Quantitative Assessment of Uncertainty in the Optimization of Metabolic Pathways

Spas B. Petkov, Costas D. Maranas

Department of Chemical Engineering, The Pennsylvania State University, 112A Fenske Laboratory, University Park, Pennsylvania, 16802; telephone: 814-863-9958; fax: 814-865-7846; e-mail: cdm8@psu.edu

Abstract: The S-System formalism provides a popular, versatile and mathematically tractable representation of metabolic pathways. At steady-state, after a logarithmic transformation, the S-System representation reduces into a system of linear equations. Thus, the maximization of a particular metabolite concentration or a flux subject to physiological constraints can be expressed as a linear programming (LP) problem which can be solved explicitly and exactly for the optimum enzyme activities. So far, the quantitative effect of parametric/experimental uncertainty on the S-model predictions has been largely ignored. In this work, for the first time, the systematic quantitative description of modeling/experimental uncertainty is attempted by utilizing probability density distributions to model the uncertainty in assigning a unique value to system parameters. This probabilistic description of uncertainty renders both objective and physiological constraints stochastic, demanding a probabilistic description for the optimization of metabolic pathways. Based on notions from chance-constrained programming and statistics, a novel approach is introduced for transforming the original stochastic formulation into a deterministic one which can be solved with existing optimization algorithms. The proposed framework is applied to two metabolic pathways characterized with experimental and modeling uncertainty in the kinetic orders. The computational results indicate the tractability of the method and the significant role that modeling and experimental uncertainty may play in the optimization of networks of metabolic reactions. While optimization results ignoring uncertainty sometimes violate physiological constraints and may fail to correctly assess objective targets, the proposed framework provides quantitative answers to questions regarding how likely it is to achieve a particular metabolic objective without exceeding a prespecified probability of violating the physiological constraints. Trade-off curves between metabolic objectives, probabilities of meeting these objectives, and chances of satisfying the physiological constraints, provide a concise and systematic way to guide enzyme activity alterations to meet an objective in the face of modeling and experimental uncertainty. © 1997 John Wiley & Sons, Inc. *Biotechnol Bioeng* 56: 145–161, 1997.
Keywords: metabolic pathways; parametric uncertainty; chance-constrained programming; nonlinear optimization

INTRODUCTION

Cell metabolism is a complex structure of biochemical reactions characterized by several levels of regulation. Recent studies have shown that cell metabolism is close to an optimal state aimed at satisfying a set of particular objectives, for example minimum energy production (Savinell and Palson, 1992), minimum NADH synthesis (Savinell and Palson, 1992), enhanced oxygen transport (Bailey et al., 1990), etc. This optimal state has been implicitly achieved through the evolutionary process aimed at maximizing the chances of survival with respect to the current environmental conditions. However, in a biochemical industrial process, under defined medium and controlled environment, the full potential of the catalytic activity of the utilized microorganism is frequently not achieved because the wild strains have not adapted to their new objective (Bailey et al., 1990). This new objective is maximum production or selectivity of a desired biochemical product. Recent developments and the new understanding of the regulatory mechanism of cell metabolism promises a way for a manifold increase in the productivity of virtually every biotechnological process by employing genetically engineered microoragnisms. The cellular metabolic controlling mechanism is two-fold: It involves translational regulations where the amount of an enzyme present in the system is controlled, and reaction level regulations where the activity of an enzyme is modified, by a present modulator (Sanwal et al., 1971; Stephanopoulos and Valino, 1991). A remarkable characteristic of cell metabolism is the "complex" response to simple changes in the control mechanism. For example, a single change in only one controlling step can result in a decline of its flux control and in increase in the flux control of other steps, which limits the extent to which the flux can be increased. In such cases, "aggregated" modeling of the system under consideration is preferred, as opposed to looking at each controlling step separately (Bailey et al., 1990).

There has been considerable effort spent in the mathematical modeling of biochemical processes. For a long time, the dominant models used to describe enzyme kinetics in complex metabolic pathways have been the Michaelis-Menten formalism and various linear representations. The

latter is appealing because of its simplicity and mathematical tractability, while the former fairly accurately describes the kinetics of enzymatic reactions carried out in vitro. Both of these models fall into the category of Conventional Mass Action (CMA) models (Shiraishi and Savageau, 1992a). While the precision of the linear representation can be insufficient away from the steady-state, Michaelis-Menten kinetics, even though nonlinear, may fail to account for allosteric effects (Shiraishi and Savageau, 1992a; Savageau, 1995), and can be mathematically intractable (Shiraishi and Savageau, 1992a).

Synergetic or S-Systems were developed based on the Biochemical Systems Theory (BST) formalism introduced by Savageau and co-workers (Savageau 1969a, 1969b, 1970, 1976; Savageau et al., 1987a, 1987b). In this mathematical formalism, the change in each metabolite is represented by two competing power-law functions describing aggregation and consumption. The advantage of this representation is that, while it preserves some of the nonlinear characteristics of the real system at steady-state, after a logarithmic transformation it reduces into a linear system of equations which can be solved efficiently for the unique solution. This enables enzyme level optimization by the means of linear programming (Voit, 1992; Regan et al., 1993; Torres et al., 1996) and recently regulatory control structure optimization by mixed-integer linear programming (MILP) (Hatzimanikatis et al., 1996a, 1996b).

While mathematically appealing, the predictive power of the S-System representation of metabolic pathways gradually diminishes away from the nominal steady-state, because the values for the model coefficients, calculated at the nominal steady-state, change. An example of such a case is the application of the S-System formalism to the optimization of the citric acid production by *Aspergillus niger* (Torres et al., 1996). The authors observed that when the predicted optimal steady-state deviated significantly from the nominal one, the predictions of the mathematical model became unrealistic. To overcome this difficulty, the authors proposed the introduction of additional constraints based on stoichiometric relations between particular fluxes. However, this action does not provide a complete answer, since the expressions for these fluxes are still calculated using the S-System parameters evaluated at the nominal steady-state. Therefore, when enzyme levels and their regulatory structure are changed over a wide range, it becomes important to account for uncertainty (imprecision) in the optimization model. Furthermore, model parameters may be subject to considerable experimental error due to either (i) our inability to measure certain quantities with sufficient precision; (ii) the varying nature of the measured quantity, or (iii) because the measurements are taken under different environmental conditions. For example, in the analysis of the tricarboxylic acid cycle in *Dictyostelium discoideum* detailed kinetic data have been compiled (Wright et al., 1992a, 1992b) and alternative kinetic formalisms for integrated biochemical systems have been examined (Shiraishi and Savageau, 1992a, 1992b, 1992c, 1992d). However, despite

the availability of a detailed kinetic description, Shiraishi and Savageau (1992c) concluded that the model is ill-defined because of the unrealistically narrow interval in which steady-state can be achieved. This example shows that considering only "best-estimate" values without quantifying their "reliability" may not always provide a complete and useful description.

While the presence of model and experimental uncertainty and its effect on the predictive power of the S-System model has been recognized (Torres, 1994a; Hatzimanikatis et al., 1996a), its quantitative description has not yet been addressed. It is the objective of this paper to quantitatively describe the effect of model/experimental uncertainty of the kinetic orders of the S-System on the optimization of metabolic pathways. We propose to describe this uncertainty with probability density distributions. The resulting probabilistic metabolic pathway optimization problem is then transformed into an equivalent problem that can be solved with available optimization algorithms. Finally, two examples are included to highlight the proposed framework of analysis and optimization under kinetic orders uncertainty.

THE S-SYSTEM FORMALISM

Background

The S-System formalism is based on BST which proposes the use of power-law functions to describe the nonlinear nature of biochemical processes Savageau (1976, 1990). Under the S-System formalism, the change in every metabolite x_i is represented as a sum of "accumulation" flux, V_i^+, and of "consumption" flux, V_i^-, expressed as an aggregated power-law function of the metabolite concentrations x_i and the enzyme activities y_k, where i or $j = 1, \ldots, N$ denote the metabolites, and $k = 1, \ldots, M$ denote the enzymes catalyzing the metabolic network at hand.

$$\frac{dx_i}{dt} = V_i^+ - V_i^-$$

$$V_i^+ = a_i \prod_{j=1}^{N} x_j^{g_{ij}} \prod_{k=1}^{M} y_k^{g'_{ik}}$$

$$V_i^- = b_i \prod_{j=1}^{N} x_j^{h_{ij}} \prod_{k=1}^{M} y_k^{h'_{ik}}$$

Note that, after specifying the enzyme activities y_k, the metabolite concentrations are uniquely specified by the solution of system of ODE's. Therefore the metabolite concentrations x_i are denoted as *dependent* variables and the enzyme activities y_k as *independent*. The model parameters g_{ij}, h_{ij}, g'_{ik}, and h'_{ik} are the *kinetic orders* and the constants a_i and b_i are the *rate constants*. The kinetic orders are defined as,

$$g_{ij} = \frac{x_j}{V_i^+} \frac{\partial V_i^+}{\partial x_j}$$

$$g'_{ik} = \frac{y_k}{V_i^+} \frac{\partial V_i^+}{\partial y_k}$$

$$h_{ij} = \frac{x_j}{V_i^-} \frac{\partial V_i^-}{\partial x_j}$$

$$h'_{ik} = \frac{y_k}{V_i^-} \frac{\partial V_i^-}{\partial y_k}$$

and are closely related to the elasticity coefficients of the enzymes in the metabolic control analysis (MCA) framework (Savageau et al., 1987a). The rate constants a_i, b_i can be calculated from flux data. The S-System representation is obtained by representing all functional relationships (i.e., rate laws, conservation relations, and fluxes) by first-order Taylor expansion in logarithmic space (Savageau, 1990). The advantage of Taylor expansion in logarithmic space as opposed to Taylor expansion in linear space is that the inherent nonlinearity of the system is somewhat preserved and this seems to extend the range over which the S-System representation is valid (Voit, 1990). The kinetic orders can be evaluated from kinetic data in vitro (Savageau, 1976; Voit et al., 1991), from steady-state experiments in situ (Voit et al., 1991), or from dynamic experiments (Voit et al., 1991; Sorribas et al., 1993).

At steady-state (i.e., $dx_i/dt = 0$) the S-System reduces to a set of algebraic nonlinear equations:

$$\sum_{j=1}^{N} (g_{ij} - h_{ij}) \ln(x_j) + \sum_{k=1}^{M} (g'_{ik} - h'_{ik}) \ln(y_k) = \ln\left(\frac{b_i}{a_i}\right),$$
$$i = 1, \ldots, N$$

By performing the following logarithmic variable transformations,

$$\hat{x}_j = \ln(x_j), j = 1, \ldots, N \text{ and } \hat{y}_k = \ln(y_k), k = 1, \ldots, M$$

validated by the monotonicity of the logarithmic function, the nonlinear system of equations can be recast in a linear form:

$$\sum_{j=1}^{N} (g_{ij} - h_{ij})\hat{x}_j + \sum_{k=1}^{M} (g'_{ik} - h'_{ik})\hat{y}_k = \ln\left(\frac{b_i}{a_i}\right), i = 1, \ldots, N$$

This implies that optimization studies under the logarithmically transformed S-System formalism yield linear optimization formulations (Voit, 1992):

$$\max \quad f(x_i, y_k)$$

$$\text{subject to} \quad \sum_{j=1}^{N} (g_{ij} - h_{ij})\hat{x}_j + \sum_{k=1}^{M} (g'_{ik} - h'_{ik})\hat{y}_k =$$
$$\ln\left(\frac{b_i}{a_i}\right), i = 1, \ldots, N$$

$$\ln(x_i^L) \leq \hat{x}_i \leq \ln(x_i^U)$$

$$\ln(y_k^L) \leq \hat{y}_k \leq \ln(y_k^U)$$

These formulations involve the maximization of an objective function $f(x_i, y_k)$ which can be a particular metabolite concentration, a flux, or any other metabolic property that is linear or expressible in BST format, as a function of metabolite and enzyme concentrations which are subject to steady-state flux balances and to lower and upper bounds on metabolite concentrations and enzyme activities. Linear optimization problems of the above form have been previously used to solve metabolic pathway optimization problems (Voit, 1992; Regan et al., 1993; Torres et al., 1996). Lower and upper bounds on metabolite concentrations are typically determined from physiological considerations (e.g., toxicity, etc.). Bounds of ±10–20% around the nominal steady-state values have been suggested as reasonable (Voit, 1992; Torres et al., 1996). Niederberger et al. (1992) suggested that ranges between 0.1 and 50 times the basal enzyme activity are feasible. In addition, upper and lower bounds may be placed on some fluxes, bearing in mind that fluxes cannot be increased indefinitely. In particular, fluxes diverging from the pathway under consideration must be kept close to their basal values to avoid disturbing the rest of the cell metabolism. Finally, Torres et al. (1996) described additional constraints in the form of relationships between metabolic fluxes aimed at preserving the stoichiometry of the process.

Based on this linear metabolic pathway representation at steady-state, the optimization of metabolic pathways has been studied by several researchers. Voit (1992) utilized linear programming (LP) to optimize the yield in xanthine monophosphate and guanosine monophosphate production. Regan et al. (1993) utilized LP for the optimization of the toluene degradation pathway. Later, Dervakos and Dean (1994) formulated the metabolic system addressed earlier in (Voit, 1992) as a mixed-integer linear programming (MILP) problem by constraining the number of enzyme levels that can be manipulated. The optimal enzyme manipulations were derived depending on how many enzyme levels were allowed to change. Recently, Torres et al. (1996) applied linear programming techniques to the optimization of citric acid by *Aspergillus niger*. Finally, Hatzimanikatis et al. (1996a, 1996b) proposed a novel mixed-integer linear programming approach to optimize not only the enzyme levels but also the entire control superstructure of metabolic pathways. The encouraging results from all these contributions indicate that manifold improvements are achievable, usually by the modification of several enzyme levels and control structures. However, there still exists doubt about how reliable these estimations are and whether the model predictions will be validated in practical applications.

Sources of Uncertainty in the S-System Formalism

A major concern in optimization studies of S-Systems is that this formalism is a *local representation*. This is a con-

sequence of the fact that the S-System representation is a first-order Taylor expansion in logarithmic space around the original (nominal) steady-state of the real system. Therefore, while the values for the S-System parameters g_{ij}, h_{ij} are "exact" at the original steady-state, their precision tends to deteriorate away from the original operating point. Due to the employed aggregation procedure, pathways with different control or regulatory structures may sometimes yield the same S-System representation (Cascante et al., 1991), suggesting that system information is sometimes lost during the formulation of the S-System model. In addition, kinetic orders are sometimes directly determined from experimental data which are subject to experimental error (Sorribas et al., 1993). Especially when the metabolite concentrations are very low, the uncertainty associated with the employed measuring techniques is typically very high. For example, the reported concentration $2.9 \cdot 10^{-4}$mM (Torres, 1994a) of the key metabolite oxaloacetate in *Aspergillus niger* involves an experimental scatter as high as 25 to 30%.

This suggests that any result obtained by S-System modeling is typically subject to *modeling* and/or *experimental* uncertainty. Several attempts have previously been made to evaluate the performance of the S-System as applied to metabolic pathway modeling. Voit and Savageau (1987) compared the performance of different representations within BST as opposed to the Generalized Mass Action (GMA) of metabolic pathways described by the Hill equation or Michaelis-Menten kinetics. It was concluded that the S-System is valid over a wider range of metabolite concentrations. Cascante et al. (1991) utilized a second-order Taylor expansion to address the problem of having different metabolic systems sharing the same S-System representation. The problem of the S-System accuracy away from the original state is troubling, considering that after genetic engineering of some enzyme levels, metabolite concentrations and fluxes might differ by orders of magnitude.

In summary, the following are some of the sources of uncertainty in the optimization of metabolic pathways represented by the S-System formalism:

1. Model uncertainty away from the steady-state (Torres et al., 1996);
2. Identical S-System representation for different control or regulatory structures (Cascante et al., 1991);
3. Experimental error in the kinetic parameters of the S-System or the associated non-linear kinetic model (Sorribas et al., 1993);
4. Experimental error in the measured metabolite concentrations in the original steady-state (Torres, 1994a).

Failure to consider these sources of uncertainty in the mathematical description of the metabolic process may render the predictions of the optimization process unreliable and in disagreement with the attributes of the real system after the genetic manipulation.

OPTIMIZATION OF METABOLIC PATHWAYS ACCOUNTING FOR UNCERTAINTY

Description

The S-System representation is based on the assumption that kinetic orders *remain unchanged* for different metabolite concentrations and enzyme activities (local representation). However, discrepancies between S-model predictions and actual experimental data imply that this assumption is only partially correct and that in reality there are always some differences between experimental measurements and model predictions. These discrepancies can be reconciled by recognizing that the values of the kinetic orders are not constant (independent of metabolite concentrations and enzyme activities), but may vary around some nominal values calculated at the nominal steady-state. This viewpoint implies that the kinetic order parameters may sample more than one value for different metabolite concentrations and enzyme activity levels. In this paper, this intuitive abstraction is expressed mathematically by utilizing probability density distributions to describe the likelihood of different realizations of the kinetic orders.

Mathematical Representation

The S-System representation, as derived above, is:

$$\sum_{j=1}^{N} (g_{ij} - h_{ij}) \ln (x_j) + \sum_{k=1}^{M} (g'_{ik} - h'_{ik})$$

$$\ln(y_k) = \ln \left(\frac{b_i}{a_i} \right), i = 1, \ldots, N$$

After subtracting the nominal steady-state description,

$$\sum_{j=1}^{N} (g_{ij} - h_{ij}) \ln (x_j^{ss}) + \sum_{k=1}^{M} (g'_{ik} - h'_{ik})$$

$$\ln(y_k^{ss}) = \ln \left(\frac{b_i}{a_i} \right), i = 1, \ldots, N$$

we obtain

$$\sum_{j=1}^{N} A_{ij}X_j + \sum_{k=1}^{M} B_{ik}Y_k = 0, i = 1, \ldots, N$$

where

$$A_{ij} = g_{ij} - h_{ij}$$
$$B_{ik} = g'_{ik} - h'_{ik}$$
$$X_j = \ln(x_j/x_j^{ss})$$
$$Y_k = \ln(y_k/y_k^{ss})$$

and the superscript *ss* stands for nominal steady-state value. In the latter representation, the reaction coefficients are no longer present and the metabolite concentrations and enzyme activities are represented by the natural logarithm of their steady-state scaled values.

This representation involves a number of important features:

1. The metabolite flux balance (i.e., $V_i^+ = V_i^-$) at the nominal steady-state is satisfied for any realization of A_{ij} and B_{ik} maintaining internal consistency of the model.
2. The reaction rate coefficients a_i and b_i which are strongly correlated with the reaction orders and steady-state fluxes (Sorribas et al., 1993) are eliminated.
3. The metabolite concentrations and enzyme activities are expressed as logarithmic deviations from their steady-states values. This provides a natural scaling and decouples the problem of modeling uncertainty from uncertainty due to errors in the experimental measurements of the steady-state concentrations.

Assuming that modeling and experimental uncertainties are mutually statistically independent, the variance of each kinetic order, g_{ij}, h_{ik}, g'_{ij}, or h'_{ik}, can be represented as the sum of the modeling uncertainty and experimental uncertainty. For example, the uncertainty of the kinetic order g_{ij} is,

$$Var(g_{ij}) = Var^{mod}(g_{ij}) + Var^{exp}(g_{ij})$$

where $Var^{mod}(g_{ij})$ is the variance due to modeling and $Var^{exp}(g_{ij})$ due to experimental uncertainty. Thus, the variances of the elements $A_{ij} = g_{ij} - h_{ij}$ and $B_{ik} = g'_{ik} - h'_{ik}$ are equal to

$$Var(A_{ij}) = Var(g_{ij}) + Var(h_{ij}) - 2Cov(g_{ij}, h_{ij})$$
$$Var(B_{ik}) = Var(g'_{ik}) + Var(h'_{ik}) - 2Cov(g'_{ik}, h'_{ik})$$

where $Cov(g_{ij}, h_{ij})$ and $Cov(g'_{ik}, h'_{ik})$ are the covariances between g_{ij}, h_{ij} and g'_{ik}, h'_{ik} respectively. Finally, covariances between different elements of the two-dimensional arrays A_{ij} and B_{ik} are given by:

$$Cov(A_{ij}, A_{i'j'}) = Cov(g_{ij}, g_{i'j'}) + Cov(h_{ij}, h_{i'j'})$$
$$- Cov(g_{ij}, h_{i'j'}) - Cov(h_{ij}, g_{i'j'})$$

$$Cov(B_{ik}, B_{i'k'}) = Cov(g'_{ik}, g'_{i'k'}) + Cov(h'_{ik}, h'_{i'k'})$$
$$- Cov(g'_{ik}, h'_{i'k'}) - Cov(h'_{ik}, g'_{i'k'})$$

$$Cov(A_{ij}, B_{i'k}) = Cov(g_{ij}, g'_{i'k}) + Cov(h_{ij}, h'_{i'k})$$
$$- Cov(g_{ij}, h'_{i'k}) - Cov(h_{ij}, g'_{i'k})$$

Stochastic Formulation

The probabilistic description of uncertainty in kinetic orders described above renders both metabolite concentrations and fluxes stochastic. Therefore, unless a particular realization of the uncertain parameters A_{ij}, B_{ik} is specified, questions regarding the satisfaction of physiological constraints and performance objectives cannot be answered. While the *a priori* identification of the realization of the uncertain parameters A_{ij}, B_{ik}, and consequently of metabolite concentrations and fluxes, is impossible, the evaluation of the probability of meeting a performance objective or a physiological constraint is computable, if the uncertain parameters A_{ij}, B_{ik} assumes values from some known probability density

distribution. This probabilistic description yields the following stochastic metabolic pathway optimization problem (SMOP) under probabilistically described uncertainty of the kinetic orders:

$$\max \quad f^t$$

$$\text{subject to} \quad Pr(f \geq f^t) \geq \alpha$$

$$\sum_{j=1}^{N} A_{ij}X_j + \sum_{k=1}^{M} B_{ik}Y_k = 0, \, i = 1, \ldots, N$$

$$Pr(X_j^L \leq X_j \leq X_j^U) \geq \beta_{x_j}, \, j = 1, \ldots, N$$

$$Pr(V_i^{\pm,L} \leq V_i^{\pm} \leq V_i^{\pm,U}) \geq \beta_{v_i}, \, i = 1, \ldots, N$$

$$Y_k^L \leq Y_k \leq Y_k^U, \quad k = 1, \ldots, M$$

Formulation SMOP involves a set of constraints imposing a probability of at least α satisfying the performance objective, and of the probabilities β_{X_j} and β_{V_i} maintaining the metabolite concentrations and fluxes between some lower and upper bounds. These constraints are called *chance-constraints*. Apart from the chance-constraints, a steady-state flux equilibrium equality is included which must be satisfied for every realization of the uncertain parameters A_{ij}, B_{ik} and X_j. In addition, lower and upper bounds on the enzyme activities Y_k (deterministic variables) are imposed.

Formulation SMOP identifies the maximum value of the performance target f^t—that the stochastic performance objective f, which is a function of X_i, Y_k, A_{ij}, and B_{ik}—can meet with probability of at least α (e.g., 90%), and at the same time maintain all metabolite concentrations and fluxes within their respective lower and upper bound, with probability greater than or equal to β_{X_j} and β_{V_i}, respectively. Therefore, the solution of SMOP will have at least an α chance of meeting the performance objective and at least a β_{X_j} chance of satisfying all imposed physiological constraints. For example, for $\alpha = 0.5$, there is a fifty-fifty chance of meeting the performance objective f^t. Higher values of α reflect a more conservative attitude. Note that the values of β_{X_j} can be specified separately for each constraint based on the importance of maintaining feasibility. For example, upper bounds on toxic intermediate concentrations must be satisfied with a high probability, while lower and upper bounds on metabolites which are not involved in adjacent pathways or regulatory superstructures may be "relaxed" by specifying a smaller value for β. By solving SMOP for different values of α and β, trade-offs between the performance objective target f^t, the probability α of meeting this performance target, and the probability β_{X_j} of satisfying all physiological requirements can readily be established. By manipulating the values of the probability levels α and β the relative importance of meeting a performance target as opposed to satisfying the physiological constraints can be adjusted.

Deterministic Equivalent Representation

The solution of SMOP requires the evaluation of probability terms whose calculation for each realization of the deterministic variables Y_k requires the integration of multivariate probability density distributions. Many integration methods exist, but in general, they all exact a heavy computational penalty due to the employed multidimensional discretization (Watanabe and Ellis, 1994), and thus so far have been restricted to optimization problems with only a few uncertain parameters. However, the number of uncertain parameters in optimization of metabolic pathways is as high as N $(N + M)$. For example, a metabolic pathway problem involving $N = 10$ metabolites and $M = 5$ enzymes involves as many as 150 uncertain parameters, without counting the metabolite concentrations. In this study, to deal with such a high number of possibly correlated uncertain kinetic orders, the transformation of the original stochastic constants into equivalent deterministic ones is sought without having to rely on computationally intensive discretization of the uncertain parameters. To this end, the deterministic equivalent representation of the chance-constrained formulations based on the ideas pioneered by Charnes and Cooper (1959, 1960, 1963) is pursued.

Based on the work of Kataoka (1963), assuming that the uncertain parameters A_{ij} and B_{ik} follow *stable* probability density distributions, (e.g., normal, Poisson, Chi-square, binomial, Cauchy, Levy, etc.), chance-constraints which are linear in terms of the uncertain parameters can be transformed into equivalent deterministic constraints (Vajda, 1970; Allen et al., 1974). Note that a probability density distribution Φ is stable if it can be completely specified with up to two parameters and the convolution of any two distribution functions is of the same form. For example, the convolution of the two parameter stable distributions Φ $(x - u_1/v_1)$ and Φ $(x - u_2/v_2)$ has to be again of the form Φ $(x - u/v)$ (Vajda, 1970). Due to its widespread acceptance and simplicity, the normal distribution ($u = \mu$ and $v = \sigma$) is selected in this work to describe the probability associated with each realization of the kinetic orders g_{ij}, h_{ik}, g'_{ij} and h'_{ik}. Nevertheless, the proposed framework is general enough to account for any stable distribution. Because A_{ij} and B_{ik} are equal to the difference of two kinetic orders they are also described by normal probability distributions.

First, a linear chance-constraint is addressed, to highlight the basic steps of the deterministic equivalent representation of chance-constraints:

$$Pr\left(\sum_i^n a_i x_i \leq 0 \right) \geq \alpha$$

Here a_i denotes the uncertain (stochastic) parameters (i.e., A_{ij}, B_{ik}) *and* x_i are deterministic variables (such as Y_k), not to be confused with the metabolite concentrations. Let $\mu(a_i)$ denote the expected value of a_i, $Var(a_i)$ the variance of a_i and $cov(a_i, a_{i'})$ the covariance between uncertain parameters a_i and $a_{i'}$. By subtracting the mean and dividing by the square root of the variance of $\sum_{i=1}^n a_i x_i$ the chance-constraint under consideration can equivalently be written as

$$Pr\left(\frac{\sum_i^n a_i x_i - \mu\left(\sum_i^n a_i x_i\right)}{\left[Var\left(\sum_i^n a_i x_i\right)\right]^{1/2}} \leq \frac{-\mu\left(\sum_i^n a_i x_i\right)}{\left[Var\left(\sum_i^n a_i x_i\right)\right]^{1/2}} \right) \geq \alpha$$

Because the normal distribution is stable, the left hand side of the inequality under the probability sign is a normally distributed random variable with a mean of zero and a variance of one (standardized form). Thus, if Φ is the standardized normal cumulative density distribution, then the chance-constraint can be replaced by the following deterministic equivalent expression

$$\Phi\left(\frac{-\mu\left(\sum_i^n a_i x_i\right)}{\left[Var\left(\sum_i^n a_i x_i\right)\right]^{1/2}} \right) \geq \alpha$$

By applying the inverse of the cumulative normal distribution function Φ^{-1}, which is a monotonically increasing function, on both sides of the last relation upon rearranging we get:

$$\mu\left(\sum_i^n a_i x_i \right) + \Phi^{-1}(\alpha)\left[Var\left(\sum_i^n a_i x_i \right) \right]^{1/2} \leq 0$$

After applying the definition for the mean and the variance of the linear combination of stochastic parameters we have

$$\sum_i^n \mu(a_i) x_i + \Phi^{-1}(\alpha)$$

$$\left[\sum_{i=1}^n Var(a_i) x_i^2 + 2 \sum_{i=1}^n \sum_{i'=i+1}^n x_i Cov(a_i, a_{i'}) x_{i'} \right]^{1/2} \leq 0$$

The inverse of the cumulative density normal distribution Φ^{-1}, for different probability levels α, is tabulated and approximated with an analytical formula in Abramowitz and Stegun (1972). As α goes to zero or one, the inverse of the cumulative normal distribution goes to minus or plus infinity, respectively. It is a strictly monotonically increasing function of α, negative for α less than 0.5 and positive for α greater than 0.5. Inspection of the deterministic equivalent constraint reveals that it is composed of the mean of the original constraint augmented by the square root of its variance times $\Phi^{-1}(\alpha)$. Typically, α is greater than 0.5 and thus

$\Phi^{-1}(\alpha) \geqslant 0$. This implies that the variance term penalizes the deterministic constraint, making it more restrictive than the original constraint. This is consistent with the intention to satisfy the original constraint not only at the mean point but for other realizations with probability greater or equal to α. Note that as the imposed probability α of satisfying the constraint increases, $\Phi^{-1}(\alpha)$ increases as well, implying that the stricter (more conservative) the imposed probability levels are, the more difficult it will be to satisfy the chance-constraint. In the limit of $\alpha = 1$, $\Phi^{-1}(\alpha)$ diverges to plus infinity and the chance-constraint becomes rigorously infeasible for any values of the deterministic variables x_i. Kataoka (1963) showed that the squared root of the variance,

$$Var\left(\sum_i^n a_i x_i\right)$$

is a convex function in x. Therefore, the deterministic equivalent constraint is convex for $\alpha \geqslant 0.5$ and concave for $\alpha \leqslant 0.5$. Convexity of the deterministic equivalent representation of the chance-constraint carries great significance in optimization studies because it greatly facilitates the identification of the optimum solution.

DETERMINISTIC EQUIVALENTS FOR SMOP FORMULATION

The SMOP formulation involves the following sets of constraints having uncertain parameters:

(1) $\quad \sum_{j=1}^{N} A_{ij} X_j + \sum_{k=1}^{M} B_{ik} Y_k = 0, i = 1, \ldots, N$

(2) $\quad Pr(f \geqslant f') \geqslant \alpha$

(3) $\quad Pr(X_j^L \leqslant X_j \leqslant X_j^U) \geqslant \beta_{xj}, j = 1, \ldots, N$

(4) $\quad Pr(V_i^{\pm,L} \leqslant V_i^{\pm} \leqslant V_i^{U,\pm}) \geqslant \beta v_i, i = 1, \ldots, N$

Each one of these constraints demands a different course of action for transforming it into an equivalent deterministic form.

Metabolite Flux Balance Constraints

Constraint set (1) is unique because it must be satisfied with probability one for every realization of the uncertain parameters A_{ij}, B_{ik}, X_j if the consumption/production fluxes for each metabolite are to match at steady-state. Constraint set (1) forms a linear square system of equations with unknowns the metabolite concentrations X_j. The solution of this system yields

$$X_j = \frac{\det A^j}{\det A}, j = 1, \ldots, N$$

where $\det A = \begin{vmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ A_{N1} & A_{N2} & \cdots & A_{NN} \end{vmatrix}$

and $\det A^j =$

$$\begin{vmatrix} A_{11} & \cdots & A_{1,j-1} & -\sum_{k=1}^{M} B_{1k} Y_k & A_{1,j+1} & \cdots & A_{1N} \\ A_{21} & \cdots & A_{2,j-1} & -\sum_{k=1}^{M} B_{2k} Y_k & A_{2,j+1} & \cdots & A_{2N} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ A_{N1} & \cdots & A_{N,j-1} & -\sum_{k=1}^{M} B_{Nk} Y_k & A_{N,j+1} & \cdots & A_{NN} \end{vmatrix}$$

Here A is a square matrix with elements A_{ij} and A^j is equal to the matrix A with the jth column substituted with the elements of the vector

$$-\sum_{k=1}^{M} B_{ik} Y_k$$

Both $\det A$ and $\det A^j$ involve nonlinear products of the uncertain parameters A_{ij}, which cannot be handled by the deterministic equivalent representation described in the previous subsection. To remedy this situation the linear approximation of $\det A$ and $\det A^j$ is sought with respect to the uncertain parameters A_{ij}. The expansion of $\det A$ (ignoring products between A_{ij}'s) around the nominal steady-state yields (Babbar, 1955),

$$\det A \cong \det \hat{A} + \sum_{i=1}^{N} \sum_{j=1}^{N} \Delta A_{ij} Cof_{ij}(\hat{A})$$

Where $\hat{A}_{ij}$ is the nominal steady-state (mean) value of A_{ij},

$$\Delta A_{ij} = A_{ij} - \hat{A}_{ij}$$

and $\hat{A}$ is the matrix of the mean values $\hat{A}_{ij}$ of A_{ij}, and $Cof_{ij}(\hat{A})$ is the ij^{th} cofactor of $\hat{A}$

$$Cof_{ij}(\hat{A}) = (-1)^{i+j} \det M_{ij}$$

Matrix M_{ij} is derived from the original matrix $\hat{A}$ after erasing the ith row and j^{th} column. This approximation is valid only if the deviations ΔA_{ij} are small compared with the mean values $\hat{A}_{ij}$. For example, if an error term is 10%, then the product of two errors is only 1% of the product of the

means. The appropriate expansion of det A^j in cofactors yields:

$$\det A^j = -\sum_{i=1}^{N} \left(-\sum_{k=1}^{M} B_{ik}Y_k \right) Cof_{ij}A$$

After applying the same rules to the expansion of the cofactors $Cof_{ij}A$ we obtain

$$Cof_{ij}A \cong Cof_{ij}\hat{A} + \sum_{\substack{i=1 \\ l \neq i}}^{N} \sum_{\substack{m=1 \\ m \neq j}}^{N} \Delta A_{lm} Cof_{lm}^{ij}\hat{A}$$

Where $Cof_{lm}^{ij}\hat{A}$ denote the cofactors of matrix $\hat{A}$ after eliminating row i and column j. After combining the last two results we get,

$$\det A^j = \sum_{i=1}^{N} \left(-\sum_{k=1}^{M} B_{ik}Y_k \right)$$
$$\left[Cof_{ij}\hat{A} + \sum_{\substack{i=1 \\ l \neq i}}^{N} \sum_{\substack{m=1 \\ m \neq j}}^{N} \Delta A_{lm} Cof_{lm}^{ij}\hat{A} \right]$$

which after rearranging yields:

$$\det A^j = \sum_{i=1}^{N} \left(-\sum_{k=1}^{M} B_{ik}Y_k \right) Cof_{ij}\hat{A}$$
$$+ \sum_{l=1}^{N} \sum_{\substack{m=1 \\ m \neq j}}^{N} \left[\sum_{\substack{i=1 \\ i \neq l}}^{N} \left(-\sum_{k=1}^{M} \hat{B}_{ik}Y_k \right) Cof_{lm}^{ij}\hat{A} \right] \Delta A_{lm}$$

This expression can be simplified by introducing the auxiliary variables F_{lm} and δ_{mj} as follows:

$$F_{lm} = \sum_{\substack{i=1 \\ i \neq l}}^{N} \left(-\sum_{k=1}^{M} \hat{B}_{ik}Y_k \right) Cof_{lm}^{ij}\hat{A}, \quad \delta_{mj} = \begin{cases} 0 & m = j \\ 1 & m \neq j \end{cases}$$

The expression for det A^j then further simplifies to

$$\det A^j = \sum_{i=1}^{N} \left(-\sum_{k=1}^{M} B_{ik}Y_k \right) Cof_{ij}\hat{A} + \sum_{l=1}^{N} \sum_{m=1}^{N} F_{lm}\delta_{mj}\Delta A_{lm}$$

After applying the mean and variance operatores to det A and det A^j we obtain:

$$\mu (\det A) = \det \hat{A}$$

$$Var (\det A) = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{i'=1}^{N} \sum_{j'=1}^{N} (Cof_{ij}\hat{A})(Cof_{i'j'})Cov(A_{ij}, A_{i'j'})$$

$$\mu (\det A^j) = \sum_{i=1}^{N} \left(-\sum_{k=1}^{M} \hat{B}_{ik}Y_k \right) Cof_{ij}\hat{A}$$

$$Var (\det A^j) =$$
$$\sum_{i=1}^{N} \sum_{k=1}^{M} \sum_{i'=1}^{N} \sum_{k'=1}^{M} (Y_k Cof_{ij}\hat{A})(Y_{k'} Cof_{i'j}\hat{A})Cov(B_{ik}, B_{i'k'})$$

$$+ \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{l'=1}^{N} \sum_{m'=1}^{N} F_{lm}F_{l'm'}\delta_{mj}\delta_{m'j'}Cov(A_{lm}, A_{l'm'})$$

$$+ \sum_{i=1}^{N} \sum_{k=1}^{M} \sum_{l=1}^{N} \sum_{m=1}^{N} (Y_k Cof_{ij}\hat{A})(F_{lm}\delta_{mj})Cov(A_{lm}, B_{ik})$$

The quality of the Babbar (1995) approximation is verified for the determinant of matrix A in Example 1 Monte Carlo (MC) simulation (Tong, 1990) was used to generate random normal variates A_{ij} with means $\hat{A}_{ij}$ and variances of 1%, 2.5% and 5% times the mean values respectively. This allowed the derivation of the "exact" probability distribution of det A. The comparison between the Babbar (1955) approximation and the MC simulation exact results (see Table I) demonstrates that the Babbar (1955) approximation describes reasonably well the cumulative probability distribution of det A.

Metabolic Objective Chance-Constraint

In this work, two different metabolic objectives are considered: (i) maximization of the steady-state concentration of a specific metabolite, and (ii) maximization of a particular flux. The first objective involves $(f = X_j, f^t = X_j^t)$ and yields the following chance-constraint,

$$Pr (X_j \geqslant X_j^t) \geqslant \alpha$$

where j is the metabolite whose concentration needs to be maximized and X_j^t is the metabolite concentration target to be met with probability of at least α. After substituting the approximation for X_j and rearranging we get,

$$Pr(X_j^t \det A - \det A^j \leqslant 0) \geqslant \alpha$$

Because uncertain parameters appear linearly in this chance-constraint (X_j^t is deterministic), based on the analysis presented in subsection 3.4, its deterministic equivalent is,

$$\mu(X_j^t \det A - \det A^j) + \Phi^{-1}(\alpha) [Var(X_j^t \det A - \det A^j)]^{1/2} \leqslant 0$$

After substituting the expressions for the means of the linear approximations of det A and det A^j we have:

$$\mu(X_j^t \det A - \det A^j)$$

Table I. Comparison of the cumulative probability distribution values of det A obtained by Monte Carlo (MC) simulation and estimated with the approximation of Babbar (1955).

$\sigma = 1\%$		$\sigma = 2.5\%$		$\sigma = 5\%$	
MC	Appr.	MC	Appr.	MC	Appr.
0.442	0.577	0.599	0.651	0.544	0.622
0.737	0.720	0.805	0.758	0.683	0.693
0.921	0.913	0.923	0.844	0.885	0.814
0.985	0.960	0.996	0.948	0.970	0.901

$$= X_j^t \, \mu \, (\det A) - \mu \, (\det A^j)$$

$$= X_j^t \det \hat{A} - \sum_{i=1}^{N} \left(-\sum_{k=1}^{M} \hat{B}_{ik} Y_k \right) Cof_{ij} \hat{A}$$

The substitution of the variance expressions yields:

$$Var \, (X_j^t \det A - \det A^j)$$

$$= \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{l'=1}^{N} \sum_{m'=1}^{N} (X_j^t Cof_{lm} \hat{A} - F_{lm} \delta_{mj})$$

$$(X_j^t Cof_{l'm'} \hat{A} - F_{l'm'} \delta_{m'j}) Cov(A_{lm}, A_{l'm'})$$

$$+ \sum_{i=1}^{N} \sum_{k=1}^{M} \sum_{i'=1}^{N} \sum_{k'=1}^{M} (Y_k Cof_{ij} \hat{A})(Y_{k'} Cof_{i'j} \hat{A}) Cov(B_{ik}, B_{i'k'})$$

$$+ 2 \sum_{i=1}^{N} \sum_{k=1}^{M} \sum_{l=1}^{N} \sum_{m=1}^{N} (Y_k Cof_{ij} \hat{A})(X_j^t Cof_{lm} \hat{A} - F_{lm} \delta_{mj})$$

$$Cov(A_{ik}, B_{lm})$$

Note that despite the complexity of the expressions presented above, the deterministic equivalent representation of the chance-constraint objective, is convex for $\alpha \geqslant 0.5$ as shown in (Kataoka, 1963).

The second type of a metabolic objective involves the maximization of a metabolite flux, $(f = V_i^{\pm}, f^t = V_i^{\pm,t})$ which either consumes (–) or produces (+) metabolite i. Based on the S-System representation we have,

$$V_i^+ = \alpha_i \prod_{j=1}^{N} x_j^{g_{ij}} \prod_{k=1}^{M} y_k^{g'_{ik}}$$

$$V_i^- = \alpha_i \prod_{j=1}^{N} x_j^{h_{ij}} \prod_{k=1}^{M} y_k^{h'_{ik}}.$$

After performing the logarithmic transformation and subtracting the nominal steady-state flux balance we have:

$$\ln \left(\frac{V_i^+}{V_i^{+,ss}} \right) = \sum_{j=1}^{N} g_{ij} X_j + \sum_{k=1}^{M} g'_{ik} Y_k \text{ or}$$

$$\ln \left(\frac{V_i^-}{V_i^{-,ss}} \right) = \sum_{j=1}^{N} h_{ij} X_j + \sum_{k=1}^{M} h'_{ik} Y_k$$

After substituting in the chance-constraint designating the metabolic objective,

$$Pr[V_i^{\pm} \geqslant V_i^{\pm,t}] \geqslant \alpha$$

the previous expressions we obtain:

$$Pr \left[\sum_{j=1}^{N} g_{ij} X_j + \sum_{k=1}^{M} g'_{ik} Y_k \geqslant \ln \left(\frac{V_i^{t,+}}{V_i^{+,ss}} \right) \right] \geqslant \alpha \text{ or}$$

$$Pr \left[\sum_{j=1}^{N} h_{ij} X_j + \sum_{k=1}^{M} h'_{ik} Y_k \geqslant \ln \left(\frac{V_i^{t,-}}{V_i^{-,ss}} \right) \right] \geqslant \alpha$$

After replacing the metabolite concentrations with their linear approximation and rearranging we have:

$$Pr \left[\sum_{j=1}^{N} g_{ij} \det A^j + \sum_{k=1}^{M} (g'_{ik} \det A) Y_k \geqslant \ln \left(\frac{V_i^{t,+}}{V_i^{-,ss}} \right) \det A \right]$$

$$\geqslant \alpha \text{ or}$$

$$Pr \left[\sum_{j=1}^{N} h_{ij} \det A^j + \sum_{k=1}^{M} (h'_{ik} \det A) Y_k \geqslant \ln \left(\frac{V_i^{t,-}}{V_i^{-,ss}} \right) \det A \right]$$

$$\geqslant \alpha$$

After replacing the nonlinear products $g'_{ik} \det A$ and $h'_{ik} \det A$ with their linear approximations,

$$g'_{ik} \det A \cong \hat{g}'_{ik} \det \hat{A} + \hat{g}'_{ik} (\det A - \det \hat{A})$$
$$+ \det \hat{A}(g'_{ik} - \hat{g}'_{ik})$$

$$h'_{ik} \det A \cong \hat{h}'_{ik} \det \hat{A} + \hat{h}'_{ik} (\det A - \det \hat{A})$$
$$+ \det \hat{A}(h'_{ik} - \hat{h}'_{ik})$$

the resulting chance-constraint is linear in the uncertain parameters and deterministic variables; thus its deterministic equivalent representation is convex for $\alpha \geqslant 0.5$ and can be obtained as shown earlier.

Lower and Upper Bounds on Metabolite Concentrations

Constraint set (3) implies that the metabolite concentrations must be within some lower and upper bounds with some designated probability β_{X_i}:

$$Pr(X_j^L \leqslant X_j \leqslant X_j^U) \geqslant \beta_{X_i}$$

Unlike constraint (2) which imposes a probability target on a single constraint, constraint set (3) imposes a limit on the joint probability of feasibility of two constraints. The probabilistic restriction on the chance-constraint set (3) can be decoupled as follows. Because,

$$Pr(X_j^L \leqslant X_j \leqslant X_j^U) = 1 - Pr(X_j \leqslant X_j^L) - Pr(X_j \geqslant X_j^U)$$

and

$$Pr(X_j \leqslant X_j^L) = 1 - Pr(X_j^L \leqslant X_j), \, Pr(X_j \geqslant X_j^U)$$
$$= 1 - Pr(X_j \leqslant X_j^U)$$

we have

$$Pr(X_j^L \leqslant X_j \leqslant X_j^U) = Pr(X_j^L \leqslant X_j) + Pr(X_j \leqslant X_j^U) - 1$$

Therefore, the original joint probability on the left hand side of the chance-constraint decomposes into the sum of two non-jointed probability expressions:

$$Pr(X_j^L \leqslant X_j) + Pr(X_j \leqslant X_j^U) \geqslant 1 + \beta_{X_j}$$

By defining,

$$\beta_j^L = Pr(X_j^L \leqslant X_j) \text{ and } \beta_j^U = Pr(X_j \leqslant X_j^U)$$

and performing the following transformation of variables,

$$t_j^L = \Phi^{-1}(\beta_j^L) \text{ and } t_j^U = \Phi^{-1}(\beta_j^U)$$

the deterministic equivalent representation of the "decoupled" chance-constraint on the metabolite bounds is:

$$\mu(X_j^L - X_j) + t_j^L[Var(X_j)]^{1/2} \leq 0, j = 1, \ldots, N$$

$$\mu(X_j - X_j^U) + t_j^U[Var(X_j)]^{1/2} \leq 0, j = 1, \ldots, N$$

$$\Phi(t_{X_j}^L) + \Phi(t_{X_j}^L) \geq 1 + \beta_{X_j}$$

However, this representation is nonconvex due to the presence of the products $t_j^L[Var(X_j)]^{1/2}$ and $t_j^U[Var(X_j)]^{1/2}$. Although multiple local optima may exist in theory, this was not observed in practice after employing multiple starting points. Of course, this is only indicative and not a rigorous proof for the existence of a unique optimum. Note that by utilizing a global optimization algorithms, such as αBB (Androulakis et al., 1995) the identification of the global optimum can be guaranteed. Alternately, by considering the lower and upper bounds separately,

$$Pr(X_j^L \leq X_j) \geq \beta_{X_j} \text{ and } Pr(X_j \leq X_j^U) \geq \beta_{X_j}$$

the nonconvexities are eliminated.

Lower and Upper Bounds on Metabolite Fluxes

Constraint set (4) imposes lower and upper bounds on metabolite fluxes (+ or −) to be satisfied with probability of at least β_{V_i}. Based on the analysis presented in earlier subsections, constraint set (4) can be recast in terms of metabolite concentrations and enzyme activities:

$$Pr\left[\ln\left(\frac{V_i^{L,+}}{V_i^{+,ss}}\right) \leq \sum_{j=1}^{N} g_{ij}X_j + \sum_{k=1}^{M} g'_{ik}Y_k \right.$$

$$\left. \leq \ln\left(\frac{V_i^{U,+}}{V_i^{+,ss}}\right) \right] \geq \beta_{V_j^+}$$

$$Pr\left[\ln\left(\frac{V_i^{L,-}}{V_i^{-,ss}}\right) \leq \sum_{j=1}^{N} h_{ij}X_j + \sum_{k=1}^{M} h'_{ik}Y_k \right.$$

$$\left. \leq \ln\left(\frac{V_i^{U,-}}{V_i^{-,ss}}\right) \right] \geq \beta_{V_j^-}$$

Following the same line of analysis presented in the previous subsection the deterministic equivalent representation can be obtained.

Modeling Discrete Enzyme Expression Levels

So far we have assumed that the enzyme levels can take any value between a lower and an upper bound. This requires a currently nonexistent level of precision for the genetic manipulations. A more realistic description of the enzyme levels can be accomplished by assuming that they can assume only a finite number of values (Hatzimanikatis et al., 1996b). Therefore, enzyme k may assume only $1, \ldots, L$ different levels denoted by y_{kl}. For example, if the level of enzyme k can be changed by only $\pm20\%$ or $\pm50\%$, the set of feasible enzyme level manipulations is $y_{kl} \in \{0.5, 0.8, 1, 1.2, 1.5\}$. This discrete enzyme level manipulation can be expressed mathimatically with the following constraints,

$$Y_k = \sum_{l}^{L} \ln(y_{kl})z_{kl}, k = 1, \ldots, M$$

$$\sum_{l}^{L} z_{kl} = 1, k = 1, \ldots, M$$

$$z_{kl} = \{0,1\}$$

where z_{kl} is a binary variable assuming values of only zero or one. The second constraint requires that only one of the binary variables $z_{k1}, z_{k2}, \ldots, z_{kL}$ is equal to one. Thus, the logarithmic deviation of the enzyme levels, Y_k, can assume only values equal to one of the discrete levels. Note that the addition of these constraints in the SMOP model transforms it into a mixed integer nonlinear programming (MINLP) problem which, however, can be solved using existing MINLP solvers.

COMPUTATIONAL RESULTS

Two metabolic pathway optimization studies are addressed to highlight the proposed framework of analysis and optimization of metabolic pathways under experimental and/or modeling uncertainty. First, an introductory example is considered where reasonable values for the variance of the kinetic orders are selected to reflect the experimental uncertainty in the S-System kinetic orders. The characteristics of the model and the significance of the obtained results are briefly discussed. Next, the framework of analysis and optimization is applied to a reference pathway described by known Michaelis-Menten kinetics. The sources of model uncertainty in the S-System kinetic orders are identified and a comparison between the deterministic S-System optimization, the nonlinear optimization using the known rate expressions, and the proposed framework is drawn. All optimization models have been implemented with General Algebraic Modeling System (GAMS) (Brooke et al., 1988) and run on a RS6000-43P-132 IBM workstation.

Example 1: Yield Optimization in XMP and GMP Production

In this example, the sum x_4 of the concentrations of xanthine monophosphate (XMP) and guanosine monophosphate (GMP) is maximized. A schematic of the simplified metabolic pathway is shown in Figure 1. A complete description of the metabolic pathway can be found in (Voit, 1992). The S-System representation of the pathway is:

$$\frac{dx_1}{dt} = 900x_3^{-0.5}x_4^{-0.5}y_1 - 10x_1^{0.5}x_2^{-0.1}x_3^{-0.2}x_4^{-0.2}y_2$$

$$\frac{dx_2}{dt} = 10x_1^{0.5}x_2^{-0.1}x_3^{-0.1}x_4^{-0.5} y_2y_4y_5 - 100x_2^{0.5}x_3^{-0.5}x_4^{-0.5} y_3y_6$$

$$\frac{dx_3}{dt} = 200x_2^{0.5}x_3^{-0.5}y_3 - 10x_1^{0.1}x_3s_4^{-1}y_4$$
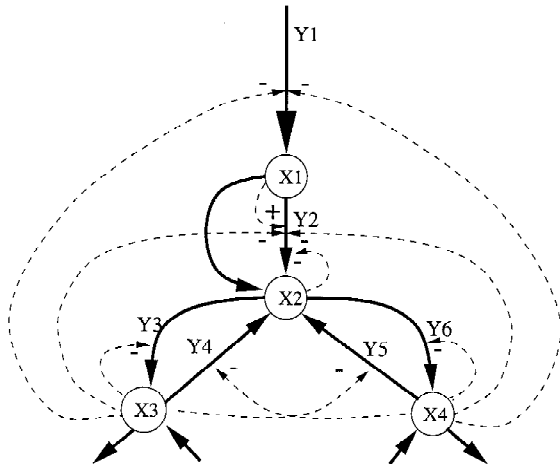
Figure 1. Diagram of the simplified pathway of XMP and GMP synthesis. Solid lines denote reactions and dashed lines denote regulatory interactions.

Figure 2. Objective value target for different specifications of α and β.

$$\frac{dx_4}{dt} = 30x_2^{0.5}x_4^{-0.5}y_6 - 100x_3^{-1}x_4y_5$$

The nominal steady-state concentrations are $x_1^{ss} = 5.42$, $x_2^{ss} = 213$, $x_3^{ss} = 2{,}417$ and $x_4^{ss} = 482$. The nominal enzyme concentrations, y_i, and their kinetic orders are equal to one. The optimization problem to be solved is defined as:

1. Objective: Maximize the concentration x_4.
2. Constraints:
   (a) The enzyme activities can be varied between 0.2 and 5 times their nominal steady-state values.
   (b) The concentrations of x_1, x_2, and x_3 must remain within ±10% of their nominal steady-state values.

Based on the analysis procedure described in the previous section, the deterministic equivalent representation of the chance constrained formulation is first obtained. A number of optimization runs were performed for different values of the probability levels α and β to vary the relative importance of meeting the objective target as opposed to satisfying the physiological constraints. The variance of the parameters A_{ij} and B_{ik} is assumed to be 2% of their mean values. The mean values are taken from Voit (1992). A 2% variance for a normally distributed uncertain parameter implies that 99% of the times the realization of the uncertain parameter will be, $(\pm 0.02 \cdot 2.58 \cdot \text{mean})$, away from the mean value (Kreyszig, 1993). Figure 2 illustrates the optimization results. Clearly, the higher the value of β (more likely to satisfy the physiological constraints), the lower is the value of the achievable with probability α target on x_4. Note that if uncertainty is ignored, a maximum concentration of $x_4 = 3{,}972$ is predicted. However, for probability levels as low as $\alpha = \beta = 0.5$, this value appears to be unachievable (violates physiological constraints more than 50% of the time). As β approaches unity, the achievable objective target approaches the original steady-state value. This is consistent with the fact that *unless a finite level of*
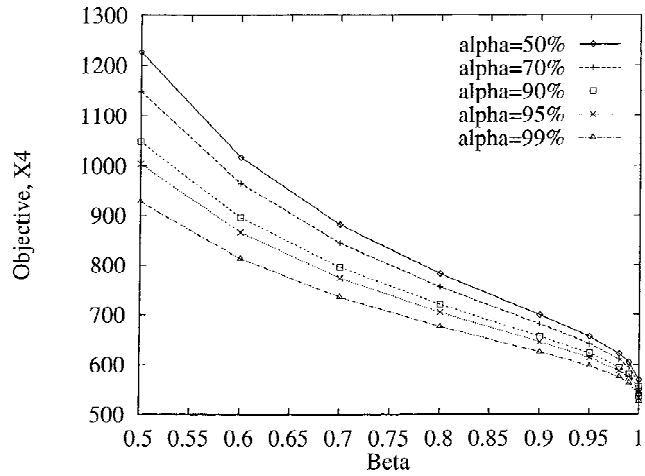
uncertainty can be tolerated the only "feasible" choice is the nominal steady-state. The effect of the probability level α on the achievable objective target is not quite as pronounced as the effect of β, but it acts in the same direction. This implies that the effect of uncertainty is to penalize the prediction of the derterministic model.

Plotted in Figure 3 are the optimal enzyme activities for different values of β with $\alpha = 0.5$. Clearly, the optimum enzyme activity profile is "strongly" dependent on the choice of β, implying that *the answer to the question of what are the best enzyme activity levels depends on how often physiological constraints are allowed to be violated.* As β moves towards one, all of the enzyme activities tend to "aggregate" towards their nominal steady-state values. The "speed" of this aggregation for different enzymes illustrates the relative sensitivity of physiological constraints on different enzymes. Figure 4 addresses the sensitivity of enzyme activities for different values of the probability level α while keeping $\beta = 0.95$. Clearly, the optimal enzyme activity profile appears to be largely unaffected (for $\alpha \leqslant$
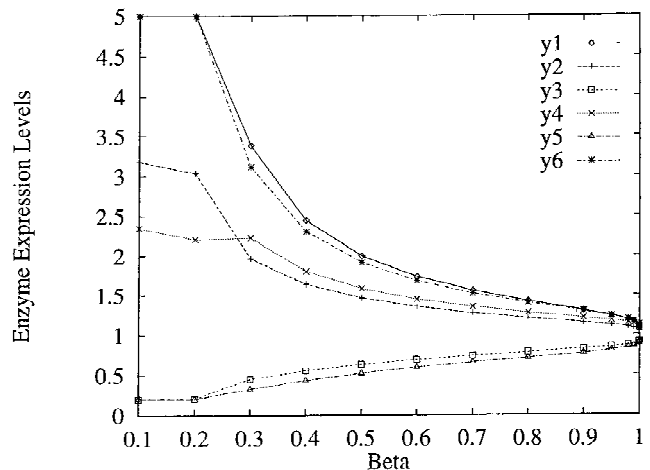
**Figure 3.** Optimal enzyme levels for different values of β with $\alpha = 50\%$.
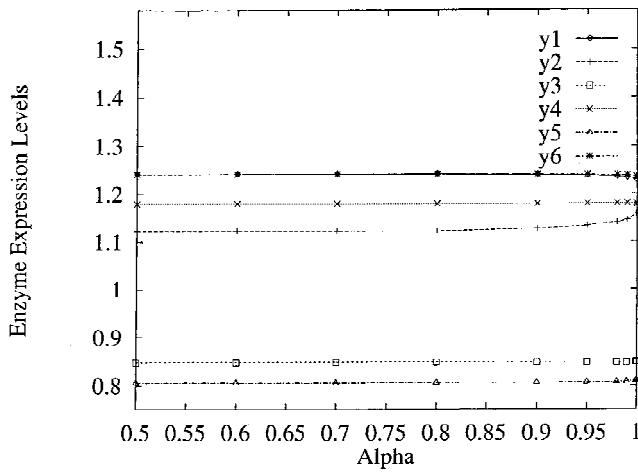
Figure 4. Optimal enzyme levels for different values of α with $\beta = 95\%$.

0.9999) by the value of the probability level α. This means that physiological constraints are more strongly affected by uncertainty. Figure 5 illustrates the effect of the extent of the uncertainty on the objective target by considering uncertain parameter variances equal to 1,2,5 and 10% of the mean values. The results indicate, as expected, that the quantitative effect of the extent of parameter uncertainty on achievable objectives is very pronounced and changes rapidly in the 1–2% range.

The optimization runs described above indicate that even a 2% variance in the S-System parameters may have a profound effect on the reliability of the obtained optimization results with respect to satisfying physiological constraints and meeting the objective target.

Example 2: Optimization under Modeling Uncertainty

In this example, a reference metabolic pathway proposed by (Sorribas et al., 1993) is examined. The diagram of the
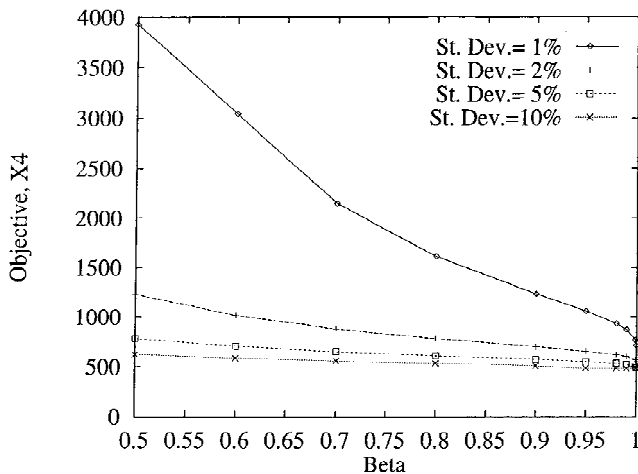
Figure 5. Objective value target for different variance values of the S-System parameters.

metabolic pathway is shown in Figure 6. The predictions of the proposed stochastic framework are compared with those ignoring modeling uncertainty. To check the validity of both predictions, we assume that the irreversible Michaelis-Menten kinetics model provides the "true" description of the metabolic pathway. The predictions of the S-System with and without considering uncertainty are thus contrasted against the Michaelis-Menten representation. The relevant kinetic parameters and nominal steady-state properties are given in Table II. The S-System representation of this metabolic pathway yields:

$$\frac{dx_1}{dt} = \alpha_1 x_5^{g_{15}} y_1^{g'_{11}} - b_1 x_1^{h_{11}} x_4^{h_{14}} y_3^{h'_{13}}$$

$$\frac{dx_2}{dt} = \alpha_2 x_6^{g_{26}} y_2^{g'_{22}} - b_2 x_2^{h_{22}} x_4^{h_{24}} y_4^{h'_{23}}$$

$$\frac{dx_3}{dt} = a_3 x_1^{g_{31}} x_2^{g_{32}} x_4^{g_{34}} y_3^{g'_{33}} y_4^{g'_{34}} - b_3 x_3^{h_{33}} y_5^{'_{35}}$$

$$\frac{dx_4}{dt} = a_4 x_3^{g_{43}} y_5^{g'_{45}} - b_4 x_4^{h_{44}} y_6^{h'_{46}}$$

The parameters of the S-System were calculated from the rate equations based on the approach outlined in (Torres, 1994a). By following the analysis described in Appendix A the S-System modeling uncertainty, represented by the variance-covariance terms of the kinetic orders a_{ij}, b_{ik}, is estimated in Appendix B.

The optimization task considered is defined as:

1. Objective: Maximize the flux (V_4^+ or V_4^-)
2. Constraints:
   (a) Maintain enzyme activities between 0.1 and 10 times their nominal steady-state values.
   (b) Keep all metabolite concentrations within 20% of their nominal steady-state values.
   (c) Fix the activity of y_4 at the nominal steady-state value.

The last restriction was introduced to break the symmetry of the pathway and eliminate some of the multiple solutions. This multiplicity of solutions is also observed elsewhere (Dervakos and Dean, 1994; Hatzimanikatis et al., 1996b) and is due to the imposed lower and upper bounds on the enzyme levels.

First, the nonlinear programming (NLP) problem based on the Michaelis-Menten rate expressions is optimized assuming that the latter provides a "perfect" description of the
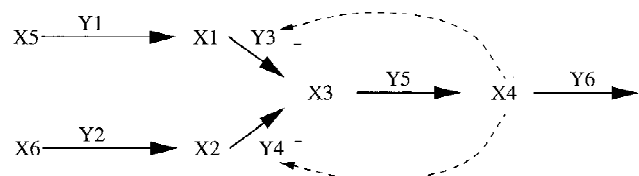
Figure 6. Reference metabolic pathway of Example 2.

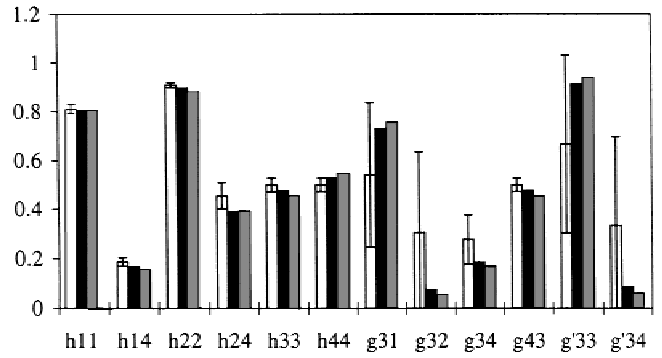**Table II.** Nominal steady-state values and kinetic parameters for Example 2.

x_i	Conc.	y_i	V_{max}	K_m	K_{i4}
x_1	2	y_1	3	100	—
x_2	3	y_2	3	250	—
x_3	1	y_3	5.33	6.67	6.67
x_4	2	y_4	5.50	15	2
x_5	50	y_5	3	1	—
x_6	50	y_6	3	2	—

Figure 7. Magnitude of the S-System parameters of Example 2: (white)-the S-System parameters at the original steady-state; (black)-the same parameters estimated at the new steady-state using second-order Taylor expansion; (grey)-the S-System parameters recalculated at the new steady-state.

metabolic pathway. These solutions are summarized in Table III. The tabulated ranges of values, instead of single values, for some of the enzyme activities and metabolite concentrations, imply that the optimal objective value is unaffected by changing the enzyme levels or metabolite concentrations within these ranges. The values of the S-System kinetic orders are (i) recalculated and (ii) estimated using second-order Taylor expansion at the new optimal steady-state obtained from the optimization of the Michaelis-Menten representation. These results are plotted in Figure 7. The first bar (white), denotes the values of the S-System kinetic orders at the original steady state. The second bar (black), gives the values of the S-System kinetic orders at the new steady-state estimated with a second-order Taylor expansion (see Appendices A,B). The third bar (grey), represents the values of the S-System kinetic orders recalculated at the new steady-state. Clearly, the new values of some of the S-System kinetic orders significantly differ from the ones at the original steady-state. The greatest difference is observed for the kinetic orders which depend on enzyme levels (i.e., g_{31}, g_{32}, g_{34}, g'_{33}, and g'_{34}). This is expected because the range within the enzyme concentrations are allowed to vary is much wider than the one for the metabolite concentrations. Error bars denote the 99% interval of confidence for the second-order Taylor expansion estimates of the kinetic orders. Note that the recalculated values of the kinetic orders at the new steady-state are within these confidence intervals.

To measure the effect of neglecting modeling error on the optimization of S-Systems the metabolic pathway was optimized based on the S-System representation at the original steady-state. The resulting optimum objective function involved a value of 7.62 for the flux V_4^+. However, after substituting the obtained optimal enzyme activities in the Michaelis-Menten description and solving for the metabo-

lite concentrations, a number of physiological constraints are violated. These results, shown in Table IV, indicate that some of the physiological restrictions are violated. Specifically, the values derived for x_2, x_3, and x_4 are 50%, 173%, and 40% higher than their respective imposed upper bounds. Torres et al. (1996) suggested adding stoichiometric constraints for the dependent fluxes to reduce the S-System errors away from the steady-state. After introducing the stoichiometric constraints $1.5V_1^- \geq V_3^-$ and $1.5V_2^- \geq 0.5V_3^-$ in the model, the new results are shown in Table V. While the new optimal solution is more conservative, the physiological constraints are again violated by as much as 17%. Note that the stability of all solutions was confirmed by eigenvalue analysis of the steady-state description based on the Michaelis-Menten rate expressions.

Next the performance of the proposed stochastic optimization of the S-System representation is evaluated. The results from the solution of the SMOP formulation for the expected objective values and corresponding enzyme expression levels at $\alpha = 0.5$ and several β values are shown in Table VI. The predicted optimal enzyme levels (see Table VI) are then substituted back into the Michaelis-Menten kinetic description to check whether the proposed approach successfully hedged against uncertainty. The results, which

Table III. Optimal enzyme levels and metabolite concentrations for Example 2 based on the Michaelis-Menten formalism.

Objective flux	Enzyme levels	Metabolite conc.
$V_4^+ = 10.64$	$y_1 = 10.0y_1^{ss}$	$x_1 = [0.95 - 1.20]x_1^{ss}$
	$y_2 = 1.29y_2^{ss}$	$x_2 = 1.20x_2^{ss}$
	$y_3 = [8.34 - 10.0]y_3^{ss}$	$x_3 = [0.80 - 1.20]x_3^{ss}$
	$y_4 = 1.00y_4^{ss}$	$x_4 = 0.80x_4^{ss}$
	$y_5 = [6.51 - 7.94]y_5^{ss}$	
	$y_6 = 7.99y_6^{ss}$	

Table IV. Optimal enzyme levels obtained by solving the S-System optimization problem without accounting for modeling uncertainty. The listed objective value and metabolite concentrations are estimated by substituting the enzyme levels in the Michaelis-Menten description of the pathway.

Objective value	Enzyme levels	Metabolite conc.
$V_4^+ = 10.65$	$y_1 = 10.0y_1^{ss}$	$x_1 = 1.16x_1^{ss}$
	$y_2 = 1.31y_2^{ss}$	$x_2 = 1.80x_2^{ss}$
	$y_3 = 10.0y_3^{ss}$	$x_3 = 3.28x_3^{ss}$
	$y_4 = 1.00y_4^{ss}$	$x_4 = 1.67x_4^{ss}$
	$y_5 = 4.64y_5^{ss}$	
	$y_6 = 5.68y_6^{ss}$	

Table V. Optimal enzyme levels and resulting metabolite concentrations obtained for the S-System optimization with stoichiometric constraints.

Objective value	Enzyme levels	Metabolite conc.
$V_4^+ = 3.57$	$y_1 = 2.91 y_1^{ss}$	$x_1 = 1.25 x_1^{ss}$
	$y_2 = 1.31 y_2^{ss}$	$x_2 = 1.29 x_2^{ss}$
	$y_3 = 2.41 y_3^{ss}$	$x_3 = 1.40 x_3^{ss}$
	$y_4 = 1.00 y_4^{ss}$	$x_4 = 0.91 x_4^{ss}$
	$y_5 = 2.03 y_5^{ss}$	
	$y_6 = 2.49 y_6^{ss}$	

were proven to be stable, are shown in Table VII. These results demonstrate that while the predicted optimal objective values are more conservative than those for the deterministic S-System, all physiological constraints are satisfied if a high enough value of β is selected. Specifically, all physiological constraints are satisfied for β greater than 0.8. Variation of α, as in the previous example, did not affect the optimal enzyme levels significantly. While no general conclusions can be drawn from this example, the results are indicative that in certain cases the optimization procedure carried out without considering model uncertainty may lead to misleading results. However, by modeling kinetic order uncertainty and selecting a large enough value for β, feasibility of all physiological constraints in the face of kinetic order uncertainty can be guaranteed.

SUMMARY AND CONCLUSIONS

In this paper, a probabilistic description of the S-System coefficients was introduced for quantitatively evaluating and optimizing metabolic pathway S-System models associated with modeling and/or experimental uncertainty. The selection of the S-System formalism was based on its relative popularity and mathematical tractability. Based on notions from chance-constrained programming and mathematical statistics, a novel approach was introduced for transforming the original stochastic formulation into a deterministic one which can be solved with existing optimization algorithms. The deterministic transformation was based on two assumptions: (i) the approximation introduced by Babbar (1955) which neglects second and higher order products of error terms in the calculation of the metabolite concentrations, and (ii) the description of uncertain parameters (kinetic orders) with stable probability distributions. While optimization results ignoring uncertainty may violate physiological constraints and yield unrealistically optimistic estimates for the objective targets, the proposed framework provides quantitative answers to questions regarding how likely it is to achieve a particular objective without violating any physiological constraints. Trade-off curves between metabolic objectives, probabilities of meeting these objectives, and chances of satisfying physiological constraints provide a concise and systematic way to guide enzyme activity alterations to meet an objective in the face of modeling and experimental uncertainty.

The implementation of metabolic optimization based on the S-System formalism, or for that matter any other mathematical formalism, faces several difficulties. The first one is the scarcity of experimental data. Though kinetic information on many biochemical reactions exists, it is either obtained in vitro or the enzyme activity is characteristic for a particular microorganism. When the definition of a metabolic pathway is compiled from data from different sources, the discrepancies due to different experimental techniques and different environments may lead to ill-defined models. In addition, some of the experimental measurements are prone to significant experimental error. The systematic optimization of metabolic pathways based on mathematical modeling, however, provides a valuable tool for identifying promising directions of enzyme activity alterations in metabolic pathways which are typically characterized by highly complex responses. It was the objective of this work to enhance the practical value of these predictions by coupling these optimization studies with a quantitative description of uncertainty.

While the approach so far has been focused on continuous enzyme activity alterations within the S-System representation formalism, it is general enough to handle discrete enzyme level manipulation and/or the optimization of the regulatory structure of enzymatic pathways (Hatzimanikatis et al. (1996a,b)) as well as other mathematical formalisms used to describe metabolic pathways.

APPENDIX A: ESTIMATION OF THE VARIANCES AND COVARIANCES OF THE KINETIC ORDERS

Given sufficient experimental data on metabolite concentrations, enzyme activities, and fluxes, rigorous regression techniques can be applied to derive unbiased estimates of means, variances and covariances as described in (Maranas, 1997) in the context of optimal molecular design. However, for most metabolic pathways such detailed information is not available. Therefore, a second-order Taylor expansion around the nominal steady-state is employed to estimate the model uncertainty in the kinetic orders within the imposed bounds of the metabolite concentrations and enzyme levels. Note that the S-System representation stops at the first-order Taylor expansion. A second-order Taylor expansion was utilized by Cascante et al. (1991) to improve predicting the response of a metabolic system to perturbation. The second-order Taylor expansion of the flux V_i^+ in log-log space yields:

$$
\ln(V_i^+) = \ln(V_i^{ss+}) + \sum_j \left(\frac{x_j}{V_i^+} \frac{\partial V_i^+}{\partial x_j} \right)_{ss} X_j
$$
$$
+ \sum_k \left(\frac{y_k}{V_i^+} \frac{\partial V_i^+}{\partial y_k} \right)_{ss} Y_k
$$
$$
+ \frac{1}{2!} \sum_j \sum_{j'} \left(\frac{x_j x_{j'}}{V_i^+} \frac{\partial^2 V_i^+}{\partial x_j \partial x_{j'}} \right)_{ss} X_j X_{j'}
$$

Table VI. Optimal solutions of the (SMOP) formulation for $\alpha = 0.5$ and a number of values for β.

β	V_4^+	y_1	y_2	y_3	y_4	y_5	y_6
0.50	5.4319	18.110	3.914	26.613	5.500	10.974	12.146
0.60	4.2483	12.871	3.707	19.351	5.500	8.583	9.214
0.70	3.5337	9.948	3.570	15.192	5.500	7.140	7.507
0.80	3.0406	8.051	3.472	12.435	5.500	6.144	6.361
0.90	2.6474	6.615	3.395	10.312	5.500	5.349	5.470
0.95	2.4498	5.921	3.357	9.272	5.500	4.950	5.031

$$+\frac{1}{2!}\sum_j\sum_k\left(\frac{x_j y_k}{V_i^+}\frac{\partial^2 V_i^+}{\partial x_j \partial y_k}\right)_{ss} X_j Y_k$$

$$+\frac{1}{2!}\sum_j\sum_k\left(\frac{x_j y_k}{V_i^+}\frac{\partial^2 V_i^+}{\partial x_j \partial y_k}\right)_{ss} X_j Y_k$$

$$+\frac{1}{2!}\sum_k\sum_{k'}\left(\frac{y_k y_{k'}}{V_i^+}\frac{\partial^2 V_i^+}{\partial y_k \partial y_{k'}}\right)_{ss} Y_k Y_{k'}$$

After recasting V_i^+ in the usual power-law form the kinetic orders are no longer functions only of the steady-state but also depend on the deviations of the metabolic concentrations and enzyme levels from the steady-state values:

$$g_{ij} = g_{ij}^{ss} + \frac{1}{2!}\sum_{j'}\left(\frac{x_j x_{j'}}{V_i^+}\frac{\partial^2 V_i^+}{\partial x_j \partial x_{j'}}\right)_{ss} X_{j'}$$

$$+\frac{1}{2!}\sum_k\left(\frac{x_j y_k}{V_i^+}\frac{\partial^2 V_i^+}{\partial x_j \partial y_k}\right)_{ss} Y_k$$

$$g_{ik}' = g_{ik}'^{ss} + \frac{1}{2!}\sum_j\left(\frac{y_k x_j}{V_i^+}\frac{\partial^2 V_i^+}{\partial y_k \partial x_j}\right)_{ss} X_j$$

$$+\frac{1}{2!}\sum_{k'}\left(\frac{y_k y_{k'}}{V_i^+}\frac{\partial^2 V_i^+}{\partial y_k \partial y_{k'}}\right)_{ss} Y_{k'}$$

$$h_{ij} = h_{ij}^{ss} + \frac{1}{2!}\sum_{j'}\left(\frac{x_j x_{j'}}{V_i^-}\frac{\partial^2 V_i^-}{\partial x_j \partial x_{j'}}\right)_{ss} X_{j'}$$

$$+\frac{1}{2!}\sum_k\left(\frac{x_j y_k}{V_i^-}\frac{\partial^2 V_i^-}{\partial x_j \partial y_k}\right)_{ss} Y_k$$

$$h_{ij}' = h_{ij}'^{ss} + \frac{1}{2!}\sum_j\left(\frac{y_k x_j}{V_i^-}\frac{\partial^2 V_i^-}{\partial y_k \partial x_j}\right)_{ss} X_j$$

$$+\frac{1}{2!}\sum_{k'}\left(\frac{y_k y_{k'}}{V_i^-}\frac{\partial^2 V_i^-}{\partial y_k \partial y_{k'}}\right)_{ss} Y_{k'}$$

In this work the S-System parameters are described as normal random variables whose variability results from the fact that the magnitudes of the enzyme and metabolite concentrations at the optimal state will be different from those at the nominal (original) steady-state. Thus, the variance of the S-System parameters may be described as the sum of the allowable deviation of these quantities. For example, in Example 2 the enzyme levels are allowed to vary between 0.1 and 10 times the nominal steady-state. Thus, variable Y_k will range between $\ln(0.1) = -2.3026$ and $\ln(10) = 2.3026$. If this range is assumed to represent the 99.99% confidence interval, the variance contribution of Y_k to the S-System parameters which depend on it will be $(2.3026/3.981)^2 = 0.3345$, because the quantile of the double sided probability interval of 99.99% is equal to 3.981 (see Kreyszig (1993)).

Based on the second-order Taylor expansion, the kinetic orders are expressed as linear functions of the deviations X_j and Y_j. This enables the evaluation of the covariances between the S-System parameters. For example, the covariance between g_{ij} and $h_{i'g'}$ is:

$$Cov(g_{ij}, h_{i'j'})$$

$$=\frac{1}{4}\sum_l^N\left(\frac{x_j x_l}{V_i^+}\frac{\partial^2 V_i^+}{\partial x_j \partial x_l}\right)_{ss}\left(\frac{x_{j'} x_l}{V_{i'}^-}\frac{\partial^2 V_{i'}^-}{\partial x_{j'} x_l}\right)_{ss} Var(X_l)$$

$$+\frac{1}{4}\sum_k^M\left(\frac{x_j y_k}{V_i^+}\frac{\partial^2 V_i^+}{\partial x_j \partial y_k}\right)_{ss}\left(\frac{x_j y_k}{V_{i'}^-}\frac{\partial^2 V_{i'}^-}{\partial x_{j'} \partial y_k}\right)_{ss} Var(Y_k)$$

Table VII. Optimization results of (SMOP) substituted into the Michaelis-Menten model and the maximum per cent violation of the imposed lower and upper bounds.

β	V_4^+	x_1	x_2	x_3	x_4	Max. viol.
0.50	6.6889	2.676	4.493	1.561	2.451	30.1%
0.60	4.9079	2.549	4.061	1.335	2.280	12.8%
0.70	3.9110	2.470	3.798	1.211	2.175	4.7%
0.80	3.2622	2.415	3.621	1.132	2.106	0.6%
0.90	2.7707	2.372	3.485	1.074	2.053	0.0%
0.95	2.5330	2.352	3.421	1.048	2.028	0.0%

The second-order derivative terms in the above equations can be evaluated from kinetic data or known kinetic laws in the form of Hill's equation or generalized Michaelis-Menten kinetics.

APPENDIX B: ESTIMATION OF KINETIC ORDER UNCERTAINTY IN EXAMPLE 2

In this appendix the estimation of kinetic order modeling uncertainty from kinetic data is described for the metabolic pathway addressed in Example 2. Following the formulations outlined in Appendix A, and after substituting the expressions for the kinetic orders, the following expansions are obtained:

$$h_{11} = 0.812 - 0.0761X_1 + 0.0175X_4$$
$$h_{14} = -0.187 + 0.0175X_1 - 0.0761X_4$$
$$h_{22} = 0.909 - 0.0413X_2 + 0.0206X_4$$
$$h_{24} = -0.455 + 0.0206X_2 - 0.1240X_4$$
$$g_{31} = 0.541 + 0.0226X_1 - 0.0821X_2 + 0.0358X_4$$
$$\quad + 0.0903Y_3 - 0.0903Y_4$$
$$g_{32} = 0.303 - 0.0821X_1 + 0.0780X_2 - 0.0201X_4$$
$$\quad - 0.1010Y_3 + 0.1010Y_4$$
$$g_{34} = -0.276 + 0.0358X_1 - 0.0201X_2 - 0.0841X_4$$
$$\quad + 0.0297Y_3 - 0.0297Y_4$$
$$h_{33} = 0.5 - 0.1250X_3$$
$$g_{43} = 0.5 - 0.1250X_3$$
$$h_{44} = 0.5 - 0.1250X_4$$
$$g'_{33} = 0.667 + 0.0903X_1 - 0.1010X_2 + 0.0297X_4$$
$$\quad + 0.1111Y_3 - 0.1111Y_4$$
$$g'_{34} = 0.333 - 0.0903X_1 + 0.1010X_2 - 0.0297X_4$$
$$\quad - 0.1111Y_3 + 0.1111Y_4$$

The variances of X_j and Y_k estimated based on the 99.99% confidence interval are:

$$\text{Var}(X_j) \approx (ln(0.8)/3.981)^2 = 0.0031$$
$$\text{Var}(Y_k) \approx (ln(0.1)/3.981)^2 = 0.3345$$

Based on these values, and after applying the variance operator, the variance of the kinetic orders can be estimated. For example, the variance of h_{11} is equal to

$$\text{Var}(h_{11}) = (-0.0761)^2\text{Var}(X_1) + 0.0175^2\text{Var}(X_2)$$
$$= 0.00001915$$

and thus the variance of A_{11} is

$$\text{Var}(A_{11}) = \text{Var}(g_{11}) + \text{Var}(h_{11}) = 0.000019156$$

Furthermore, the covariances between any two kinetic orders A_{ij} and B_{ik} can be calculated. For example,

$$\text{Cov}(A_{11}, A_{22})$$
$$= \text{Cov}(g_{11}, g_{22}) + \text{Cov}(h_{11}, h_{22}) - \text{Cov}(g_{11}, h_{22})$$
$$\quad - \text{Cov}(h_{11}, g_{22})$$
$$= (0.0175)(0.0206)\text{Var}(X_1) = 0.000001132$$

The remaining elements of the variance-covariance arrays are calculated in the same manner using the relations for calculating the variances and covariances of linear functions of random variables (Wilks, 1962).

References

Abramowitz, M., Stegun, I. E. 1972. Handbook of mathematical functions. 10th edition. National Bureau of Standards, Washington D.C.

Allen, F. M., Braswell, R. N., Rao, P. V. 1974. Distribution-free approximations for chance constraints. Oper. Res. **22**: 610–621.

Androulakis, I. P., Maranas, C. D., Floudas, C. A. 1995. αBB: a global optimization method for general constrained nonconvex problems. J. Global Optimiz. **7**: 337–343.

Babbar, M. 1955. Distributions of solutions of a set of linear equations (with an application of linear programming). JASA. **50**: 854–869.

Bailey, J., Birnbaum, S., Galazo, J., Khosla, C., Shanks, J. 1990. Part I. Manipulation and analysis of metabolic pathways: strategies and challenges in metabolic engineering. Ann. N.Y. Acad. Sci. **589**: 1–15.

Brooke, A., Kendrick, D., Meeraus, A. 1988. GAMS: a user's guide. Scientific Press, Palo Alto, CA.

Cascante, M., Sorribas, A., Franco, R., Canela, E. I. 1991. Biochemical systems theory: increasing predictive power by using second-order derivatives measurements. J. Theor. Bio. **149**: 521–535.

Charnes, A., Cooper, W. W. 1959. Chance-constrained programming. Manag. Sci. **6**: 73–79.

Charnes, A., Cooper, W. W. 1960. Chance-constraints and normal deviates. J. Am. Statis. Assoc. **55**: 134–148.

Charnes, A., Cooper, W. W. 1963. Deterministic equivalents for optimizing and satisfying under chance constraints. Oper. Res. **11**: 18–39.

Dervakos, G., Dean, J. 1994. Optimization of metabolic pathway networks. In: Institution of Chemical Engineers Symposium Series, Rugby Eng. Inst. Chem. Eng.

Hatzimanikatis, V., Floudas, C. A., Bailey, J. E. 1996a. Optimization of regulatory architectures in metabolic reaction engineering. Biotechnol. Bioeng. **52**: 485–500.

Hatzimanikatis, V., Floudas, C. A., Bailey, J. E. 1996b. Analysis and design of metabolic reaction networks via mixed-integer linear optimization. AIChE J. **42**: 1277–1292.

Kataoka, S. 1963. A stochastic programming model. Econometrica **31**(1–2): 181–196.

Kreyszig, E. 1993. Advanced engineering mathematics. 7th edition. John Wiley & Sons, Inc., New York.

Maranas, C. D. 1997. Optimal molecular design under property prediction uncertainty. AIChE J. **43**:1250–1264.

Niederberger, P., Prasad, R., Miozzari, G., Kascar, H. 1992. A strategy for increasing an in vivo flux by genetic manipulations: the tryptophan system of yeast. Biochem. J. **287**: 473–479.

Regan, L., Bogle, I. D. L., Dunnill, P. 1993. Simulation and optimization of metabolic pathways. Comput. Chem. Eng. **17**: 627–637.

Sanwal, B. D., Kapor, M., Duckworth, H. W. 1971. The regulation of branched and converging pathways. Curr. Top. Cell. Reg. **3**: 1–115.

Savageau, M. A. 1969a. Biochemical systems analysis, I: some mathematical properties of the arte law for the component enzymatic reactions. J. Theor. Bio. **25**: 365–369.

Savageau, M. A. 1969b. Biochemical systems analysis, II: the steady-state solutions for an n-pool system using a power-law approximation. J. Theor. Bio. **25**: 370–379.

Savageau, M. A. 1970. Biochemical systems analysis, III: Dynamic solutions using a power-law approximation. J. Theor. Bio. **26**: 215–226.

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

Savageau, M. A. 1990. Biochemical systems theory: alternative views of metabolic control. In: E. Cornish-Bowden, M. L. Cardenas, (eds.), Control of metabolic processes. NATO ASI Series–Plenum Press, New York.

Savageau, M. A. 1995. Enzyme kinetics in vitro and in vivo: Michaelis-Menten revisited: In: E. Bittar, N. Bittar, eds., Principles of medical biology: cell chemistry and physiology: part I. JAI Press, Greenwich, CN.

Savageau, M. A., Voit, E. O., Irvine, D. H. 1987a. Biochemical systems

theory and metabolic control theory: 1: fundamental similarities and differences. Math. Biosci. **86**: 147–169.

Savageau, M. A., Voit, E. O., Irvine, D. H. 1987b. Biochemical systems theory and metabolic control theory: 2: the role of summation and connectivity relationships. Math. Biosci. **86**: 147–169.

Savinell, J., Palson, B. 1992. Network analysis of intermediary metabolism using linear optimization. I: development of mathematical formalism. J. Theor. Bio. **154**: 421–454.

Shiraishi, F., Savageau, M. A. 1992a. The tricarboxylic acid cycle in *dictyostelium discoideum* I: formulation of alternative kinetic representations. J. Bio. Chem. **267**: 22912–22918.

Shiraishi, F., Savageau, M. 1992b. The Tricarboxylic acid cycle in *dictyostelium discoideum* II: evaluation of model consistency and robustness. J. Biol. Chem. **267**: 22919–22925.

Shiraishi, F., Savageau, M. 1992c. The tricarboxylic acid cycle in *dictyostelium discoideum* III: analysis of steady state and dynamic behavior. J. Biol. Chem. **267**: 22926–22933.

Shiraishi, F., Savageau, M. 1992d. The tricarboxylic acid cycle in *dictyostelium discoideum* IV: resolution of discrepancies between alternative methods of analysis. J. Biol. Chem. **267**: 22934–22943.

Sorribas, A., Samitier, S., Canela, E. I., Cascante, M. 1993. Metabolic pathway characterization from transient response data obtained. In Situ: parameter estimation in S-System models. J. Theor. Bio. **162**: 81–102.

Stephanopoulos, G., Valino, J. J. 1991. Network rigidity and metabolic engineering in metabolite overproduction. Science **252**: 1675–1681.

Tong, Y. L. 1990. The multivariate normal distribution. Springer-Verlag, New York.

Torres, N. V. 1994a. Modeling approach to control of carbohydrate metabolism during citric acid accumulation by *aspergillus niger:* I: model definition and stability of the steady state. Biotechnol. Bioeng. **44**: 104–11.

Torres, N. V. 1994b. Modeling approach to control of carbohydrate metabolism during citric acid accumulation by *aspergillus niger:* II: sensitivity analysis. Biotechnol. Bioeng. **44**: 112–118.

Torres, N. V., Voit, E. O., Gonzales-Alcon, C. 1996. Optimization of nonlinear biotechnological processes with linear programming: application to citric acid production by *aspergillus niger.* Biotechnol. Bioeng. **49**: 247–258.

Vajda, S. 1970. Stochastic programming. In: J. Abadie, (ed.), Integer and nonlinear programming. North Holland Publishing Co., Amsterdam.

Voit, E. O. 1990. Comparison of accuracy of alternative models for biochemical pathways. In: E. Cornish-Bowden, M. L. Cardenas, (eds.), Control of metabolic processes. NATO ASI Series–Plenum Press, New York.

Voit, E. O. 1992. Optimization of integrated biochemical systems. Biotechnol. Bioeng.**40**: 572–582.

Voit, E. O., Savageau, M. A. 1987. Accuracy of alternative representations for integrated biochemical systems. Biochemistry, **26**: 6869–6880.

Voit, E. O., Savageau, M. A., Irvine, D. H. 1991. Introduction to S-Systems. In: E. O. Voit (ed.), Canonical nonlinear modeling: S-System approach to understanding complexity, Van Nostrand Reinhold, New York.

Watanabe, T., Ellis, H. 1994. A joint chance–constrained programming model with row dependence. Eur. J. Oper. Res. **77**: 325–343.

Wilks, S. S. 1962. Mathematical statistics. John Wiley & Sons, New York.

Wright, B. A., Butler, M. H., Albe, K. R. 1992a. Systems analysis of the tricarboxylic acid cycle in *dictyostelium discoideum* I: The basis for model construction. J. Bio. Chem. **267**: 3101–3105.

Wright, B. A., Butler, M. H., Albe, K. R. 1992b. Systems analysis of the tricarboxylic acid cycle in *dictyostelium discoideum* II: control analysis. J. Bio. Chem. **267**: 3106–3114.