Copyright © 2002 IFAC
15th Triennial World Congress, Barcelona, Spain

SENSITIVITY FUNCTION BASED MODEL REDUCTION: A BACTERIAL GENE EXPRESSION CASE STUDY

Ilse Smets [*] **Kristel Bernaerts** [*] **Jun Sun** [**]
Kathleen Marchal [***] **Jos Vanderleyden** [**]
Jan Van Impe [*]

[*] *BioTeC–Katholieke Universiteit Leuven,*
B-3001 Leuven (Belgium)
Fax: +32-16-32.19.60 e-mail: jan.vanimpe@agr.kuleuven.ac.be
[**] *CMPG–Katholieke Universiteit Leuven,*
B-3001 Leuven (Belgium)
[***] *ESAT–Katholieke Universiteit Leuven,*
B-3001 Leuven (Belgium)

Abstract:
In the area of genetically engineered micro-organisms grown in bioreactors, mathematical modeling usually results in balance type models involving (*i*) a (rather) large number of state variables and, (*ii*) complicated kinetic expressions containing a large number of parameters. Therefore, a generic methodology is developed to reduce the model complexity at the level of the kinetics, while maintaining high prediction power. As a case study to illustrate the method and results obtained, the influence of the dissolved oxygen concentration on the *cytN* gene expression in the bacterium *Azospirillum brasilense* Sp7 is modeled.

Keywords: modelling, model reduction, sensitivity functions, biotechnology, continuous systems

1. INTRODUCTION

To qualify and quantify the influence of external signals on bacterial gene expression, continuous culture steady-state experiments have been performed throughout the past (Chao *et al.*, 1997; Kasimoglu *et al.*, 1996). These costly, labor-intensive and time consuming experiments can be reduced to a minimum with the aid of a mathematical model that describes the intrinsic properties of the dynamic bioprocess.
Although the advantages of model based optimization and control of *fermentations* (e.g., baker's yeast production processes and biological wastewater treatment systems) are well established, the introduction of mathematical modeling in the field of genetic engineering is fairly recent. The scarce, knowledge based models that have been developed are usually characterized by complex kinetic expressions involving a high number of parameters. Within the context of quantitatively assessing the influence of the dissolved oxygen concentration (DO) on the *cytN* gene expression in the bacterium *Azospirillum brasilense* Sp7, this paper presents a generic model reduction technique based on sensitivity function analysis. The genetic engineering details, which are out of the scope of this contribution, are elaborated in (Sun *et al.*, 2001).
The organization of this paper is as follows. In Section 2, the bacterial gene expression system is introduced and a primary, knowledge based, model is developed. The parameters of the primary model are identified in Section 3. Section 4 presents a model reduction procedure based

on sensitivity function analysis which is applied to the primary model. The main conclusions are summarized in Section 5.

2. MATERIALS AND METHODS

2.1 *Bacterial reporter gene system*

To characterize the behavior of genes of which the expression cannot be screened directly, it is common practice to link these genes to a gene of which the expression is readily monitored. This genetic construction is called a *reporter gene system*. In *Azospirillum brasilense* Sp7 the *cytN* gene is linked to the *gusA* gene of which the induced *β-glucuronidase* activity is used to (indirectly) monitor the *cytN* gene expression. The construction of the translational *cytN-gusA* fusion is extensively described in (Sun *et al.*, 2001).

2.2 *Fermentation strategy*

A. brasilense Sp7 containing plasmid pFAJ873 (with the *cytN-gusA* fusion) was cultivated in a batch fermenter until the end of the exponential growth phase was reached (after approximately 13 hours). Subsequently, the continuous fermentation started in a 2-liter DO-stat fermentor. Consecutive small dissolved oxygen shifts were applied at regular intervals of about 1.5 hrs, before a new steady-state was reached. Samples were taken just before a new DO-shift to determine the β-glucuronidase activity and cell density. During the continuous fermentation, the carbon source malate was assigned as the limiting growth factor.

2.3 *Mass balance equations*

Changes in β-glucuronidase activity of a hybrid gene reporter in the presence of an altering external signal indicate that this external signal initiates transcriptional activation. However, in the experimental set-up under study, not only alterations in transcriptional activation of the plasmid encoding *cytN-gusA* fusion induced by the DO-shifts, but also accumulation and turnover of the fusion protein (i.e., the *cytN-gusA* gene product) can account for the measured β-glucuronidase activity. Therefore, only the *specific* expression rate of the fusion protein (i.e., the amount of fusion proteins expressed per cell and per hour), being independent of the experimental design, can reflect the influence of dissolved oxygen on the expression of the target gene. In order to derive the specific expression rate of the fusion protein from the measured β-glucuronidase activity, the following general dynamic mathematical model based on mass balances was applied (Van Impe and Bastin, 1995):

$$\frac{dC_X}{dt} = \mu C_X - DC_X \qquad (1)$$

$$\frac{dC_S}{dt} = -\sigma C_X - DC_S + DC_{S,in} \qquad (2)$$

$$\frac{dC_P}{dt} = \pi C_X - DC_P - kC_P \qquad (3)$$

where C_X [g cells/L] denotes the concentration of the biomass, C_S [g malate/L] is the concentration of the carbon source and C_P [g protein/L] is the concentration of the fusion protein. $C_{S,in}$ is the limiting carbon source concentration in the feed flow, D [1/hr] is the dilution rate, μ [1/hr] is the specific growth rate of cells, σ [g malate/g cells/hr] is the specific consumption rate of the carbon source, π [g protein/g cells/hr] is the specific expression rate of the fusion protein and k [1/hr] is the in vivo degradation rate of the fusion protein. According to the definition of the GUS activity (Miller, 1972), the value of β-glucuronidase activity is assumed to be proportional to the amount of the fusion protein (denoted by P [g protein]) per cell (denoted by X [g cells]):

$$P = \alpha U X \qquad (4)$$

where U [Miller Unit] is the β-glucuronidase activity [Miller Unit stands for GUS enzyme activity/g cells/hr] and α [g protein/g cells/Miller Unit] is a proportionality constant. By combining equation (4) with equations (1) and (3), the following equation can be deduced:

$$\frac{dU}{dt} = \beta - \mu U - kU \qquad (5)$$

where β equals π/α [Miller Unit/hr] and is defined as the apparent specific expression rate of the fusion protein, reflecting the direct influence of the external signal on the transcriptional activation of the hybrid gene fusion.

To complete the model, the following kinetic expressions are proposed inspired by available knowledge. A double Haldane type model is applied to describe the specific growth rate of cells, μ, as a function of two substrates: malate and dissolved oxygen DO [%]. The apparent specific expression rate of the fusion protein, β, is described as function of the carbon substrate (malate) with a Monod type model and as function of dissolved oxygen with a Haldane-like model (in which the background expression of the fusion protein is introduced since constitutive background expression of the fusion protein has been observed under anaerobic conditions during preliminary experiments). Since the dependency of the degradation of the fusion protein on DO has been observed in the experiments carried out in test tubes (data not shown), the degradation rate k of the fusion pro-

tein is expressed as function of DO in the frame of the Monod model. In Table 1 the proposed kinetic equations with their corresponding parameters are listed. In addition, the correlation between the specific reaction rates μ and π is expressed by the specific substrate consumption rate σ and the yield coefficients.

In order to assess the influence of DO on the specific expression level of the *cytN-gusA* fusion, the measured experimental values have been fed to the model to identify the appropriate parameters. Once identified, the complete model (parameters and model structure) will be used to predict the behavior of the hybrid fusion protein as function of the external variables.

The simulation and parameter identification of the model described above are performed using Matlab 5.3 (The MathWorks, Inc., Natick) on a Linux platform.

3. RESULTS

3.1 *Parameter identification*

The profile of the dissolved oxygen concentration for the parameter identification experiment is presented in Figure 1(A). The 14 parameters in the mathematical model were identified by minimizing the following cost function $\mathcal{J}$:

$$\mathcal{J} = \sum_{j=1}^{n} \frac{\sum_{i=1}^{m} \left(\frac{Y_{s,ij} - Y_{e,ij}}{\widetilde{Y}_{e,j}} \right)^2}{\sigma_{sj}^2} \qquad (10)$$

where i denotes the sampling time and j denotes the components. $Y_{s,ij}$ is the data set of the simulation results, $Y_{e,ij}$ is the data set of the experimental results and $\widetilde{Y}_{e,j}$ is the average value of the components. σ_{sj} is the standard deviation of the experimental data. The identified initial values (t=0) are: $C_S(0)$=5.2087 g malate/L; $C_X(0)$=0.0554 OD_{578} and $U(0)$=41.9123 Miller Unit. The carbon source concentrations in the feed flow is equal to $C_{S,in}$=5.0075 g malate/L. The results are shown in Figure 1(B, C, D) (solid lines) while the values of the identified parameters are summarized in Table 1. Because cell density rather than dry weight of cells was used to monitor the cell growth, the yield coefficient Y_{XS} as defined in equation (9) is expressed in [OD_{578}/g malate] instead of [g cells/g malate]. The agreement between the simulation results and the experimental data is remarkable.

3.2 *Model validation*

In order to further validate the applicable range of the model, a continuous fermentation with a totally different DO profile (Figure 2(A)) was

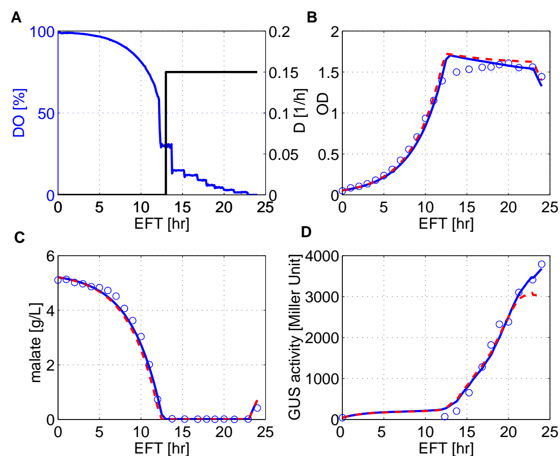

Fig. 1. Identification experiment. A is the profile of on-line measured values of DO and the applied dilution rate profile D. In B, C and D, open symbols: experimental data; solid lines: simulation general model; dashed lines: simulation reduced model. EFT = Elapsed Fermentation Time.

performed. The DO concentration was kept at 10% during the batch fermentation and subsequently shifted from low to high values during the continuous fermentation. The continuous fermentation started at 12 hours with a dilution rate of 0.1193 1/hr. The initial values for validation are based on the first experimental measurements: $C_S(0)$=5.3812 g malate/L; $C_X(0)$=0.038 OD_{578}; $U(0)$=126.84 Miller Unit. For the carbon source concentrations in the feed flow, the measured value was used: $C_{S,in}$=5.308 g malate/L. The simulation results (solid line) and the experimental data (open symbols) are shown in Figure 2(B,C,D). The agreement between simulated and experimental results corroborates the generality of the model.

4. SENSITIVITY FUNCTION BASED MODEL REDUCTION

When examining the general model structure represented by Equations (1),(2),(5) and (6) to (9) and given the large range in the order of magnitude of the 14 identified parameters summarized in Table 1, a legitimate question is whether a similar high quality fit of the experimental data can be obtained with a simplified model including less parameters. In order to mathematically investigate this statement, a thorough sensitivity analysis has been performed.

Hereto, the dilution rate D and the dissolved oxygen signal DO are defined as the system inputs u_1 and u_2, respectively, and the biomass concentration X, the malate concentration S and the GUS activity U are defined as the system outputs y_1, y_2 and y_3, respectively. The parameters are denoted p_j with j ranging from 1 to 14.

Table 1. Kinetic expressions of the original model with the corresponding parameters.

Kinetic expressions	Identification exp	Validation exp
$\mu = \mu_{max} \frac{C_S}{(KM_{XS} + C_S + \frac{C_S^2}{KI_{XS}})} \frac{DO}{(KM_{XG} + DO + \frac{DO^2}{KI_{XG}})}$ (6) $\beta = \beta_{max} \frac{C_S}{(KM_{PS} + C_S)} \frac{(DO + KB_{PG})}{(KM_{PG} + DO + DO^2/KI_{PG})}$ (7) $k = k_{max} \frac{DO}{K_k + DO}$ (8) $\sigma = \frac{\mu}{Y_{XS}} + \frac{\beta}{Y_{US}}$ (9)	$D = 0.15$ 1/h $C_S(0) = 5.2087$ g mal/L $C_X(0) = 0.0554$ OD_{578} $U(0) = 41.9123$ M.U. $C_{S,in} = 5.0075$ g mal/L	$D = 0.1193$ 1/h $C_S(0) = 5.3812$ g mal/L $C_X(0) = 0.038$ OD_{578} $U(0) = 126.84$ M.U. $C_{S,in} = 5.308$ g mal/L

Parameter values

Y_{XS}	yield coefficient of biomass on malate	[OD_{578}/g malate]	0.3196 $\rightarrow p_{01}$
Y_{US}	yield coefficient of fusion protein on malate	[Miller Unit/g cells]	2.7957×10^4 $\rightarrow p_{02}$
μ_{max}	maximal growth rate of cells	[1/hr]	0.2766 $\rightarrow p_{03}$
β_{max}	maximal apparent expression rate of the fusion protein	[Miller Unit/hr]	9.3150×10^3 $\rightarrow p_{04}$
k_{max}	maximal degradation rate of the fusion protein	[1/hr]	0.1194 $\rightarrow p_{05}$
KM_{XS}	saturation constant of malate to cell growth	[g malate/L]	1.2531×10^{-2} $\rightarrow p_{06}$
KI_{XS}	inhibition constant of malate to cell growth	[g malate/L]	1.0483×10^3 $\rightarrow p_{07}$
KM_{XG}	saturation constant of DO to cell growth	[%]	5.2341×10^{-2} $\rightarrow p_{08}$
KI_{XG}	inhibition constant of DO to cell growth	[%]	1.0219×10^4 $\rightarrow p_{09}$
KM_{PS}	saturation constant of malate to the fusion protein expression	[gmalate/L]	1.2056×10^{-3} $\rightarrow p_{10}$
KB_{PG}	background expression constant of the fusion protein	[%]	0.2766 $\rightarrow p_{11}$
KM_{PG}	saturation constant of DO to the fusion protein expression	[%]	11.6485 $\rightarrow p_{12}$
KI_{PG}	inhibition constant of DO to the fusion protein expression	[%]	0.8135 $\rightarrow p_{13}$
K_k	saturation constant of DO to decay of the fusion protein	[%]	0.8941 $\rightarrow p_{14}$

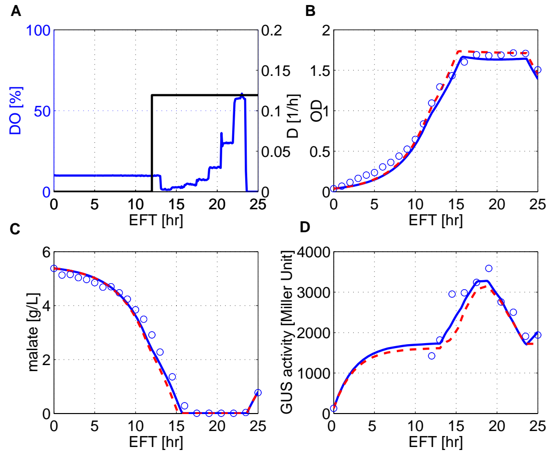

Fig. 2. Validation experiment. A is the profile of on-line measured values of DO and the applied dilution rate profile D. In B, C and D, open symbols: experimental data; solid lines: simulation general model; dashed lines: simulation reduced model. EFT = Elapsed Fermentation Time.

The 3×14 sensitivity functions $\frac{\partial y_i}{\partial p_j}(t)$ are then computed (see Figure 3). These sensitivity functions represent the sensitivity of each output y_i to (small) variations in each model parameter p_j. To make comparison possible each sensitivity function is rescaled by multiplying with the parameter value under study. The system (or at least one of the outputs) is said to be sensitive to a certain parameter if a change in the parameter's value significantly affects the predictive quality of the model. In other words, the fit of the experimental data becomes worse, if that parameter value is changed, and the parameter can, accordingly, be classified as *essential*. Hence, how can the *essential* parameters be selected on the basis of the sensitivity functions plotted in Figure 3? Hereto, each column of subplots, i.e., the sensitivity of one output with respect to all 14 parameters, is to be considered separately and the different orders of magnitude have to be compared. As mentioned before, due to the rescaling this comparison is justified. If the order of magnitude of a certain sensitivity function is substantially larger than the average order of magnitude, then the corresponding parameter is retained. Following this line of reasoning, parameters p_{01} and p_{03}, i.e., Y_{XS} and μ_{max}, respectively, are considered essential for the biomass concentration output y_1, since their order of magnitude ($\mathcal{O}(10^1)$) is significantly larger than the average value of $\mathcal{O}(10^{-2})$. As for the malate concentration output y_2, the same parameters are selected. Finally, with respect to the GUS activity output y_3, following parameters with $\mathcal{O}(10^3)$ (compared to $\mathcal{O}(10^1)$ or $\mathcal{O}(10^2)$) will be retained: p_{04}, p_{05}, p_{12} and p_{13}, i.e., β_{max}, k_{max}, KM_{PG} and KI_{PG}, respectively.

For more details on sensitivity function related procedures, reference is made to, e.g., (Bernaerts *et al.*, 2000).

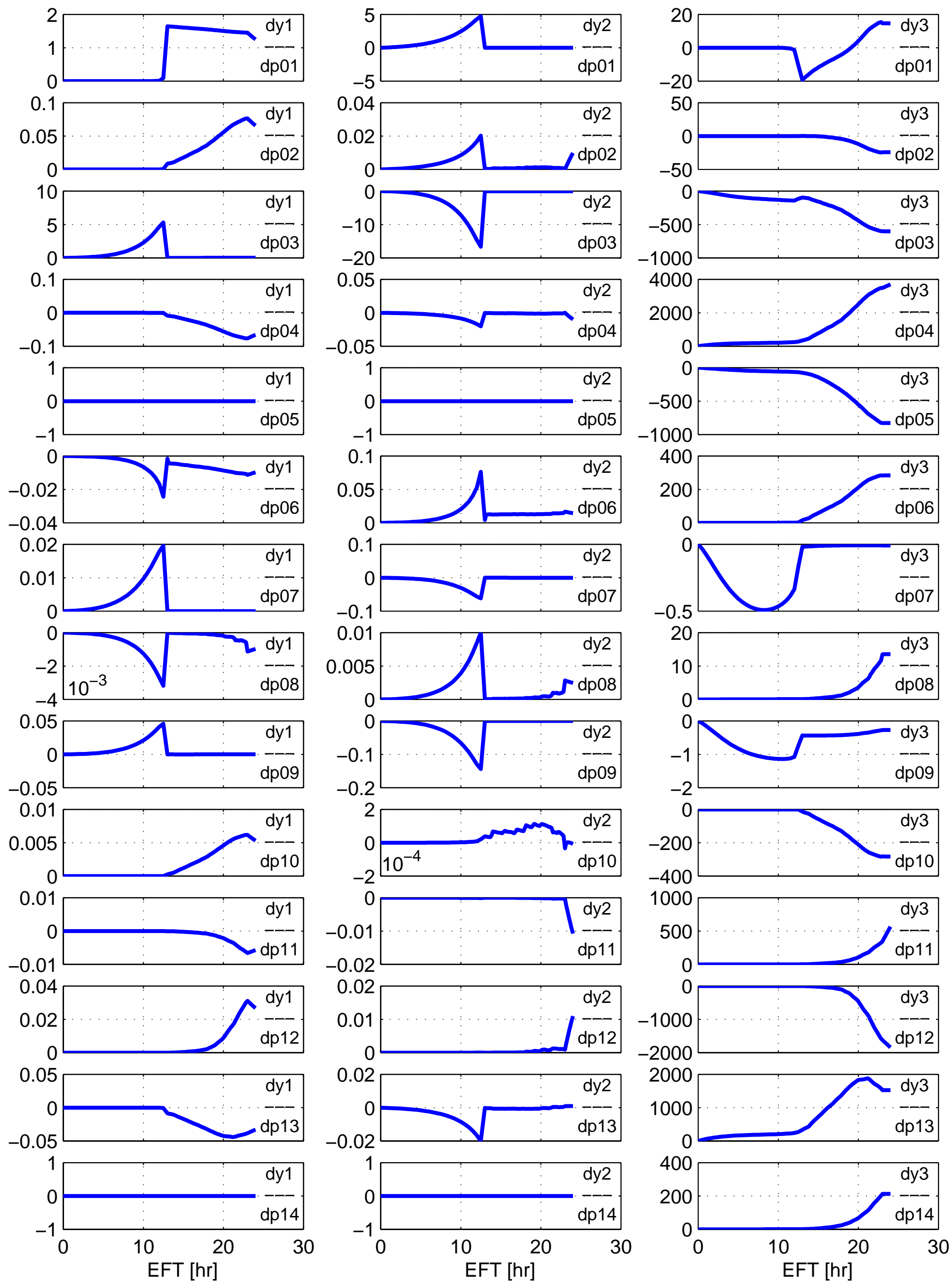

Fig. 3. Evolution of the 3×14 (rescaled) sensitivity functions with respect to time (EFT).

Based on this sensitivity analysis and given the experimental data, the (kinetic expressions of the) above introduced, primary model is substantially simplified to the model presented in Table 2. The number of model parameters has been reduced to 6. As can be seen from Table 2 most of the saturation constants (all but one classified as *non-essential*) are replaced by ϵ_1 or ϵ_2, reflecting the *switch* characteristic of the remaining term which has to prevent the predicted concentrations to become negative. ϵ_1 is of the same order of magnitude as the residual substrate concentration, i.e., 10^{-2}, and induces a 50% reduction of the specific growth rate when the malate concentration is low. ϵ_2 is a very small number, e.g., 10^{-6}, inducing a switch from the maximum specific rate when substrate or dissolved oxygen is present, to a rate equal to zero when they are lacking. Notice that, without any loss of accuracy, the switch factor in terms of the malate concentration in the kinetic expression for β, can be omitted. Furthermore, the inhibition constants (i.e., KI_{XS} and KI_{XG}) that have been excluded can be considered as replaced by an infinitely large (positive) number corresponding to a non-inhibition situation. Note, however, that, from a mechanistic point of view, it cannot be claimed that the growth of this species is *not* inhibited by high substrate or dissolved oxygen concentration. It is merely concluded that an inhibition effect cannot be inferred from the

Table 2. Kinetic expressions of the simplified model with the corresponding parameters.

Kinetic expressions

$$\mu = \mu_{max} \frac{C_S}{\epsilon_1 + C_S} \cdot \frac{DO}{\epsilon_2 + DO}$$
$$\beta = \beta_{max} \frac{C_S}{\epsilon_2 + C_S} \cdot \frac{DO}{KM_{PG} + DO + DO^2/KI_{PG}}$$
$$k = k_{max} \frac{DO}{\epsilon_2 + DO}$$
$$\sigma = \frac{\mu}{Y_{XS}}$$

Parameter values

μ_{max}	[1/hr]	0.2766
β_{max}	[Miller Unit/hr]	9.3150×10^3
KM_{PG}	[%]	11.6485
KI_{PG}	[%]	0.8135
k_{max}	[1/hr]	0.1194
Y_{XS}	[OD_{578}/g malate]	0.3196

Constants

$\epsilon_1 = 10^{-2}$ $\quad$ $\epsilon_2 = 10^{-6}$

available experimental data. Also, the yield coefficient Y_{US} is assumed to be infinitely large as to reflect the negligible contribution of the product formation to the substrate consumption rate σ.

Figures 1 (identification) and 2 (validation) illustrate that the descriptive quality of this simplified model (dashed line) is as good as the descriptive quality of the original 14 parameter model (full line). Observe that there was apparently no need to re-optimize the 6 parameters in the simplified model (see Tables 1 and 2).

Further, in the simplified model, the background GUS activity (represented by KB_{PG}), is neglected. This explains why the simplified model predicts a stagnation instead of a persisting increase of GUS activity at the end of the experiments, i.e., under (early) anaerobic conditions.

5. CONCLUSIONS

In this paper it is illustrated that sensitivity function analysis is a powerful tool to reduce the complexity of a knowledge based model.

As a vehicle to present the model reduction methodology and the results obtained, a bacterial gene expression case study is considered in which the influence of dissolved oxygen concentration on the expression of the *cytN* gene in *Azospirillum brasilense* Sp7 is modeled.

In a first approach available physiological knowledge is incorporated into a mass balance equation model with 3 states and 14 parameters.

The large differences in order of magnitude of the identified parameter values, is a clear indication that not all these parameters are significant. A careful sensitivity function analysis revealed that a reduced model with only 6 parameters is almost as accurate as the original model.

A detailed exploration of the significance of model parameters in the general as well as in the reduced model are the subject of ongoing research. Hereto, optimal experiments (complemented with parameter uncertainty analysis) will be designed to check whether the model features present in the general model, but omitted in the simplified model (e.g., inhibition of the specific growth rate at high substrate or dissolved oxygen concentrations or presence of background gene expression in the absence of dissolved oxygen), are truly needed or not.

ACKNOWLEDGMENTS

Ilse Smets is a research assistant with the Fund for Scientific Research Flanders. Kristel Bernaerts is a research assistant with the Institute for the Promotion of Innovation by Science and Technology in Flanders. Work supported in part by grants from the Flemish Government (GOA) and the Fund for Scientific Research Flanders, Project OT/99/24 of the Research Council of the Katholieke Universiteit Leuven and the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility is assumed by its authors.

6. REFERENCES

Bernaerts, K., K.J. Versyck and J.F. Van Impe (2000). On the design of optimal dynamic experiments for parameter estimation of a Ratkowsky-type growth kinetics at suboptimal temperatures. *Int. J. Food Microbiol.* **54**(1-2), 27–38.

Chao, G., J. Shen, C. P. Tseng, S.-J. Park and R. P. Gunsalus (1997). Aerobic regulation of isocitrate dehydrogenase gene (*icd*) expression in *Escherichia coli* by the *arcA* and *fnr* gene products. *J. Bacteriol.* **179**, 4299–4304.

Kasimoglu, E., S.-J. Park, J. Malek, C. P. Tseng and R. P. Gunsalus (1996). Transcriptional regulation of the proton-translocating ATPase (*atpIBEFHAGDC*) operon of *Escherichia coli*: control by cell growth. *J. Bacteriol.* **178**, 5563–5567.

Miller, J.H. (1972). *Experiments in molecular genetics*. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.

Sun, J., I. Smets, K. Bernaerts, J. Van Impe, J. Vanderleyden and K. Marchal (2001). Quantitative analysis of bacterial gene expression by using the *gusA* reporter gene system. *Appl. Envir. Microbiol.* **67**, 3350–3357.

Van Impe, J.F. and G. Bastin (1995). Optimal adaptive control of fed-batch fermentation processes. *Control Eng. Pract.* **3**(7), 939–954.