Data Mining Based Hybrid Modelling and Experimental Design for Bioprocess Optimisation

R. Guthke, W. Schmidt-Heck
Hans Knöll Institute for Natural Products Research
Beutenbergstr. 11a, 07745 Jena, Germany
rguthke@pmail.hki-jena.de

R. Berkholz
BioControl Jena GmbH
Wildenbruchstr. 15, 07745 Jena, Germany
biocontrol@t-online.de

Abstract

The dynamics of growth and hyaluronidase production of *Streptococcus agalactiae* were formulated by a hybrid model. The fuzzy rules implemented in the model describing the inhibition of growth by lactate were generated from fermentation process data. The model was used for experimental design to minimize the model parameter estimation error as well to maximize the productivity.

Keywords: Fuzzy rule induction; hybrid modeling; sequential experimental design; optimisation.

1 Introduction

Models used for bioprocess optimisation describe certain but limited aspects of reality. Because of the complexity of microbial metabolism, models for bioprocess optimisation can not be valid under all kinds of cultivation conditions. They can be valid under certain cultivation conditions, in particular those aiming at potentially productive processes. Model structure and measurement noise may also lead to problems with regard to model parameter estimation. Therefore, experiments for model based bioprocess optimisation have to be optimal with respect to process performance as well as parameter estimation accuracy. Established experimental design methods led to experiments that may either informative (e.g. [1, 2]) *or* productive (e.g. [3, 4]). But they have to be productive *and* informative. The novel Λ-optimal experimental design introduced by Berkholz et al. [5] fulfils these requirements. Using this approach it is possible to design productive and informative experiments at the same time. The procedure was successfully applied to the optimisation of a hyaluronidase fermentation by *Streptococcus agalactiae* in simulation studies and real fermentation experiments [6]. The models used for the Λ-optimal experimental design may be deterministic or hybrid ones containing mechanistic, neural network and fuzzy components. This contribution presents simulation studies using a fuzzy-hybrid model for the Λ-optimal experimental design to optimise the hyaluronidase fermentation.

2 Material and Methods

The hyaluronidase fermentation was run in a fed-batch fermenter on a complex medium, using glucose dosage, pH control via NaOH dosage and temperature control. During the fermentation the concentrations of the 8 state variables

- biomass dry weight c_X,
- product hyaluronidase c_P
- by-product lactate c_L,
- three substrates (glucose, ammonia, phosphate)
- protein
- activity of proteases

were determined off-line all hours. The four state variables dissolved oxygen concentration pO_2, pH and the weights of the NaOH reservoir w_{NaOH} and the glucose reservoir w_G were measured on-line. The cultivation volume V was calculated by mass balance considering the dosages and the volume loss due to sampling. Further details about cultivation conditions and analytical methods used can be found in Berkholz *et al.* [6].

The kinetics of all 12 state variables as well as of 12 derived variables (rates and specific rates of growth, substrate consumption, product and by-product formation) were clustered by Fuzzy-C-means clustering method into three classes each. One, two or three conditional fuzzy rules IF A [AND B AND C] THEN D were generated where the conclusion D are the classes of the kinetics of biomass, product or by-product as well the classes of the derived rates and specific rates. The rules were rated and selected using the "relevance index" as criterion. More details of the data mining procedure used are described by Guthke et al. [7]. Rules with the highest and positive relevance index were selected, evaluated using general microbiologcial and engineering knowledge and used for formulation of the reaction rates in dependence on the state variables or the dynamic model.

3 Results

A dynamic model that is suitable for optimisation of the product yield J_P expressed as the mass of the product m_P at the end of the fermentation t_E:

$$J_P = m_P(t_E) = c_P(t_E) \cdot V(t_E) \tag{1}$$

has to formulate both, the product formation rate and the dilution of product concentration c_P by the total feeding rate F_T The correlation analysis showed that the product formation rate is proportional to the biomass concentration c_X

$$\frac{dc_P}{dt} = \pi_P \cdot c_X - \frac{F_T}{V} \cdot c_P \tag{2}$$

and that the specific hyaluronidase formation rate π_P is proportional to the specific growth rate $:\mu = c_X^{-1}$ dc_X/dt:

$$\pi_P = \alpha_P \cdot \mu \tag{3}$$

i.e. the product formation is growth associated. The model has also to describe the dynamics of the state variables biomass c_X and culture volume V involved in eq. (2):

$$\frac{dc_X}{dt} = \mu \cdot c_X - \frac{F_T}{V} \cdot c_X \tag{4}$$

$$\frac{dV}{dt} = F_T = F_{NaOH} + F_G + F_S(\tau) \tag{5}$$

F_S describes the dosage of complex medium discussed below. To formulate the dosage rates F_{NaOH} and F_G two facts were considered:

(a) NaOH were fed for pH control. No other pH changing compounds than lactate with the concentration c_L are produced, i.e. the NaOH dosage rate F_{NaOH} is only a function of the lactate formation rate $\pi_L \cdot c_X$

$$F_{NaOH} = \pi_L \cdot c_X \cdot V \cdot \frac{M_{NaOH}}{M_L \cdot c^F_{NaOH}} \tag{6}$$

$$\frac{dc_L}{dt} = \pi_L \cdot c_X - \frac{F_T}{V} \cdot c_L \tag{7}$$

(b) Lactate formation is homofermentative, i.e. 1 mol of glucose is metabolized into 2 mol of lactate, therefore, to sustain an almost constant glucose concentration, the glucose dosage rate F_G has to be the same as the NaOH dosage rate F_{NaOH} with set NaOH and glucose feed concentrations of $c_{NaOH}{}^F$=200 g/l and $c_G{}^F$=500 g/l, respectively

$$F_G = F_{NaOH} \tag{8}$$

The rules generated as described above were used to explain the declining growth and product formation rates after about 8 hours of fermentation. It was found that the growth and product formation rates were not limited by the substrates measured, i.e. glucose, ammonia, phosphate and oxygen. (The hypothesis of limitation of a non-measured amino acid were falsified experimentally by dosage of various amino acids, e.g. glutamine.) The rules generated led to the hypothesis that

(c) Lactate formation is inhibited by lactate, i.e. the specific lactate formation rate π_L is a declining function of the lactate concentration c_L

$$\pi_L = \beta_L \cdot \exp\left(-\frac{c_L{}^2}{K_I{}^2}\right) \tag{9}$$

(d) Growth of the micro-organism is inhibited by lactate, i.e. the specific growth rate μ is a declining function of the lactate concentration c_L

To model this function $f(c_L)$ three alternative approaches were applied: mechanistic growth modelling, growth modelling by Artificial Neural Networks (ANN) and fuzzy growth modelling. In this paper for the hyaluronidase fermentation investigated a simple fuzzy sub-model was applied to express the specific growth rate μ as a declining function $f(c_L)$

$$\begin{aligned}&\text{IF } c_L \text{ is } low \text{ THEN } \mu \text{ is } high\\&\text{IF } c_L \text{ is } high \text{ THEN } \mu \text{ is } low\end{aligned} \tag{10}$$

The membership functions of the fuzzy growth model (10) are shown in Figure 1.

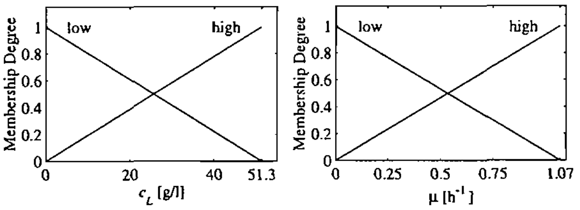

Figure 1: Membership functions for the lactate concentration c_L (left) and the specific growth rate μ (right)

By sensitivity analysis it was found that the number of linguistic terms have to be small, i.e. two ("low" and "high") and not three ("low", "medium", high"). For fuzzy inference the MIN-MAX method and for defuzzification the modified center-of-gravity method were applied. Using the known molecular weights M_{NaOH}=40 g/mol and M_L=90 g/mol and by help of the Sequential Quadratic Programming (SQP, [4]) method the following model parameters were identified: $(\alpha_P, \beta_L, K_I, c_X(0))$ = (17.0, 4.05, 60.6, 0.00807).

Figure 2 shows the trajectories of the variables of the fuzzy hybrid model (2) to (11) and the measured values of a single fermentation run. The fuzzy hybrid model is able to fit the data quite well.

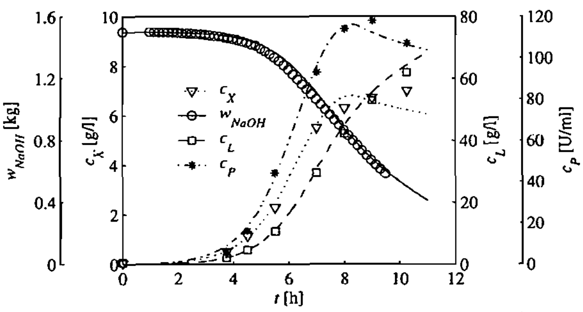

Figure 2: Trajectories of the variables of the fuzzy hybrid model and measured values

The parameter estimation accuracy J_S can be calculated using the condition number of the Fisher information matrix **F**

$$J_S = \frac{1}{\mathrm{cond}(\mathbf{F})} = \frac{\lambda_{min}(\mathbf{F})}{\lambda_{max}(\mathbf{F})} \tag{11}$$

where λ_{min} and λ_{max} are the minimum and the maximum eigenvalue of **F**, respectively ([1]). The objective function for the Λ-optimal experimental design is the weighted sum of J_P and J_S [5]:

$$J_\Lambda = \omega \cdot J_P^* + (1-\omega) \cdot J_S^* \rightarrow \max \tag{12}$$

The stars (*) indicate the normalisation of J_P and J_S within the interval [0,1] to account for their different orders of magnitude. By changing the weight factor ω the focus of the experimental design can be shifted between process performance and parameter estimation accuracy. At the beginning of a fermentation optimisation a low value of ω should be chosen. In this case the experimental design focuses on the estimation accuracy of the model parameters which are still unknown in this early stage of the optimisation procedure. The weight factor ω can be increased when the parameter values converge during successive experiments. Thereby the focus of the experimental design is primarily set on the process performance.

Previous hyaluronidase fermentation experiments have indicated that an additional pulse-like dosage F_S of the complex substrate leads to improved values of J_P and J_S. With τ as the dosage time, i.e. the pulse start time, the following experimental design problem has to be solved:

$$\tau_{opt} = \arg\max_{\tau} J_\Lambda(\tau) \tag{13}$$

Figure 3 (left) shows the normalised values of J_P and J_S as functions of the dosage time τ. The two objective functions reach maximum at different dosage times. Figure 3 (right) shows different trajectories of J_Λ. Different weight factors ω result in different values of the optimal dosage time τ_{opt}.

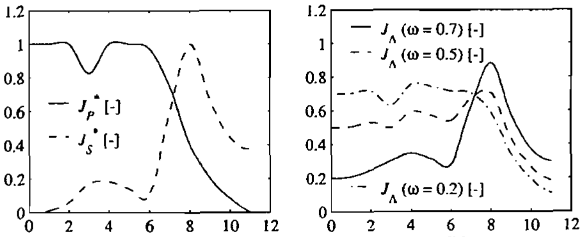

Figure 3: Trajectories of the objective functions J_S^*, J_P^* (left) and trajectories of the Λ-criterion J_Λ (right) for different weight factors ω.

For $\omega = 0.2$ the optimal dosage time $\tau_{opt} = 8$ h was calculated by maximising J_Λ as shown in Figure 3. The fit of the fuzzy hybrid model according to eqs. (2)-(10) to data of a fermentation run with dosage of complex medium at $\tau = 8$ h is shown in Figure 4. This model fitted to this second experiment could be used to design a third experiment and so on.

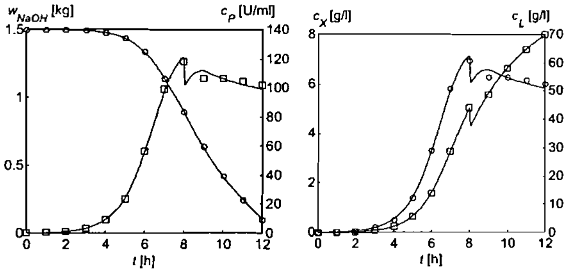

Figure 4: Variables of the fuzzy hybrid model of a second fermentation run with dosage of complex medium at $\tau = 8$ h

4 Discussion

Sequential experimental design that uses the novel Λ criterion introduced by Berkholz et al. [5] was applied to a hybrid, i.e. mechanistic and fuzzy model of the hyaluronidase fermentation of *Streptococcus agalactiae*. This concept of hybrid model based experimental design reflects a typical situation of bioprocess development. The knowledge available may be mechanistic for instance concerning the stoichiometry and the volumetric fluxes and it may be fuzzy for instance concerning regulatory effects of growth and product formation. The mechanistic knowledge may be formulated from natural laws, analogies to related processes and general engineering by differential equations. A major disadvantage of mechanistic modelling is the requirement of detailed process information. If there is, however, incomplete or inaccurate process understanding ANN or fuzzy models should be preferred. Fuzzy hybrid modelling is a suitable approach if process data is available and if there is enough or increasing process knowledge to formulate qualitative relations between measurable variables and unknown reaction rates. The fuzzy knowledge formulating the reaction rates in dependence e.g. on substrates and/or (by-) products may be formulated by the skilled engineer on the basis of his experience or computer-aided be suitable software modules from experimental data.

Acknowledgements

We gratefully acknowledge the financial support from the Thuringian Ministery of Science, Research and Art (TMWFK).

References

[1] Munack, A. (1989): Optimal feeding strategy for identification of Monod-type models by fed-batch experiments. In: Fish, N.M., Fox, R.I., Thornhill, N.F. (eds.): Computer Applications in Fermentation Technology. pp. 195-204. Elsevier, London, New York.

[2] Versyck, K.J., Claes, J.E., van Impe, J.F. (1997): Practical Identification of Unstructured Growth Kinetics by Application of Optimal Experimental Design. Biotechnol. Prog. 13, 524-531.

[3] Glassey, J., Montague, G.A., Ward, A.C., Kara, B.V. (1994): Artificial Neural Network Based Experimental Design for Enhancing Fermentation Development. Biotechnol. Bioeng. 44, 397-405.

[4] Galvanauskas, R., Simutis, R., Volk, N., Lübbert, A. (1998): Model Based Design of a Biochemical Cultivation Process. Bioprocess Eng. 18, 227-234.

[5] Berkholz, R. and Guthke, R. (2001): Model Based Sequential Experimental Design for Bioprocess Optimization - An Overview. Focus in Biotechnology, Vol. IV, Kluwer Academic Press, in press.

[6] Berkholz, R., Guthke, R. and Röhlig, D. (2000): Data and Knowledge Based Experimental Design for Bioprocess Optimization. Enzyme and Microbial Technology 27, 784-788.

[7] Guthke, R., Schmidt-Heck, Pfaff, M. (1998): Knowledge acquisition and knowledge based control in bioprocess engineering. J. Biotechnol. 65, 37-46.

[8] Powell, M.J.D. (1981): A Fast Algorithm for Nonlinear Constrained Optimization Calculations. Numerical Analysis. ed. Watson, G.A.: Lecture Notes in Mathematics, Springer Verlag, Vol. 630.