SYSTEMATIC STRUCTURE AND PARAMETER IDENTIFICATION FOR BIOLOGICAL REACTION SYSTEMS SUPPORTED BY A SOFTWARE-TOOL

Jana Leifheit and Rudibert King [1]

Measurement and Control Group
Institute of Process and Plant Technology
Berlin University of Technology
Hardenbergstr. 36a, 10623 Berlin, Germany

Abstract: The most time consuming and difficult step in applying model-based concepts for synthesis, control and supervision of bioprocesses is the iterative, manual, and therefore time-consuming model building process. In this contribution a software tool TAM-B is introduced which provides methods for automating the main steps of this iterative procedure. Several qualitative tests are developed to efficiently generate only such models that are suitable for the kinetic problem considered. It will be shown that using the modelling tool not only supports the human modeller in many cases, it can also accelerate the model building process. The main emphasis of the paper will be on new rule-based methods for structure selection.

Keywords: modeling, qualitative analysis, rule based reasoning, system identification, S.cerevisiae

1. INTRODUCTION

Models of biological reaction systems play an important role for the application of model-based methods in supervision and control of biotechnological processes. Apart from the large number of necessary experiments, the limiting factor in using these methods is the time-consuming model building process.

A human modeller will carry out most steps of this process manually in an iterative way. For describing the system, the necessary state variables and a reaction scheme have to be postulated by the modeller first. Afterwards, for every single reaction step a kinetic description has to be assumed. The values of the model parameters will then be calculated by a numerical identification. In most cases, the modeller will not get a suitable model by doing this only once. Rather the reaction scheme and the kinetic model have to be modified many times, and each time a quantitative identification must be performed.

For the last step, the parameter identification, a variety of supporting software is available. But only a few approaches are reported for automating the identification of the reaction scheme and the kinetic model as well. In (Soo, 1989), (Ludewig, 1999), (Ruenglertpanyakul, 1996) and (Brendel and Marquardt, 2003) several systems and methods are introduced for supporting the identification of reaction kinetics. For the identification of

[1] Corresponding author, e-mail: rudibert.king@tu-berlin.de, web: http://mrt.tu-berlin.de

the reaction scheme, Boegarts and van de Wouver have published a method (Bogaerts and Vande Wouver, 2001). However, the latter approach only works under very limited conditions.
With TAM-B (Tool for the Automatic Modelling of Biological reaction systems) a software tool was developed which combines methods for adaptation of the reaction scheme together with methods for an automatic identification of the structure and parameters for the reaction rates. Therefore, this tool can support the user during the whole modelling process. In this contribution some newly developed methods in the tool will be described and an example is given, how a reaction system was modelled supported by TAM-B.

2. SUPPORTING THE MODEL BUILDING PROCESS WITH TAM-B

In the following, the modelling process supported by TAM-B (see Fig. 1) will be introduced in general, whereas for reasons of space, only some central parts and new developments will be demonstrated in detail. More information about the other methods can be found in (King *et al.*, 2002), (Schaich *et al.*, 2001), and (Leifheit and King, 2004).

Before a model identification is started, the user has to input the description of the experiments, the measurement data and a list of necessary state variables. On the basis of a qualitative description of the data, typical biological phenomena can be detected by a rule-based data analysis. Now, the user will input a possible reaction scheme. During this input the user will get advises, which reactions or components have to be added to take the prior detected phenomena into account. Afterwards, it can be checked by an efficient qualitative simulation whether the obtained scheme is largely consistent with the experimental data.

A model generator will create possible kinetic structures. Here, also results of the rules-based analysis can be used to avoid the building of unreasonable model candidates. Next, by qualitative identification methods some of these model candidates can be ruled out in an efficient manner, so that the final parameter identification performed for the remaining candidates is possible within a manageable time.

3. DESCRIPTION OF SOME METHODS USED IN TAM

3.1 Qualitative description of data

The basis of most of the methods used for model verification and adaptation in TAM is a qualitative analysis of data, sometimes combined with an

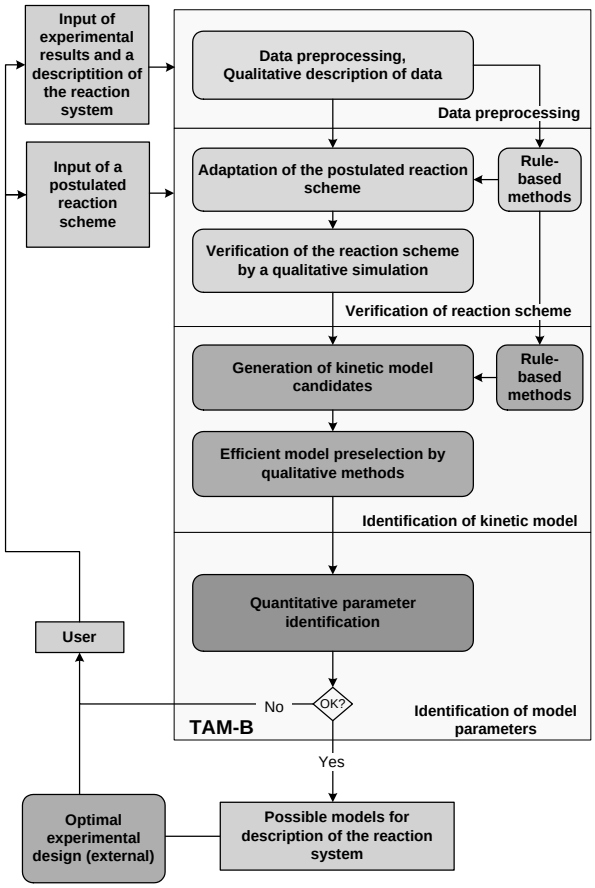

Fig. 1. Scheme of TAM-B

interpretation of domain knowledge. For analysing data on a qualitative level, a qualitative description has to be generated first. This is done on the basis of smoothed data curves for the measured quantities, e.g. for concentration measurements for nutrients, products, and biomass. Different spline methods are available in TAM-B for that purpose. Afterwards, the time plots are divided into sections of the same qualitative behaviour. Such sections could be characterized by the sign of the first derivative of the smoothed data curve, see for example Fig. 2. This is a simpler version of a formerly used description which was based on an idea of (Cheung and Stephanopoulos, 1990). The position where a new qualitative section begins is called phase transition. Because for most of the measured quantities only a few data are available, the times for these phase transitions will often be determined incorrectly. In such cases, online measurements, like pO2- or exhaust gas CO2-values, could be used to locate the transition times more exactly.

3.2 Rule-based generation of information about the reaction model

Metabolic processes in biotechnological reaction systems are mainly characterized by the occurrence of typical biological phenomena, like sub-

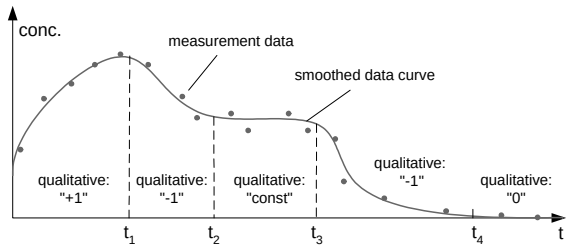

Fig. 2. Possible qualitative description of a measured quantity

strate limitation, product inhibition, or diauxic effects. In the model these phenomena will be described by separate reactions, like regulatory enzyme reactions, or by appropriate kinetic terms, like an inhibition term for example. To detect such phenomena and to take them into account for model building, rules were implemented in TAMB.

By use of these rules, the qualitative representations of measured data will be analysed systematically, qualitative dependencies will be detected, and accordingly, occurring phenomena will be found. For example, a completed growth of biomass after depletion of a nutrient will be an indicator for a growth limitation by this nutrient. In this case the appropriate rule is formulated as:

IF

substance A (nutrient) is depleted (qualitative "0" after qualitative "-1")

AND

substance B (biomass) becomes constant (qualitative "const" after qualitative "+1")

THEN

a limitation from B by A exists (**limitation(B,A)**)

More typical relations will be found in Table 1. If a phenomenon was detected in at least one experiment, it will be added to a list of possible phenomena. But because the number of measured data for some substances is typically very low, the qualitative description of these curves may be inaccurate. That's why some detected phenomena are not really true for the system. Another reason could be that a relation between two substances was caused by a second effect, which wasn't taken into account. Therefore, the phenomena will be checked in all experiments by further rules. A rule for checking **limitation(B,A)** looks as follows:

IF

in the i-th experiment there is at least one time period (2 percent of whole length of the experiment)

WHERE

substance A is depleted

AND

substance B is growing (qualitative "'+1"')

THEN

limitation(B,A) is disproved by the i-th experiment.

After such a kind of automatic detection and verification of phenomena by rules, the obtained information will be announced to the user. Now, the user may decide which phenomena are really true and which phenomena could be rejected.

3.3 Consideration of detected phenomena during the input of a reaction scheme

In order to input a reaction scheme into the modelling tool, the user will define single reaction steps, for which the reaction components like nutrients, products, inhibitors, and catalysts will be defined. One substance, like biomass for example, could exist as a catalyst as well as a product in the same reaction step. During this manual input the user will get advises, how the phenomena identified before could be described by this reaction scheme. For example, if a growth inhibition by a substance takes place, the user will get the suggestion that he should add this inhibitor, whenever he defines a reaction step, where biomass is a product.

Moreover, the user can be informed that new reactions have to be added. This can be the case if phenomena could be described by regulatory enzyme reactions, like diauxic effects for example. After a manual confirmation, such reactions will be added automatically.

3.4 Qualitative simulation of a postulated reaction scheme

The qualitative simulation of a reaction scheme is a simple, but efficient approach for checking whether the postulated reaction scheme is largely consistent with the data courses. Because at this stage there is no or just little information about the kinetic dependencies in the model, a quantitative simulation cannot be performed. Instead, a model will be generated automatically, which describes the qualitative evolution of the reactions. By using this model, the postulated reaction network can be simulated on a qualitative level. A more detailed description of this qualitative simulation will be found in (Leifheit and King, 2004).

As a result, it can be detected if there are discrepancies between postulated reactions and the measured data. For example, it can be discovered that

Substance A	**Substance B**	**Further condition**	**Biol. phenomenon**
consumption	finished growth		limitation of substance B by A
consumption	decomposition after growth		limitation of substance B by A (B is an intermediate)
formation	finished growth		inhibition of substance B by A
consumption	beginning decomposition	no decomp. of B while A available	diauxic effect
consumption	decomposition after growth	B is biomass	death of biomass
no consumption	finished growth	A is nutrient B is biomass	inhibtion or limitation by a not considered substance
no consumption	decomposition after growth	A is nutrient B is biomass	inhibition or limitation by a not considered substance
consumption of an essential nutrient	no reaction (continued growth)	B is biomass qualitative "+1"	not considered nutrient or internal storage

Table 1. Typical biological phenomena characterized by qualitative dependencies of measured quantities

a reaction doesn't take place, though all necessary reaction components, like nutrients and catalysts, are available in the medium. Here, an inhibition could be the reason. If such or similar model deficits are detected, the user will be informed and suggestions will be made, how the reaction scheme could be adapted.

It should be pointed out that the expert could perform all of the above steps as well manually. However, TAM-B not only speeds up this process. It also performs the same tests when many experiments are available. In such situations, a human modeller tends to concentrate only on a subset of experiments, thereby postulating models which could explain this subset, but possibly contradict other experiments.

3.5 Generation of kinetic models for the reaction system

The modelling tool TAM-B offers the possibility to identify the mathematical structures for describing the concentration-dependencies of the reaction rates automatically. Therefore, a mathematical model is constructed inside TAM-B in a symbolic fashion, where different structures can be inserted during the model identification. This is realized by integrating so-called *sockets* in the model where successively different kinetic terms are plugged in.

For every reaction component (reactants, inhibitors and catalysts as well) a separate *socket* is created for describing the influence of its concentration on the reaction rate. Moreover, *sockets* for nutrient-catalyst-combinations are added, in which for example a *Contois*-term [2] could be inserted. The rate of this reaction is finally the product of all *sockets*. During model identification, all those kinetic terms will be plugged in, which where enabled for identification for this socket. In this manner, all combinatorial possibilities of model structures are created.

To hold the number of the thereby created models as small as possible, only those structures should be enabled which lead to meaningful models from a biological point of view. Hence, for every socket a list of preselected kinetics is prepared inside TAM. Depending on the type of the reaction component (nutrient, catalyst etc.), only such kinetics are added which are typical for describing the belonging influence. If due to the rule-based analysis biological phenomena were found, the lists of kinetics can be adapted accordingly. If, for example, a product inhibition is detected, only inhibition kinetics for this product will be enabled. Additionally, the lists can be restricted further by the user.

3.6 Qualitative model verification and quantitative identification of the model parameters

For all generated models the model parameter values will be identified by a numerical optimisation. Under certain conditions, an efficient check of models by a so-called qualitative interval algebra can be performed before. This method is described in detail in (King *et al.*, 2002). The advantage of this method is that a variety of models can be ruled out automatically in a small amount of time.

Only for the remaining candidates the time-consuming quantitative identification is necessary. For this identification several optimisation methods are available in an external FORTRAN program (e.g. a Simplex or a SQP-solver) which is connected to TAM-B via a file interface.

Often, the user will get more than one model with similar quality values. Here, it is possible

[2] typical biological term $c_S/(K_S c_X + c_S)$ with c_x and c_s as concentrations of biomass and nutrient, respectively

to use an external optimal experimental design software (Heine, 2004) for determining new optimal experiments. The results of these experiments complete the existing measurements and allow the verification of the best model candidates.

4. AN EXAMPLE

The functionality of TAM-B will be demonstrated with an example with experimental data. The microbial strain *Saccharomyces cerevisiae* (bakers yeast) was cultivated in a 10 litre pH-controlled fermenter. A defined medium was used which is described in (Rieger *et al.*, 1983). Glucose was the only carbon and energy source. For a first identification, 6 different experiments were used, 5 batch and one fed-batch experiment. The initial concentrations of the batch experiments laid between 5 and 30 g/l of glucose. The concentrations of the other nutrients were identical. Biomass concentrations at t=0 differed slightly due to variations in the precultures.

After the input of the substance list (Glc-glucose, X-biomass and Eth-ethanol) and the experimental conditions, the rule-based detection of biological phenomena was started. A limitation of biomass by ethanol, a limitation of ethanol by glucose and a diauxic effect were found. The diauxic effect was described by 2 regulatory enzyme reactions, so that now the following reaction scheme was proposed manually:

$$\begin{aligned} Glc + X &\xrightarrow{r_2} \nu_2 X + \nu_3 Eth \\ Eth + X &\xrightarrow{r_3} \nu_4 X \\ Enz &\xrightarrow{r_4} EnzAct \\ EnzAct &\xrightarrow{r_5} Enz \end{aligned}$$

As a further reaction

$$Glc + X \xrightarrow{r_1} \nu_1 X$$

was added. This reaction describes the Crabtree effect known from literature (Barford and Hall, 1981), which is typical for yeasts. However, this relation couldn't be detected on the basis of the available experiments.

For the two enzyme reactions a definite kinetic structure was inserted. The conversion of the enzyme (Enz) to the activated enzyme (EnzAct) is inhibited by glucose and limited by the enzyme itself, see r_4, whereas the backward reaction is limited by the activated enzyme and by glucose. Furthermore, the enzyme inhibits ethanol degradation, so that this reaction won't take place until the enzyme concentration is very low. Therefore, r_3 was restricted to kinetics containing an inhibition with respect to the enzyme. For r_1 and r_2 TAM-B could insert all kinetic terms with the exception of inhibition terms with respect to ethanol. All parameter p_i were assumed to be unknown. The parameters comprise parameters inside the reaction rates r_i as well as the stoichiometric coefficients ν_i.

Altogether 2304 models were formulated inside TAM automatically after this users input. As a result about 20 models were found with nearly the same model quality measured by the sum of squared errors. It was concluded that the experiments didn't contain enough information to identify some kinetic structures exactly. That's why more fermentations were planned by the external optimal experimental design (OED) using the most plausible model of this list. After adding these new experimental results, a new kinetic identification run was performed, were, after analysing the results of the first identification, more restrictions concerning possible kinetic terms were formulated to limit the search space. These restrictions were that an inhibition by glucose in r_2 and r_3 was excluded. The structure of the best-identified model now looks as follows:

$$\begin{aligned} r_1 &= p_1 \frac{c_{Glc} c_X}{p_2 + c_{Glc} + p_3 c_X} \\ r_2 &= p_4 \frac{c_{Glc} c_X}{p_5 + c_{Glc}} \\ r_3 &= p_6 \frac{c_{Eth} c_X}{(p_7 + c_{Eth})(p_8 + c_{Enz})} \\ r_4 &= p_9 \frac{c_{Enz}}{(p_{10} + c_{Enz})(p_{11} + c_{Glc})} \\ r_5 &= p_{12} \frac{c_{EnzAct} c_{Glc}}{(p_{10} + c_{EnzAct})(p_{11} + c_{Glc})} \end{aligned}$$

Simulation results for some of the experiments are shown in figure 3.

Without the time for executing the experiments, preparing the identification and creating the reaction scheme, the automatic model identification took about 68 hours using a personal computer with an Athlon 2000+ processor. The overall time needed is significantly lower compared to the classical approach. Moreover, the search space used, could not have been covered in a manual approach, thereby increasing the chance of finding a more adequate model.

5. CONCLUSIONS

Today an experienced human modeller identifies models for biological reaction systems mainly manually in an iterative and time-consuming way. Only for the parameter identification a variety

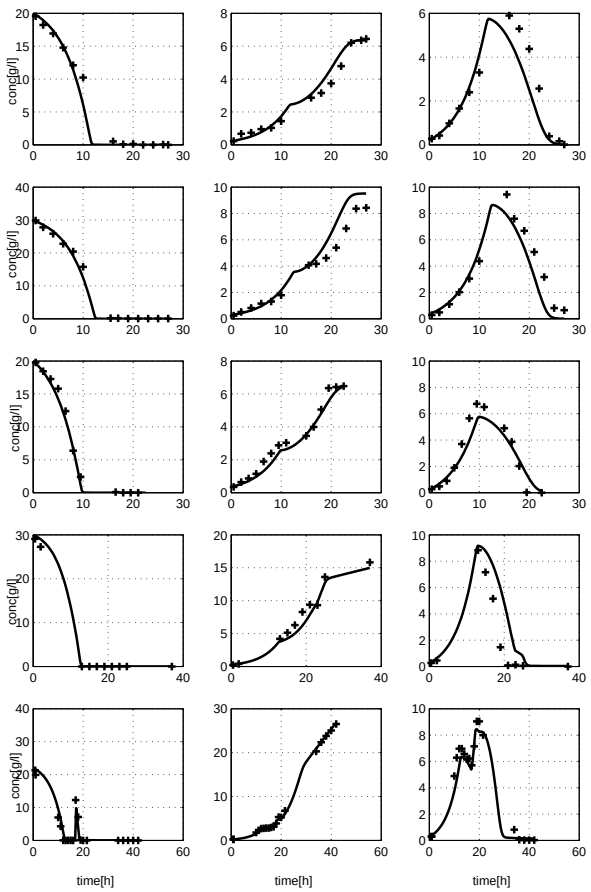

Fig. 3. Results of the model identification for some of the used experiments, - simulated results, + experimental values of biomass (x), glucose (glc), and ethanol (eth). The last two experiments were planned by the OED.

of supporting software is available. TAM-B additionally automates parts of reaction network identification and the model structure selection. By this, the user is supported during the whole modelling process. Consequently, inexperienced modeller, but as well modelling experts, can build up models for biological reaction systems with minor efforts in a manageable amount of time.

Acknowledgement: Part of this work was financed by the Deutsche Forschungsgemeinschaft, Grant Ki 679/4-1.

REFERENCES

Barford, J.P. and R.J. Hall (1981). A mathematical model for aerobic growth of saccharomyces cerevisiae with a saturated respiratory capacity. *Biotechnology and Bioengineering* **23**, 1735–1762.

Bogaerts, Ph. and A. Vande Wouver (2001). Systematic generation of identifiable macroscopic reaction schemes. In: *8th Int. Conference on Computer Applications in Biotechnology CAB8* (D. Dochain and M. Perrier, Eds.). IFAC. Quebec. pp. 13–18.

Brendel, M.; Mhamdi, A.; Bonvin D. and W. Marquardt (2003). An incremental approach for the identification of reaction kinetics. In: *Preprints of the ADCHEM 2003, International Symposium on Advanced Control of Chemical Processes*. HongKong.

Cheung, J.T.-Y. and G. Stephanopoulos (1990). Representation of process trends – part I. A formal representation framework. *Computers and Chemical Engineering* **14**, 495–510.

Heine, T. (2004). Modellgestützte Überwachung und Führung von Fed-Batch-Prozessen. PhD thesis. Technische Universität Berlin.

King, R., J. Leifheit and S. Freyer (2002). Automatic identification of mathematical models of chemical and biochemical reaction systems. In: *CHISA 2002*. Prag. pp. 495–510.

Leifheit, J. and R. King (2004). (Semi-)automatic modeling of biological reaction systems with TAM-B. In: *9th Int. Conference on Computer Applications in Biotechnology CAB9*. IFAC. Nancy(France).

Ludewig, D. (1999). Expertensystem zur Entwicklung von Prozessmodellen für biotechnologische Prozesse. PhD thesis. University of Hannover.

Rieger, M., O. Kaeppeli and A. Fiechter (1983). The role of limited respiration in the incomplete oxidation of glucose by saccharomyces cerevisiae. *Journal of General Microbiology* **129**, 653–661.

Ruenglertpanyakul, W. (1996). Development of an expert system for modeling of bioprocesses. PhD thesis. University of Hannover.

Schaich, D., R. Becker and R. King (2001). Qualitative modelling for automatic identification of mathematic models of chemical reaction systems. *Control Engineering Practice* **9**, 1373–1381.

Soo, V.-W. (1989). Automating the expert reasoning in postulating enzym kinetic models. *Artificial Intelligence in Scientific Computation* **2**, 215–220.