ON DERIVING A HYBRID MODEL FOR CARBOHYDRATE UPTAKE IN ESCHERICHIA COLI

Eric Bullinger * **Thomas Sauter** ** **Frank Allgöwer** *
Ernst-Dieter Gilles **

* *Institut für Systemtheorie technischer Prozesse, Universität Stuttgart, 70550 Stuttgart, Germany*
`{bullinger,allgower}@ist.uni-stuttgart.de`
** *Institut für Systemdynamik und Regelungstechnik, Universität Stuttgart, 70550 Stuttgart, Germany*
`{sauter,gilles}@isr.uni-stuttgart.de`

Abstract: In the past years, several large dynamical models of cellular metabolism have been published. Even more are going to come on the way towards models of the complete cellular metabolism. As the models are nonlinear, analyzing and understanding them is a not trivial task. This paper proposes a method to finding approximate models that explicitly take into account the switching behavior inherently present in several parts of the cellular metabolism. First, the states are separated into functional units based on similar "activity" along suitable trajectories. Important are the choices of sufficiently stimulating trajectories and of when a state is called active. Secondly, an automata model is build using the result of the first step. A model of carbohydrate uptake and the glycolysis is used to demonstrate the feasibility of the proposed method.

Keywords: model reduction, hybrid system, biochemical modeling.

1. MOTIVATION

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 (Lengeler *et al.*, 1999). 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 (Kremling *et al.*, 2000; Hood, 2001).

In the past years, the models of certain cellular pathways have grown to several dozen states (see e.g. Xu *et al.*, 1999; Ko *et al.*, 1994; Fussenegger *et al.*, 2000; Kremling *et al.*, 2001; Lee, 1984; Rizzi *et al.*, 1993; Rizzi *et al.*, 1997; Schöberl *et al.*, 2002; Shu and Shuler, 1989; Spiro *et al.*, 1997; Stelling and Gilles, 2000; Wang, 2001). These models can be used for simulations, but for analyzing their dynamic behavior, they are usually much too complex. Smaller models can be derived by modeling the system less precisely, e.g. by combining intermediate products to a single state, which requires a-priori knowledge. Or, the complete model can be used to find a reduced model using mathematical techniques where less a-priori knowledge is required.

Classically, biologists divide the cellular metabolism into functional units. For example, the glycolysis is the part responsible of breaking up carbohydrates, production of energy and of precursors fur synthesis. Several functional units of a cell behave in a switching manner.
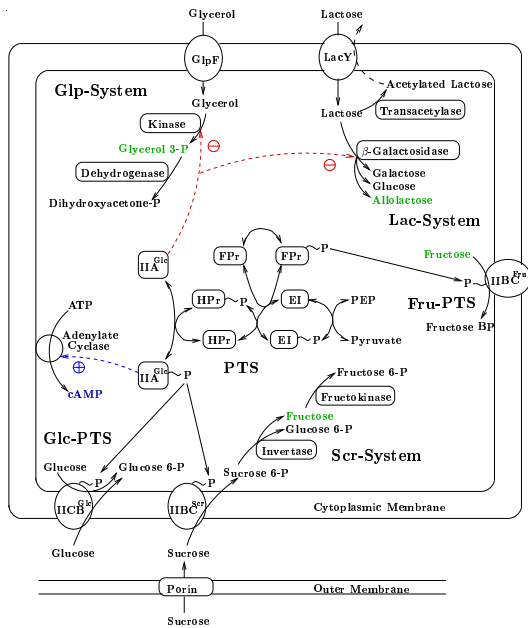
Fig. 1. Pathways of the uptake of glycerol, glucose, sucrose and lactose in *Escherichia coli* (Wang, 2001).

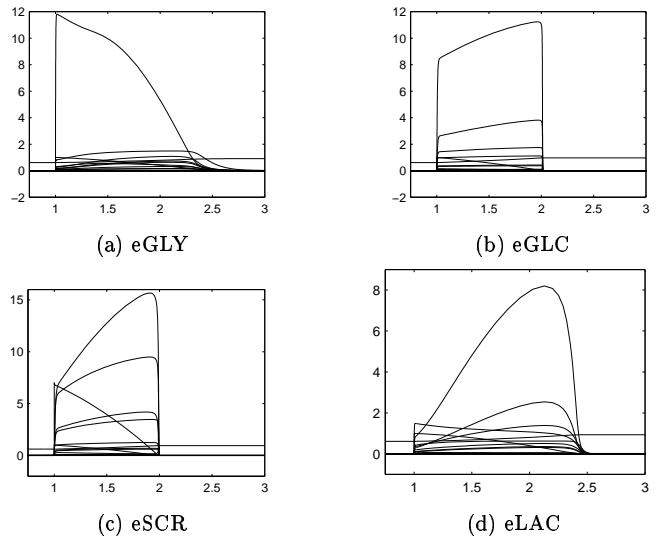
(a) eGLY

(b) eGLC

(c) eSCR

(d) eLAC

Fig. 2. Simulation of the cell with sudden increase from 0 to 1 g/l of glycerol (eGLY), glucose (eGLU), sucrose (eSCR) and lactose (eLAC) at 1h. The simulations show that certain metabolites react quickly to the external sugars. Also at the end of the uptake, fast decreases can be observed.

Typical examples are those involved in the cell cycle (see Keener and Sneyd, 2001) and the units responsible of the transport of certain substrates into the cell. (Varma *et al.*, 1993) uses a flux balance approach for determining optimal metabolic performance and shows that changes in metabolic pathway utilization occur under various oxygenation rates. A model of the uptake of the substrates glucose, saccharose, lactose and glycerol (Wang, 2001) consisting of 32 states (see Appendix A) is used to demonstrate the feasibility of the proposed approach. Figure 1 shows a sketch of the pathways involved. Figure 2 shows the response of the states of the cell's model to sudden increases of each one of the modeled external carbohydrates. It can be seen that certain proteins are quickly produced in response to the sudden presence of the carbohydrates at 1h. After the carbohydrate has been completely taken up, quick decreases are also observable, see e.g. the simulation with sucrose at 2h. A closer look reveals that each carbohydrate triggers a particular subset of the modeled states. This will be used to partition the states for getting a hybrid model.

The proposed approach is as follows. First, typical trajectories are simulated which should stimulate the system sufficiently. Then, linkage-methods from cluster-analysis subdivide the set of modeled state variables. This requires the definition of a suitable measure of distance. The result is a tree structure (dendrogram) 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. These groups are then used to generate a hybrid model in which inactive

states are kept constant. The paper shows that this approach leads to very satisfactory results.

This paper first presents in Section 2 a method for partitioning states into groups that behave similarly. Then Section 3 shows that this grouping can be used to extract reduced-order models that describe the cell's behavior at a certain mode of operation. The paper concludes with a discussion and outlook. Appendix A lists the states and functional units of the model of Wang (2001).

2. PARTITIONING OF THE STATES INTO FUNCTIONAL UNITS

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 differential-algebraic (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 (Ederer, 2001). Therefore the assignment of gene products was not possible.

Another approach is to use typical trajectories where all parts of the model are simulated. By correlating when the different variables are active, a good structuring of the model (Wang, 2001) was obtained (Ederer, 2001). The crucial parts of this algorithm are the choice of typical trajectories and the definition of activity. In this
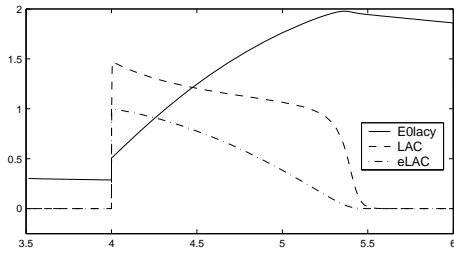
Fig. 3. Typical trajectories of P (LAC), I$_+$ (E0lacy) and I$_-$ (eLAC) variables; abscissa in [h], ordinate in [μmol/gDCW].
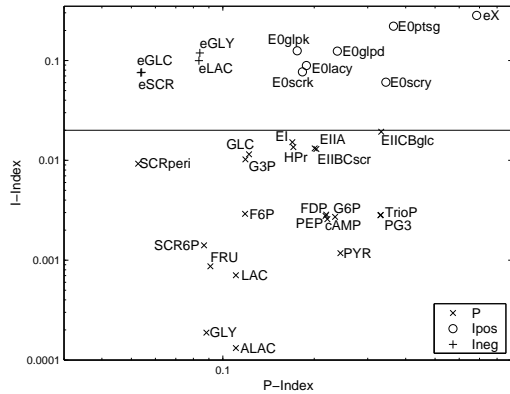

Fig. 4. From the simulated trajectories, P- and I-activities are calculated. They correspond to the proportion a certain state would be called "active" if it where a P- or an I-variable, respectively.

Fig. 6. Cross-correlation plot of the activity of the states. Black shows identical activity, white strongly different activity.

paper, similarly to (Ederer, 2001), three types of variables are defined. The variables are separated into three classes corresponding to their typical trajectories: I$_+$-variables (from integral) typically show flat ramps with different gradients in the observed time intervals (Fig. 3). They are considered to be active whenever its value is increasing. This is equivalent to the time derivative of the variable being positive, see Figure 3. Gene products can usually well be described by this type. Similarly, I$_-$-variables are active whenever they decrease. In contrast P-variables (from proportional) typically show steps, respectively steep ramps. Variables of this type are said to be active whenever their value is larger than a certain threshold around a nominal value. This is motivated by them being constant most of the time in the considered model, see Fig. 3. Metabolites seem to fit into this category. Other definitions of activity might be necessary, for example for signal substances which are active if their concentration is low. The time a certain variable is active in the P or I$_+$ sense is plotted in Fig. 4. All variables having a I-activity larger than 0.02 are then considered to be I-variables. In the discussed model, the external substrates are the only I$_-$-variables. All other I-variables are I$_+$-variables.

Using this definitions of activity, the distance between two states is calculated depending on how often the two states are "active" together considering a typical
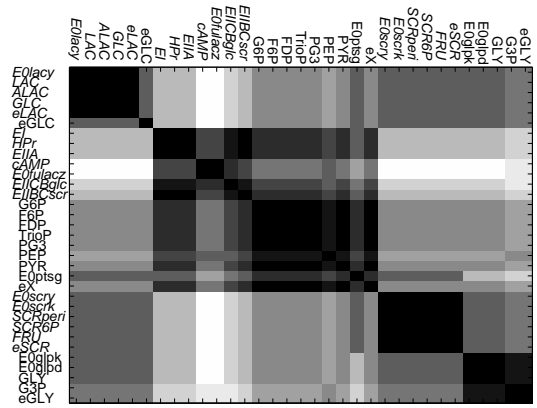
trajectory. Based on this measure of distance the tree structure can be calculated by the already mentioned linkage-methods. For a model of the transport systems of four carbohydrates (glucose, sucrose, lactose, glycerol) and the glycolysis (Wang, 2001) the resulting tree structure is shown (Fig. 5, page 4). All biologically defined functional units (phosphotransferase system, *glp* regulon, glycolysis, *lac* operon, *scr* regulon), see also Appendix A, were found by this formal algorithm. For example for the PTS the appertaining modeled states EI, HPr, EIIA EIICBglc, EIIBCscr are linked by U-shaped lines of very small height. The height of the U-shaped lines represent the distance between the linked states. The various PTS-proteins are therefore found to be very close to each other and clearly separated from the other units. The same can be seen for the other units.

Another way of analyzing the distance between the activities of the different states is to plot their cross-correlation, see Figure 6: The darker an entry, the more the two corresponding states behave similarly.

Both the dendrogram (Figure 5) and the correlation plot (Figure 6) show that the metabolites can be split into two main parts, the PTS and glycolysis on the one hand and the substrate transport units on the other. In the dendrogram, PTS and glycolysis form the lower half of the diagram, in the correlation plot their cross-correlation is the central dark square. The analysis results are very close to classical grouping into functional units, but for example the total concentration of glucose transporter PtsG (E0ptsg) seems to behave more like the states of the glycolysis than like the external glucose concentration (eGLU).

This new approach offers an opportunity to review biologically motivated structures. The algorithm will be tested on other models. In further applications it can be used to structure complex metabolic networks with unknown or uncertain structure and to make such networks amenable to modular modeling. A similar approach was used for clustering of cell cycle gene products of *Caulobacter crescentus* based on gene expres-
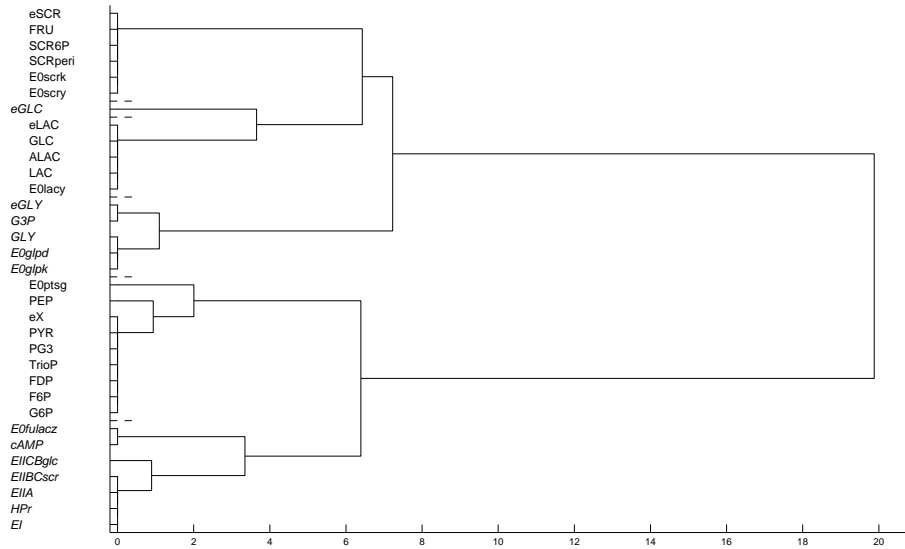
Fig. 5. Dendrogram generated by cluster analysis of model of the transport systems of four carbohydrates (glucose, sucrose, lactose, glycerol) and the glycolysis. The groups are separated by a short horizontal line and furthermore in italic or normal font shape. Some results, e.g. the assignment of the total concentration of the glucose transporter PtsG (E0ptsg) to glycolysis are unusual in respect of the traditional classification and are subjected to further investigations.
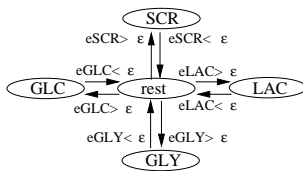

Fig. 7. Automata part of the hybrid model of substrate uptake. The transitions occur whenever the concentration of a carbohydrate crosses the level ϵ.

sion data received by two-dimensional gel electrophoresis (Grünenfelder *et al.*, 2001). As shown, this kind of clustering does not depend on using experimental data but can also be done based on simulated data (typical trajectories). The application of the concept of activity (on/off-behavior) and of different types of variables, as used in this section, enables more accurate structuring.

These structural information can be used for model reduction, as shown in the following section.

3. HYBRID MODEL

Hybrid dynamical systems can be described by a combination of an automata model and a continuous-time state-space model (van der Schaft, 2000). The continuous-time part can be a differential-algebraic (DAE) model.

If the presence of several external carbohydrates together is not taken into account, the cell model of Wang (2001) has five characteristic modes: the presence of glucose, saccharose, lactose or glycerol or of none of them. This is depicted in Figure 7. This figure depicts the modes of the used automata model. The transitions from

one mode to another occur whenever the concentration of the external metabolite crosses a level ϵ.

The dynamics corresponding to each mode of the automata are as follows (see Appendix A for the abbreviations):

rest:	All states are kept constant
GLC:	eGLC, PTS and Glycolysis are on, the others constant;
SCR:	eSCR, scr-regulon, PTS and Glycolysis are on, the others constant;
GLY:	eGLY, glp-regulon, PTS and Glycolysis are on, the others constant;
LAC:	eLAC, lac-operon, PTS and Glycolysis are on, the others constant;

This results both from the analysis of Section 2 as well as of biological knowledge as the gene expression (the activation of regulons and operons) is regulated by proteins of the corresponding pathways. For example, the glp-regulon is activated by molecular inductor Glycerol-3-P which is correlated to internal glycerol.

The simulation of the hybrid model with the same stimulations as in Figure 2 is shown in Figure 8. It looks very similar to the simulation of the full model, Figure 2. The differences in the first phase of the carbohydrate uptake are due to the fact that the hybrid model assumes the concentration of enzymes to remain constant while the full model includes decay rates for them, see Figure 9. The large differences at the end of the uptake periods are mainly due to small differences in the timing of when the P-variables go back to their nominal value, see especially the glucose and sucrose simulations.

These differences are due to small changes in states the hybrid model keeps constant. These difference are relevant in the initial phases of the carbohydrate uptake

(a) eGLY   **(b) eGLC**
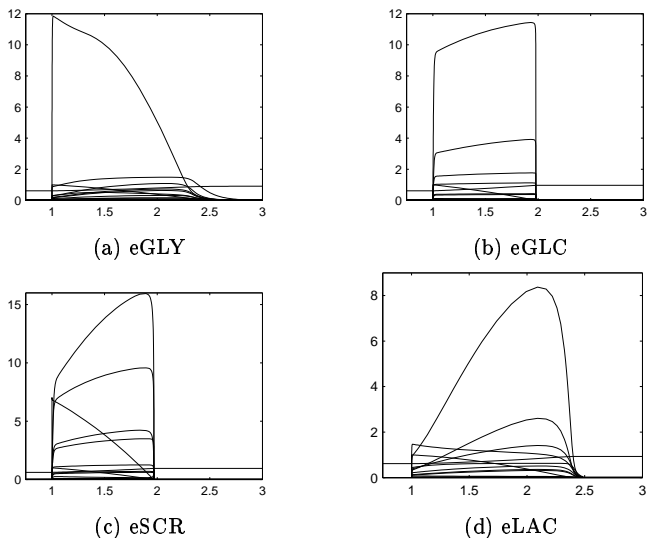
(c) eSCR   **(d) eLAC**

Fig. 8. Simulation of the hybrid model of the cell with sudden increase of glycerol, glucose, sucrose and lactose at 1h; abscissa in [h], ordinate in $[\mu\text{mol/gDCW}]$.

(a) eGLY   **(b) eGLC**
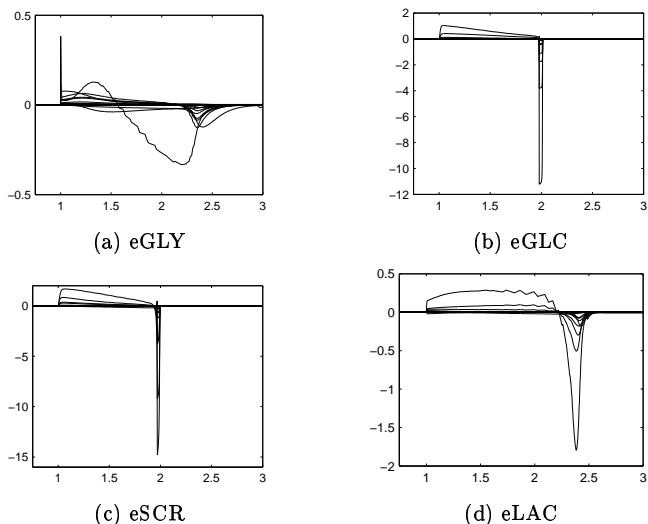
(c) eSCR   **(d) eLAC**

Fig. 9. Differences between the simulations of the full and of the hybrid model, Figure 2 and 8; abscissa in [h], ordinate in $[\mu\text{mol/gDCW}]$

and when other states have almost converged to low values. Then the small errors in enzyme concentration influence at what moment the metabolites of the glycolysis decrease, causing the sharp peaks in Figure 9.

Besides demonstrating the switching character of the cell model, the hybrid model can be used as a reduced model. The table below shows that during each period, more than a third of the states have been neglected:

Input	Neglected states
Glycerol (GLY)	13 of 32
Glucose (GLC)	17 of 32
Sucrose (SCR)	12 of 32
Lactose (LAC)	13 of 32

The small approximation error (Figure 9) shows that the neglected states do not play an important role for the modes their functional unit is not needed.

4. DISCUSSION

The proposed hybrid model shows that "unused" functional units of carbohydrate transport can be neglected. In this paper, only one sort of carbohydrate was presented to the cell at a time. The case of several simultaneous carbohydrates can easily be incorporated by introducing corresponding states in the automata model. Another improvement would be to replace that "unused" enzymes are constant with some nominal decay rate.

The dynamic models used in the hybrid simulation still have about 2/3 of the original states. Further reduction can be achieved by performing a model reduction of the remaining dynamics. Possibly, a separate reduced model of the common part (glycolysis and PTS) for each mode increases the quality of the approximation.

5. CONCLUSION

Many algorithms cannot be used for systems that are too large or that are nonlinear. Also, analyzing and understanding large systems is not an easy task. In this paper, an approach towards finding reduced hybrid models approximating continuous-time dynamical systems has been presented. In a first step, functional units are determined based on similar behavior of states along suitably chosen trajectories. These trajectories need to stimulate all parts of the systems sufficiently. An important point of this phase is the definition of an activity. States are active if either they are larger than a certain threshold (P-variable) or if their derivative is larger (I_+) or smaller (I_-) than a threshold. Using the activity over time instead of the trajectories themselves increases quality of the separation into functional units. In a second step, a hybrid model is build using the fact that certain functional units are only active when an external carbohydrate is present. Other functional units are needed independent of what carbohydrate is present. The simulation results show that the hybrid model is a good approximation of the system.

ACKNOWLEDGEMENT

Appendix A. FUNCTIONAL UNITS

The states of the model by (Wang, 2001) discussed in this paper are usually partitioned into the following groups (PTS contains only non-phosphorylated species):

Glycolysis:	• glucose-6-phosphate (G6P) • fructose-6-phosphate (F6P) • fructose-1,6-diphosphate (FDP) • triose-phosphate (TrioP) • 3-phospho-glycerate (PG3) • phosphoenolpyruvate (PEP) • pyruvate (PYR)
Phospho-transferase (PTS):	• proteinkinase enzyme I (EI) • phospho-carrier (HPr) • glucose-specific phospho-carrier (EIIA) • glucose-specific PTS permease (EIICBglc) • sucrose-specific PTS permease (EIIBCscr)
glp-regulon:	• glycerol kinase (E0glpK) • glycerol-3-phosphate dehydrogenase (E0glpD) • glycerol (GLY) • glycerol-3-phosphate (G3P)
lactose operon:	• galactosidase permease (E0lacY) • lactose (LAC) • allolactose (ALAC) • glucose (GLC)
sucrose regulon:	• sucrose porin (E0scrY) • sucrose kinase (E0scrK) • periplasmic sucrose (SCRperi) • sucrose-6-phosphate (SCR6P) • fructose (FRU)
Extracellular:	• glycerol (eGLY) • lactose (eLAC) • sucrose (eSCR) • glucose (eGLC)
Varia:	• biomass (eX) • adenosine cyclic-monophosphate (cAMP) • glucose-specific PTS permease (E0ptsg)

Appendix B. REFERENCES

Ederer, M. (2001). Gliederung des zellulären Metabolismus in Funktionseinheiten mittels mathematischer Methoden. Master's thesis. Universität Stuttgart.

Fussenegger, M., J. E. Bailey and J. Varner (2000). A mathematical model of caspase function in apoptosis. *Nat. Biotechnol.* **18**(7), 768–774.

Grünenfelder, B., G. Rummel, J. Vohradsky, D. Röder, H. Langen and U. Jenal (2001). Proteomics analysis of the bacterial cell cycle. *Proc. Nat. Acad. Sci. USA* **98(8)**, 4681–86.

Hood, L. (2001). Computing life: The challenges ahead. *IEEE Engineering in Medicine and Biology Magazine* **20**(4), 20.

Keener, J. and J. Sneyd (2001). *Mathematical Physiology*. Vol. 8 of *Interdisciplinary Applied Mathematics*. second ed.. Springer-Verlag. New York.

Ko, Y.-F., W.E. Bentley and W.A.Weigand (1994). A metabolic model of cellular energetics and carbon flux during aerobic *Escherichia coli* fermentation. *Biotech. Bioeng.* **43**, 847–55.

Kremling, A., K. Bettenbrock, B. Laube, K. Jahreis, J.W. Lengeler and E.D. Gilles (2001). The organisation of metabolic reaction networks: III. Application for diauxic growth on glucose and lactose. *Metabol. Eng.* **3**(4), 362–379. Accepted.

Kremling, A., K. Jahreis, J.W. Lengeler and E.D. Gilles (2000). The organization of metabolic reaction networks: II. A signal-oriented approach to cellular models. *Metabol. Eng.* **2**(3), 190–200.

Lee, S.B. und J.E. Bailey (1984). Genetically structured modells for lac promoter-operator function in the *Escherichia coli* chromosome and in multicopy plasmids: lac operator function. *Biotech. & Bioeng.* **26**, 1372–82.

Lengeler, J.W., Schlegel, H.G. and Drews, G., Eds. (1999). *Biology of the Prokaryotes*. Thieme Verlag, Stuttgart, Blackwell Science Inc., Oxford.

Rizzi, M., M. Baltes, U. Theobald and M. Reuss (1997). In vivo analysis of metabolic dynamics in saccharomyces cerevisiae: II. mathematical model. *Biotech. & Bioeng.* **55**, 592–608.

Rizzi, M., U. Theobald, M. Baltes and M. Reuss (1993). Bioreactor and bioprocess fluid dynamics. In: *Measurements and Modelling of the dynamic glucose response of Saccharomyces cerevisiae in the time window of mixing times* (A.W. Nienow, Ed.). pp. 401–12. Mechanical Engineering Publications Ltd. London.

Schöberl, B., C. Eichler-Jonsson, E.D. Gilles and G. Müller (2002). Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.* Accepted.

Shu, J. and M.L. Shuler (1989). A mathematical model for the growth of a single cell of E. coli on a glucose/glutamine/ammonium medium. *Biotech. & Bioeng.* **33**, 1117–26.

Spiro, P.A., J.S. Parkinson and H.G. Othmer (1997). A model of excitation and adaptation in bacterial chemotaxis. *Proc. Natl. Acad. Sci. USA* **94**, 7263–7268.

Stelling, J. and E.D. Gilles (2000). Modular mathematical modeling of complex regulatory networks: Application to cell cycle regulation in saccharomyces cerevisiae. In: *Proceedings of the 4th Congress in Biochemical Engineering, Stuttgart, Germany*. Fraunhofer IRB Verlag. pp. 43–47.

van der Schaft, Arjan (2000). *An Introduction to Hybrid Dynamical Systems*. Lecture Notes in Control and Information Sciences. Springer. London.

Varma, A., B. Boesch and B. Palsson (1993). Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl. Env. Microbiol.* **59**, 2465–73.

Wang, J. (2001). A Systematic Modeling Approach for Cellular Systems and its Application on Transport and Catabolism of Four Carbohydrates in *Escherichia coli*. PhD thesis. Universität Stuttgart.

Xu, B., M. Jahic and S.-O. Enfors (1999). Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli. Biotechnol. Prog.* **15**, 81–90.