10th International IFAC Symposium on
Computer Applications in Biotechnology
Proceedings Vol.1, June 4-6, 2007, Cancún, Mexico

PRODUCT FORMATION KINETICS IN A RECOMBINANT PROTEIN PRODUCTION PROCESS

Stefan Gnoth[1], Marco Jenzsch[1], Rimvydas Simutis[2], Andreas Lübbert[1,*]

[1]Center of Bioprocess Engineering, Martin-Luther-University Halle-Wittenberg, D-06120 Halle/Saale, Germany
[2]Institute of Automation and Control Systems, Kaunas University of Technology, LT-3028 Kaunas, Lithuania
** Author for correspondence (Tel: +49-345-5525-942; Fax: +49-345-5527-260;*
E-mail: andreas.luebbert@biochemtech.uni-halle.de

Abstract: Protein formation in recombinant protein production cannot yet be modeled in a way sufficiently accurate for process supervision and control. Here we propose using a new hybrid approach based on mass balances for the state variables involved, where the kinetics are represented by artificial neural networks (ANN). We first demonstrate by means of simulations that this method works well even when the networks are trained on noisy process data. Then, secondly, we show that the method is applicable to real fermentation data. As an accompanying example we use an E.coli culture that produces a recombinant protein, namely the green fluorescent protein GFP, which remains dissolved within the cytoplasm. For this case the ANN resulted in a concrete relationship between the specific product formation rate π, the specific growth rate μ and the specific product concentration p/x. The $\pi(\mu)$-part of the relationship confirms what was obtained with a conventional approach and the additional information about the influence of the specific product concentration characterizes the metabolic load of the cell.

Keywords: Fermentation processes, Process models, Neural networks, Identification Soft sensing

1. INTRODUCTION

Bacterial fermentation is a major workhorse for producing recombinant therapeutic proteins; hence, it is very desirable to derive optimal fermentation control strategies.

As the mass of product that can finally be purified from the culture depends of the amount of biomass x employed and their performance, represented by their specific product formation rate π, one is interested in high cell density cultivations with well performing cells (Lee 1996, Riesenberg and Guthke 1999). In most industrial production systems, both factors determining product mass are primarily dependent on the specific biomass growth rate μ. This may be trivial for x, but is in most cases also valid for π: The growth rate that a particular fermentation medium supports, determines the physiological state of the cells and particularly the cell's protein-synthesizing machinery, and in most industrially relevant cases, recombinant protein production is under growth rate control (Neidhardt et al. 1990). Consequently, much work has been devoted to controlling the specific biomass growth rate in fermentation processes (Shioya 1992, Yoon et al. 1994, Levisauskas et al. 1996, Kim et al. 2004, Picó-Marco et al. 2005, Jenzsch et al. 2005, and 2006a, Soons et al. 2006). Numerical exploitable models of fermentation processes for recombinant protein manufacturing thus need a sufficiently accurate submodel relating the specific growth rate μ to the specific product formation rate π, the so-called π-μ-relationship (Pirt 1993).

Traditionally, optimal process trajectories have been obtained from mechanistic models of the processes under consideration (e.g. Levisauskas et al. 2003).

The latter can be derived step-by-step where the actual model version is used to compute the optimal process procedure, e.g. in terms of the productivity with respect to the product, and improvements of a model are deduced from the deviations between the predicted values and those measured in a validation experiment (Galvanauskas et al. 1997, 2004).

Here we propose a new alternative to this basic approach which is purely data-driven. It has the disadvantage of needing much data, but the decisive advantage of not being restricted by unproven model assumptions. At running production plants the supply of many data records is not a problem at all, hence, in these cases, the advantages clearly prevail.

The method proposed is based on artificial neural networks describing the more or less insufficiently known process kinetics within a well known set of basic mass balance equations. Since such hybrid modeling usually suffers from the fact that there are no directly measurable data for the key variables, e.g. μ and π, we must train the artificial neural networks depicting the really interesting kinetic relationships indirectly, extending the work of Simutis and Lübbert (1997). We solved this problem by a stepwise training of neural networks using online measured variables and, additionally, corresponding off-line values for the amount of biomass x, and total product mass p. The result of this training procedure is a $\pi(\mu)$-profile which can be used for process simulation, and finally in process supervision and control.

Validation of the model was performed at the example of E.coli fermentations, where the soluble GFP, the green fluorescence protein was produced in its active form within the cells' cytoplasm.

2. STRUCTURE OF THE DATA-DRIVEN MODEL

2.1 General idea behind the model.

The backbone of the process model is a classical system of mass balance equations for all species, the masses of which are changing significantly during the cultivation process. The components considered here are total biomass x, and total product mass p.

The first step in modeling the kinetics is representing the specific growth rate μ. It can be determined using nonlinear relationships in form of an ANN with important process variables such as carbon dioxide production rate (CPR), total biomass x, time after induction t_{ai} , etc.. Also, other online variables can be used to strengthen this relation, e.g., the oxygen uptake rate, as well as the base fed into the reactor during pH control.

This specific growth rate representation can directly be used within balance equations determining the amount of biomass. In the upper part of Figure 1 this procedure is schematically shown. It can be interpreted as an ANN-aided software sensor estimating the total biomass x. Once this ANN is trained, it can supply μ(t)-values for training the a second artificial neural network computing π. This procedure is shown in the lower part of Figure 1.

2.2 Training of the artificial neural network system

Simple feedforward networks are used that map the input variables across a hidden layer of 5 nodes (hyperbolic tangent) onto a single output variables μ or π respectively. As already mentioned, online measurements data (CPR, t_{ai}, ...) are used as inputs together with biomass x and product p, estimated in the time step before. For network training we used offline measurement data for biomass x as well as total product mass p from previously performed experiments.

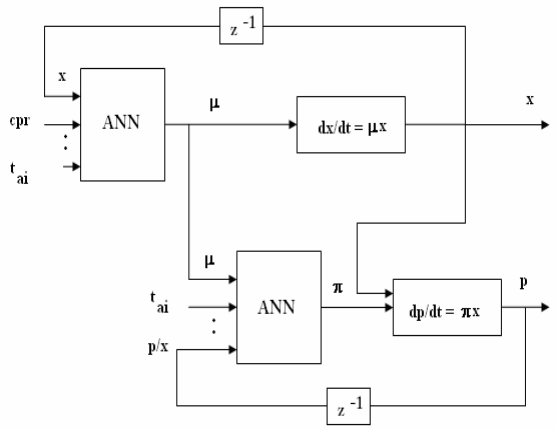

Figure 1: Scheme of the proposed procedure for identification of the $\pi(\mu)$-relationship. The artificial neural networks (ANN) are feedforward networks with a single hidden layer.

Network training was based on the sensitivity equations approach (e.g., Schubert et al., 1994). This can be applied to train neural networks, which are incorporated into differential equation systems of the form

$$\frac{dy}{dt} = f(y(t), W) \qquad (1)$$

where W are weights of the neural network, and y are process state variables. The training is essentially a fit of this equation to experimental offline measurement data for biomass x and product p. Its efficiency can be improved if the gradients $\partial y / \partial W$ can be exploited. These gradients satisfy an ordinary differential equation that can easily be derived from equation (1) by partial derivatives with respect to the weights W.

$$\frac{d}{dt}\frac{\partial y}{\partial W} = \frac{\partial f}{\partial y}\cdot\frac{\partial y}{\partial W} + \frac{\partial f}{\partial W}, \; with \; \left.\frac{\partial y}{\partial W_p}\right|_{t=0} = 0 \qquad (2)$$

Equation (2) is referred to as the sensitivity equation. With the solutions, the $\partial y/\partial W$ values, the well known neural networks training procedures (backpropagation, gradient methods, cf. e.g., Rumelhard and McClelland 1986) can be applied to train the neural network.

The sensitivity equation approach for specific growth rate estimation appears when y is replaced by x, the biomass and equation (1) is specified by the equation defining the specific growth rate μ