Putting the "Control" in Metabolic Control Analysis

Christopher Rao, Herbert Sauro, and Adam Arkin

November 17, 2003

1 Introduction

2 Background

We assume that the dynamics of the biochemical reaction network are described by the following differential equation

$$\dot{s}(t) = \mathbf{N}v(s(t), p), \tag{1}$$

where the vector $s \in \mathbb{R}^n$ contains the concentrations for n chemical species, the matrix $\mathbf{N} \in \mathbb{R}^{n \times m}$ defines the stoichiometry of the reaction network, the function $v(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^p \mapsto \mathbb{R}^m$ describes the rates for m chemical reactions, and the vector $p \in \mathbb{R}^p$ contains the relevant kinetic parameters for these rates such as enzyme concentrations and catalytic efficiencies.

As the inclusion of some species in the model (1) leads to redundant equations, one commonly removes these species to create a reduced stoichiometric matrix $\mathbf{N_p}$ with linearly independent rows. We can recover the full matrix by introducing the link matrix $\mathbf{L}$:

$$\mathbf{N} = \mathbf{L}\mathbf{N_p}$$

Let $s_i(t)$ denote the concentration of the reduced set of species. Then, we can recover $s(t)$ from $s_i(t)$ using the relation

$$s(t) = \mathbf{L}s_i(t) + t$$

with the appropriate choice for the constant vector t. The constant vector t accounts typically for mass conservation. For example, the mass constraint $a(t)+b(t) = 1$ yields the link matrix $\mathbf{L} = \begin{bmatrix} 1 & -1 \end{bmatrix}^T$ and constant vector $t = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$ with $s(t) = \begin{bmatrix} a(t) & b(t) \end{bmatrix}^T$ and $s_i(t) = a(t)$.

MCA is concerned with how the properties of the network change when the parameters p are perturbed. The sensitivity function or **control coefficient** is defined as

$$x(t) \triangleq \frac{ds_i(t)}{dp}.$$

Using the chain rule for differentiation, one can show that the control coefficient $x(t)$ satisfies the following differential equation

$$\dot{x}(t) = \left(\mathbf{N_p}\frac{\partial v(s(t), p)}{\partial s(t)}\mathbf{L}\right) x(t) + \mathbf{N_p}\frac{\partial v(s(t), p)}{\partial p}. \tag{2}$$

Note that equation (2) is linear: it is simply the linearization of equation (1). If the reaction network (1) is at a steady state s_{ss}, then the sensitivity equation (2) for infinitesimal perturbations about the steady state satisfies the linear, time-independent, differential equation

$$\dot{x}(t) = (\mathbf{N_p}\epsilon_s\mathbf{L})\, x(t) + \mathbf{N_p}\epsilon_p, \tag{3}$$

where

$$\epsilon_x \triangleq \frac{\partial v(s_{ss}, p)}{\partial s}, \quad \epsilon_p \triangleq \frac{\partial v(s_{ss}, p)}{\partial p}.$$

The matrices $\epsilon_x \in \mathbb{R}^{m\times n}$ and $\epsilon_p \in \mathbb{R}^{m\times p}$ are called the **elasticity coefficients**. They provide a measure for how strongly a single reaction in isolation is changed by infinitesimal perturbations either to the concentrations s or parameters p. It is possible to directly measure the elasticity coefficients from experiments without needing to know the rate laws $v(\cdot)$ explicitly.

As equation (3) is linear and time invariant, we can take the Laplace transform and obtain the frequency response for the control coefficients:

$$X(j\omega) = (j\omega I - \mathbf{N_p}\epsilon_x\mathbf{L})^{-1}\mathbf{N_p}\epsilon_p \tag{4}$$

where $X(j\omega)$ is the Laplace transform of $x(t)$ and j is the complex number $\sqrt{-1}$. While the elasticity coefficient ϵ_p is typically constant, it is also possible to consider time-varying perturbations $\epsilon_p(t)$. To explore frequency variations, we replace ϵ_p with $\epsilon_p(\omega)$ in equation (4). At steady state ($\omega = 0$), the sensitivity equation (4) become

$$X(0) = -(\mathbf{N_p}\epsilon_x\mathbf{L})^{-1}\mathbf{N_p}\epsilon_p,$$

assuming the inverse exists. This equation is known as the **connectivity theorem** as it relates the control coefficients $X(0)$ to the elasticity coefficients ϵ_p at steady state. It is a cornerstone of MCA.

In addition to the concentrations $s(t)$, MCA is concerned with perturbations to the flux through the network. We define the flux control coefficient as

$$y(t) \triangleq \frac{dv}{dp}.$$

The flux control coefficient $y(t)$ is related to the control coefficient $x(t)$ by the affine relation

$$\begin{aligned} y(t) &= \frac{\partial v}{\partial p} + \left(\frac{\partial v}{\partial s}\mathbf{L}\right) x(t), \\ &= \epsilon_p + (\epsilon_x\mathbf{L})\, x(t), \end{aligned}$$

where the second relation holds for perturbations about the steady state s_{ss}. Note the difference between y and ϵ_p.

If $\mathbf{A} \triangleq (\mathbf{N_p}\epsilon_x\mathbf{L})$, $\mathbf{B} \triangleq \mathbf{N_p}$, $\mathbf{C} \triangleq \epsilon_x$ and $\mathbf{D} \triangleq \mathbf{I}$, then we represent the connectivity equations using the following state-space form:

$$\dot{x} = \mathbf{A}x + \mathbf{B}\epsilon_p, \tag{5a}$$

$$y = \mathbf{C}x + \mathbf{D}\epsilon_p, \tag{5b}$$

where ϵ_p is the input.

In the remainder of the paper, we explore the connectivity equation (3) and demonstrate how some elementary tools from control analysis can be applied to this equation to explore regulation in biochemical networks.

3 Signal-Flow Graphs

Molecular biologist commonly draw complex diagrams describing intracellular pathways in *cartoon* format. While these diagrams sucessfully enumerate the players and interactions within the pathway, they are unable to convey any information regarding dynamic or regulatory properties.

When analyzing and designing control systems, engineers often employ signal-flow graphs. Signal-flow graphs diagram the feedback structure using unambiguous notation. Signal-flow diagrams are also able to convey dynamic information. While signal-flow graphs are less relevant for design now with numerous CAD procedures, they are still indispensable for analysis, especially for linear systems.

As we demonstate, it is possible to diagram the regulatory structure of pathways characterized by the control and elasticity coefficients using signal-flow graphs. While we would be too optimistic if we expect molecular biologists will adopt a similar convention, our pragmatic goal is to bridge biological net-

work analysis with control theory, and in the process leverage the tools, theories, and intuition developed over the last century gained with artificial systems for biology. What better way than to start with a control diagram!

3.1 Linear Cascades

Consider the following linear cascade of reactions

$$() \xrightarrow{\mathrm{v}_1} \mathrm{s}_1 \xrightarrow{\mathrm{v}_2} \mathrm{s}_2 \xrightarrow{\mathrm{v}_3} () \tag{6}$$

We assume that the reaction mechanisms are elementary to the degree that

$$v^1 = v^1(s_1, p), \;\; v^2 = v^2(s_1, s_2, p), \;\; v^3 = v^3(s_3, p).$$

The kinetic equations for this mechanism are of the form

$$\begin{aligned} \dot{s}_1 &= v_1(s_1, p) - v_2(s_2, s_3, p), \\ \dot{s}_2 &= v_2(s_1, s_2, p) - v_3(s_2, p), \\ \dot{s}_3 &= v_3(s_2, p). \end{aligned}$$

Let

$$\epsilon_j^i = \frac{\partial v^i(\cdot)}{\partial s_j},$$

where s_j denotes the j^{th} element of the n-dimensional vector s.

The sensitivity equations for this reaction mechanism are

$$\dot{x}_1(t) = \left(\epsilon_1^1 - \epsilon_1^2\right) x_1(t) - \epsilon_2^2 x_2(t) + \epsilon_p^1 - \epsilon_p^2 \tag{7a}$$

$$\dot{x}_2(t) = \left(\epsilon_2^2 - \epsilon_2^3\right) x_2(t) + \epsilon_1^2 x_1(t) + \epsilon_p^2 - \epsilon_p^3 \tag{7b}$$

If we take the Laplace transform, then we can recast the sensitivity equations (7) in the frequency domain

$$\begin{aligned} j\omega X_1(\omega) &= \left(\epsilon_1^1 - \epsilon_1^2\right) X_1(\omega) - \epsilon_2^2 X_2(\omega) \\ &\quad + \epsilon_p^1(\omega) - \epsilon_p^2(\omega), \\ j\omega X_2(\omega) &= \left(\epsilon_2^2 - \epsilon_2^3\right) X_2(\omega) + \epsilon_1^2 X_1(\omega) \\ &\quad + \epsilon_p^2(\omega) - \epsilon_p^3(\omega). \end{aligned}$$

We can represent the sensitivity equations in standard, block diagram format for the single perturbation ϵ_p^1 (Figure 1), where the elasticity coefficients are viewed as transfer functions. Alternatively we can equivalently represent these equations graphically using a signal flow diagram (Figure 2). For aesthetic reasons, we prefer the signal flow diagram to the transfer function representation.

Evident from the diagram (Figure 2) simple reaction result in local feedback loops, negative and possibly positive. For example, a first-order degradation reaction is equivalent to a negative feedback loop:

$$\dot{s} = p - ks$$

where p is the rate of production and k is the degradation rate. In the example, the reactions v_2 and v_3 act as negative feedback loops of s_1 and s_2 respetively ($-\epsilon_1^2$ and $-e_1^3$). For reversible reactions, the product negatively regulates the reactant ($-\epsilon_2^2$). Likewise, if the reactions v_1 and v_2 are reversible, then the product inhibits its synthesis. This inhibition is equivalent to a negative feedback loop ($\epsilon_1^1 < 0$ and $e_2^2 < 0$). If these reactions are autocatalytic, then the product enhances its synthesis and process is equivalent to a positive feedback loop ($\epsilon_1^1 > 0$ and $e_2^2 > 0$). The perturbations ϵ_p are equivalent to additive disturbances.

We did not need to know the specific rate equations in order to construct the signal flow diagram. All we needed to know was the general structure of the equations and elasticity coefficient if we wish to numerically evaluate the network. Note also that reactions form local feedback loops on adjacent species in the graph. These loops result from mass conservation and reversible reactions.

For a signal-flow graph, the transfer function T_{ij} between i^{th} and j^{th} node is given by Mason's rule

$$T_{ij} = \frac{\sum_k P_{ijk} \Delta_{ijk}}{\Delta}$$

where the summation is taken over all possible paths from node i to j and

$$\begin{aligned} P_{ijk} &= k^{\text{th}} \text{ path from node } i \text{ to } j, \\ \Delta &= \text{ determinant of graph,} \\ \Delta_{ijk} &= \text{ cofactor of the path } P_{ijk}. \end{aligned}$$

Numerous algorithms exist for applying Mason's rule to generic signal-flow networks.

Applying Mason's rule we obtain the following transfer function relating ϵ_p^1 to the control coefficients $X_1(\omega)$ and $X_2(\omega)$:

$$X_1(\omega) = \frac{\left(j\omega - (\epsilon_2^2 - \epsilon_2^3)\right)\epsilon_p^1 + \epsilon_2^2\epsilon_p^2}{-\omega^2 - ((\epsilon_1^1 - \epsilon_1^2) + (\epsilon_2^2 - \epsilon_2^3))\, j\omega + \epsilon_1^1\epsilon_2^2 - \epsilon_1^1\epsilon_2^3 + \epsilon_1^2\epsilon_2^3}$$

$$X_2(\omega) = \frac{\epsilon_2^1\epsilon_p^1 + \left(j\omega + (\epsilon_1^1 - \epsilon_1^2)\right)\epsilon_p^2}{-\omega^2 - ((\epsilon_1^1 - \epsilon_1^2) + (\epsilon_2^2 - \epsilon_2^3))\, j\omega + \epsilon_1^1\epsilon_2^2 - \epsilon_1^1\epsilon_2^3 + \epsilon_1^2\epsilon_2^3}$$

Note that both equations have the same denominator and the perturbations ϵ_p arise only in the numerator. This fact is well known in control: the denominator is the **sensitivity function** and is characterized by the network structure irrespective of the perturbations.

At steady state, we have the sensitivity equations

$$\begin{aligned} X_1(0) &= \frac{-(\epsilon_2^2 - \epsilon_2^3)\epsilon_p^1 + \epsilon_2^2\epsilon_p^2}{\epsilon_1^1\epsilon_2^2 - \epsilon_1^1\epsilon_2^3 + \epsilon_1^2\epsilon_2^3} \\ X_2(0) &= \frac{\epsilon_2^1\epsilon_p^1 + (\epsilon_1^1 + \epsilon_1^2)\epsilon_p^2}{\epsilon_1^1\epsilon_2^2 - \epsilon_1^1\epsilon_2^3 + \epsilon_1^2\epsilon_2^3} \end{aligned}$$

These equations are the result of the connectivity theorem.

3.2 Linear Cascades With Feedback Inhibition

Consider the following set of reactions with endproduct inhibition

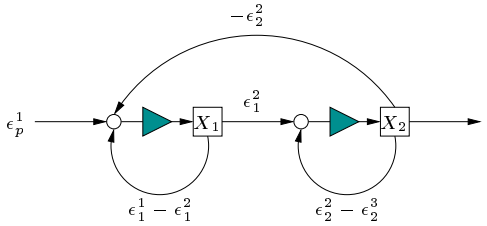
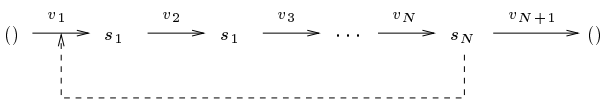


Figure 2: A signal-flow representation of the dynamic connectivity equations. The open circle denotes a summer and the shaded triangle denotes an integrator: $\frac{1}{j\omega}$. For aesthetic reasons, we prefer these graphical representations to transfer functions.



The kinetic equations for this mechanism are

$$\begin{aligned} \dot{s}_1 &= v_1(s_1, s_N, p) - v_2(s_1, s_2, p), \\ \dot{s}_2 &= v_2(s_1, s_2, p) - v_3(s_2, s_3, p), \\ &\vdots \\ \dot{s}_N &= v_N(s_{N-1}, s_N) - v_{N+1}(s_N, p). \end{aligned}$$

For this example, the connectivity equations become

$$\begin{aligned} \dot{x}_1(t) &= \left(\epsilon_1^1 - \epsilon_1^2\right) x_1(t) - e_2^2 S_2(t) + \epsilon_N^1 x_N + \epsilon_\theta^1 \\ \dot{x}_2(t) &= \left(\epsilon_2^2 - \epsilon_2^3\right) x_2(t) + \epsilon_1^2 x_1(t) - \epsilon_3^3 x_3(t) \\ &\vdots \\ x_N(t) &= (\epsilon_N^N - \epsilon_{N+1}^N) x_N + \epsilon_{N-1}^N X_{N-1} \end{aligned}$$

Figure 3 diagrams the frequency response of the connectivity equations as a signal-flow graph. As with the previous example, the reactions result in local feedback loops. However, there is also the global negative feedback loop resulting from endproduct inhibition ($-\epsilon_5^1$). This example illustrates how it is difficult

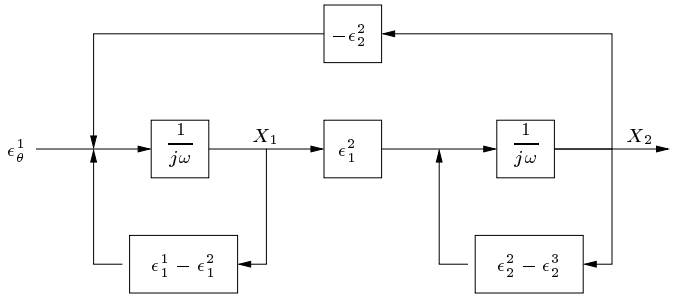


Figure 1: A transfer function representation of the dynamic connectivity equations. As evident from the diagram, reversible reactions introduce feedback loops as the concentration of the downstream species regulates the concentration of the upstream species.

to distinguish the process dynamics from the regulation and the regulatory dynamics can be embedded directly in the process.

3.3 Tryptophan Biosynthesis

The *trp* operon encodes five genes that synthesize tryptophan from chorismate, the common precursor for the aromatic amino acids phenyalanine, tyrosine, and trypotophan. As with many biosynthetic pathways, the process is subject to a hierarchy of regulatory feedback loops. At the level of the gene, expression of the operon is negatively regulated by tryptophan: tryptophan binds the trp repressor and also inhibits transcription. Typtophan, in the form of charged tRNAs, inhibits transcription through the process of translation attenuation. At the metabolic level, tryptophan allosterically inhibits the enzymes in the pathway.

A simple model for tryptophan biosynthesis is described by the following differential equations

$$\begin{aligned} \dot{m} &= \underbrace{v^1(p)}_{\text{transcription}} - \underbrace{v^2(m)}_{\text{mRNA deg.}} \\ \dot{e} &= \underbrace{v^3(m)}_{\text{translation}} - \underbrace{v^4(e)}_{\text{prot. deg.}} \end{aligned}$$

$$\dot{p} = \underbrace{v^5(e,p)}_{\text{trp synthesis}} - \underbrace{v^6(p)}_{\text{trp deg.}} - \underbrace{v^7(p)}_{\text{trp consump.}}$$

The sensitivity equations for this model take the following form

$$\begin{aligned} \dot{x_1} &= \epsilon_3^1 x_3 - \epsilon_1^2 x_1 \\ \dot{x_2} &= \epsilon_1^3 x_1 + \epsilon_3^3 x_3 - \epsilon_2^4 x_2 \\ \dot{x_3} &= \epsilon_2^5 x_2 + (\epsilon_3^5 - e_3^6 - \epsilon_3^7) x_3 \end{aligned}$$

The signal diagram (4) illustrates the signaling hierarchy in the tryptophan example. The global loop ($-\epsilon_3^1$) results from the genetic regulatory mechanisms. The loops ($\epsilon_3^5 < 0$, $-\epsilon_3^6$, and $-\epsilon_3^7$) result from allosteric inhibition by tryptophan, degradation, and consumption respectively. In terms of the connectivity equations and local dynamics, they are equivalent, though obviously they each have unique role.

3.4 Comments

The aim in the preceding sections was to demonstrate that there is a one-to-one map between MCA and control theory. In fact, the previous discussion extends MCA which until recently was limited to steady-state analysis (cite Ingalls). For brevity, we

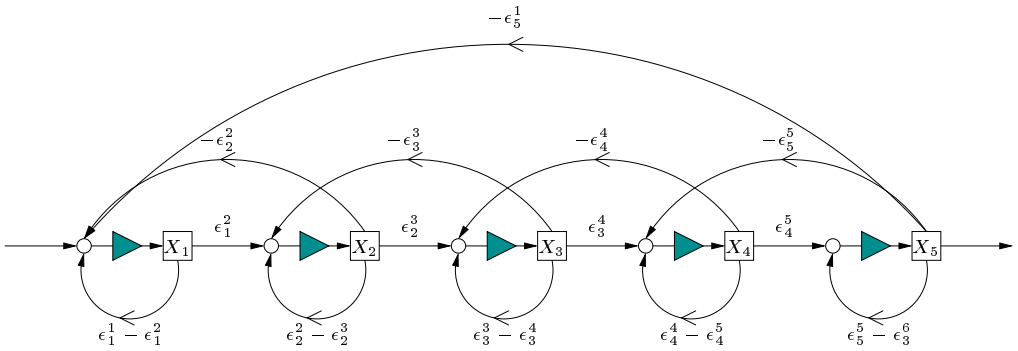


Figure 3: Signal-flow diagram for a linear cascade with feedback inhibition ($N = 5$). The open circle denotes a summer and shaded triangle denotes the integrator $\frac{1}{j\omega}$.

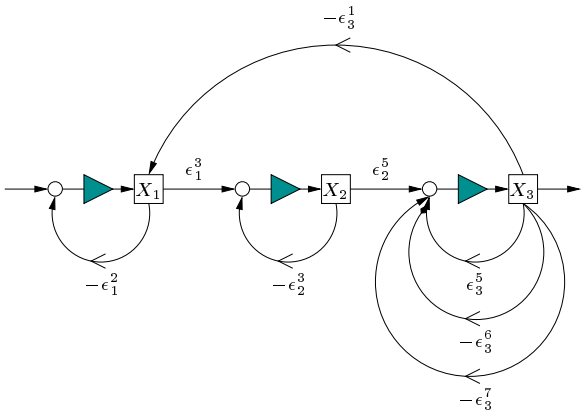


Figure 4: Tryptophan Example: Signal-flow diagram for the tryptophan pathway. Shaded triangles denotes the integrator $\frac{1}{j\omega}$.

have avoided explicit rate expressions and numerical values. However, one can imagine how tools from control theory can be applied to the analysis of biochemical networks (at least to first order approximation). Possible questions include controllability, robustness, and the influence of noise.

4 Controllability

In drug design and metabolic engineering, the goal is to perturb the properties of the network typically by the addition of a small molecule inhibitor or genetic manipulation. In strongly connected networks, the perturbations may propagate through the network and lead to unanticipated consequences. As we demonstrate this problem is identical to problem of controllability.

In designing a drug or genetic intervention, we seek to simultaneously upregulate and downregulate a subset of species while leaving the others unaffected. This amounts to defining a concentration target x^* or flux target y^*. The question is whether there exists a perturbation ϵ_p, possible time-varying though preferably constant, that can reach this target. This

problem is similar to issue of steady-state controllability.

The steady-state gain for the sensitivity equations is

$$\begin{aligned} x &= -(\mathbf{N}_p\epsilon_x\mathbf{L})^{-1}\mathbf{N}_p\epsilon_p, \\ &\triangleq \mathbf{C^x}\epsilon_p, \end{aligned}$$

and for the flux is

$$\begin{aligned} y &= \left(I - (\epsilon_x\mathbf{L})\,(\mathbf{N}_p\epsilon_x\mathbf{L})^{-1}\mathbf{N}_p\right)\epsilon_p, \\ &\triangleq \mathbf{C^y}\epsilon_p, \end{aligned}$$

where $\mathbf{C^x}$ and $\mathbf{C^Y}$ are known as the **concentration control matrix** and **flux control matrix**. Note that

$$\mathbf{C^x}\epsilon_x\mathbf{L} = -\mathbf{I}, \qquad \mathbf{C^y}\epsilon_x\mathbf{L} = 0.$$

These equalities are known as the **summation theorem**. Also note that $\mathbf{C^y}\,\mathbf{C^y} = \mathbf{C^y}$.

We can formulate the design problem as the following control problem. Does there exist ϵ_p such that

$$x^* = C^x\epsilon_p \qquad \text{or} \qquad y^* = C^y\epsilon_p$$

The answer is yes (to a first-order approximation) if x^* and y^* are contained in the range of C^x and C^y. Even if the answer is yes, not all perturbations are experimentally feasible nor advised. In many ways, this problem is similar to the steady-state analysis performed in multi-variable control.

4.1 Flux Analysis: Linear Pathway

Consider again the linear pathway

$$() \xrightarrow{\mathrm{v_1}} \mathrm{s_1} \xrightarrow{\mathrm{v_2}} \mathrm{s_2} \xrightarrow{\mathrm{v_3}} () \tag{8}$$

We assume for simplicity that the reactions are effectively irreversible. The sensitivity equations are

$$\dot{x} = \left[\begin{array}{cc} -\epsilon_1^2 & 0 \\ \epsilon_1^2 & \epsilon_2^3 \end{array}\right] x + \left[\begin{array}{c|c|c} 1 & -1 & 0 \\ 0 & 1 & -1 \end{array}\right] \left[\begin{array}{c} \epsilon_p^1 \\ \epsilon_p^2 \\ \epsilon_p^3 \end{array}\right]$$

Analysis of the controllability matrix indicates that both perturbations to reactions v_1 and v_2 can independently control sensitivity. However, perturbations to v_3 can only effect x_2.

The flux sensitivity equations are

$$\begin{aligned} y^1 &= \epsilon_p^1 \\ y^2 &= \epsilon_p^2 + \epsilon_1^2 x_1 \\ y^3 &= \epsilon_p^3 + \epsilon_2^3 x_2 \end{aligned}$$

At steady state, we have the following connectivity relations

$$\begin{aligned} x_1 &= \frac{\epsilon_p^1 - \epsilon_p^2}{\epsilon_1^2} \\ x_2 &= \frac{\epsilon_p^1 - \epsilon_p^2}{\epsilon_2^3} \end{aligned}$$

and flux relations

$$y^1 = \epsilon_a^1, \;\; y^2 = \epsilon_a^1, \;\; y^3 = \epsilon_a^1$$

It is immediate from the these equations that enzyme a can change the flux. This result is a simple illustration of a pathway with a rate-limiting step.

4.2 Branched Pathway

Consider the branched pathway illustrated in Figure 4.2 where the goal is to produce substrate s_1 without effecting a key metabolic substrate s_3. The sensitivity equations assuming irreversible reactions are

$$\dot{x} = \left[\begin{array}{ccc} -(\epsilon_1^2 + \epsilon_1^3) & 0 & 0 \\ \epsilon_1^2 & -\epsilon_1^4 & 0 \\ \epsilon_1^3 & 0 & -\epsilon_3^5 \end{array}\right] x + \left[\begin{array}{cc} 1 & -1 \\ 0 & 1 \\ 0 & 0 \end{array}\right] \left[\begin{array}{c} \epsilon_p^1 \\ \epsilon_p^2 \end{array}\right]$$

The control equations are

$$x_1 = \frac{\epsilon_p^1 - \epsilon_p^2 - \epsilon_p^3}{\epsilon_p^2 + \epsilon_p^3}$$

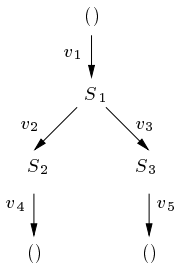


Figure 5: Branched Pathway

$$x_2 = \frac{1}{\epsilon_2^4}\left[\frac{\epsilon_1^2}{\epsilon_1^2+\epsilon_1^3}\left(\epsilon_p^1-\epsilon_p^2\right)+\epsilon_p^2\right]$$
$$x_3 = \frac{1}{\epsilon_3^5}\left[\frac{\epsilon_1^2}{\epsilon_1^2+\epsilon_1^3}\left(\epsilon_p^1-\epsilon_p^2\right)\right]$$

Simple inspection of these equations indicates that both reactions v_1 and v_2 need to be increased in order to satisfy the design objective. This problem and it's solution is well-known in metabolic engineering.

5 Conclusion