1 (1971).
Savageau, M. A., *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*, Addison-Wesley, Reading, MA (1976).
Schlosser, P. M., and J. E. Bailey, "An Integrated Modelling-Experimental Strategy for the Analysis of Metabolic Pathways," *Math. Biosci.*, **100**, 87 (1990).
Schlosser, P. M., T. Holcomb, and J. E. Bailey, "Determining Metabolic Sensitivity Coefficients Directly from Experimental Data," *Biotechnol. Bioeng.*, **41**, 1027 (1993).
Schuster, S., and R. Heinrich, "Minimization of Intermediate Concentrations as a Suggested Optimality Principle for Biochemical Networks: I. Theoretical Analysis," *J. Math. Biol.*, **29**, 425 (1991).
Shuler, M. L., and M. M. Domach, "Mathematical Models of the Growth of Individual Cells—Tools for Testing Biochemical Mechanisms," *ACS Symp. Ser.*, **207**, 93 (1983).
Starbuck, C., and D. A. Lauffenburger, "Mathematical Model for the Effects of Epidermal Growth-Factor Receptor Trafficking Dynamics on Fibroblast Proliferation Responses," *Biotech. Prog.*, **8**, 132 (1992).
Stephanopoulos, G., and J. J. Vallino, "Network Rigidity and Metabolic Engineering in Metabolite Overproduction," *Science*, **252**, 1675 (1991).
Torres, N. V., "Application of Transition Time of Metabolic Systems as a Criterion for Optimization of Metabolic Processes," *Biotechnol. Bioeng.*, **44**, 291 (1994).
Varma, A., and B. O. Palsson, "Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use," *Bio/Technol.*, **12**, 994 (1994).
Voit, E. O., "Optimization in Integrated Biochemical Systems," *Biotechnol. Bioeng.*, **40**, 572 (1992).

Appendix A

The logarithmic transformation used for the linear description of metabolic systems will be detailed. Consider the nonlinear dynamical system

$$\frac{dx}{dt} = f(x;p)$$

where x is the n-dimensional independent variable vector and p is the s-dimensional parameter vector. Let x_o be a steady state, or one of the steady states in case of steady-state multiplicity, that corresponds to the parameter vector values p_o, and with nonzero, positive elements. Linearization around this steady state will result in the linear system

$$\frac{d(x-x_o)}{dt} = \left.\frac{\partial f}{\partial x}\right|_{x_o,p_o}(x-x_o) + \left.\frac{\partial f}{\partial p}\right|_{x_o,p_o}(p-p_o)$$

if we define the matrices X_o and P_o to be the diagonal matrices with diagonal elements $X_{o,ii} = x_{o,i}$ and $P_{o,ii} = p_{o,i}$, respectively. Then, for the linear system we can write

$$X_o^{-1}\frac{d(x-x_o)}{dt} = X_o^{-1}\left.\frac{\partial f}{\partial x}\right|_{x_o,p_o} X_o X_o^{-1}(x-x_o) + X_o^{-1}\left.\frac{\partial f}{\partial p}\right|_{x_o,p_o} P_o P_o^{-1}(p-p_o)$$

In the above equation we can redefine the following vectors for the scaled variables and parameters

$$z = X_o^{-1}(x-x_o) \Rightarrow z_i = \frac{x_i - x_{o,i}}{x_{o,i}}$$

$$q = P_o^{-1}(p-p_o) \Rightarrow q_i = \frac{p_i - p_{o,i}}{p_{o,i}}$$

Given the above definitions we can also observe that

$$\left.\frac{\partial f}{\partial x}\right|_{x_o,p_o} X_o = \left.\frac{\partial f}{\partial (x-x_o)}\right|_{x_o,p_o}$$

$$X_o \Rightarrow \left.\frac{\partial f_i}{\partial (x_j - x_{o,j})} x_{o,j}\right|_{x_o,p_o} = \left.\frac{\partial f_i}{\partial z_j}\right|_{x_o,p_o}$$

$$\left.\frac{\partial f}{\partial p}\right|_{x_o,p_o} P_o = \left.\frac{\partial f}{\partial (p-p_o)}\right|_{x_o,p_o}$$

$$P_o \Rightarrow \left.\frac{\partial f_i}{\partial (p_j - p_{o,j})} p_{o,j}\right|_{x_o,p_o} = \left.\frac{\partial f_i}{\partial q_j}\right|_{x_o,p_o}$$

On the other hand, for any logarithmic function of variable y we can write for up to first-order approximation the Taylor series around a reference value y_o

$$\ln(y) = \ln(y_o) + \frac{y-y_o}{y_o} \Rightarrow \ln\left(\frac{y}{y_o}\right) = \frac{y-y_o}{y_o}$$

Therefore, for the scaled variables z and the scaled parameters q we can write

$$z_i = \ln\left(\frac{x_i}{x_{o,i}}\right)$$

and

$$q_i = \ln\left(\frac{p_i}{p_{o,i}}\right)$$

Finally, by defining

$$g_i = \frac{f_i}{x_{o,i}}$$

we can write the linearized system for the logarithmic deviations, noting that $[z_o, q_o] = [0, 0]$

$$\frac{dz}{dt} = \left.\frac{\partial g}{\partial z}\right|_{0,0} z + \left.\frac{\partial g}{\partial q}\right|_{0,0} q$$

In Eq. 3 for example

$$\left.\frac{\partial g}{\partial z}\right|_{0,0} = N\epsilon + \kappa$$

and

$$\left.\frac{\partial g}{\partial q}\right|_{0,0} = N\Pi + \Lambda$$

We also have three expressions that account for the incorporation of the amino acids into biomass

$$\upsilon'_1 = 54\mu, \quad \upsilon'_2 = 131\mu, \quad \upsilon'_3 = 176\mu$$

and six expressions that account for the dilution brought about by increases in the biomass

$$\upsilon'_4 = \mu[DAHP], \quad \upsilon'_5 = \mu[CHR], \quad \upsilon'_6 = \mu[PHP]$$
$$\upsilon'_7 = \mu[PHE], \quad \upsilon'_8 = \mu[TYR], \quad \upsilon'_9 = \mu[TRP]$$

The growth function μ used is

$$\mu = 0.014\frac{Y[PHE][TYR][TRP][PEP]}{(0.25+Y)(18+[PHE])(13+[TYR])(5+[TRP])(0.005923+[PEP])}$$

Appendix B

The rate expressions for the aromatic amino acid pathway are taken from Schlosser and Bailey (1990). Here we consider only the aromatic amino acid biosynthesis reactions as an isolated, model subsystem of overall metabolism. In particular the glucose catabolic reactions considered by Schlosser and Bailey (1990) are not included here. The parameter values used for the dissociation constants are the same as in Schlosser and Bailey (1990) where the references for the estimation of these parameters can be found. The values for $\upsilon_{j,\max}$ have been adjusted to give steady-state values similar to those found in bacterial cells for $[G6P] = 0.8\ mM$, $[PEP] = 0.1\ mM$, $[ATP] = 2.5\ mM$, $[ADP] = 0.4271\ mM$, and $[AMP] = 0.0729\ mM$. The rate expressions for the 6 enzymatically catalyzed reactions in the pathway are

$$\upsilon_1 = \upsilon_{m,1}\frac{\dfrac{0.79}{1+[PHE]/53} + \dfrac{0.2}{1+[TYR]/40} + \dfrac{0.01}{1+[TRP]/16}}{\left(\dfrac{0.0002}{[PEP][G6P]} + \dfrac{0.006}{[PEP]}\right)(1+50[DAHP]) + \dfrac{0.1}{[G6P]} + 1}$$

$$\upsilon_2 = \upsilon_{m,2}\frac{[DAHP][PEP][ATP]}{(2+[DAHP])(0.00867+[PEP])(0.9281+[ATP])}$$

$$\upsilon_3 = \upsilon_{m,3}\frac{[CHR]}{(2+[CHR])(1+[PHE]/50)(1+[TYR]/40)}$$

$$\upsilon_4 = \upsilon_{m,4}\frac{[PHP]}{(1+[PHP])(1+[PHE]/50)}$$

$$\upsilon_5 = \upsilon_{m,5}\frac{[PHP]}{(1+[PHP])}$$

$$\upsilon_6 = \upsilon_{m,6}\frac{[G6P][CHR][ATP]}{(1.269+[G6P])(2+[CHP])(0.9821+[ATP])(1+[TRP]/16)}$$

where

$$Y = \frac{[ATP]+0.5[ADP]}{[ATP]+[ADP]+[AMP]}$$

The mass balance equations for each of the metabolites in the aromatic amino acid pathway is given by

$$\begin{aligned}
DAHP:&\ 0 = \upsilon_1 - \upsilon_2 - \upsilon'_4\\
CHR:&\ 0 = \upsilon_2 - \upsilon_3 - \upsilon'_6 - \upsilon'_5\\
PHP:&\ 0 = \upsilon_3 - \upsilon_4 - \upsilon_5 - \upsilon'_6\\
PHE:&\ 0 = \upsilon_4 - \upsilon'_3 - \upsilon'_7\\
TYR:&\ 0 = \upsilon_5 - \upsilon'_2 - \upsilon'_8\\
TRP:&\ 0 = \upsilon_6 - \upsilon'_1 - \upsilon'_9
\end{aligned}$$

where

$$V_m^T = [710,\ 22,\ 474,\ 64,\ 10.5,\ 28]$$

The stable steady state at which the linear model was constructed is

$$x_o^T = [3.41404,\ 30.3097,\ 0.609277,\ 263.660,\ 321.496,\ 82.6866]$$

and the values for the rates at these values are

$$v_o^T = [9.340355 \quad 9.310362 \quad 7.843243 \quad 3.862561 \quad 3.975329]$$

$$v_o'^T = [1.200839 \quad 0.474409 \quad 1.150880 \quad 1.546221 \quad 0.029994 \quad 0.266281 \quad 0.005353 \quad 2.316340 \quad 2.824449 \quad 0.726430]$$

For the matrix N in Eq. 1 we have

$$N = \begin{bmatrix} 2.735867 & -2.727082 & 0 & 0 & 0 & 0 \\ 0 & 0.307175 & -0.258770 & 0 & 0 & -0.039619 \\ 0 & 0 & 12.873042 & -6.339585 & -6.524671 & 0 \\ 0 & 0 & 0 & 0.014650 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.012365 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.014523 \end{bmatrix}$$

The elasticity matrices at the steady state are

$$\epsilon^s = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.369410 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.061901 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.621397 & 0 & 0 & 0 \\ 0 & 0 & 0.621397 & 0 & 0 & 0 \\ 0 & 0.061901 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$\epsilon^r = \begin{bmatrix} -0.899843 & 0 & 0 & -0.705837 & -0.126183 & -0.008709 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.840591 & -0.889349 & 0 \\ 0 & 0 & 0 & -0.840591 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.837871 \end{bmatrix}$$

For the matrix $\mathfrak{K}$ in Eq. 3 we have

$$\mathfrak{K} = \begin{bmatrix} -0.008785 & 0 & 0 & -0.000561 & -0.000341 & -0.000501 \\ 0 & -0.008785 & 0 & -0.000561 & -0.000341 & -0.000501 \\ 0 & 0 & -0.008785 & -0.000561 & -0.000341 & -0.000501 \\ 0 & 0 & 0 & -0.009721 & -0.000569 & -0.000835 \\ 0 & 0 & 0 & -0.000790 & -0.009266 & -0.000705 \\ 0 & 0 & 0 & -0.000928 & -0.000564 & -0.009613 \end{bmatrix}$$

We consider the enzymes of the first six reactions as the manipulated variables. Moreover, the three precursor metabolites are also treated as manipulated variables. Therefore, we can write for the manipulated variables the vector

$$p^T = [v_{m,1}, v_{m,2}, v_{m,3}, v_{m,4}, v_{m,5}, v_{m,6}, [G6P], [PEP], [ATP]]$$

and for the matrices Π and Λ we have

$$\Pi = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0.046747 & 0.905114 & 0 \\ 0. & 1 & 0 & 0 & 0 & 0 & 0 & 0.079740 & 0.270733 \\ 0. & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0. & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0. & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0. & 0 & 0 & 0 & 0 & 1 & 0.613340 & 0 & 0.270733 \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000491 & -0.001119 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000491 & -0.001119 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000491 & -0.001119 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000819 & -0.001865 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000691 & -0.001574 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.000812 & -0.001849 \end{bmatrix}$$

The parameters in Eq. 9 that correspond to specific growth rate are

$$\Xi_\mu = \mathbf{0}$$

$$\mathfrak{K}_\mu = [0, 0, 0, 0.063907, 0.038864, 0.057021]$$

$$\Theta_\mu = [0, 0, 0, 0, 0, 0, 0, 0.055918, 0.127322]$$

Manuscript received Mar. 17, 1995, and revision received Aug. 18, 1995.