Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

Modelling and control of cell reaction networks

S. Jha, J.H. van Schuppen

REPORT PNA-R0116 SEPTEMBER 30, 2001

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3711

Modelling and Control of Cell Reaction Networks

Siddhartha Jha

Indian Institute of Technology
IIT Kanpur, 208016 Kanpur, India
sjha@iitk.ac.in

Jan H. van Schuppen

CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
J.H.van.Schuppen@cwi.nl

ABSTRACT
The project aims at a study of the nonlinear systems arising in the biochemical processes occuring inside a cell. The cellular regulation has been formulated in the more familiar framework used in control and system theory in terms of inputs as the variables which can be influenced externally. A graph-theoretic approach has been taken to elicit the rich structure of the dynamical systems represented by metabolic pathways inside the cell. Problems of realization, analysis, and control of cell reaction networks are described.

2000 Mathematics Subject Classification: 92C40, 92C45, 93C15, 93A13.
Keywords and Phrases: Molecular biology, kinetics, enzymes, metabolic pathways, control, cell reaction networks, graphs, hierarchichal systems.

1 Introduction

The aim of this project was to study the metabolic processes occuring inside the cell from a system theoretic perspective by considering them as a network of nonlinear dynamical systems.

Biologists have been reasonably successful in understanding control in metabolic network using Metabolic Control Analysis. A motivation for undertaking this project was to investigate whether tools and concepts of control and system theory could be used to further this insight and help tackle pressing issues of model reduction and structural complexity in a more concrete and regular framework. A need was felt to formulate suitable models for this purpose and aim at simplifying the control analysis by decomposing large networks to smaller systems of low complexity.

Towards this purpose dynamical systems were derived from biochemical models of enzyme catalysed reaction kinetics. Several types of kinetic systems were studied and some properties generic to this class were identified. An attempt was made to investigate the structural aspect of such a network by its graph theoretic representation and to see if some of the relevant questions could be meaningfully formulated in this framework. Also the problem of system reduction for the general class of biochemical systems was looked into and an attempt was made to analyse this problem with a graph-theoretic approach.

Emphasis throughout this project was on identifying new problems for research and attempting to answer at least some of the questions that were raised in the context, in the short time that it lasted. Though the aim initially was to investigate if regular control and system theoretic approach would be useful to modelling the cell reaction networks, a lot of interesting and new control problems also arose during the investigation. Controllability and observability questions for large positive nonlinear networks are still open for research. In this context it would be interesting from a theoretical perspective to study how the cell solves the problem for the metabolic pathways. This can aid understanding the control of large networks of systems arising in other contexts. Control of cellular regulation by enzymes, and the problem of rational drug design can be looked upon as control synthesis issues and studied in this framework. Such an effort would be rewarding not only for the biologists but also aid understanding of control of large networks of dynamical systems for control and system theorists, by taking inspiration from nature.

In brief the outline of this report is as follows: In the next section suitability of the state space model in the context of cellular regulation has been discussed, the steps required to transform the metabolic system into a state space model have been explained with the help of examples. In Section 3 the model of a cell network in a graph theoretic framework is described followed by a discussion of the properties of the systems studied. The issue of model reduction has been looked into and finally, in Section 5 problems for further research have been framed.

This report is written for readers with a background either in biology or in control and system theory. Therefore it contains parts which are new to readers of one research community but well known to readers of the other research community. The understanding of the readers for this will be appreciated.

The investigative effort was as an internship project during which the first author was affiliated to CWI as a research trainee in the system theory and control research group of the second author. The duration of the project was from 13th May, 2001 to 24th July, 2001.

2 Modelling

2.1 Modelling

Modelling the cell as a complex interconnected web of dynamical systems is a rather formidable task. The difficulty in any such attempt arises not just because of the multitude of interactions involved but also due to the nonlinearity of the individual interaction itself. The cell has about $4000 - 10000$ reactions going on inside at any time.These reactions are organised in metabolic pathways in which every reaction is linked to the next through a common metabolite. Each of these reactions is catalysed by a different enzyme which is generally specific to the reaction.

Biologists have developed a way to analyse the control of this complex system through Metabolic Control Analysis (MCA) which has been useful in understanding the influence exerted by each of these basic units on the global variables. However the motivation for such a model is mostly to gain an insight into the working of the cell. We are now in a position to go further and ask whether it is possible to exert control on the function of the cell system from outside. This question can perhaps be better posed in the framework of conventional system theory than in the language of MCA. Such a formulation might also be very useful if we aim at model reduction or abstraction of the complex system into smaller, more tractable units. However to set up such a model with a state space approach we must clearly specify the state variables and the inputs to the system.

In the model that we have considered the metabolite concentrations are the state variables and the enzymes the inputs to the system. The enzyme activity in the cell can be influenced in two ways:

1. inhibition by a suitable compound,

2. regulating the gene expression.

The cell itself uses inhibition in a metabolic pathway to ensure stability. Use of feedback to produce oscillations which have biological significance is also seen in certain sections of the metabolic pathways. The Goodwin oscillator is a standard model in which this has been investigated [9]. Considerable interest has been generated in regulating the gene expression over the past few years specifically after the discovery of gene sequence of many organisms. In this report we consider the possibility of formulatng a control law through either of these methods to transfer the state of the system to a desired position in state space. In this report we have also tried to explore the structural properties of the cell system by modelling the interactions in a graph theoretic framework. However the usefulness of this as regards investigating the controllability and observability properties remains limited due to the nonlinearities involved. Inspite of the complexity arising due to the nonlinearities we know exactly what kind of nonlinear interactions exist between the state variables and hence the general class of non linear systems that need to be investigated.

A fundamental rate law for enzyme catalysed reactions is Michaelis Menten kinetics which arises in the reaction shown in the general scheme below:

$$E + S_1 \rightleftharpoons ES \rightleftharpoons E + S_2$$

Here ES represents the intermediate enzyme-substrate complex concentration Using the quasi-steady-state assumption $\dot{ES} = 0$ one obtains the enzyme kinetic rate law

$$v(S_1, S_2) = \frac{V_m^+ S_1/K_{m1} - V_m^- S_2/K_{m2}}{1 + S_1/K_{m1} + S_2/K_{m2}}.$$

The phenomenological constants V_m^+ and V_m^- denote the maximal activities of the forward and reverse reactions respectively. K_{m1} and K_{m2} are the Michaelis constants for S_1 and S_2 respectively. Similarly the reaction rates for the cases with reversible and irreversible inhibition can be expressed. These have been described in Section 2.4.

A general reference on biology is [3] and a reference on regulation of cell reaction networks is [6].

2.2 Transformations

In this subsection a mathematical model of the biochemical processes of the cell will be transformed to a dynamic system in state space form. The resulting state space form will be used in the remainder of the report.

In system theory a dynamic system is represented in a particular way. One distinguishes inputs, outputs, and state variables. Inputs are signals or inflows from outside the model. In some cases inputs can be set by a controller but in other cases, for example sunshine, these cannot be directly influenced by a controller. Outputs are variables which can be observed and measured. Outputs will not be considered in this report. During special experiments, concentrations of chemicals in the cell can be measured at the cost of influencing the process of the cell. Finally there are the state variables. These are mainly the concentrations of the chemical compounds in the cell.

In the following a dynamic system is formulated which is a mathematical model for part of the biochemical processes of the cell. This part does not include the nucleus and neither the production of enzymes.

The following notation will be used to describe a dynamic system for the biochemical processes of the cell.

$$
\begin{aligned}
T = [t_0, \infty), &\quad \text{the time index set,} \\
n \in \mathbb{N}, &\quad \text{the number of chemical substances,} \\
n_v \in \mathbb{N}, &\quad \text{the number of chemical reactions,} \\
n_{en} \in \mathbb{N}, &\quad \text{the number of enzymes,} \\
n_{ex} \in \mathbb{N}, &\quad \text{the number of chemical substances from outside the cell,} \\
n_{in} \in \mathbb{N}, &\quad \text{the number of chemical substances from other parts of the cell,} \\
u_{en} : T \to \mathbb{R}^{n_{en}}, &\quad \text{input of enzymes to the model,} \\
u_{ex} : T \to \mathbb{R}^{n_{ex}}, &\quad \text{input from outside the cell to the model,} \\
u_{in} : T \to \mathbb{R}^{n_{in}}, &\quad \text{input from other parts of the cell but not the enzyme input,} \\
x : T \to \mathbb{R}^n = X &\quad \text{the state vector.}
\end{aligned}
$$

Often, though not always, the state set is the set of the positive real numbers, $X = \mathbb{R}^n_+$.

According to the general modelling of the biochemistry of the cell, see [6, Ch. 2], the dynamic system may be written as,

$$
\begin{aligned}
\dot{x}(t) \;=\;& N\, v(x(t), u_{en}(t), u_{ex}(t), u_{in}(t)),\ x(t_0) = x_0, \text{ where,} \\
& N \in \mathbb{R}^{n \times n_v}, \text{ represents the stoichiometric matrix of the reactions,} \\
& v : X \times \mathbb{R}^{n_{en}} \times \mathbb{R}^{n_{ex}} \times \mathbb{R}^{n_{in}} \to \mathbb{R}^{n_v}, \text{ represents the rate function,} \\
& n_v \in Z_+, \text{ denotes the number of reactions.}
\end{aligned}
$$

Below dynamic systems are derived for special cases of reaction rates.

2.3 Examples

Example 2.1 The following small example illustrates the transformation of a mathematical model of the biochemical processes of a cell to a dynamic system. The model is derived from the example described in [13, p. 32]. The inputs and state variables are:

$$
\begin{aligned}
n \;&=\; 5,\ n_v = 6,\ n_{en} = 6,\ n_{ex} = 2, \\
u_{en,1} \;&=\; e_1,\ \ldots,\ u_{en,6} = e_6,\ u_{ex,1} = x_0,\ u_{ex,2} = x_6,\ x_1 = s_1,\ \ldots\ x_5 = s_5.
\end{aligned}
$$

The stoichiometric matrix and the rate functions are:

$$
\begin{aligned}
N &= \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}, \\
v_1 &= e_1 r_1 = \frac{e_1(10x_0 s_1 - s_2 s_3)}{1 + x_0 s_1 + s_2 s_3} \\
&= u_{en,1}\left[\frac{10u_{ex,1}x_1}{1 + u_{ex,1}x_1 + x_2 x_3} - \frac{x_2 x_3}{1 + u_{ex,1}x_1 + x_2 x_3}\right] \\
&= u_{en,1}[v_{1,+}(x, u_{ex}) - v_{1,-}(x, u_{ex})], \\
v_2 &= e_2 r_2 = u_{en,2}\left[\frac{10x_2}{1 + x_1 + x_2} - \frac{x_1}{1 + x_1 + x_2}\right], \\
v_3 &= e_3 r_3 = u_{en,3}\left[\frac{5u_{ex,2}}{1 + x_3 + u_{ex,2}} - \frac{x_3}{1 + x_3 + u_{ex,2}}\right], \\
v_4 &= e_4 r_4 = u_{en,4}\left[\frac{10x_3}{1 + x_3 + x_4} - \frac{x_4}{1 + x_3 + x_4}\right], \\
v_5 &= e_5 r_5 = u_{en,5}\left[\frac{10x_4}{1 + x_4 + x_5} - \frac{x_5}{1 + x_4 + x_5}\right], \\
v_6 &= e_6 r_6 = u_{en,6}\left[\frac{10x_5}{1 + x_5}\right], \\
v &= \begin{pmatrix} v_1 \\ \vdots \\ v_6 \end{pmatrix}, \; u_{en} = \begin{pmatrix} u_{en,1} \\ \vdots \\ u_{en,6} \end{pmatrix}, \; u_{ex} = \begin{pmatrix} u_{ex,1} \\ u_{ex,2} \end{pmatrix}, \\
\mathrm{diag}(r) &= \begin{pmatrix} r_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & r_6 \end{pmatrix}.
\end{aligned}
$$

The resulting dynamic system is then,

$$
\begin{aligned}
\dot{x}(t) &= N\ v(x(t), u_{en}(t), u_{ex}(t)), \\
\dot{x}_i(t) &= \sum_{j=1}^{n_v} N_{i,j} v_j(.) \\
&= \sum_{j=1}^{n_v} (N^+_{i,j} - N^-_{i,j})\left[v_{j,+}(x(t), u_{ex,j}(t), u_{en}(t)) - v_{j,-}(x(t), u_{ex,j}(t), u_{en}(t))\right], \; x(t_0) = x_0, \\
\dot{x}(t) &= N\mathrm{diag}(r(x(t))\ u_{en}(t), \quad x(t_0) = x_0.
\end{aligned}
$$

Note that for each $i \in Z_6 = \{1, 2, 3, 4, 5, 6\}$, $v_i : X \times \mathbb{R}^{n_{ex}} \to \mathbb{R}$ is a rational function of which the numerator and the denominator degrees are equal. Moreover, the denominator is always strictly positive. For example,

$$
\begin{aligned}
v_2(x, u_{ex}) &= \frac{10x_2}{1 + x_1 + x_2} - \frac{x_1}{1 + x_1 + x_2}, \\
v_3(x, u_{ex}) &= \frac{5u_{ex,2}}{1 + u_{ex,2} + x_3} - \frac{x_3}{1 + u_{ex,2} + x_3}.
\end{aligned}
$$

The dynamic system of this example will be called a Michaelis-Menten system below because the rate functions without the enzymes are as they occur in the Michaelis-Menten kinetics.

Example 2.2 *Glycolis pathway* [1].
In the following example the model for trypansome enzyme kinetics has been transformed into the state space formulation. The model and the rate equations can be found in [1, p.31, Ch.2]. The inputs and the state variables are:

$$
\begin{aligned}
n &= 10,\ n_{en} = 9, n_{ex} = 3,\\
u_{en,GPO} &= e_1,\\
u_{en,HK} &= e_2,\\
u_{en,GAPDH} &= e_3,\\
u_{en,PGK} &= e_4,\\
u_{en,GDH} &= e_5,\\
u_{en,GK} &= e_6,\\
u_{en,PFK} &= e_7,\\
u_{en,PYK} &= e_8,\\
u_{en,ALD} &= e_9.\\
x_1 &= Glc_{in},\\
x_2 &= [hexose - P],\\
x_3 &= [Fru - 1,6 - BP]_g,\\
x_4 &= triose - P,\\
x_5 &= [1,3 - BPGA]_g,\\
x_6 &= N,\\
x_7 &= [PYR]_c,\\
x_8 &= [NADH]_g,\\
x_9 &= P_g,\\
x_{10} &= P_c.
\end{aligned}
$$

The stoichiometric matrix set up from the rate expressions in [1] is :

$$N = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{pmatrix},$$

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_{12} \end{pmatrix}, \; r = \begin{pmatrix} r_1 \\ \vdots \\ r_{12} \end{pmatrix}, \; u_{en} = \begin{pmatrix} 1 \\ u_{en,1} \\ \vdots \\ u_{en,9} \\ 1 \\ 1 \end{pmatrix}, \; u_{ex} = \begin{pmatrix} u_{ex,1} \\ u_{ex,2} \\ u_{ex,3} \end{pmatrix},$$

$$\mathrm{diag}(r) = \begin{pmatrix} r_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & r_{12} \end{pmatrix}.$$

The 1's in the u_{en} vector correspond to those processes which are not enzyme catalysed and hence cannot be directly influenced by any enzyme inputs. The resulting dynamic system in the state space form is then:

$$\begin{aligned} \dot{x}(t) &= N\ v(x(t), u_{en}(t), u_{ex}(t)), \\ \dot{x}_i(t) &= \sum_{j=1}^{n_v} N_{i,j} v_j(.) = \sum_{j=1}^{n_v} N_{i,j} u_{en,j}(t) r_j(x(t), u_{ex,j}(t)) \\ &= \sum_{j=1}^{n_v} N_{i,j}\ \mathrm{diag}(r_j(x(t), u_{ex}(t))\ u_{en,j}(t), \\ \dot{x}(t) &= N\ \mathrm{diag}(r(x(t), u_{ex}(t)))\ u_{en}(t),\ x(t_0) = x_0. \end{aligned}$$

Here the Glycolysis pathway has been considered as a dynamic system with the enzymes forming an input vector and separating out linearly in the rate expression. The readers are requested to consult [1] for a complete description and details regarding this example.

2.4 Definitions

In this subsection several classes of cell systems are defined.

Definition 2.3 *Denote the* class of rational functions of degree $(k_1, k_2) \in \mathbb{N}^2$ *by*

$$\begin{aligned} F_{rat,\mathbb{R}_+}(k_1, k_2) &= \left\{ \begin{array}{l} f : \mathbb{R}^s \to \mathbb{R} | f(x) = p(x)/q(x), \\ p, q : \mathbb{R}^s \to \mathbb{R}\ \textit{polynomials of degrees}\ k_1, k_2\ \textit{respectively} \\ \textit{and coefficients in}\ \mathbb{R}_+ \end{array} \right\}, \\ F_{rat,\mathbb{R}_+}(k) &= F_{rat,\mathbb{R}_+}(k, k). \end{aligned}$$

Denote a polynomial $p : \mathbb{R}_+^n \to \mathbb{R}_+$ or $p \in \mathbb{R}_+[x]$ by

$$p(x) = \sum_{k_1,\dots,k_n=0}^{m} c(k_1,\dots,k_n) x_1^{k_1} \dots x_n^{k_n}, \text{ where, } c(k_1,\dots,k_n) \in \mathbb{R}_+.$$

Definition 2.4 *A* kinetic system *is a dynamic system described by the first-order ordinary differential equation,*

$$\begin{aligned}
\dot{x}_i(t) &= \sum_{j=1}^{n_v} N_{i,j} v_j(x(t), u_{en,j}(t), u_{ex}(t), u_{in}(t)) \\
&= \sum_{j=1}^{n_v} (N_{i,j}^+ - N_{i,j}^-) \left[v_{j,+}(x(t), u_{en,j}(t), u_{ex}(t), u_{in}(t)) \right. \\
&\quad \left. - v_{j,-}(x(t), u_{en,j}(t), u_{ex}(t), u_{in}(t)) \right] \\
&= \sum_{j=1}^{n_v} (N_{i,j}^+ - N_{i,j}^-) \left[r_{j,+}(x(t), u_{ex}(t), u_{in}(t)) - r_{j,-}(x(t), u_{ex}(t), u_{in}(t)) \right] \\
&\quad \times u_{en,j}(t) \; x_i(t_0) = x_{i,0}, \; \forall \; i \in \mathbb{Z}_n, \; \textit{where,} \\
n \in \mathbb{Z}_+, &\quad \textit{denotes the number of chemical substances,} \\
n_v \in \mathbb{Z}_+, &\quad \textit{denotes the number of reactions,} \\
n_{en} \in \mathbb{Z}_+, &\quad \textit{denotes the number of enzymes,} \\
n_{in} \in \mathbb{Z}_+, &\quad \textit{denotes the number of inputs from inside the cell,} \\
n_{ex} \in \mathbb{Z}_+, &\quad \textit{denotes the number of inputs from outside the cell,} \\
N \in \mathbb{N}^{n \times n_v}, &\quad \textit{denotes the stoichiometric matrix,} \\
N_{i,j}^+ &= \begin{cases} N_{i,j}, & \textit{if } N_{i,j} \geq 0, \\ 0, & \textit{otherwise,} \end{cases} \quad N_{i,j}^- = \begin{cases} -N_{i,j}, & \textit{if } N_{i,j} < 0, \\ 0, & \textit{otherwise,} \end{cases} \\
N_{i,j} &= N_{i,j}^+ - N_{i,j}^-, \; \forall i \in \mathbb{Z}_n, \; \forall j \in \mathbb{Z}_{n_v}., \\
X &= \mathbb{R}_+^n, \; U_{ex} = \mathbb{R}_+^{n_{ex}}, \; U_{in} = \mathbb{R}_+^{n_{in}}, \\
&\quad v_{j,+} : X \times U_{ex} \times U_{in} \to \mathbb{R}_+^n, \; v_{j,-} : X \times U_{ex} \times U_{in} \to \mathbb{R}_+^n.
\end{aligned}$$

The reaction $j \in \mathbb{Z}_{n_v}$ is said to be reversible *if if both $v_{j,+} \neq 0$ and $v_{j,-} \neq 0$. It is said to be* irreversible *if either $v_{j,+} = 0$ or $v_{j,-} = 0$ but not both.*

Define the following special cases of a kinetic system:

- *A* power-law kinetic system *if*

$$\begin{aligned}
v_{j,+}(x, u_{ex}, u_{in}) &= k_{j,+} \prod_{m=1}^{n} x_m^{N_{m,j}^-}, \\
v_{j,-}(x, u_{ex}, u_{in}) &= k_{j,-} \prod_{m=1}^{n} x_m^{N_{m,j}^+}, \; \forall j \in Z_{n_v}.
\end{aligned}$$

- *A* bilinear kinetic system *if it is a power-law kinetic system such that $\forall i \in \mathbb{Z}_n$ there exist $j_1, j_2 \in \mathbb{Z}_{n_v}$, $j_1 \neq j_2$, such that $N_{i,j_1}^+ = 1$, $N_{i,j_2}^+ = 1$, and $\forall j \in \mathbb{Z}_n$, $j \neq j_1$, $j \neq j_2$, $N_{i,j}^+ = 0$; and there exist $k_1, k_2 \in \mathbb{Z}_{n_v}$, $k_1 \neq k_2$, such that $N_{i,k_1}^- = 1$, $N_{i,k_2}^- = 1$, and $\forall k \in \mathbb{Z}_n$, $k \neq k_1$, $k \neq k_2$, $N_{i,k}^- = 0$;*

- *A* reversible Michaelis-Menten kinetic system, *for short a RM^2 kinetic system, if*

$$\begin{aligned} v_{j,+}(x, u_{ex}, u_{in}) &= \frac{p_+(x, u_{ex}, u_{in})}{q(x, u_{ex}, u_{in})} \in F_{rat,\mathbb{R}_+}(k_1, k_2), \\ \text{or } v_{j,-}(x, u_{ex}, u_{in}) &= \frac{p_-(x, u_{ex}, u_{in})}{q(x, u_{ex}, u_{in})} \in F_{rat,\mathbb{R}_+}(k_3, k_2), \ \forall j \in \mathbb{Z}_n. \end{aligned}$$

- *An* irreversible Michaelis-Menten kinetic system *if, for example,*

$$v_{j,+}(x) = \frac{c_1 x}{c_2 + x}, \ v_{j,-}(x) = \frac{c_3 x}{c_2 + x}.$$

- *An* Hill kinetic system *if, for example,*

$$v(x) = k\frac{(x/k_s)^{n_H}}{1 + (x/k_s)^{n_H}},$$

where n_H is called the Hill coefficient. This rate formula was proposed by A.V. Hill in a paper published in 1910 to model the sigmoid character of a rate formula. Various extensions of this formula have been proposed, see [6, pp. 23-24].

The interpretation of the above formulas in terms of biochemical reactions is that $v_{j,+}$ is the rate of the forward reaction, $v_{j,-}$ is the rate of the backward reaction, and the values of the stoichiometric matrix indicate the multiplicity of the molecules in the reaction.

The power-law kinetic system defined above is modeled following [6, (2.8), (2.10)] as follows. The general formula for the vector of chemical compounds is in terms of the notation of that reference,

$$\begin{aligned} \dot{S}(t) &= Nv(S(t)), \\ \dot{S}_i(t) &= \sum_{j=1}^{n_v} N_{i,j} v_j(S(t)) = \sum_{j=1}^{n_v} (N^+_{i,j} - N^-_{i,j}) \left[k_{j,+}(\prod_{m=1}^{n} S_m^{N^-_{m,j}}) - k_{j,-}(\prod_{m=1}^{n} S_m^{N^+_{m,j}}) \right]. \end{aligned}$$

With the substitutions $x_i = S_i$ for $i \in \mathbb{Z}_n$ the formula of the above definition for the power-law kinetic system follows.

E.D. Sontag in the paper [16] has defined a slightly different representation of a power-law kinetic system. That representation and the one of Definition 2.4 are related as clarified below.

Proposition 2.5 *Consider a power-law kinetic system as defined in Definition 2.4,*

$$\dot{x}_i(t) = \sum_{j=1}^{n_v} (N^+_{i,j} - N^-_{i,j}) \left[k_{j,+}(\prod_m x_m^{N^-_{m,j}}) - k_{j,-}(\prod_m x_m^{N^+_{m,j}}) \right], \ \forall i \in \mathbb{Z}_n.$$

This system may be rewritten as,

$$\begin{aligned} \dot{x}_i(t) &= \sum_{r=1}^{2n_v} \sum_{j=1}^{2n_v} (B_{i,r} - B_{i,j}) A_{rj} (\prod_m x_m^{B_{m,j}}), \ \forall i \in \mathbb{Z}_n, \ \textit{where,} \\ B &= \begin{pmatrix} N^+_{\cdot,1} & N^-_{\cdot,1} & \dots & N^+_{\cdot,n_v} & N^-_{\cdot,n_v} \end{pmatrix} = \begin{pmatrix} b_1 & \dots & b_{2n_v} \end{pmatrix}, \\ A &= \text{Block-diag.}(A_1, \dots, A_{n_v}), \ A_j = \begin{pmatrix} 0 & k_{j,+} \\ k_{j,-} & 0 \end{pmatrix}. \end{aligned}$$

The system is therefore a special case of that defined in [16] except that the matrix A is not irreducible but reducible to block-diagonal form with irreducible blocks on the diagonal.

Proof Note that

$$\begin{aligned}
\dot{x}_i(t) &= \sum_{j=1}^{n_v}(N_{i,j}^+ - N_{i,j}^-)\left[k_{j,+}(\prod_m x_m^{N_{m,j}^-}) - k_{j,-}(\prod_m x_m^{N_{m,j}^+})\right] \\
&= \left[\sum_{j=1}^{n_v}(N_{i,j}^+ - N_{i,j}^-)k_{j,+}(\prod_m x_m^{N_{m,j}^-})\right] + \left[\sum_{j=1}^{n_v}(N_{i,j}^- - N_{i,j}^+)k_{j,-}(\prod_m x_m^{N_{m,j}^+})\right] \\
&= \sum_{j=1}^{2n_v}(B_{i,2j-1} - B_{i,2j})A_{2j-1,2j}(\prod_m x_m^{B_{m,2j}}) \\
&\quad + \sum_{j=1}^{2n_v}(B_{i,2j} - B_{i,2j-1})A_{2j,2j-1}(\prod_m x_m^{B_{m,2j-1}}) \\
&= \sum_{r=1}^{2n_v}\sum_{j=1}^{2n_v}(B_{i,r} - B_{i,j})A_{rj}(\prod_{m=1}^{n} x_m^{B_{m,j}}).
\end{aligned}$$

□

Example 2.6 *Reversible Michaelis-Menten system.* Consider the reversible Michaelis-Menten kinetics as presented in [6, (2.20), p. 17],

$$\begin{aligned}
\dot{S}(t) &= Nv(S(t)), \\
\dot{S}_i(t) &= \sum_{j=1}^{2} N_{i,j}v_j(S(t)), \\
v_j(S) &= \frac{V_j^+ S_1/K_{j1} - V_j^- S_2/K_{j2}}{1 + S_1/K_{j1} + S_2/K_{j2}}, \quad V_j^+ = k_2 E_T,\ V_j^- = k_{-1}E_T.
\end{aligned}$$

With the substitutions $x_1 = S_1$, $x_2 = S_2$, $u_{en} = E_T$ the equation becomes,

$$\begin{aligned}
\dot{x}_i(t) &= \sum_{j=1}^{2}(N_{i,j}^+ - N_{i,j}^-)\left[r_{j,+}(x(t)) - r_{j,-}(x(t))\right]u_{en,j}(t),\ x_i(t_0) = x_{i,0},\ \forall i \in \mathbb{Z}_+, \\
v_{j,+}(x) &= \frac{c_{j,+}x_1}{1 + d_{j1}x_1 + d_{j2}x_2},\ v_{j,-}(x) = \frac{c_{j,-}x_1}{1 + d_{j1}x_1 + d_{j2}x_2}, \\
c_{j,+} &= \frac{k_1k_2}{k_{-1} + k_2} \in (0,\infty), \quad c_{j,-} = \frac{k_{-1}k_{-2}}{k_{-1} + k_2} \in (0,\infty), \\
d_{j1} &= \frac{k_1}{k_{-1} + k_2} \in (0,\infty), \quad d_{j2} = \frac{k_{-2}}{k_{-1} + k_2} \in (0,\infty), \text{ because,} \\
\frac{1}{K_{j1}} &= d_{j1},\ \frac{1}{K_{j2}} = d_{j2}, \\
\frac{V_j^+}{K_{j1}} &= \frac{k_2E_T}{K_{j1}} = \frac{k_1k_2}{k_{-1} + k_2}u_{en,j} = c_{j,+}u_{en,j}, \\
\frac{V_j^-}{K_{j2}} &= \frac{k_{-1}E_T}{K_{j2}} = \frac{k_{-1}k_{-2}}{k_{-1} + k_2}u_{en,j} = c_{j,-}u_{en,j}.
\end{aligned}$$

3 Cell reaction networks

3.1 Introduction

There are several ways of modelling a large scale dynamical system. When one is dealing with a system as complex as a cell no particular model may be sufficient for all our purposes. Here we choose to investigate the suitability of a graph theoretic approach to modelling the cell reaction network. As explained earlier the hundreds of reactions taking place inside the cell are organised into metabolic pathways, chains of intricately interlinked steps each catalysed by a specific enzyme. The inspiration for modelling the reaction network as a graph comes from the natural structure of the metabolic pathways. We hope to address issues of structural complexity and model reduction through this approach. However choosing a graph as a model for the reaction network is not new. Trees and networks as biological models have been investigated extensively in the past, see [9]. Here some models for enzyme networks have also been discussed. We hope to develop this model further and examine its suitability in investigating structural properties of the complex dynamic system.The model here is very closely linked to the state space formulation discussed in the previous sections. General reference on the use of graphs for control and system theory are [10, 12].

3.2 Graphs of cell networks

In the state space model discussed above we have already seen that the number of state variables is very large. The metabolic pathways can themselves be seen as some kind of a flow-graph, however to study the associated complexity properly the picture needs to be described in a more formal manner. This becomes all the more important as there are state variables (metabolites) which might be present in more than one reaction while this may not be explicit from the picture of the metabolic pathway.

First we take a look at the most general graphical representation of the metabolic pathway. Each metabolite or state variable (in the state space notation discussed above) is represented by a node x_i in the graph. Nodes x_i and x_j are linked by a directed edge from j to i if the following condition is fulfilled:

$$\frac{\partial f_i(\mathbf{x})}{\partial x_j} \neq 0$$

where $f_i(\mathbf{x})$ is as denoted in Section 2.3.

So if the node x_i can be modulated by the node x_j there is a directed edge from j to i. The entire graph can thus be constructed from the state space description given in 2.3.

This representation can be improved further to include allosteric regulation, inhibition or other kind of influences from one variable to the other which is not apparent from the picture of the metabolic pathway. Enzymes in this representation would then be nodes from which directed edges lead to metabolites participating in the reactions catalysed by them. This is still a very elementary model, in the sense that we have not incorporated the kinetics of the reactions or the extent of influences exterted by one varable on the other. It can be done by ascribing suitable weights to the edges of the digraph. However even without this we can formulate some meaningful questions and try to find their answers with a graph theoretic approach. As we do this it also become clear as to why the formulation needs to be improved further to be more useful. The digraph description contains less information than the equations in the state space model, however it reflects the structure of the dynamical system very adequately. The entire development in this section is based on the approach discussed in [12].

To analyse the cell network with this approach we define the structure matrix for the non-linear system under consideration. The elements of the structure matrix A are given by

$$[a_{i,j}] \quad = \quad \begin{cases} L, \text{ if } \frac{\partial f_i(\mathbf{x})}{\partial x_j} \neq 0, \\ 0, \text{ if} \frac{\partial f_i(\mathbf{x})}{\partial x_j} = 0, \end{cases}$$

where $f_i(\mathbf{x})$ is as defined before. The digraph will have an edge from x_j to x_i if $a_{i,j} \neq 0$. We shall denote the digraph corresponding to the structure matrix $\mathbf{A}$ by $G(\mathbf{A})$. With this equivalent representation of the system by a digraph and the corresponding structure matrix we can investigate the structural properties of the metabolic network.

3.3 Decomposition of the graph

The idea of the structure matrix is suitable to work with so long as we are not concerned with the extent of influences or the weights on the edges of the digraph. For large dynamical systems there is a high degree of sparsity in the structure matrix reflected very clearly in the graph. The graph $G(\mathbf{A})$ is invariant with respect to permutation transformations of $\mathbf{A}$, and an appropriate reordering of the vertices can sometimes be very meaningful. For this we decompose the graph $G(\mathbf{A})$ into subgraphs based on connectability properties. An interesting subgraph to look at while studying large systems is a *strongly connected component*.

Definition 3.1 *Two nodes x_i and x_j are said to be* strongly connected *if a path exists from node x_i to x_j as well as a path from x_j to x_i.The subset of nodes strongly connected to a node x_i form an equivalence class within the set of all the nodes of $G(\mathbf{A})$. Each such equivalence class of nodes with all the edges incident only with these nodes form a subgraph $G(\mathbf{Q})$ corresponding to an irreducible square submatrix $\mathbf{Q}$ of $\mathbf{A}$.*

A concept of order among equivalence classes is introduced by enumerating them in such a way that a transition from a lower to a higher equivalence class is not possible. Finally the reordered structure matrix $\tilde{\mathbf{A}}$ is obtained from $\mathbf{A}$ by a permutation transformation

$$\tilde{\mathbf{A}} \quad = \quad \mathbf{P}'\mathbf{A}\mathbf{P}$$

where $\mathbf{P}$ is a permutation matrix.The transformed matrix $\tilde{\mathbf{A}}$ is a quasi upper-triangular matrix with irreducible diagonal blocks on the main diagonal. An algorithm for computing the transformed matrix and thus decomposing the entire graph into strongly connected components is given in [12]. An attempt to decompose the structure matrices of the graphs of examples discussed in section 2.3 using the above algorithm did not yield any positive results as these matrices for the examples considered were irreducible. Strongly connected components are interesting mathematical structures to have in a large scale dynamical system with each node belonging to one such subgraph having an influence on all other nodes of that particular equivalence class. In the metabolic network that we are dealing with, a set of ordered equivalence classes would mean a hierarchical ordering of the reactions based on their scope of influence. However we find that in many examples of interest the structure matrices are irreducible themselves and cannot be further decomposed into strongly connected components. The reason for this is the predominance of reversible interactions in the metabolic network.

It might be interesting to see what irreducibility implies for stability. Also here it is good to ask the question that if strongly connected components are not common in the graph representing the metabolic network then what kind of structures are? It happens that the subsystems that we should be interested in from a biological perspective are *monofunctional units*. Here we reproduce the definition from [13]

Definition 3.2 *In intracellular metabolic networks, it is useful to recognise subsystems in which the metabolites are only produced or consumed by reactions within that subsystem or by a limited number of fluxes crossing the borders of that subsystem. In many cases such subsystems function as units with respect to their effect on the remainder of the system.*

The analogue of the approach of monofunctional units is not developed in control and system theory as far as the authors know but, if done, it is not well known to researchers in the field.

A monofunctional unit satisfies the following three criteria:

- the reactions outside the subsystem are not affected directly by metabolites belonging to the subsystem.
- there are no conservation relations linking the subsystem to the rest
- the subsystem is linked to the remainder of the system only via one degree of freedom in fluxes.

The three properties as listed above appear to be structural properties of the subsystem and hence it should be possible to replace them with their graph theoretic analogues and look for corresponding submatrices by an appropriate partitioning scheme of $\mathbf{A}$.

The same purpose is achieved by using Metabolic Control analysis to partition the stoichiometric matrix using the concept of co-response coefficients [13]. The equivalent structural properties of the corresponding subgraph would be:

- Only one vertex of the subgraph shares an edge with the rest of the graph
- There are no conservation relations on variables associated with each of the nodes of the subgraph linking them to the rest of the graph.

However the above conjecture needs to be carefully looked into.

3.4 Discussion

The above approach of investigating the structural properties of the metabolic system using the corresponding graph and the structure matrix seems promising, however the power of the method can be enhanced further by including the kinetics of the interactions using suitable weights on the edges of the digraph. That would allow us to judge the magnitude of influence of particular paths of the metabolic network on some of the state-variables (nodes) of interest and help elicit substructures interesting from a control perspective in a relatively simple manner. In a metabolic pathway the main interest lies in the chain of reactions which lead from a starting substrate to the product of interest. Usually the side reactions are ignored if the mass percentage of the variable (substrate) of interest involved is not significant. Similarly the picture does not give a good idea of many other regulatory influences that might be there. However from a control point of view these insignificant side interactions might be important if they are very sensitive to changes. Such interactions can be modelled in a formal manner in this framework. Also the number of interactions inside a cell is much larger than the number of substrates. This implies that the same substrates might take part in a large number of interactions. This further justifies working with the above model if the aim is to gain insight into the regulatory aspect of the network.

However there are disadvantages of such a kind of modelling too. The metabolic pathways give a very clear qualitative picture of the chemical reactions which is lost in the graph. Also the segments of metabolic pathways like glycolysis pathway etc. offer us very insightful

reduced order picture if the concern is chemical flux and the important intermediate products of the chains of reactions. In eukaryotic cells many chemical reactions are spatially localised in organelles. This means that a particular interaction might affect only a part of the substrate which is present in a specific location inside the cell leaving the same unchanged in other parts. The graph picture as it has been described above does not take this into account. Also information about many of the interactions and regulatory influences inside a cell is not completely known and this poses a serious handicap to understand functional and regulatory structures in the framework discussed above.

Graphs are a neat way of looking at large dynamical systems. The above approach can also be found in the recent interesting report by A. Carbone and M. Gromov [4], however there the authors have focussed chiefly on graphs of gene regulatory network. An alternate graphical representation of the metabolic network is also possible where a node represents a reaction instead of a metabolite and the edge represents the metabolite linking two reactions. Infact this is the way in which graphs of metabolic pathways have been discussed in [4]. However since the number of interactions are much larger compared to the number of substrates we prefer the other representation to have the structure matrix $\mathbf{A}$ of a smaller dimension. A graphical representation as described above also helps us to analyse concepts of structural controllability and observability of a large scale dynamical system using the structure matrix. This has been discussed extensively in [12], however the concepts are applicable mostly to linear systems and hence of limited use here. This issue has been discussed further in section 4.

3.5 Closure of class of systems with respect to interconnections

Is the class of RM^2 systems closed with respect to series and feedback connections? This question is of interest to control and system theory, in particular for the study of cell networks.

Proposition 3.3 *Consider two RM^2 systems with representations*

$$\begin{aligned}
\dot{x}_{1,i}(t) &= \sum_{j=1}^{n_{1,v}} (N^+_{1,i,j,} - N^-_{1,i,j}) \left[r_{1,j,+}(x_1(t), u_{1,in}(t)) - r_{1,j,-}(x_1(t), u_{1,in}(t)) \right] \\
&\quad \times u_{1,en,j}(t), \quad x_{1,i}(t_0) = x_{1,i,0}, \ \forall i \in \mathbb{Z}_{n_1}, \\
\dot{x}_{2,i}(t) &= \sum_{j=2}^{n_{2,v}} (N^+_{2,i,j,} - N^-_{2,i,j}) \left[r_{2,j,+}(x_2(t), u_{2,in}(t)) - r_{2,j,-}(x_2(t), u_{2,in}(t)) \right] \\
&\quad \times u_{2,en,j}(t), \quad x_{2,i}(t_0) = x_{2,i,0}, \ \forall i \in \mathbb{Z}_{n_2}.
\end{aligned}$$

Assume that the substances represented by x_1 and x_2 are different.

(a) Series connection with positive linear control law. *Consider the linear control law for a series connection,*

$$u_{2,in}(t) = G x_1(t), \ G \in \mathbb{R}_+^{n_{2,in} \times n_1}. \tag{1}$$

Then the series connection is again an RM^2 system with representation,

$$\begin{aligned}
\dot{x}_i(t) &= \sum_{j=1}^{n_v}(N_{i,j}^{+} - N_{i,j}^{-})\left[r_{j,+}(x(t), u_{1,in}(t)) - r_{j,-}(x(t), u_{1,in}(t))\right] u_{en,j}(t), \\
&\quad x_i(t_0) = x_0,\ \forall i \in \mathbb{Z}_n,\ \textit{with,} \\
&\quad n = n_1 + n_2,\ n_v = n_{1,v} + n_{2,v}, n_{en} = n_{1,en} + n_{2,en},\ n_{in} = n_{1,in}, \\
x &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^n,\ N = \begin{pmatrix} N_1 \\ N_2 \end{pmatrix} \in \mathbb{N}^{n \times n_v},\ u_{en} = \begin{pmatrix} u_{1,en} \\ u_{2,en} \end{pmatrix} \in \mathbb{R}^{n_{en}}, \\
& r_{j,+}(x, u_{1,in}) \\
&= \begin{pmatrix} r_{1,j,+}(x_1, u_{1,in}) \\ r_{2,j,+}(x_2, Gx_1) \end{pmatrix} \in F_{rat,\mathbb{R}_+}(k_1, k_2), \\
&\quad \textit{and similarly } v_{j,-}(x, u_{1,in}).
\end{aligned}$$

(b) Series connection with a positive rational control law. *The same conclusion as in (a) holds in case the control law equals,*

$$u_{2,in}(t) = g(x_1(t)),\ g \in F_{rat,\mathbb{R}_+}(m_1, m_2).$$

However, the formulas are slightly different in this case.

(c) Feedback connection - Linear control law. *Consider the feedback connection of the system defined above,*

$$\begin{aligned}
u_{2,in} &= G_1 x_1, \quad G_1 \in \mathbb{R}_+^{n_{2,in} \times n_1}, \\
u_{1,in} &= G_2 x_2 + u, \quad G_2 \in \mathbb{R}_+^{n_{1,in} \times n_2}.
\end{aligned}$$

The feedback connection is again an RM^2 system with representation as the standard representation of an RM^2 system with

$$r_{j,+}(x, u) = \begin{pmatrix} r_{1,j,+}(x_1, G_2 x_2 + u) \\ r_{2,j,+}(x_2, G_1 x_1) \end{pmatrix} \in F_{rat,\mathbb{R}_+}(m_1, m_2).$$

(d) Feedback connection with a positive rational control law. *The same conclusion as in (c) holds if the control law equals,*

$$u_{2,in}(t) = g(x_1(t)),\ g \in F_{rat,\mathbb{R}_+}(m_1, m_2).$$

However, the formulas are also slightly different in this case.

Proof (a) The series connection results in the formulas stated above with the following details.

$$\begin{aligned}
r_{2,j,+}(x_2, u_{2,in}) &= r_{2,j,+}(x_2, Gx_1) = \frac{p_{2,j,+}(x_2, Gx_1)}{q_{2,j,+}(x_2, Gx_1)} \\
&= \frac{p_{3,j,+}(x)}{q_{3,j,+}(x)} \in F_{rat,\mathbb{R}_+}(k_1, k_2), \text{ because,} \\
p_{2,j,+}(x_2, Gx_1) &= \sum_{k_1,\dots,k_{n_2}=0}^{m_1} \sum_{r_1,\dots,r_{n_1}=0}^{m_2} c(k_1, \dots, k_{n_2}, r_1, \dots, r_{n_1}) \\
&\quad x_{2,1}^{k_1} \dots x_{2,n_2}^{k_{n_2}} (Gx_1)_1^{r_1} \dots (Gx_1)_{n_3}^{r_{n_3}}, \\
(Gx_1)_i^{r_i} &= (\sum_{s=1}^{n} G_{is} x_{1,s})^{r_i} = \sum_{q_1,\dots,q_{n_1}=0}^{m_3} d(q_1, \dots, q_{n_1}) x_{1,1}^{q_1} \dots x_{1,n_1}^{q_{n_1}}, \\
p_{2,j,+}(x_2, Gx_1) &= \sum_{k_1,\dots,k_{n_2}=0}^{m_1} \sum_{q_1,\dots,q_{n_1}=0}^{m_3} \\
&\quad ie(k_1, \dots, k_{n_2}, q, \dots, q_{n_1})\ x_{2,1}^{k_1} \dots x_{2,n_2}^{k_{n_2}} x_{1,1}^{q_1} \dots x_{1,n_1}^{q_{n_1}} \\
&= \sum f(h_1, \dots, h_n) x_1^{h_1} \dots x_n^{h_n} \in \mathbb{R}_+[x].
\end{aligned}$$

(b) The proof in this case is analogous to that of (a) but with the following changes.

$$\begin{aligned}
r_{2,j,+}(x_2, u_{2,in}) &= r_{2,j,+}(x_2, g(x_1)) = \frac{p_{2,j,+}(x_2, g(x_1))}{q_{2,j,+}(x_2, g(x_1))} \in F_{rat,\mathbb{R}_+}(m_1, m_2), \\
p_{2,j,+}(x_2, g(x_1)) &= \sum c(.) x_{2,1}^{k_1} \dots x_{2,n_2}^{k_{n_2}} g(x_1)_1^{r_1} \dots g(x_1)_{n_3}^{r_{n_3}} \\
&= \sum d(.) x_{2,1}^{k_1} \dots x_{2,n_2}^{k_{n_2}} p_g(x_1)^{r_1} \dots p_g(x_1)_{n_3}^{r_{n_3}} / q_g(x_1)^r.
\end{aligned}$$

(c) and (d) These parts of the proof are analogous to those of (a) and (b). □

The same conclusions as the above Proposition holds for a power-law kinetic system.

The network of a cell system is determined by the rate functions $v_{j,+}, v_{j,-}$. A characteristic of a rate function is the set of state variables on which it depends. The state variables represent concentrations of chemical substances. In a system for one reaction the rate functions $v_{j,+}, v_{j,-}$ depend only on the states involved in the reaction and in a reversible reaction the number of such state variables can be two or higher. In case there is an input from inside the cell and there is a series connection, then the rate functions depend also on the states of the upstream reaction. However, in case of a feedback connection to an internal input function, the rate functions depend also on states of several reactions downstream or elsewhere in the network. The subclasses of cell systems obtained by feedback connections require further investigation.

3.6 System reduction

The biologist Hans Westerhoff has asked the authors to investigate system reduction of cell systems. *System reduction* or *model reduction* is the procedure by which a dynamic system is transformed into a second dynamic system with the same input-output sets such that the following *system reduction criteria* are satisfied:

1. *approximation*: the input-output trajectories of the first system are approximated by those of the second system; and

2. *complexity*: the second system is of lower complexity than the first system.

This definition requires definition of an approximation criterion and of a complexity criterion. As an example, the reader may think for the approximation criterion of the L_2 norm on input-output functions and for the complexity criterion of the number of state variables or the number of reactions.

There is a body of theory about system reduction for finite-dimensional linear systems. Known approximation criteria are the L_2 norm and the Hankel norm. The complexity measure is of the dimension of the state space of the reduced system. There are initial approaches to system reduction of nonlinear system but for this problem much more research is needed.

System reduction of cell systems is a new subject as far as the authors can determine. The purposes of reduced systems are many. A realistic system of a complete cell, say E. Coli, has about 1000 reactions. Even numerical simulation of such a system is hazardous. It seems likely that the realistic behavior of several cell variables can be described by a reduced system of low complexity. The existence of feedback loops in cell networks may restrict the dynamics such that a system of low complexity is a good approximate of the complete cell system. These conjectures need investigation.

A first approach to system reduction of cell systems is to restrict attention to a linearized system and to apply system reduction techniques for finite-dimensional linear systems. The resulting reduced systems should be investigated as well as the values of the corresponding approximation and complexity criteria. System reduction for nonlinear positive systems requires the development of new theory. Should the reduced system be in a particular class of kinetic systems so as to allow a physical interpretation?

Two problems of system reduction are formulated below. Consider a series network of kinetic systems with representation,

$$x_i(t) = \sum_{j=1}^{n_v} (N_{i,j}^+ - N_{i,j}^-) \left[r_{j,+}(x(t), u_{in}(t)) - r_{j,-}(x(t), u_{in}(t)) \right] u_{en,j}(t), \;\; x_i(t_0) = x_{i,0}.$$

As established in Subsection 3.5 particular subclasses of kinetic systems are closed with respect to series connections. Therefore system reduction of a series network is in those cases equivalent to system reduction of a single system. A particular question is: Can the number of reactions of a kinetic system be reduced? If so then the number of state variables can possibly be reduced also. A conjecture is that the reduced system may be obtained by retaining in the series network only the slowest reactions; of course, with due account of the reversibility of the reactions. An approach to a time scale decomposition and a system reduction is not described here.

System reduction of feedback connections can probably also be handled as that of a series connection.

These system reduction problems require urgent investigation.

3.7 Hierarchical system for cell reaction network

In this subsection will be sketched how a hierarchical system for a cell network may be constructed and be used for understanding and control of cell functions.

Hierarchy is a well known modelling and structuring technique in nature and in man-made engineering systems. To manage large numbers of units of anything requires imposition of a control structure. Complexity is the term used in control theory and, in a slightly more formal meaning, in computer science for the analysis of large structures. Considering the large number of reactions in cells, hierarchical models and hierarchical systems are to be

considered as models. The Ph.D. thesis of Rohwer, [13], contains an approach to functional submodels of the cell that is within the hierarchical approach discussed. Hierarchical models of engineerings systems with which the authors are familiar are communication networks and traffic control networks.

A cell is classified as either Eukaryotic or prokaryotic depending on whether it contains organelles inside the cell. An organelle is a subunit of the cell in which functions of the cell are concentrated. Examples of organelles are: nucleus, mitochondria etc.

As a first approach it seems useful to distinguish the following levels in a hierarchical model of the cell:

1. Cell.
2. Organelles of a cell. Biologists have already defined organelles.
3. Metabolic pathways. Biologists have recognized such pathways.
4. Reaction networks. A reaction network consist of several reactions and it forms part of a metabolic pathway.
5. Individual biochemical reactions.

A particular level in the hierarchy is a bundle of items in the level directly below it and this bundle forms part of a bundle in the level directly above it. Thus, a reaction networks bundles several reactions and with other reaction networks it forms a metabolic pathway.

How to formulate a hierarchical model for a cell? Biological modelling will lead to the definition of organelles and metabolic pathways. For the level of reaction networks possibly a system theoretic approach may be useful. For such a system theoretic approach the concept of equivalence relations and abstractions have been formulated. It needs to be explored how useful these concepts are for the isolation of reaction networks and for metabolic pathways on the basis of an interconnection of biochemical reactions.

A hiearchical system can be used for several purposes. System reduction is best based on the existing network and hierarchical structure of the reactions in a cell. This type of approach is discussed further in the next subsection. Control of biochemical reactions is best also considered based on a hierarchical system. Control at the level of reaction networks is best restricted to feedback laws based on local information of the reaction network. Control of metabolic pathways and organelles has to be handled at a higher level of the hierarchy. It will be of interest to investigate how this is done by nature in cells. This topic is discussed further in Section 5.

3.8 Stability properties of reaction networks

A major problem area of reaction networks and of metabolic pathways is to establish their stability properties. In this short investigation little attention has been spent on this problem. Researchers with whom the authors are familiar, including E.D. Sontag and P. De Leenheer, are actively involved in investigations of stability properties of reaction networks. In Section 5 several problems for stability of reaction networks are mentioned. Stability of positive linear systems is discussed in the book [2].

4 Control of cell reaction networks

4.1 Modelling for control

In this subsection is presented a model of how inputs to a cell system can be used to control the biochemical processes of the cell.

Consider a kinetic system with representation,

$$\begin{aligned}\dot{x}_i(t) &= \sum_{j=1}^{n_v}(N_{i,j}^+ - N_{i,j}^-) \times \\ &\quad \times \left[r_{j,+}(x(t), u_{ex}(t), u_{in}(t)) - r_{j,-}(x(t), u_{ex}(t), u_{in}(t))\right] u_{en,j}(t), \\ &\quad x_i(t_0) = x_{i,0},\ \forall i \in \mathbb{Z}_n.\end{aligned}$$

The inputs to the system are:

- the enzyme input u_{en};
- the external input u_{ex} representing chemical substances which flow from outside the cell to the part of the cell modeled; and
- the internal input u_{in} representing chemical substances which flow from other parts of the cell to the part of the cell being modeled.

In principle all three inputs can be used for control though most interest is focused on enzyme input. Biochemists know best for which cell which input is feasible but it seems that much of the bioengieering remains to be explored.

The enzyme input can be used directly or indirectly in several ways:

1. addition of enzyme through supply to the cell from the outside;
2. inhibition of enzyme, discussed below; and
3. gene expression, by influencing the production of enzymes by the genes.

An enzyme is a protein which acts as a catalyst for a particular reaction. The reaction takes place on the workbench of the enzyme determined by a sequence of chemical structures. The reaction is said to be inhibited if the place on the enzyme for the reaction is occupied by another chemical substance. The reaction normally catalyzed by the enzyme cannot take place then. The degree of inhibition depends on the concentration of the inhibition substances. A formula for this is,

$$u_{en,j}(t) = \left(1 - \frac{u_{inhib,j}(t)}{\bar{u}_{en,j}(t)}\right)^+ \bar{u}_{en,j}(t),$$

where $\bar{u}_{en,j}$ denotes the actual concentration of the enzyme in the cell, $u_{inhib,j}$ denotes the concentration of the inhibitor, $u_{en,j}$ denotes the effective concentration of the enzyme which is available for reaction $j \in \mathbb{Z}_{n_v}$, and

$$x^+ = \begin{cases} x, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$$

See [6, pp. 21-22] for several types of inhibitions. According to a knowledgeable source, inhibitors are known for about 10% of the reactions of a particular cell. The effectiveness

of the known inhibitors is limited. Modes of control are the increase or decrease of natural inhibitors.

Before starting the development of control theory for cell systems it seems useful to study the way feedback is used in cell networks by nature. This requires a detailed study of cell systems of several cells. Knowledge of the system of differential equations of a complete cell is still quite limited but expected to increase rapidly in the coming years. Questions to be considered include: (1) Does feedback take place by internal inputs from other parts of the cell or through inhibition of enzymes? (2) Is feedback local, meaning does the control law depend on state variables which are relatively close in the network? (3) Can feedback of small networks by analyzed analytically?

4.2 Control problems

At the level of biology or that of bioengineering the problem is to supply an input to the cell such that the biochemical processes of the cell achieve specified control objectives. At the level of control theory the biological problem amounts to the construction of a control law such that the control objectives are achieved.

Control objectives include:

- to stimulate one or more reactions inside the cell; and
- to prevent one or more reactions inside the cell.

A cell may be diseased, which means that it does not function properly. An organism may be supplied with a drug which inhibits a particular reaction in the cell. The cell then dies and is no longer an obstacle for the organism. This procedure works properly only if it is known which substances are enzymes for which reactions. The knowledge about this is still limited. It is expected that knowledge about biochemical processes of cells will increase in the coming decade.

Control of biochemical processes of a cell leads to the following control problems.

1. *Steady states.* Calculate or compute the steady states of a kinetic system. How does the steady state depend on the inputs, the enzyme input, the external input, and the internal input.

2. *Local stability.* Is a kinetic system locally stable at a steady state, for initial conditions in a neighborhood of the steady state? Results for this problem on a class of cell networks have been provided in [16].

3. *Global stability.* Is a kinetic system globally stable at the steady state, for arbitrary initial conditions? Is the system globally asymptotically stable?

4. *Input-to-state stability with respect to enzyme inputs.* Is the steady state locally stable with respect to the enzyme input. Results on a related class of cell networks are presented in [16].

5. *Observers.* Construct an observer for the state of a kinetic system based on partial observations of the state. Results on this problem are described in [5].

6. *Control synthesis.* Determine a control law, denoted by $g : X \to \mathbb{R}_+^{n_{en}}$ in $u_{en,j}(t) = g_j(x)$, such that the closed-loop kinetic system achieves specified control objectives. A major question is what types of control laws can be effectively implemented in a cell. The class of control laws has to be restricted because the state space vector has a very

high dimension and for practical reasons a control law should depend on only a few state variables.

7. *Controllability.* Is the system controllable with respect to the enzyme input? The difficulty is that there are results for controllability of systems in $\mathbb{R}^n$ or a on a manifold, but these results do not apply directly to systems with the state set $\mathbb{R}^n_+$. The geometry of the nonlinear positive systems requires further study.

4.3 Reachability and controllability of positive nonlinear systems

For the existence of a control law the property of controllability of the corresponding control system is often a necessary and sufficient condition. A system is said to be *reachable* if from any time and any initial state it is possible to reach at a particular time any other state in the state set of the dynamic system.

Controllability of cell systems will be of interest for control of such systems but also for the general understanding of the dynamics. Cell systems are positive nonlinear systems. Controllability properties have been studied for systems whose state set is $\mathbb{R}^n$, see the books [7, 11]. Conditions for the controllability of such systems are expressed in terms of a particular Lie algebra determined by the dynamics of the system.

Reachability and controllability of a positive nonlinear system have not been studied in full generality. This remains to be done and requires a study of the geometric properties of positive nonlinear systems. The dynamic system that the authors have in mind is,

$$\dot{x}_i(t) = \sum_{j=1}^{n_v} N_{i,j} r_j(x(t))\ u_{en,j}(t),\ x_i(t_0) = x_{i,0}.$$

This system is linear in the enzyme inputs and thus fits a general subclass of nonlinear systems for which controllability properties have been characterized.

An approach is to investigate the local behavior of a positive nonlinear system. In a first approach, the system is linearized at an equilibrium point and the resulting linear system is analyzed for its stability properties. In a second approach the full positive nonlinear system is analyzed but it is assumed that the positivity constraints will not be violated. Then the controllability conditions for ordinary nonlinear systems on $\mathbb{R}^n$ can be used. The theory for this remains to be worked out. References on controllabilty of nonlinear systems as described above are [7, 1.5, 1.6] and [11].

5 Problems to be investigated

This section outlines problems and envisioned approaches to problems which are based on the investigation of this report.

1. *Modelling of cell reaction networks.*

   - Formulate in state space form examples of parts of the biochemical processes of the cell. Sources are [1, 13]. See also the E. Coli web page
     `http://ecocyc.pangeasystems.com/ecocyc/ecocyc.html`
   - Tools are needed to handle large examples, say with more than 100 to several thousands of chemical substances or reactions.
   - Subclasses of reaction systems. It will be of interest to delineate in more detail several classes of systems which frequently occur in modelling of the biochemical processes of the cell.
   - Interconnections of kinetic systems. Is a particular class of kinetic systems closed with respect to series and feedback connections and, if not, which new classes are so generated.
   - Formulate concepts to study the geometry of positive nonlinear systems. For positive linear systems the basic geometric object is a polyhedral cone. For example, the reachable set is a polyhedral cone. For positive nonlinear systems, the appropriate concept is not clear.
   - Sensitivity of the steady state with respect to changes in enzyme concentrations or with respect to changes in parameters. For this the metabolic control analysis (MCA) has been developed by H.V. Westerhoff and colleagues, see [8, 14, 15, 18, 17, 19]. It will be of interest to work out the formulas for the sensitivity coefficients based on the state-space formulas for dynamic systems.

2. *System structure - Graphical decomposition, hierarchical systems, and system reduction.* The motivation for this investigation is that the cell systems obtained for realistic cells will be very large, possibly with a number of reactions of the order of several 1,000 to about 30,000.

   - Decomposition of cell reaction networks into organelles, metabolic pathways, and subnetworks. Association of a cell reaction network with a graph. Decomposition of the graph in strongly connected components. Decomposition of a strongly connected component of a graph into subcomponents based on monofunctional units, see [13]. There are algorithms to decompose a graph into its strong components, see [12].
   - Decomposition of cell reaction network in hierarchical levels by equivalence relations and abstractions.
   - System reduction of series connections and of feedback connections of cell reaction network needs investigation.

3. *Control of cell systems.*

   - How is control over the cell reaction network exerted by cells? Modelling of control of cell reaction networks as it occurs in actual cells by enzymes, by inhibition of enzymes, by external chemical influences, and by electrical influences. Modelling of inhibition.

- Study the stability properties of positive nonlinear systems. Results on input-to-state stability of a particular class of kinetic systems for this problem have already been derived in [16].
- Controllability properties of cell systems with respect to enzyme inputs. Use has to be made of the graphs associated with cell reaction networks.
- Control synthesis for control by enzymes or by inhibitions.
- Control theory for small cell reaction networks.
- Control of cell processes by gene expression.

6 Concluding remarks

The investigation on which this report is based is a new research area for both authors. The research has been performed in the period 13 May 2001 till 24 July 2001 when the first author visited the research institute CWI. The aim of the project was to investigate modelling and control of cell reaction networks, primarily to formulate problems and approaches.

In Section 2 models of cell reaction networks in the form of dynamic systems are formulated. Topics studied in Section 3 are the closure of particular subclasses of systems with respect to series and feedback connections, the graphs associated with cell reaction networks, decomposition of such graphs, hiearchical systems, and system reduction. Control problems for cell reaction networks are formulated in Section 4 . Section 5 contains a list of problems of modelling, system structure, and control for cell reaction networks.

The most important topics for research in cell reaction networks are: (1) The system structure of cell reaction networks. (2) Hierarchical system modelling and system reduction. (3) Control of cell reaction networks. The authors are of the opinion that they have only recognized the surface of the important and relevant research area of control and system theory for cell reaction networks.

Acknowledgements

The authors are very grateful to Prof. Hans V. Westerhoff (Vrije Universiteit, Amsterdam) for many discussions with him on the topic of the report. The first author thanks CWI for having granted him the opportunity to work on this investigative project in the control and system theory research group of Prof. Jan H. van Schuppen.

References

[1] B.M. Bakker. *Control and regulation of glycolis in Trypanosoma brucei.* PhD thesis, Vrije Universiteit, Amsterdam, 1998.

[2] A. Berman and R.J. Plemmons. *Nonnegative matrices in the mathematical sciences.* Academic Press, New York, 1979.

[3] N.A. Campbell, J.B. Reece, and L.G. Mitchell. *Biology (Fifth Ed.).* Addison Wesley Longoman Inc., Menlo Park, 1987.

[4] A. Carbone and M. Gromov. Mathematical slices of molecular biology. Report, Institut des Hautes Études Scientifiques, Bures-sur-Yvette, 2001.

[5] M. Chaves and E.D. Sontag. State-estimators for chemical reaction networks of Feinberg-Horn-Jackson zero deficiency type. In *Proceedings NOLCOS2001*, 2001.

[6] R. Heinrich and S. Schuster. *The regulation of cellular systems.* Chapman and Hall, New York, 1996.

[7] A. Isidori. *Nonlinear control systems: An introduction*, volume 72 of *Lecture Notes in Control and Information Sciences.* Springer-Verlag, Berlin, 1985.

[8] D. Kahn and H.V. Westerhoff. Control theory of regulatory cascades. *J. Theor. Biol.*, 153:255–285, 1991.

[9] N. MacDonald. *Trees and networks in biological models.* John Wiley & Sons, Chichester, 1983.

[10] K. Murota. *Matrices and matroids for systems analysis.* Number 20 in Algorithmics and combinatorics. Springer-Verlag, Berlin, 1999.

[11] H. Nijmeijer and A.J. van der Schaft. *Nonlinear dynamical control systems.* Springer-Verlag, Berlin, 1990.

[12] K.J. Reinschke. *Multivariable control - A graph-theoretic approach*, volume 108 of *Lecture Notes in Control and Informations Sciences.* Springer-Verlag, Berlin, 1988.

[13] J.M. Rohwer. *Interaction of functional units in metabolism.* PhD thesis, Vrije Universiteit, Amsterdam, 1997.

[14] S. Schuster, D. Kahn, and H.V. Westerhoff. Modular analysis of the control of complex metabolic pathways. *Biophysical Chemistry*, 48:1–17, 1993.

[15] S. Schuster, D. Kahn, and H.V. Westerhoff. Control analysis of metabolic systems consisting of uni- and/or multifunctional units - application to modular systems and slipping enzymes. *J. Biological Systems*, 3:217–230, 1995.

[16] E.D. Sontag. Structure and stability of certain chemical networks and applications to the kinetic proof-reading model of T-cell receptor signal transduction. *IEEE Trans. Automatic Control*, 46:1028–1047, 2001.

[17] H.V. Westerhoff, J.-H. S. Hofmeyr, and B.N. Kholodenko. Getting to the inside of cells using metabolic control analysis. *Biophysical Chemistry*, 50:273–283, 1994.

[18] H.V. Westerhoff, W. van Heeswijk, D. Kahn, and D.B. Kell. Quatitative approaches to the analysis of the control and regulation of microbial metabolism. *Antonie van Leeuwenhoek*, 60:193–207, 1991.

[19] J.E. Wijker etal. Energy, control and DNA structure in the living cell. *Biophysical Chemistry*, 55:153–165, 1995.