INTEGRATING WORKBENCH FOR MODELING AND NUMERICAL ANALYSIS OF CELLULAR SYSTEMS

Andreas Kremling * Martin Ginkel * Ernst Dieter Gilles *

* *Max-Planck-Institut für Dynamik komplexer technischer Systeme, Zenit-Gebäude Leipziger Str. 44, D-39120 Magdeburg, Phone: +49 391 6117 515, E-mail: kre@mpi-magdeburg.mpg.de*

Abstract With increasing knowledge in biology and improved measurement methods it becomes possible to build detailed models of the cellular interior. In this contribution a computer framework is described which allows to build, analyze and visualize modular structured metabolic models using a abstract and general modeling methodology. With this methodology reusable modeling entities are introduced which lead to the development of modeling knowledge-bases. Therefore the modeling tool PROMOT is introduced. For the numerical analysis and identification of the resulting models the simulation environment DIVA is described. For the management of the resulting data from measurements and simulations on different models management and visualization facilities are necessary which are available in the MATLAB tool CELLVIEW. The described tools are the first steps towards an integrated computer tool, a "Virtual Laboratory", which allows experiments analogous to experiments in a real laboratory. *Copyright © 2001 IFAC*

Keywords: computer-aided process modeling, knowledge base design, metabolic modeling,

1. INTRODUCTION

The increasing knowledge in microbiology and genetics and the availability of new techniques to measure intracellular dynamics via cDNA microarrays and proteomics offers the possibility to set up mathematical models, describing in detail the processes of signal transduction and signal processing as well as metabolic fluxes. In comparison to mathematical models used for process design and control, the new class of very detailled models is characterized by a high complexity, i.e. a great number of interconnected elements. Besides a need of a suitable modeling concept, there will also be a large number of new problems in computer science concerning data handling, automatically generation of model equation, dynamical simulation, model validation and visualization of simulation and experimental results.

A framework to set up mathematical models was introduced and is based on the definition of modeling objects (submodels) (Kremling *et al.*, 2000*b*). These modeling objects cover a broad range from single enzymatic reaction steps to rather complex structures, which are called operon and modulon in bacterial genetics (Neidhardt *et al.*, 1990). This contribution deals with two computational aspects in modeling cellular systems:

(1) the generation of mathematical model equations and
(2) model validation based on parameter estimation from available measurements.

A software environment, combining three tools, namely PROMOT, DIVA and CELLVIEW is discussed in this contribution. The environement was sucessfully applied for modeling carbohydrate uptake and metabolsim in *Escherichia coli* (Kremling *et al.*, 2000*a*). The process modeling tool PROMOT (Tränkle *et al.*, 1999; Tränkle, 2000) has been designed for the computer-aided modeling of chemical processes as well as for the implementation of knowledge bases that contain reusable modeling entities .

The differential-algebraic process models created with PROMOT are added to the model library of the simulation environment DIVA (Mohl *et al.*, 1997) by calling the DIVA code generator. (Köhler *et al.*, 1997). The numerical methods provided by DIVA may be applied to the numerical analysis, dynamic and steady state simulation, and identification of model parameters. The MATLAB application CELLVIEW allows a interactive access on measurements and simulation results stored in a file system.

2. MODEL SETUP WITH PROMOT

2.1 *Abstract Modeling Concept*

For the modeling of cellular systems the **Pro**cess **M**odeling **T**ool PROMOT is used. This tool makes it possible to model complex systems with differential and algebraic equations as well as with the mechanisms of object-oriented modeling including abstraction, encapsulation, aggregation, and inheritance. In PROMOT, process models are built by aggregating structural and behavioural modeling entities. Structural modeling entities represent subsystems and their topological structure. The dynamic and steady-state behaviour of the investigated process is made up by the behavioural modeling entities. Two types of structural modeling entities are distinguished: modules and terminals. Modules represent differential-algebraic process models and their submodels. Examples for modules in bioprocess engineering are process units (e. g. fermentation reactors), balanced volumes (e. g. phases) and functional units of the metabolism (e. g. glycolysis).

To achieve encapsulation, the modules must be separated from their environment. Connections across the system boundary are modeled by attaching terminals to a module. Via these terminals, the module may exchange material, momentum, energy, and information with other modules. There are two types of behavioural modeling entities: process variables and ordinary differential as well as algebraic equations. The aggregation of process variables and equations leads to a differential-algebraic model (DAE) for the considered module.

The modeling entities in PROMOT are organized as an object-oriented class hierarchy with multiple inheritance. Every entity in this hierarchy inherits all parts and attributes from their respective superclasses. With this method abstraction is possible and more general, reusable entities can be formed.

For simulation the modeling system forms a complete DAE by aggregating all equations from the different submodules together with coupling relations to form a compact, unstructured differential algebraic equation system that can be analyzed numerically in a very efficient way.

2.2 *Methodology for Cellular Systems*

Hundreds of enzyme catalysed reactions occur simultaneously in every living cell. However, cells are able to respond very quickly to changes of the environmental conditions by turning on or off metabolic pathways. It can be concluded that a powerful but sensitive control management with well balanced actions upon metabolic fluxes is realized within the cellular interior. Therefore, in microbiology, the thinking in units (describing a subset of metabolic processes) has become very popular and has resulted in the definition of subnetworks that are under control of a common regulator protein (Neidhardt *et al.*, 1990). This allows the cell to stop or to activate the biosynthesis of a large number of enzymes belonging to this subnetwork. As an example the *crp* modulon in *Escherichia coli* is considered. This well investigated subnetwork for transport and metabolism of carbohydrates is under control of the protein Crp. This protein is able to activate or to inhibit initiation of transcription of the corresponding genes.

This subnetwork can be decomposed into metabolic and regulatory functional units down to elementary modules. Important elementary modules are substance storages and substance transformers for the metabolic part of the reaction network and signal transformers for the regulatory network. Substance storages are used to represent two classes of substances: on one hand intermediates of the metabolism possessing no genetic information, like precursors and amino acids and on the other hand macromolecules, like proteins, DNA and RNA which do possess genetic information. Differential equation are used for the mathematical description of storages. Two ore more storages can be connected by a substance transformer representing a biochemical reaction. Transformers are looked at two complementary aspects:

(1) the representation of the stoichiometric structure of the reaction
(2) the reaction kinetics together with the participating and controlling ligands (substrates, activators and inhibitors).

These two aspects are investigated and described seperately and due to their complementary structure can be combined to form a complete transformer.

Since the understanding of signal transduction and processing is the key for describing the overall behaviour of cellular systems, these processes are described using signal transformers. An example for a signal transducing process is the initiation of transcription (first step of protein biosynthesis). In *E.coli* the activity of the activator Crp can be varried by the alarmone cAMP.In activated form Crp is able to bind to about 40 binding sites on the DNA. The distribution of Crp to the binding sites is described with a signal transformer called "coordinator".

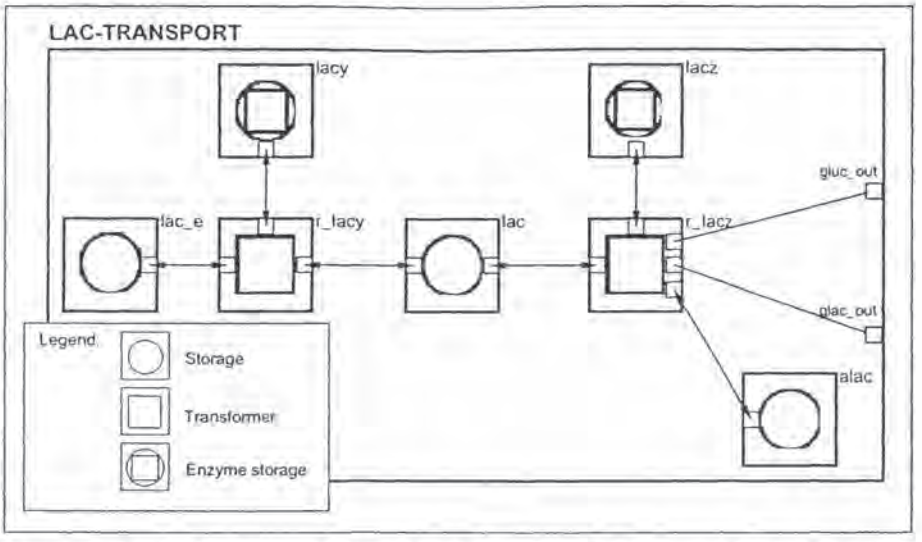

Figure 1. Structure of a simple subnetwork: lactose uptake and degradation. Lactose is taken up with the help of the enzyme lacy and is split into glucose and galactose by lacz. An important by-product is Allolactose.

Higher structured units like metabolic pathways can be aggregated from elementary modeling objects. As an example, Figure 1 shows the uptake and degradation of the carbohydrate lactose. External lactose from the medium (lacie) taken up by the enzyme LacY (lacy). Intracellular lactose (lac) is subsequently metabolized by the enzyme LacZ (lacz) to glucose and galactose. Since these metabolites belong to other pathways the terminals gluc_out and glac_out are provided to connect them to other units. The transformers r_lacy and r_lacz are named according to their enzymes. The production of the enzymes lacy and lacz is controlled by Crp and by the repressor LacI on superior levels of the metabolic network (gene expression of LacY and LacZ is not shown in the Figure 1). A by-product of this reaction is the metabolite allolactose (alac) which acts as an inducer for enzyme synthesis. This guarantees that the enzymes responsible for lactose degradation are synthesized only when lactose is present in the medium.

2.3 *Knowledge Base for Metabolic Models*

The knowledge base is designed in a bottom-up approach. At first, elemental modules like substance storages and aspects of transformers are defined. As mentioned above, the structure and the kinetics of transformers are represented separately by two distinct specialization hierarchies of modules in the knowledge base. The different kinds of modules can be combined by multiple inheritance in order to form many different transformers with specific reaction kinetics. E.g. the transformer TRANS2A_M1A, shown in Figure 2 is the combination of a stoichiometric structure TRANS2A with the kinetic module M1AT. This transformer is used in the example above (Figure 1) for r_lacy. The module TRANS2A represents

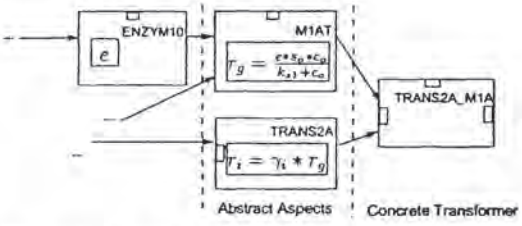

Figure 2. Part of the inheritance hierarchy for the transformer TRANS2A_M1A (most special class on the right side)

a transformer for two substances with stoichiometric calculations for consumption and production of the substances while M1AT is a kinetic module which calculates the reaction rate using a Michaelis-Menten kinetic under the control of one enzyme. There are other tranformer modules in the knowledge base which use an identical structure but adi fferent kinetic aspect. The proposed design of the knowledge base facilitates the integration of other kinetic aspects by recombining the structural aspect with another kinetic. Due to the stability of the interface defined by the terminals of the structural aspect module, models for transformers in a larger subnetwork can be easily replaced by a different version.

2.4 *Implementation of the modeling tool*

As mentioned above, PROMOT is amodel ing tool for equation-based and object-oriented modeling in the field of (bio-)process engineering. It provides a special modeling language as well as a graphical user interface (GUI) for interactive modeling. As shown in Figure 3 the kernel of the system is implemented as an modeling server in object-oriented Common Lisp.

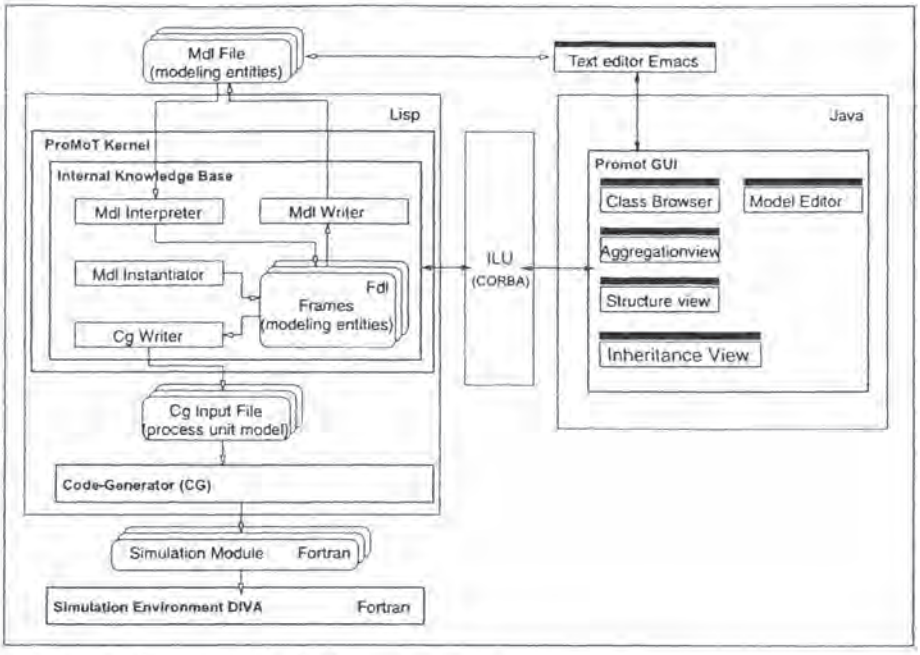

Figure 3. Software Architecture of PROMOT

The GUI is a client that is implemented in Java using the Java Foundation Classes (Swing).

The modeling language MDL (Model Description Language) of PROMOT is a declarative, object-oriented language that allows a symbolic implementation of variables and equations rather than the programming of imperative code. The system is able to read and also to write MDL thus it is used as the storage format for models and modeling knowledge bases. Because every aspect of the model can be described within MDL, the modeling language is the most powerful way to model in PROMOT. With the GUI users can explore and and manipulate the modeling entities by their graphical representation. Therefore views of the inheritance hierarchiy and the topology of submodules and their connections can be presented. These graphical representaions are very important especially for the communication in interdisciplinary teams, to have a common notion of the considered modeling entity. Besides that also graphical editing of the topological structure of modeling entities can be done interactively in a drawing-like form. In this way new higher structured modules can be created easily.F or changes on behavioural modeling entities the GUI launches a text editor in order to change the MDL source code of a single modeling entity. Thus modeling language and graphical editor can be used alternatively to change modeling entities from the user interface.

As mentioned above for the numerical analysis a compact DAE is generated from the structured representation within PROMOT. Prior to this step the designated model is checked for consistency and optimized to eliminate some redundant equations from the equation system. These equations occur as an overhead due to the connections in the structured model representation. Then PROMOT generates FORTRAN modules that can be used within the simulation environment DIVA. Therefore the Code Generator is invoked which translates the symbolic representaion of PROMOT to assignments in FORTRANand prepares the initialization of the sparse matrix numerics of DIVA.

3. NUMERICAL MODEL ANALYSIS

The numeric anlaysis of the models is done with the simulation environment DIVA (Kröner *et al.*, 1990; Mohl *et al.*, 1997). Within DIVA many different numerical computations are possible. For metabolic models 3 methods are of special interest:

(1) Dynamic simulation of the model (integration)
(2) Parameter analysis with respect to experimental data
(3) Identification of parameters

For the use in DIVA the linear implicit form form of the equation system reads:

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

$$\mathbf{x}(t_0) = \mathbf{x}_0 \quad (2)$$

$$\mathbf{y} = \mathbf{H}\cdot\mathbf{x} \quad (3)$$

In this notation $\mathbf{x}$ is the vector of states, $\mathbf{p}$ the vector of parameters, $\mathbf{u}$ the input vector, $\mathbf{y}$ the output vector and t is the simulation time. The matrix $\mathbf{B}$ is the descriptor matrix of the DAE and with the matrix $\mathbf{H}$ ($h_{i,j} \in \{0,1\}$) some of the states of the system are mapped as outputs of the system. For simulating the dynamic

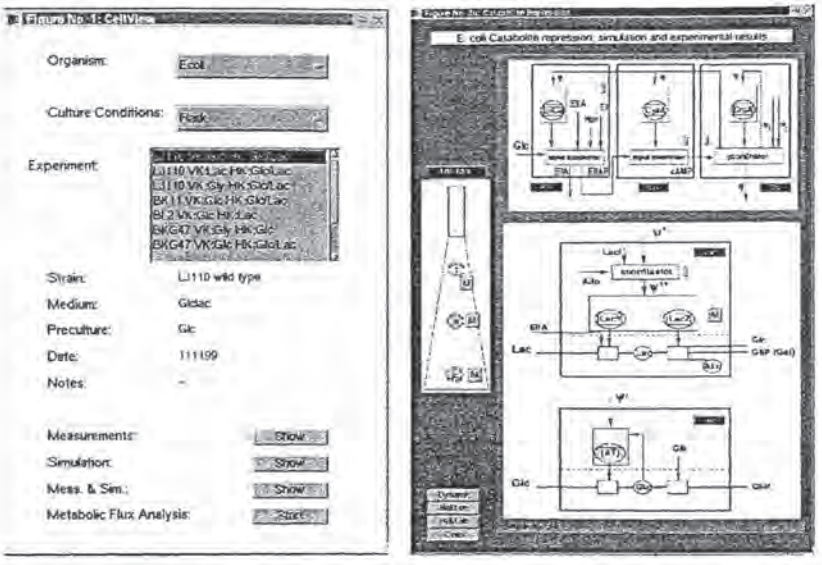

Figure 4. Left: Initial window of CELLVIEW to choose organim and experiment. Right: Individual window for *E. coli* catabolite repression. A click on the small buttons opens a windows with the correspond time course.

behaviour of the metabolic network under consideration a number of integrators for differential-algebraic equation systems can be used. The sensitivity analysis (Majer,19 98) is based on a version of the integration algorithm DDASAC, which calculates sensitivities of parameters with respect to some given states $S = \frac{\partial x}{\partial p}$ during integration.

To analyze the parameters a method using the Fisher information matrix (Ljung, 1999), which was introduced by Posten and Munack for biotechnological processes, is applied (Posten and Munack, 1990). With this method it is possible to determine such parameters which could be estimated according to the available mesurements. In the case of the model for carbohydrate uptake in *E. coli*, 12 from 80 parameters could be identified from a set of 9 experiments.

The parameter identification in DIVA is an optimization routine that is based on a sequential quadratic programming algorithm from the NAG numeric library. With this method unknown parameters of the model can be identified based on measured values from experiments. And in some cases the identification algorithm unveils, that it is inpossible to fit the parameters to the measurements and the structure of the model has to be revised.

4. VISUALIZATION OF EXPERIMENTAL AND SIMULATION DATA

CELLVIEW allows an interactiveacc ess on experimental and simulation data. All files are stored in a filesystem. The filesystem is organized by the following characteristics of an experiment: (i) organism (ii) culture condition (e.g. flask or bioreactor) (iii) strain (wild type and mutants) (iv) carbohydrate mixture in the medium and (v) preculture used. If e. g. an experiment with a LacI mutant was performed, using glucose and lactose in the medium and glucose as preculture, the data file has the following loacation **/ecoli/flask/LacI/glclac/glc010100.txt**. The experimental data are stored in a ASCII matrix where each column represents one measurement. Since the simulation environement provides data for use in MATLAB the corresponding file is named **glc010100_sim.mat** CELLVIEW possesses an interactive graphical user interface. It starts with a menue to choose an experiment according to the structure discribed above. Furthermore the user has the possibility to analyze experimental and/or simulation data, separately (see Figure 4).

To show the data, for every organism under investigation an individual window opens. The window shows the structure of the model by using the symbols introduced in the modeling framework. If measurement data for a substance storage are available the button is marked with a "M". In the example of carbon catabolite repression in *E. coli* four states could be measured: biomass, glucose and lactose in the medium and intracellular LacZ. For some ohter states in the model, a button to show the simulation results is also available. Another part of the in the described workbench is a tool for metabolic flux analysis that is also implemented in MATLAB (Klamt *et al.*, 2000). With this tool an analysis of the stoichiometric properties of metabolic networks is possible. The input for this tool can also be generated from PROMOT models.

5. CONCLUSIONS AND FUTURE WORK

In this contribution an overview was given for a workbench of software tools that support modeling and

numerical analysis of cellular systems. The modeling tool PROMOT provides an approved methodology and the possibility to use ready made modeling entities out of knowlegde-bases. Efficient modeling is supported by the use of a graphical user interface and a modeling language. Sophisticated methods for the numerical analysis of the resulting models are provided by the simulation environment DIVA. Besides dynamic simulation the identification of parameters and the calculation of sensitivites are possible. For managing and visualizing the large amounts of data produced in the virtual and real laboratory the prototype of an data mangement system was implemented which allows easy access and comparison of the results produced by measurements and simulation experiments.

The depicted goal of the described software tools is a "Virtual Laboratory" containing facilities for modeling, simulating and visualizing parts of intracellular metabolism. Therefore other computer aided methods have to be integrated. For raising amounts of knowledge the use of databases is necessary which allow the sharing of discoveries between workgroups at different locations connected by the internet.

For effective development of complex metabolic models the use of standardized modeling entities is only one aspect. Facilities which allow the visualization of different aspects of the resulting special models and support especially the debugging of models are under construction.

6. REFERENCES

Klamt, S., A. Kremling and E. D. Gilles (2000). Flux-analyzer: A graphical interface for stoichiometric and quantitative analysis of metabolic networks. Submitted for CAB8.

Köhler, R., S. Räumschüssel and M. Zeitz (1997). Code generator for implementing differential algebraic models used in the process simulation tool DIVA. In: *15th IMACS World Congress* (A. Sydow,Ed .). Vol. 3. Wissenschaft und Technik Verlag. Berlin, Germany. pp. 621–626.

Kremling, A., K. Bettenbrock, B. Laube, K. Jahreis, J.W. Lengeler and E.D. Gilles (2000*a*). The organization of metabolic reaction networks: III. Application for diauxic growth on glucose and lactose. *Metabolic Engineering*. submitted.

Kremling, A., K. Jahreis, J.W. Lengeler and E.D. Gilles (2000*b*). The organization of metabolic reaction networks: A signal-oriented approach to cellular models. *Metabolic Engineering*. In press.

Kröner, A., P. Holl, W. Marquardt and E.D. Gilles (1990). DIVA – An Open System for Dynamic Process Simulation. *Comput. Chem. Engng.* **14**, 1289–1295.

Ljung, L. (1999). *System Identification – Theory for the user*. Prentice Hall PTR. Upper Saddle River, New Jersey.

Majer, Clemens (1998). Parameterschätzung, Versuchsplanung und Trajektorenoptimierung für verfahrenstechnische Prozesse. PhD thesis. Universität Stuttgart. Düsseldorf.

Mohl, K. D., A. Spieker, R. Köhler, E. D. Gilles and M. Zeitz (1997). DIVA - A simulation environment for chemical engineering applications. In: *Informatics, Cybernetics and Computer Science (ICCS-97), Collected Volume of Scientific Papers*. pp. 8–15. Donetsk State Technical University, Donetsk, Ukraine.

Neidhardt, F.C., J.L. Ingraham and M. Schaechter (1990). *Physiology of the bacterial cell: A molecular approach*. Sinauer Associates. Sunderland, Massachusetts.

Posten, C. and A. Munack (1990). On-line application of parameter estimation accuracy to biotechnical processes. In: *Proceedings of the American Control Conference*. Vol. 3.

Tränkle, F., A. Kienle, K. D. Mohl, M. Zeitz and E. D. Gilles (1999). Object-Oriented Modeling of Distillation Processes. In: *ESCAPE-9, Suppl. to Comput. Chem. Engng.*. Vol. 23. Budapest, Hungary. pp. 743–746.

Tränkle, Frank (2000). Rechnergestützte Modellierung verfahrenstechnischer Prozesse für die Simulationsumgebung DIVA. PhD thesis. Universität Stuttgart. Düsseldorf.