

**Tutorial summary:**

# Modeling and Simulation with the simulation environment PROMOT/DIVA

Tutorial given by: Martin Ginkel\* and Andreas Kremling<sup>†</sup>  
Estimated duration: 2 hours

Max-Planck-Institute for Dynamics of complex technical systems,  
Magdeburg, Germany

Email: {ginkel|kre}@mpi-magdeburg.mpg.de

August 7, 2001

## Introduction

The increasing knowledge in biology and improved measurement methods allow to build detailed models of the cellular interior. Due to the complexity of these models, tools to support researchers to set up models and to analyze them are needed. The tutorial aims to introduce the simulation environment PROMOT/DIVA (see also [5]), developed at the Max-Planck-Institute for Dynamics of complex technical systems, Magdeburg, and the University of Stuttgart, Germany. It is not only intended for experts in the field of systems biology but also as an introduction for biologists with some basic experience in mathematical modeling. The tutorial is divided in four parts:

Part 1 introduces at first a very simple example of a metabolic pathway. This example is used to explain the function of the modeling and simulation tool and the ideas behind the modeling concept and the modeling knowledge bases. The model setup is described either using a modeling language or graphical user interface. Part 2 focuses on the facilities to analyze the model and the available measurements. Besides dynamical simulation, an analysis of the sensitivities and the parameters is also possible. If measurements are available, parameters showing a high sensitivity could be estimated. The third part shows a more complex model describing carbohydrate uptake and metabolism for the bacterium *Escherichia coli* which is used in Part 4 to identify parameters. Furthermore an experimental design procedure offers the possibility to design experiments in such a way that yet insensitive parameters could be estimated.

## 1 Setup of a simple model with Promot

Within the modeling tool PROMOT [3] dynamic models of cellular networks can be constructed. Therefore modeling *knowledge bases* have been prepared, which contain basic and common complex modeling entities to let the

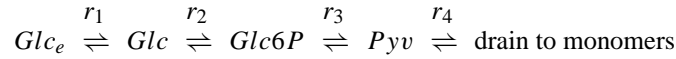
---

\*Computer scientist, working on the modeling tool and the modeling knowledge bases.

<sup>†</sup>Control engineer, developing math. models for *E. coli* and working on methods from system theory for application on biological systems

user build up such models more efficiently. The user can also extend the knowledge bases with his own generalized modeling entities. The modeling tool uses an object-oriented modeling methodology known from computer science. As in computer science it is used to prepare comprehensible and flexible modules that can be easily combined to form complex models.

First a simple example is shown and set up that will be analyzed in the later parts of the tutorial. The following reaction mechanism is considered:



Glucose in the liquid phase  $Glc_e$  is taken up by a transport step  $r_1$  and is metabolized in the subsequent reaction steps  $r_2, r_3$  and  $r_4$ . The growth rate  $\mu$  is calculated depending on the glucose uptake.

For all reaction rates  $r_i$  a Michaelis-Menten kinetic is chosen. The organisms are growing in a bioreactor (volume  $V_l$ ), which offers the possibility to control the system states with inflow rate  $q_{in}$  and substrate concentration  $Glc_{in}$  in the feed. The model comprises differential equations for all components as well as for biomass  $X$ . The effective equation system to calculate the system reads as:

$$\begin{aligned} \dot{Glc}_e &= \frac{q_{in}(Glc_{in} - Glc_e)}{V_l} - r_{1e} & r_{1e} &= r_1 X \frac{mw_{Glc}}{V_l} \\ \dot{X} &= \mu X - \frac{q_{in}X}{V_l} & \mu &= r_1 y_{Glc} \frac{mw_{Glc}}{V_l} \\ \dot{V}_l &= q_{in} - q_{out} \\ \dot{Glc} &= r_1 - r_2 - \mu Glc & r_1 &= v_{max1} \frac{Glc_e}{Glc_e + k_{s1}} \\ \dot{Glc6p} &= r_2 - r_3 - \mu Glc6p & r_2 &= v_{max2} \frac{Glc}{Glc + k_{s2}} \\ \dot{Pyv} &= r_3 - r_4 - \mu Pyv & r_3 &= v_{max3} \frac{Glc6p}{Glc6p + k_{s3}} \\ & & r_4 &= v_{max4} \frac{Pyv}{Pyv + k_{s4}}, \end{aligned}$$

with molecular weight for glucose  $mw_{Glc}$ , yield coefficient  $y_{Glc}$  and kinetic parameters  $v_{max_i}, k_{s_i}$ .

Goal of this tutorial part is to construct this simple model using the prepared modeling entities from the knowledge base. For an introduction the basic modules in the knowledge base and their respective interfaces will be discussed. This modules are substance storages, reactions (substance transformers) and signal transformers (see [7, 5] for details about the modeling ideas). For some simple models the attached behavioural (mathematical) model will be explained and also the way how these mathematical models are connected structurally to form differential algebraic simulation models.

With the help of the graphical user interface the substance storages and transformers for this simple pathway are selected from the knowledge base and then connected via their interfaces. Further a prepared container model for the biophase is set up that calculates growth rate  $\mu$  and biomass. To specify the behaviour of the mathematical model, these modules have to be parametrized, i.e. the initial state values and parameter values for the different states and parameters are set, using the modeling language MDL of PROMOT.

## 2 The Simulation Environment DIVA

PROMOT generates models for the simulation environment DIVA. Models for DIVA are formulated as standardized FORTRAN routines which can be used by a couple of numerical methods e. g.:

1. Dynamic simulation of the model with different integration algorithms suited for different model structures. A very efficient algorithm for DAE systems is DDASAC which is based on a paper of Caracotsios and Stewart [2]. Besides the system equations the algorithm also solves the sensitivity equations  $\frac{\partial x_i}{\partial p_j}$  for states  $x_i$  to parameter  $p_j$ .
2. Sensitivity analysis for parameters, also with respect to experimental data. The procedure is divided into two parts. (i) Detecting a set of parameters which has a maximal influence on selected states. Here, a method provided by Hearne [4] is applied. (ii) Based on these results available measurements have to be analyzed to check if they contain enough information for parameter estimation. An analysis based on the calculation of the Fisher information matrix is used to group such parameters which could be estimated together [8, 10].
3. Parameter identification according to experimental data. DIVA provides optimization algorithms from the NAG library [9].
4. Model-based experimental design. Often experiments are performed in such a way that they do not contain enough information to estimate parameters. This is the case in batch experiments when kinetic constants, e.g.  $K_s$  values for transport systems, have to be estimated. An experiment which shifts the sensitivities of the parameters must be designed. Different criteria are introduced in the literature. Here, we use the E-optimal design [1]. This criterion tries to minimize  $\lambda_{max}/\lambda_{min}$  with  $\lambda$  are the eigenvalues of the Fisher information matrix.

With the simple model introduced above some exercises are carried out, using the simulation tool DIVA and Matlab for the visualization of the simulation results. First a dynamic simulation is performed. The sensitivity of the model for these parameters then can be computed using the sensitivity calculating integration algorithms of DIVA. Using a set of given measurements it can be analyzed how sensitive the parameters of this model are with respect to the measured values i. e. which parameters can be identified best with the given measurements. When this parameters are sorted out the parameter identification itself can be performed, using the sequential quadratic programming algorithms of DIVA.

Normally the procedure is iterative: starting values of the parameters are often too far from their correct values. Hence, after a first run of the methods it is expected that only a small number of parameters could be estimated. So, two and more runs must be performed to get the maximal number of parameters which could be estimated.

## 3 Presentation of a Model for Catabolite Repression in *E. coli*

A more complex model describing carbohydrate uptake and metabolism for the bacterium *Escherichia coli* is introduced (for further information on the model equations see [6]). The complete model comprises 23 differential equations and has about 80 parameters.

Using this more complex example the possibilities of model exploration and documentation with the help of the modeling tool are presented. The central part is the regulation of the *crp* modulon that controls the uptake of glucose, lactose and glycerole. Regarding gene expression the model has a hierarchical structure where the expression of the transporter proteins for glucose and lactose is controlled by the regulator Crp and the small signaling molecule cAMP. Parts of the central catabolism, and a simple model for the drain to monomer synthesis are also covered by the model. The regulation inside the *crp* modulon determines that *E. coli* prefers glucose over lactose and shows a diauxic growth behaviour in presence of both sugars. For this more complex example some

higher structured modules are used to model signal transduction and regulation. An overview on the modular structure of the model is given and some important detail models, especially the main glucose uptake system (Phosphotransferase system PTS) will be shown within the graphical views of PROMOT.

## 4 Application of computational methods to the model for *Escherichia coli*

The methods for sensitivity and parameter analyses are applied for the model for two sets of data. The first set comprises 4 measurements (biomass, external glucose and lactose and the lactose splitting enzyme LacZ) while the second one comprises 6 measurements (in addition intracellular EIIA protein, extracellular cAMP). Based on the available measurement, a number of parameters for glucose and lactose uptake can be estimated.

Measurements are performed with a number of isogenic mutant strains. Each mutant is missing a gene in one of the relevant signal transduction pathways. For the following strains measurements are available: Wild type strain LJ110 (modified W3110), *Cya*<sup>-</sup>, *LacI*<sup>-</sup> and *PtsG*<sup>-</sup>.

## References

- [1] 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.
- [2] M. Caracotsios and W. E. Stewart. Sensitivity analysis of initial value problems with mixed odes and algebraic equations. *Computers and Chemical Engineering*, 9(4):350–365, 1985.
- [3] M. Ginkel, A. Kremling, F. Tränkle, M. Zeitz, and E. D. Gilles. Application of the process modeling tool promot to the modeling of metabolic networks. In I. Troch and F. Breitenecker, editors, *IMACS Symposium on Mathematical Modeling*, volume 2, pages 525–528, Vienna, Austria, 2000. ARGESIM.
- [4] J. W. Hearne. Sensitivity analysis of parameter combinations. *Appl. Math. Modelling*, 9:106–108, 1985.
- [5] J. Stelling, A. Kremling, M. Ginkel, K. Bettenbrock, and E. D. Gilles. Towards a virtual biological laboratory. In H. Kitano, editor, *Foundations of Systems Biology*. MIT Press, 2001. *in press*.
- [6] A. Kremling and E. D. Gilles. The organization of metabolic reaction networks: II. Signal processing in hierarchical structured functional units. *Metabolic Engineering*, 3(2):138–150, 2001.
- [7] A. Kremling, K. Jahreis, J. W. Lengeler, and E. D. Gilles. The organization of metabolic reaction networks: A signal-oriented approach to cellular models. *Metabolic Engineering*, 2(3):190–200, 2000.
- [8] L. Ljung. *System Identification – Theory for the user*. Prentice Hall PTR, Upper Saddle River, New Jersey, second edition, 1999.
- [9] NAG Ltd., Oxford, England, UK. *NAG Fortran Library Manual.*, 1993.
- [10] 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.