

Biosystems Engineering: Applying methods from systems theory to biological systems

A. Kremling*
MPI for Dynamics of Complex
Technical Systems
Magdeburg, Germany
kre@mpi-
magdeburg.mpg.de

T. Sauter
Institute for System Dynamics
and Control Engineering
University of Stuttgart,
Stuttgart,
Germany
sauter@isr.uni-
stuttgart.de

E. Bullinger
Institute for Systems Theory in
Engineering
University of Stuttgart,
Stuttgart,
Germany
eric@ist.uni-stuttgart.de

ABSTRACT

Using methods and tools from systems theory will offer new possibilities to analyze and design biological systems. The intention of this contribution is twofold: giving an overview on different areas where such methods can be applied. The main focus in the first part will be the set up of mathematical models. Here two complementing ways to come to a suitable model will be discussed. The second part will describe possibilities to validate mathematical models and to design experiments while the third part will discuss methods to analyze complex systems. Besides the overview the second intension is to present also some new ideas for the application of methods from systems theory.

1. INTRODUCTION

In the last years, interdisciplinary cooperation between different autonomous research disciplines, namely biology, informatics and engineering science have let to a dramatic change in life science. Mainly in biotechnology and biomedical engineering, the application of tools and methods from different disciplines succeed in new techniques and products to improve health and supply with foodstuff. The analysis of sequence data and the utilization of database systems to structure the large amount of data has leveraged bioinformatics to became a very popular science. Moreover the generation of these data by high throughput methods seems not possible without ingenious application of automation and robotics.

Starting last year with a meeting in Tokyo a new field, namely systems biology, has emerged to emphasize the system character of biological systems. Here, systems biology “aims at system-level understanding of biological systems, and to understand ‘Biological System as System’ ” [25]. Systems biology uses very different tools and methods to analyze cellular systems. The main focus lies on system structure and behavior analysis while on the long run technologies should be established which allow to design biological systems. To identify the whole structure of a cellular network, methods from genomics and proteomics will be applied. Questions in behavior analysis concern stability, ro-

*Corresponding author

bustness and adaption of cellular systems under different growth situations. Since these works require high computational effort, using different software tools, a group of researchers started to define a software platform for systems biology [6]. The heart of this platform is a standard language for cellular models which allow an exchange of models between different tools.

Nearly contemporaneously the Institute for Systems Biology was founded in Seattle, USA, by Leroy Hood. The private institute tries to occupy a niche between academia and industry, using the strength of both. The aim of the research is to produce complete mathematical models of complex biological systems. Since biology “has been studied from the perspective of analyzing individual genes and individual proteins, systems biology, on the other hand, is interested in analyzing whole systems of genes or proteins. What this means is that we use tools for capturing information from many different elements of the overall system. And we have to be able to integrate the information that’s obtained from all the different biological levels—DNA information, RNA information, protein information, protein interaction information, pathways and so forth. The ultimate objective is to use this information to write mathematical models that are capable of predicting something about the structure of the biologic system under evaluation as well as predicting something about its properties, given particular kinds of stimuli or perturbations” [21].

A further definition for systems biology is given by the company called *Beyond Genomics* [14]. The company uses the term in the sense of a systematic approach to reduce time to develop new drug targets “to bridge the gap between molecular biology and clinical science”. This will be achieved by integrating a number of data from genomics and proteomics to have a more complete picture of the molecular interactions inside the cell.

This contribution concentrate on biosystems engineering¹, a consequential application of tools and methods from systems engineering to biological systems (for an overview see Figure 1). The systematic elucidation of biological phenomena

¹The term is taken from a research project in Germany.

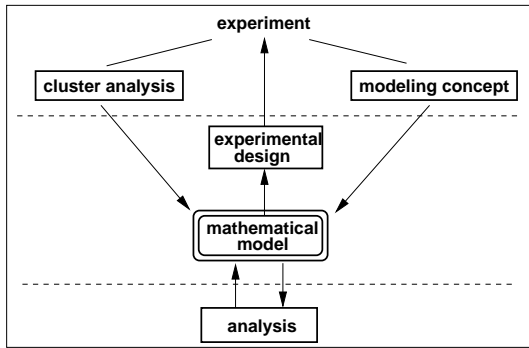


Figure 1: Overview on activities in biosystems engineering/ systems biology. The figure is divided in three parts, representing the three sections of this contribution.

and the development of effective biotechnological solutions to a series of commercial problems are both increasingly dependent on systems engineering approaches to handle the effective analysis of rapidly accumulating amounts of data [42]. The heart of this new approach is the development of mathematical models to simulate and analyze cellular systems. Here, we concentrate on continuous deterministic models, although there a number of different other types. However, Figure 1 is a very general scheme, valid also for discrete and stochastic models.

The term “mathematical model” transcends writing simply equations, e.g. for a continuous deterministic model with states \mathbf{x} , a set of parameters \mathbf{p} and inputs \mathbf{u} :

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{p}, \mathbf{u}) . \quad (1)$$

We suggest, to define also a structure, i.e. decomposing a complex system in smaller parts, which are named modules or functional units. In Figure 1 “mathematical model” have one of the following relevances

- flat model: a model without any structural elements;
- cluster model: structural elements without mathematical equations;
- structured model: model with mathematical equations for structural elements.

The first part focus on a systematic approach to build meaningful models. In previous work [29] some verbal biological motivated criteria were developed to define structure elements which are called functional units (right arrow in Figure 1 from *experiment* to *mathematical model*). However, if a flat model is given (set of equation like equation (1)), the problem is how to demarcate such submodels, i.e. which elements (metabolites, reactions) can be grouped together. These questions are also addressed in the first section. A second way, indicated by the left way in Figure 1, uses time series of mRNA (from DNA micro arrays) and proteins (quantitative proteomics) and tries directly to cluster various elements and if possible assign also equations to the defined elements.

A further emphasis is the validation of a model, based on real experiments. Two tasks will be discussed: how to check the structure of the model by formulating different hypothesis and how to estimate parameters based on the available data. To complete model validation, often some more experiments must be performed. These experiments can be designed according for the purpose of parameter identification or structure identification.

The third part will discuss possibilities of model analysis. Stationary and dynamical behavior has to be considered to describe regulation and adaptation processes. To apply models in process control and design, e.g. during a biotechnological process, they have to be reduced in order and in number of parameters.

2. MODEL SET UP

2.1 Modeling concept

Modern biotechnology will increasingly require quantitative analysis of the complex behavior of cellular systems. Even the “simple” bacterium *Escherichia coli* possesses over 4,400 genes, about 2,500 active proteins and enzymes, 50 -70 sensors in the cell membrane, and hundreds of metabolic pathways converting substrates into intermediary products and cellular structures [32]. The analysis of such complex biochemical networks becomes even more difficult due to the great number of feedback and feedforward loops also involved in cellular control [29]. An overview on mathematical modeling and analysis in biochemical engineering is given in [3]. Concerning *Escherichia coli*, a single cell model has been developed over the past two decades [8, 40]. The model relies on the macromolecular composition, on signaling molecules, and on parameters such as cell volume and cell surface. However, structural elements are not given and therefore the model is difficult to understand and model alterations can hardly be made. A second approach, “cybernetic” modeling, was developed by Ramkrishna and co-workers [43] to model both sequential and simultaneous utilization of substrates. Its advantage is the ability to predict growth behavior on different substrate sources based on parameters determined from single substrate growth experiments. The disadvantage of this approach is that neither the specific mechanisms of induction/repression nor the hierarchical structure of the regulatory network is considered.

Simulations using realistic, molecular-level models of genetic mechanisms and of signal transduction networks are needed to analyze dynamic behavior of multigene systems, to predict behavior of mutant systems and to identify design principles applicable to design of genetic regulatory circuits. When the underlying design rules for regulatory circuits are understood, it will be far easier to recognize common circuit motifs, to identify functions of individual proteins in regulation, and to redesign circuits for altered functions [34].

A new approach [29] is based on the concept that cellular metabolism is structured in functional units that could be used in modeling [31]. Complex metabolic networks can be decomposed into smaller physiologically meaningful units. This is necessary for avoiding problems with dispensable numbers of equations and (mostly) unsure parameters; this makes modeling easier and more transparent. It also sup-

ports the finding of the meaningful structures of cellular control and thus supports the holistic understanding of cellular metabolism. In [29] three verbal biological criteria that allow the demarcation of functional units were defined: (i) the presence of an enzymatic network with a common physiological task; (ii) its control at the genetic level by a common regulatory network, e.g. an operon, a regulon, and a modulon - these networks are usually organized in a hierarchical way; and (iii) the coupling of this regulatory network to the environment through a signal transduction network. These structuring principles have been applied to protein synthesis and metabolic pathways. The process of modeling is consequently understood as a progressive combination and linkage of submodels to higher aggregated model structures. Submodels organize the biological knowledge along two coordinates: a structural and a behavioral coordinate. The structural coordinate is described by the number and types of inputs and outputs (terminals). The connections between submodels are fixed according to the type of the terminal. The behavioral coordinate, in contrast, is described by mathematical equations. The coupling between the structural and the behavioral coordinate is the essential step in a quantitative description and is realized by the assignment of mathematical equations to the structural modeling object, e.g. assigning a kinetic rate equation to a modeling object called substance transformer representing an enzyme catalyzed reaction.

A new graphical method for mapping especially multimolecular complexes and protein modifications was presented by Kohn [27]. This seems to be an ideal supplementation to the concept in [29], since it allows the systematic representation of e.g. the detailed conversion of a substrate into a product during a enzym catalyzed reaction (see Fig. 2).

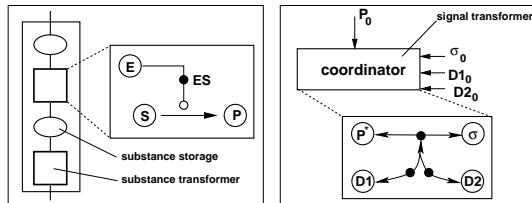


Figure 2: Right: Representation of a functional unit according to the concept in [29] with the representation of a Michaelis-Menten kinetic reaction according to Kohn [27]. Left: Modeling object “signal transformer” (Fig. 3 in [28]) and representation according to Kohn.

The procedure imply that a biochemical reaction network is known, i.e. all interactions between the ‘players’ are identified. The alternative way which leads to the set up of mathematical models is based on the analysis of experimental data from DNA micro arrays and proteomics.

2.2 Cluster analysis

A number of tools and methods were discussed based on real or simulated data [36, 7, 46]. In [16, 3] a hierarchy of mathematical structures for utilizing experimental information on gene networks is given. In [22] a very complete

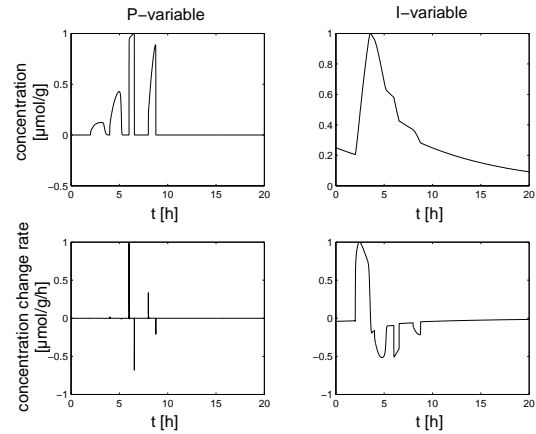


Figure 3: Normed trajectories and time derivatives of typical P- and I-Variables as used in the described algorithm considering typical trajectories.

picture of galactose utilization in the yeast *Saccharomyces cerevisia* is presented. In a very systematic manner experiments were performed to perturb the system by deleting or overexpression of genes and the alteration of the environmental conditions. The results led to an interaction network where groups of genes can be defined which are highly interconnected. A further step using singular value decomposition [19, 20, 2] makes it possible to predict future expression levels. In this work a linear relationship between different modes was calculated and compared to experimental data. In [42] the wavelet decomposition of gene expression signals is discussed allowing the correlation of features, which are localized in time or/and frequency, thus leading to a richer class of potential gene-gene interactions over periods of time or/and of specific periods.

Up to now, often enough experimental data are not available to test algorithms for cluster analysis. To manage this, simulated data from detailed models are used. In addition to the already described verbal criteria, a new formal demarcation procedure will be sketched. Starting with an already set up DAE model (differential and algebraic equations) of the biological system, different mathematical methods have been tested for finding meaningful subunits. Linkage-methods from cluster-analysis were successfully applied to subdivide the set of modeled state variables by means of a measure of distance. A tree structure (dendrogram) is received representing a hierarchical classification of the state variables. Based on this hierarchy, groups of states can be isolated which are barely connected to the environment representing the requested subunits.

Different measures of distance have been examined representing two main paths: First on basis of the Jacobian matrix of the linearized system at different points in state space and second on basis of a typical solution of the DAE-model. Using the Jacobian matrix, good results can be obtained for metabolic models excluding regulation of gene expression. The entries in the Jacobian of signals affecting gene expression were found to be too low in the considered models. Therefore the assignment of gene products was not possible.

If sensitive parameters are detected, measurement are needed to estimate them. A very common tool to check if the measurement contains enough information is the Fisher information matrix \mathbf{F} . The aim of the analysis of \mathbf{F} is to detect which parameters of the parameter vector \mathbf{p} can be estimated together with a given variance γ . The choice of γ depends on the accuracy of the measurement data and the demands on the model.

The Fisher information matrix is defined by

$$\mathbf{F} = \sum_{k=1}^N \left[\boldsymbol{\Omega}^T(t_k) \mathbf{C}(t_k)^{-1} \boldsymbol{\Omega}(t_k) \right], \quad (4)$$

with N data points taken at time points t_k and with the matrix of the sensitivities $\boldsymbol{\Omega}$ and the covariance matrix \mathbf{C} . The covariance matrix \mathbf{C} is assumed to be diagonal with the variance of the states σ_i as elements. It is assumed that σ_i do not depend from time point t_k while it is taken as a constant. In [33] the correlation between the covariance matrix of the parameters \mathbf{C}_p and \mathbf{F} is given by

$$\mathbf{C}_p \geq \mathbf{F}^{-1}. \quad (5)$$

By analyzing the eigenvalues of the Fisher information matrix a set of parameters can be determined which can be estimated with a given minimal variance γ [39]. Although there is only a statement on the lower bound of the variance of the parameter which has to be estimated, the method was applied successfully in optimal experimental design for a biotechnological process [4]: The task is to find an input \mathbf{u} to alter the Fisher information matrix in such a way that one of the following criteria are minimized:

- D-optimal design: $\det(\mathbf{F}^{-1})$
- A-optimal design: $\text{trace}(\mathbf{F}^{-1})$
- E-optimal design [4]: $\lambda_{max}/\lambda_{min}$

with λ are eigenvalues of \mathbf{F} .

An open question in this procedure is the selection of the system output (measurement). In the given references, only simple models to describe a biotechnological production process are used. To apply the procedure with detailed models, it seems necessary to incorporate the choice of the output in the method. Since very often measurement of intracellular metabolites or proteins are sumptuary, only a limited number of outputs are provided. If all states x_i are possible outputs, we suggest to formulate a MINLP (mixed-integer nonlinear programming) problem to determine such outputs y_j which contribute significant to the proposed criteria. MINLP methods were used for product-oriented, constrained optimization of metabolic reaction networks, e.g. maximization of ethanol production [17]. MINLP extends a common optimization problem by restricting some variables to an integer value. If outputs and states are connected in the following way

$$\mathbf{y} = \begin{bmatrix} c_{11} & 0 & \cdots & 0 \\ 0 & c_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & c_{nn} \end{bmatrix} \mathbf{x}, \quad (6)$$

the coefficients c_{ii} are restricted to

$$c_{ii} \in \{0, 1\}. \quad (7)$$

Moreover it is also an open question, in which way the analysis of a structure property called observability can be used in this field. This may help experimentalists designing experiments.

3.2 Model discrimination

If a model is set up, there might also be some uncertainties about the model structure, e.g. two or more version of a submodels formulated as hypothesis are given. Here, the design of an experiment results in finding an input \mathbf{u} which maximize the difference of the model output [37]:

$$\begin{aligned} \text{max } & J \\ \text{s.t. } & \mathbf{e} = \mathbf{y}_1(\mathbf{p}_1, \mathbf{u}) - \mathbf{y}_2(\mathbf{p}_2, \mathbf{u}) \\ & J = \int_0^T \mathbf{e}^T \mathbf{Q} \mathbf{e} dt \end{aligned} \quad (8)$$

where \mathbf{e} is the difference of the model outputs \mathbf{y}_1 and \mathbf{y}_2 with their respective parameter sets \mathbf{p}_1 , \mathbf{p}_2 and the common input \mathbf{u} . J is the error functional that must be maximized.

To find a possible input \mathbf{u} to distinguish two model variants, a new idea is presented whereas the system may be analyzed using linear systems theory. A common tool in control theory for systems analysis is the Bode plot which analyze the output in the form of gain and phase shift, if the system is forced by a sine signal with a selected frequency. In general, different models may describe the same stationary state with their set of parameters. A simple example with two different models is given in the appendix. Model A controls the activity of enzyme 2 while in model B the amount (gene expression) of enzyme 2 is controlled (see Figure 6). Analyzing the model equations, a set of pa-

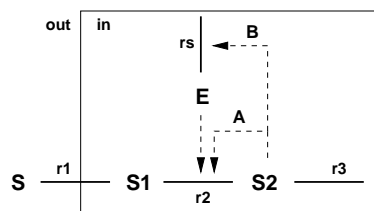


Figure 6: Two model variants: control of enzyme activity, e.g. by allosteric modification (model A) and control of gene expression, e.g. by repression (model B).

parameter for each model can be found which describes the same stationary state. The organism is assumed to grow in a bioreaction and a continuous fermentation is performed. In the following the system is linearized around the steady-state, which is given by $\mu = q_{in}/V$. System inputs are inflow rate q_{in} and feed concentration c_S^{in} . System outputs are the concentration of S1 c_{S1} , S2 c_{S2} and E c_E . Furthermore, it is assumed that a part of the model is validated and the parameters are known. Parameters K_{I1} for model A and K_{I2} for model B describing the control of the enzyme are not known.

In the following the input/output behavior is analyzed by calculating the phase shift for inputs q_{in} and c_S^{in} on outputs c_{S1} , c_{S2} and c_E varying parameters $0 < K_{I1} < K_{I1}^{max}$ and $0 < K_{I2} < K_{I2}^{max}$. Figure 7 shows the result for input q_{in} and output c_{S1} . As can be seen, there exist a small frequency span where the two models display different phase shift for all parameter combinations. Therefore an experiment should be done which forces the system with a distinct frequency from the frequency window to see if model A or model B will be true.

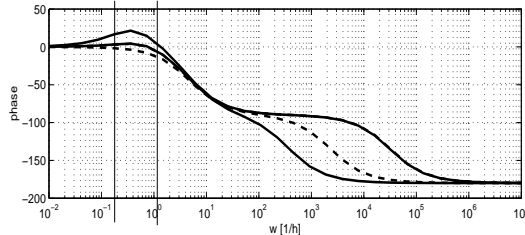


Figure 7: Phase shift for input q_{in} on output c_{S1} . Solid lines shows maximal and minimal values for model A while dashed lines shows minimal and maximal values for model B varying parameters K_{I1} and K_{I2} between 10^{-3} and 10. For the small frequency span indicated by the vertical lines, the models can clearly separated.

4. MODEL ANALYSIS

The crucial point for the benefit of biosystems engineering/ systems biology is the success in analyzing a mathematical model by detecting new correlations among the components or the structural elements. Certain cell functions can be analyzed by looking at the stationary behavior. A widely-used tool is metabolic flux analysis [38]. This analysis is only based on the structure of the stoichiometric matrix of the network and allows the detection of topological properties. Even if not enough measurements are available (the system is under-determined), some stationary fluxes are calculable [26]. Including the assumption that the organism always tries to optimize the growth rate the stationary behavior of *Escherichia coli* and a number of mutant strains could be predicted [10, 9].

Other functions are inherently dynamical. This is for example the case for adaptation which denotes an insensitivity in the long term with respect to a persistent stimulus. As was pointed out by Kitano [25] the identification and systematic storage of so called design patterns will allow a faster analysis of complex systems. To think out this idea from a control engineer's point of view, this is analogous to the identification of transfer functions in technical systems. However, a fundamental difference between man made machines and biological systems is that technical systems are composed step by step: a (unstable) plant is stabilized by a controller. On the other hand, analyzing cellular systems we are always confronted with the closed loop behavior. This implies the question which part of the system is the controller and which part is the controlled system. If we are able to assign known transfer functions to different parts of the biological system this will be a great help in the understanding of the over-

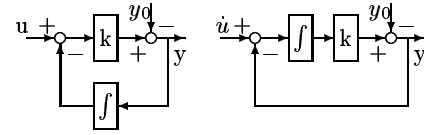


Figure 8: Integral control loop in chemotaxis: u denotes the stimulus by the chemoattractant, y the cell tumbling frequency and y_0 the difference between actual and normal cell tumbling frequency. The left diagram, taken from [45], is equivalent to the right one.

all dynamical behavior. Two examples, exemplifying these ideas are chemotaxis [45] and the regulation of the inositol triphosphate (IP_3) level in the cell [23].

4.1 Regulation and adaptation

Inositol triphosphate is an intracellular messenger which causes the release of calcium from the sarcoplasmic reticulum by an action on IP_3 receptors [5]. A step increase of calcium (Ca^{2+}) causes an initial peak in the IP_3 receptor which then decay back to its initial value [23]. Thus, the level of IP_3 is statically independent of the Ca^{2+} concentration. In control theory, disturbance attenuation is the term used to describe similar phenomena. In this context, the calcium concentration is seen as a disturbance and the level of IP_3 as the output to be controlled. Another way of describing this behavior is to say that the level of IP_3 is regulated by a controller which needs to be adapted each time the Ca^{2+} level changes.

The relatively oriented movement of a cell towards a chemoattractant is called chemotaxis. For *Escherichia coli*, which moves with the help of flagellas, the signal transduction is relatively well known [30]. In [1] it is shown that chemotaxis in *Escherichia coli* is robust against large changes in parameters. By analyzing the model of [1], [45] were able to show that this robustness is achieved via integral control: A simplified description is that the integral of the cell-tumbling frequency is compared to the stimulus by the chemoattractant, see left part in Figure 8. This is equivalent to saying that the cell is tracking the time derivative of the stimulus. For a moving cell, this is an approximation of measuring the chemoattractant's concentration at several locations at the same time for finding the optimal direction. This equivalent representation is shown in the right part of Figure 8.

Unlike classical control problems, in physiological systems like cells there is usually no clear separation between controller and controlled system. Instead, regulation is often "built into" the system [24]. Nevertheless, similar behaviors can be observed. But analyzing a system with embedded controllers is more complicated than a classical control loop.

Synthesis of simple regulatory circuits is already possible [35]. For example, [13] designed cells as thermometers and [11] oscillators which are independent of the cell cycle. Besides the pure regulatory task, the cell also needs some kind of safety mechanisms, as do man-made control systems [30]. Such mechanism can yield robustness against environmental or intercellular variability[35].

A possible application of biosystems engineering is to identify and characterize the regulatory elements present in the cell's signaling pathways by comparing their transfer function with known transfer functions from technical systems. Moreover, if the behavior of the controlled system and the overall behavior is identified, system theory will allow to identify the behavior of the controller and this may help biologist to elucidate new control schemes.

4.2 Model reduction

In the past years, the models of certain cellular pathways have grown to several dozen states, e.g. [12, 44]. These first principle models can be used for simulations, but for analysis purposes, they are often too complex to be understood. Particularly their dynamic behavior is complex due to the interlaced regulatory loops. Smaller models can be found by modeling the system less precisely, e.g. by combining intermediate products to a single state, which requires a-priori knowledge. A further idea, presented in [28] uses the hierarchical structure of the system to come to a reduced model describing gene expression in a multigen network. Or, the complete model can be used to find a reduced model using mathematical techniques. A reduced-order continuous-time or Boolean automata model can either be found by an algorithm or as a consequence of simplifying assumptions. In [41] an example for a simple gene regulatory network can be found.

In many cases, a modular approach in modeling is desired, as was pointed out in section 2. This is for example the case if a certain part should be modeled exactly with all participating metabolites, but other parts by approximate models. Using the model in [44] for analyzing the lactose uptake, the regulatory pathways of the other carbohydrates could be simplified. A possible approach to perform a modular model reduction approach is to use a cluster analysis algorithm as in Section 2.2. This yields a partitioning of the model into parts with similar activity. Each activity type could be modeled as a state of an automata and a dynamical system for modeling the dynamic behavior of the states of the full system, i.e. the integral behavior of I-states. In the case of the model of four carbohydrates [44], the automata could have as states the presence of these carbohydrates. In comparison to classical model reduction algorithms, the resulting model is relatively close to a biological first principle model.

5. CONCLUSION

For a better understanding of complex biological systems tools and methods from systems theory should be used which will offer new possibilities in analysis and design of complex biological reactions schemes. On the other hand, for systems theory, the application to a relatively new field may also be a challenge. There seems no doubt that only with suitable mathematical models a better understanding of biological systems could be achieved. Hence, actual research will concentrate in this field. The crucial point of further activities will be a substantial support of biologists work. Only if biological researchers will occupy help from systems theory in analyzing a network or designing an experiment, biosystems engineering/ systems biology will succeed.

Although the presented contribution reviews a number of applications of methods from systems theory, an extensive evaluation is still missing. The time to set up a flat model is very short in comparison with the time to set up a structured model which is validated with wild type strains and mutant strains. This shortcut is not only due to the missing of suitable models but also due to the lack of consistent quantitative experimental data. Methods described in section 3 and 4 are far from application to complex cellular systems and this might also be a challenge for systems theory.

6. ADDITIONAL AUTHORS

M. Ederer, email address: koehler6@isr.uni-stuttgart.de, F. Allgöwer, email: allgower@ist.uni-stuttgart.de and E.D. Gilles, email: gilles@mpi-magdeburg.mpg.de.

7. REFERENCES

- [1] U. Alon, M. G. Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. *Nature*, 397:168–171, 1999.
- [2] O. Alter, P. O. Brown, and D. Botstein. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci.*, 97(18):10101–10106, 2000.
- [3] J. Bailey. Mathematical modeling and analysis in biochemical engineering: Past accomplishments and future opportunities. *Biotech. Prog.*, 14:8–20, 1998.
- [4] M. Baltes, R. Schneider, C. Sturm, and M. Reuss. Optimal experimental design for parameter estimation in unstructured growth models. *Biotech. Prog.*, 10:480–488, 1994.
- [5] M. Berridge. Inositol triphosphate and calcium signalling. *Nature*, 361:315–25, 1993.
- [6] H. Bolouri, A. Finney, M. Hucka, H. Sauro, J. Doyle, and H. Kitano. The ERATO systems biology workbench: an integrated environment for multiscale & multi-theoretic simulations of molecular biology. In *Proceedings of the First International Conference on Systems Biology*, page 25, Tokyo, Japan, 2000.
- [7] T. Chen, V. Filkov, and S. S. Skiena. Identifying gene regulatory networks from experimental data. *Parallel computing*, 27:141–162, 2001.
- [8] M. Domach, S. Leung, R. Cahn, G. Cocks, and M. Shuler. Computer model for glucose-limited growth of a single cell of *Escherichia coli* B/r-A. *Biotech. Bioeng.*, 26:203–216, 1984.
- [9] J. S. Edwards, R. U. Ibarra, and B. O. Palsson. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, 19:125–130, 2001.
- [10] J. S. Edwards and B. O. Palsson. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci.*, 97:5528–5533, 2000.
- [11] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

- [12] M. Fussenegger, J. E. Bailey, and J. Varner. A mathematical model of caspase function in apoptosis. *Nature Biotech.*, 18(7):768–774, 2000.
- [13] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403:339–342, 2000.
- [14] B. Genomics. Internet: <http://www.beyondgenomics.com/>.
- [15] B. Grünfelder, G. Rummel, J. Vohradsky, D. Röder, H. Langen, and U. Jenal. Proteomics analysis of the bacterial cell cycle. *Proc. Nat. Acad. Sci. USA*, 98(8):4681–86, 2001.
- [16] V. Hatzimanikatis, L. Choe, and K. Lee. Proteomics: theoretical and experimental considerations. *Biotech. Prog.*, 15:312–18, 1999.
- [17] V. Hatzimanikatis, M. Emmerling, U. Sauer, and J. Bailey. Application of mathematical tools for metabolic design of microbial ethanol production. *Biotech. Bioeng.*, 58:154–161, 1998.
- [18] J. W. Hearne. Sensitivity analysis of parameter combinations. *Appl. Math. Modelling*, 9:106–108, 1985.
- [19] N. S. Holter, A. Maritan, M. Cieplak, N. V. Fedoroff, and J. R. Banavar. Dynamic modeling of gene expression data. *Proc. Natl. Acad. Sci.*, 98(4):1693–1698, 2001.
- [20] N. S. Holter, M. Midra, A. Maritan, M. Cieplak, J. R. Banavar, and N. V. Fedoroff. Fundamental patterns underlying gene expression profiles: Simplicity from complexity. *Proc. Natl. Acad. Sci.*, 97(15):8409–8414, 2000.
- [21] L. Hood. Unifying logic: Searching for the biggest truths in the smallest elements. Internet: http://www.geneforum.org/learnmore/articles/hoodl_200101.cfm, January 2001.
- [22] T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292:929–9334, 2001.
- [23] J. Keener and J. Sneyd. *Mathematical Physiology*, volume 8 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, Ney York, second edition, 2001.
- [24] M. C. K. Khoo. *Physiological Control Systems — Analysis, Simulation and Estimation*. IEEE Press Series in Biomedical Engineering. IEEE Press, Piscataway, NJ, 2000.
- [25] H. Kitano. Perspectives on systems biology. *New Generation Computing*, 18(3):199–216, 2000.
- [26] S. Klamt, S. S., and G. E.D. Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotech. Bioeng.*, 2001. accepted.
- [27] K. W. Kohn. Molecular interaction maps as information organizers and simulation guides. *Chaos*, 11(1):84–97, 2001.
- [28] A. Kremling and E. Gilles. The organization of metabolic reaction networks: II. Signal processing in hierarchical structured functional units. *Metabolic Engineering*, 3(2):138–150, 2001.
- [29] A. Kremling, K. Jahreis, J. Lengeler, and E. Gilles. The organization of metabolic reaction networks: A signal-oriented approach to cellular models. *Metabolic Engineering*, 2(3):190–200, 2000.
- [30] D. A. Lauffenburger. Cell signaling pathways as control modules: Complexity for simplicity? *Proc. Natl. Acad. Sci.*, 97(10):5031–5033, 2000.
- [31] J. Lengeler. *Regulation of Gene Expression in Escherichia coli*, chapter The phosphoenolpyruvate-dependent carbohydrate phosphotransferase system (PTS) and control of carbon source utilisation. R.G. Landes Co., 1995.
- [32] J. Lengeler, H. Schlegel, and G. Drews, editors. *Biology of the Prokaryotes*. Thieme Verlag, Stuttgart, Blackwell Science Inc., Oxford, 1999.
- [33] L. Ljung. *System Identification – Theory for the user*. Prentice Hall PTR, Upper Saddle River, New Jersey, second edition, 1999.
- [34] H. McAdams and A. Arkin. Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.*, 27:199–224, 1998.
- [35] H. H. McAdams and A. Arkin. Gene regulation: Towards a circuit engineering discipline. *Current Biology*, 10:R318–R320, 2000.
- [36] M. Morohashi and H. Kitano. Identifying gene regulatory networks from time series expression data by *in silico* sampling and screening. In *Proceedings of the 5th European Conference on Artificial life*, pages 477–486, Lausanne, Switzerland, 1999. Springer.
- [37] A. Munack. Some improvements in the identification of bioprocesses. In M. N. Karim and G. Stephanopoulos, editors, *Modeling and control of biotechnical processes 1992*, IFAC Symposia series, pages 89–94. IFAC, Pergamon Press, 1992.
- [38] J. Nielsen and J. Villadsen. Modelling of microbial kinetics. *Chemical Engineering Science*, 47:4225–4270, 1992.
- [39] C. Posten and A. Munack. On-line application of parameter estimation accuracy to biotechnical processes. In *Proceedings of the American Control Conference*, volume 3, pages 2181–2186, 1990.
- [40] J. Shu and M. Shuler. A mathematical model for the growth of a single cell of *E. coli* on a glucose/glutamine/ammonium medium.
- [41] R. Somogyi and C. Sniegoski. Modeling the complexity of gene networks: understanding multigenic and pleiotropic regulation. *Complexity*, 1:45–63, 1996.

- [42] G. Stephanopoulos, D. Hwang, J. Misra, W. Schmitt, and G. Stephanopoulos. Functional genomics and systems engineering. In G. Stephanopoulos, J. Lee, and E. Yoon, editors, *6th IFAC Symposium on Dynamics and Control of Process Systems (Pre-Print)*, pages 16–25, Jeju Island, Korea, June 2001.
- [43] B. Turner, D. Ramkrishna, and N. Jansen. Cybernetic modeling of bacterial cultures at low growth rates: Mixed substrate systems. *Biotechnol. Bioeng.*, 32:46–54, 1988.
- [44] J. Wang. *A Systematic Modeling Approach for Cellular Systems and its Application on Transport and Catabolism of Four Carbohydrates in Escherichia coli*. PhD thesis, Universität Stuttgart, 2001.
- [45] T.-M. Yi, Y. Hunag, M. I. Simon, and J. Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl. Acad. Sci.*, 97(9):4649–4653, 2000.
- [46] L. You and J. Yin. Patterns of regulation from mRNA and protein time series. *Metabolic Engineering*, 2(3):210–217, 2000.

APPENDIX

A. METHOD OF HEARNE

With sensitivities ω_{ij}

$$\omega_{ij} = \frac{\partial x_i p_j}{\partial p_j x_i}, \quad \Omega = \begin{bmatrix} \omega_{11} & \cdots & \omega_{1m} \\ \cdots & \cdots & \cdots \\ \omega_{n1} & \cdots & \omega_{nm} \end{bmatrix}, \quad (9)$$

the vector of sensitivities \mathbf{s}

$$\mathbf{s} = \left[\frac{\Delta p_1}{p_1} \cdots \right], \quad (10)$$

the optimization problem can be formulated as

$$\max \mathbf{s}^T \int_0^T \Omega^T \Omega dt \mathbf{s}. \quad (11)$$

With scaling of the sensitivities

$$\mathbf{s}^T \mathbf{s} = 1 \quad (12)$$

the problem can be reformulated as an Euler-Lagrange equation with the solution

$$\mathbf{G} \mathbf{s} = \lambda \mathbf{s}, \quad (13)$$

with

$$\mathbf{G} = \int_0^T \Omega^T \Omega dt. \quad (14)$$

B. MODEL A AND MODEL B

The following model is used:

$$\begin{aligned} \dot{c}_X &= (\mu - q_{in}/V) c_X && \text{biomass} \\ \dot{c}_S &= q_{in}/V (c_S^{in} - c_S) - r_1 M c_X && \text{substrate} \\ \dot{c}_{S1} &= r_1 - r_2 - \mu c_{S1} \\ \dot{c}_{S2} &= r_2 - r_3 - \mu c_{S2} \\ \dot{c}_E &= r_s - \mu c_E, \end{aligned} \quad (15)$$

with $\mu = Y_{x/s} r_1$, volume V , molecular weight M , yield $Y_{x/s}$. The following reaction rates are used for model A:

$$\begin{aligned} r_1 &= r_{1max} \frac{c_S}{K_S + c_S} \\ r_2 &= r_{2max} \frac{c_S}{K_S + c_S} \frac{K_{I1}}{K_{I1} + c_{S2}} \\ r_3 &= r_{3max} \frac{c_{S2}}{K_{S2} + c_{S2}} \\ r_s &= r'_{smax} \end{aligned} \quad (16)$$

and for model B

$$\begin{aligned} r_1 &= r_{1max} \frac{c_S}{K_S + c_S} \\ r_2 &= r'_{2max} \frac{c_S}{K_S + c_S} \\ r_3 &= r_{3max} \frac{c_{S2}}{K_{S2} + c_{S2}} \\ r_s &= r_{smax} \frac{K_{I2}}{K_{I2} + c_{S2}} \end{aligned} \quad (17)$$