Original paper

Bioprocess Biosyst Eng 25 (2003) 255–262
DOI 10.1007/s00449-002-0305-x

Model-based optimization of viral capsid protein production in fed-batch culture of recombinant *Escherichia coli*

D. Levisauskas, V. Galvanauskas, S. Henrich, K. Wilhelm, N. Volk, A. Lübbert

Abstract An optimized fed-batch cultivation process for the production of the polyoma virus capsid protein VP1 in recombinant *Escherichia coli BL21* bacteria is presented. The optimization procedure maximizing the amount of desired protein is based on a mathematical model. The model distinguishes an initial cell growth phase from a protein production phase initiated by inducer injection. A new approach to model the target protein formation rate was elaborated, where product formation is primarily dependent on the specific biomass growth rate. Lower growth rates led to higher specific protein concentrations. The model was identified from a series of fed-batch experiments designed for parameter identification purposes and possesses good prediction quality. Then the model was used to determine optimal open-loop control profiles by manipulating the substrate feed rates in both phases as well as the induction time. Feed-rate optimization has been solved using Pontryagin's maximum principle. The solution was validated experimentally. A significant improvement of the process performance index was achieved.

Keywords Model-based optimization, Model for recombinant protein production, Pontryagin's maximum principle, Fed-batch *Escherichia coli* cultivation, Virus capsid protein

List of symbols

F_1	mass flow rate in fermenter due to alkali feed, medium evaporation, and carbon loss with CO_2 in off-gas, kg/h;
F_2	sampling rate in fermenter, kg/h;
H	Hamilton function, U/h;
J	total protein amount, process performance index, U;
k_i	inhibition constant for substrate consumption, g/kg;
$k_{i\mu}$	inhibition constant in protein model, U/g;
k_m	proportionality coefficient in protein model, U/g;
k_s	Monod constant for substrate consumption, g/kg;
k_μ	Monod constant in protein model, 1/h;
m	biomass maintenance term, g/g/h;
m_{ox}	oxygen uptake parameter related to biomass maintenance, g/g/h;
OUR	oxygen uptake rate, g/kg/h;
p_{max}	maximal specific protein activity, U/g biomass;
p_x	specific protein activity, U/g biomass;
q_{px}	specific protein accumulation rate, U/g/h;
q_s	specific substrate consumption rate, g/g/h;
s	substrate concentration, g/kg;
s_f	substrate concentration in feeding solution, g/kg;
t	current process time, h;
t_f	process duration, h;
t_{ind}	induction time, h;
T	medium temperature, °C;
T_{ref}	reference medium temperature, °C;
u	substrate feed rate, control variable, kg/h;
w	culture mass, kg;
x	biomass concentration, g/kg;
X	biomass amount in fermenter, g;
Y_{ox}	oxygen uptake yield related to biomass growth, g/g;
Y_{xs}	biomass/substrate yield, g/g;
α	temperature constant, 1/°C;
μ	specific biomass growth rate, 1/h;
μ_{max}	maximal specific biomass growth rate, 1/h;
λ_p	adjoint variable, related to p_x, g;
λ_X	adjoint variable, related to X, U/g

Received: 15 July 2002 / Accepted: 11 September 2002
Published online: 16 November 2002

D. Levisauskas, V. Galvanauskas
Process Control Department, Kaunas University of Technology, Studentu 50, 3028 Kaunas, Lithuania

V. Galvanauskas, S. Henrich, K. Wilhelm, N. Volk, A. Lübbert (✉)
Institute of Bioengineering, Martin-Luther University, Halle-Wittenberg, 06099 Halle/Saale, Germany
E-mail: Andreas.Luebbert@iw.uni-halle.de
Tel.: +49-345-5525942
Fax: +49-345-5527260

Financial assistance of DAAD to DL and Roman Herzog fellowship of Alexander von Humboldt Foundation to VG are gratefully acknowledged. The experiments were performed within a project financed by the German Federal Ministry for Education and Science. The bacterial strain used was kindly provided by Dr. H. Lilie and Prof. R. Rudolph, Institute of Biotechnology, Martin Luther University Halle-Wittenberg, Germany.

1 Introduction

During the last two decades, more and more recombinant proteins have been produced. In this period, time-to-market was the main objective for process development and the processes were often not thoroughly optimized. In the near future, some of these first products will be running out of patent [1], and particularly in such cases the efficient process optimization with respect to the benefit/cost ratio will become an important issue in biopharmaceutical companies.

Straightforward approaches to process optimization and control rest on mathematical process models and their exploitation with numerical optimization methods (e.g. [2, 3]). Applications in industry require compact representations of the relations between the process variables that can be adjusted and the process performance measure. Further, fast and reliable numerical optimization procedures should be used. The models must be accurate enough to describe the relevant process features, and, finally, it should be possible to exploit them such that the optimal values of the control variables can be determined within the constraints imposed by the real processes.

The foundations of model-based process optimization have already been discussed in the literature. Ryu et al. [4] discussed model-based optimization of recombinant fermentation processes. Lim and Lee [5] discussed optimization of biotechnological cultivation processes from the control engineer's perspective and suggested Pontryagin's maximum principle [6] for determining optimal feed-rate profiles. The maximum principle leads to a rigorous analytical solution for the control profiles. It is thus preferred to nonlinear programming or stochastic optimization methods for optimization of nonlinear systems, where convergence problems may appear. However, the application of the maximum principle requires mathematical treatment of the process model.

In this paper, using the example of heterologeous virus capsid protein production in *E. coli*, we show that the relevant features of production processes for recombinant proteins can be described by relatively simple process models. The mechanistic model proposed describes the substrate consumption by the *E. coli* cells, the accumulation in the cells' biomass, and the formation of the desired viral capsid protein within the bacteria. We are not aiming at a comprehensive description of recombinant protein production. Instead, we focus on a compact accurate representation of the specific VP1 production process under consideration. The main advantage of the then simpler model is that its parameters can be identified more reliably with the limited amount of data available.

The procedure reported is used to identify the process model using data measured during a few well planned experiments, and then to use the model for calculating the optimal open-loop control strategy, which maximizes the amount of product that can be harvested within a fixed cultivation time. Pontryagin's maximum principle is employed to determine the optimal feed-rate profile directly, thus avoiding iterative or stochastic optimization methods. It is shown that the algorithm can be kept compact and fast, so that it can be used for online optimization as well. The resulting control strategy is then tested experimentally. A comparison of the measurement data obtained in this validation experiment with the model predictions was performed in order to test the extrapolation capability of the model.

The process considered is the production of the recombinant protein VP1 in *E. coli* bacteria. The three proteins forming the envelope of a polyoma virus are named VP1, VP2, and VP3 . VP1 makes up about 75% of the envelope, and can be expressed in *E. coli* cells, where it appears within the cytoplasm in soluble form [7]. Interestingly, artificial capsid envelopes can be constructed with VP1 only [8]. After incorporation of DNA material into such *in vitro* constructed virus envelopes, transfection into eukaryotic cells proved to be possible [9]. Consequently, VP1 is considered as an interesting species for constructing artificial virus envelopes for gene transfer purposes.

2
Materials and methods

2.1
Microorganism, medium and culture

The bacterial host used in this study was *E. coli* BL21 harboring a derivate of the pBR322 plasmid. This contains a gene encoding for VP1 fused with the enzyme dihydrofolreductase (DHFR, EC 1.5.1.3). Additionally it has a gene segment that confers ampicillin resistance. The VP1-DHFR gene was expressed under stringent control of the *tac* promoter. The latter was induced using 1.5 mM isopropyl-β-D-thio-galactopyranosidase (IPTG).

Pre-cultures were developed by inoculating 100 ml defined medium containing 5 g/l glucose with 2 ml of frozen culture at 37°C in shaker flasks. After the initial amount of the supplied substrate is depleted, the cells were centrifuged, washed and transferred to the main 10 l fermenter (Biostat C+, B. Braun Biotech International, Melsungen, Germany). The fermenter was operated under the local control of B. Braun's DCU, supervised by the MFCS/Win software (B. Braun Biotech International, Melsungen, Germany).

The fermenter was equipped with 2 standard T/3-Rushton turbines and 4 flow baffles in standard configuration. The stirrer speed was varied between 400 and 1200 rev/min and the aeration rate between 1 and 2 vvm in order to keep the dissolved oxygen level above 25%.

pH was controlled at 7.0 with 25% ammonium hydroxide and 20% phosphoric acid. Temperature was controlled at 37°C before induction, but afterwards temperature was reduced to 32°C, which proved favorable for soluble protein production.

Fermentations were started as batch cultures. Upon the consumption of all of the initially supplied substrate glucose, which was observed by an increase in the dissolved oxygen concentration, the operation mode was switched to fed-batch.

Besides glucose, the defined medium used during preculture as well as the fed-batch cultivations contained: KH_2PO_4 13.3 g/l, $(NH_4)_2HPO_4$ 4 g/l, citric acid 1.7 g/l, EDTA 8.4 mg/l, $CoCl_2$ $6H_2O$ 2.5 mg/l, $MnCl_2$ $4H_2O$ 15 mg/l, $CuCl_2$ $2H_2O$ 1.5 mg/l, H_3BO_3 3 mg/l, Na_2MoO_4 $2H_2O$ 2.5 mg/l, $ZnSO_4$ $7H_2O$ 2 mg/l, $FeSO_4$ $7H_2O$ 200 mg/l and 2 ml/l of trace element solution.

Composition of this trace element solution was: $CaCl_2$ $2H_2O$ 0.5 g/l, $ZnSO_4$ $7H_2O$ 0.18 g/l, $MnSO_4$ H_2O 0.1 g/l, Na_2-EDTA 20.1 g/l, $FeCl_3$ $6H_2O$ 16.7 g/l, $CuS0_4$ $5H_2O$ 0.16 g/l, CoCl $6H_2O$ 0.18 g/l.

Depending on the experiment, the initial biomass, glucose concentrations, culture broth mass, induction time, and glucose concentrations in the feed were as specified in Table 1.

Table 1. Initial conditions, induction time, and substrate concentrations in feed in experiments 1–4

Abbreviation	Value in experiment				Dimensions
	1	2	3	4	
$x(0)$	0.143	0.119	0.116	0.116	g/kg
$s(0)$	10.7	10.1	9.6	9.5	g/kg
$w(0)$	4.7	4.9	4.95	5.05	g/kg
t_{ind}	5.27	5.47	6.75	8.0	h
s_f	100.	100.	156.	151.	g/kg

2.2 Assays

Cell concentrations were obtained by measuring the absorption in the culture broth at 600 nm with a spectrophotometer (UV-2102PC, Shimadzu Scientific Instruments, Inc., Columbia, Md., USA), and recalculating it to dry cell mass concentration by means of a predetermined correlation.

Glucose was measured during the cultivation enzymatically with an analyzer (YSI 2700 Select, Yellow Springs Instruments, Springfield, Oh, USA).

As the product, the viral capsid protein VP1 of the murine polyoma virus is fused with the enzyme dihydrofolate reductase (DHFR, EC 1.5.1.3), the protein concentration could easily be measured via the activity of its enzyme moiety. For product quantification it is assumed that the measured enzyme activity is proportional to the concentration of the desired protein. The desired product is the protein construct that remained dissolved within the cytoplasm of the bacterial cell. A sample with a fixed amount of biomass is taken. The biomass is first separated by centrifugation. Then, the cells within the pellet are disrupted in a ball mill and, after a further centrifugation, the protein in the supernatant is probed.

The following DHFR-activity measurement is derived from the method of Ginkel et al. [10]. Raw enzyme extract (100 μl) was added to a sample of 850 μl. This was augmented with 50 mM potassium phosphate, 1 mM EDTA, 10 mM mercaptoethanol and 0.1 mM NADPH and 50 μl 0.1 mM dihydrofolate. In this solution, DHFR catalyses the reduction of dihydrofolate, using $NADPH/H^+$, to tetrahydrofolate and $NADP^+$. The change in the $NADPH/H^+$ concentration was determined by an extinction measurement at 340 nm wavelength. Under these conditions, the change in the extinction is directly proportional to the activity of DHFR.

3 Modeling of the physical system

3.1 Model description

Identification of the model structure and its parameters was based on experimental data obtained from fed-batch laboratory scale cultivations.

The model is based on mass balances for biomass x and substrate s. As the processes were performed in the fed-batch mode, an equation for the culture mass w is added. Additionally, an equation for the specific activity of the desired recombinant protein is considered. This is a balance around a representative cell, formulated in terms of p_x [units/cell mass], and does not refer to the protein concentration usually taken in the medium, in so far as it must be distinguished from the balances of biomass and substrate around the reactor. The variables defining the state vector $\boldsymbol{c}$ are $[x, s, p_x, w]$.

The following structure of the mathematical model was obtained:

$$\frac{dx}{dt} = \mu(s)x - (u + F_1)\frac{x}{w} \tag{1}$$

$$\frac{ds}{dt} = -q_s(s)x + u\frac{s_f}{w} - (u + F_1)\frac{s}{w} \tag{2}$$

$$\frac{dp_x}{dt} = q_{px}(\mu, p_x) \tag{3}$$

$$\frac{dw}{dt} = u + F_1 + F_2 \tag{4}$$

Note that the rates by which the culture mass w changes, is split into three mass flows: u is the main control variable, the rate by which substrate s is fed to the reactor. F_1 considers the changes in culture mass w through alkali addition, CO_2 outflow through the vent line, and evaporation. These components are computed, using separate algebraic equations and the state variables obtained during the model (1–4) integration procedure, as described by Galvanauskas et al. [11]. F_2 considers the changes due to sampling for offline analysis.

The dependency of the specific biomass growth rate μ on substrate concentration s and temperature T is described as follows:

$$\mu(s, T) = \mu_{max}\frac{s}{k_s + s}\cdot\frac{k_i}{k_i + s}\exp(\alpha(T - T_{ref})) \tag{5}$$

Besides the Monod kinetics, expression (5) also considers possible substrate inhibition effects. This expression was taken for both the biomass production as well as the protein formation phase. The difference between both phases is that the temperature during the protein production phase is lower than the one adjusted during the biomass growth phase, where it was T=T_{ref}=37°C. This reduces the maximally possible biomass growth rate during the induction phase.

The desired temperature for the protein production phase was found from exploratory batch experiments performed beforehand [12], where the temperature was changed over a relatively wide range of 17–37°C. The experimental results showed that the maximal specific protein concentration is to be expected at a cultivation temperature of T=32°C.

As the optimal values of temperature in the biomass growth and protein production phases are determined beforehand, the exponential term in model equation (5) can be considered a constant parameter, the value of which depends on the process phase. Therefore, the specific growth rate in model equations (1, 2, 3) is considered a function of the substrate concentration only.

The specific substrate consumption rate q_s is assumed to be mainly dependent on the specific biomass growth rate μ. Additionally, maintenance requirements of the cells by means of a term m are considered:

$$q_s(s) = \frac{1}{Y_{xs}}\mu(s) + m \tag{6}$$

Figure 1a–c depicts the measurement data for biomass and substrate concentrations in experiments 1–3. These experiments were performed in order to provide data for process model identification. The curves depict the corresponding modeling results obtained after fitting the model to all three data sets with the same parameter set (see Table 2).

The kinetic expression for the rate by which the specific product concentration p_x changes, which was also derived from the experimental observations, is very important to the objective of the project. In order to get reliable data about the dynamics of the product development, experiments 1–3 were designed in such a way that the specific growth rate μ was held constant at different values for periods of about 1.5–3 h (see Fig. 2a–c). Hence, the specific protein concentrations, as depicted in Fig. 2a–c, could be estimated quite reliably for the adjusted growth rates.

Separate analyses of the data from the time intervals within which the specific growth rate remained unchanged suggested that the specific concentration of the recombinant protein, which accumulates within biomass, approaches a maximal value. The approached maxima p_{max} is different at different specific biomass growth rates. Higher specific product concentrations were observed at lower specific growth rates, and vice versa, lower specific product concentrations were achieved at higher growth rates (see Fig. 2a–c).

In order to quantify these observations, the rate of change of the specific protein concentration q_{px} in Eq. (3) was described by a first-order dynamical process in the following way:

$$q_{px}(\mu, p_x) = \frac{1}{T_{px}}\left(p_{max}(\mu) - p_x\right) \tag{7}$$

where T_{px} is the corresponding time constant. This approach can be considered as a form of self-inhibition effect [13]. However, we assume that the maximal specific product concentration value p_{max}, approached asymptotically, to be dependent on μ. p_{max} was observed to decrease with higher biomass growth rates μ, but is expected to be very low when the specific biomass growth rate μ approaches zero. This function was modeled by means of an equation of the following structure:

$$p_{max}(\mu) = \frac{\mu k_m}{k_\mu + \mu + \mu^2/k_{i\mu}} \tag{8}$$

Note, that no recombinant protein is produced during the biomass growth phase. For the growth phase, Eq. (7) is replaced by q_{px}=0.

The experimental data depicted in Fig. 2a–c were used to identify the parameters in Eqs. (7) and (8). The resulting fit shows that the agreement between model and measured data is satisfactory, particularly if one takes into account the significantly different operational conditions under which the process was examined.

The conventional approach used for example by Miao and Kompala [13] and Bellgardt [14] was tested for its ability to describe the measurement data presented in this paper. This showed that our measurement data was accurate when compared with the representation proposed here. However, with respect to process optimization, the approach [13] requires additional state variables and more complex kinetics to avoid unrealistic control profiles. The latter tend to maintain a specific biomass growth rate in the protein production phase close to zero. Furthermore, the increased number of modeled, but not measured, variables and free model parameters reduce the extrapolation capabilities of the model. This is particularly important at the beginning of the process investigation providing that the number of measurement data available is quite small.

Parameter identification was performed using an evolutionary programming algorithm [15]. The resulting model parameter values identified from experiments 1–3 are given in Table 2.

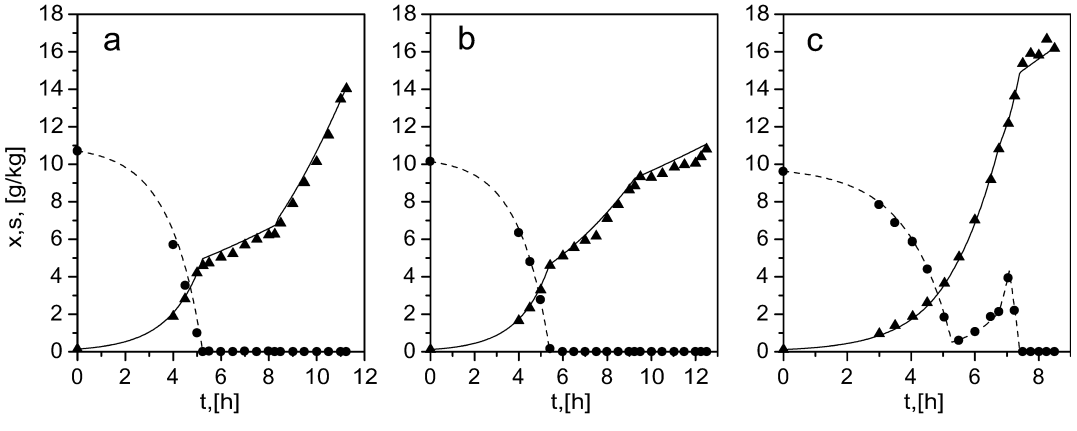

Fig. 1a–c. Biomass (*solid triangles*) and glucose (*solid circles*) concentrations measured during three fed-batch cultivations (subplots **a–c** correspond to experiments 1–3 respectively) together with the corresponding model outputs (*solid line* and *dashed line*) determined from Eqs. (1–9)

Table 2. Values of the model parameters, identified from experiments 1–3

Abbreviation	Value	Dimensions
k_i	93.8	g/kg
$k_{i\mu}$	0.0174	1/h
k_m	751	U/g
k_μ	0.61	1/h
k_s	0.00333	g/kg
m	0.0242	g/g/h
T_{px}	1.495	h
Y_{xs}	0.46	g/g
α	0.0495	1/°C
μ_{max}	0.737	1/h

3.2
Feed-rate optimization

The goal of the optimization is to maximize the total amount of recombinant protein at the end of the cultivation process. Manipulated variables are the inducer injection time t_{ind} and the feed-rate profile $u(t)$.

The mathematical model applied for solving of the optimization problem consists of the set of generalized state Eqs. (1), (2), (3), and (4) and the functional relationships (Eqs. 5, 6,, 7, and 8), described in the previous section. The model parameter set depicted in Tables 2 and 3 was used in the calculations.

3.2.1
Formulation of the optimization problem

The objective function, as quantified by:

$$J = p_x(t_f)x(t_f)w(t_f) \rightarrow \max \tag{9}$$

has to be maximized while satisfying the following constraints: the cultivation time t_f was limited to 10 h, and the feed pumps allowed feed-rate variations in the interval

$$0 \leq u \leq u_{max} \tag{10}$$

As the exploratory batch experiments [12] performed beforehand suggested that the cells become irreparably injured when they are exposed to extremely small oxygen concentrations during production, pO_2 was kept above 25%. This lead to another constraint that is discussed later in more detail.

3.2.2
Solving of the optimization problem

The optimization procedure was divided into three parts.

The first subproblem is to maximize the amount of cell biomass at a given induction time t_{ind}.

The second subproblem is to maximize the amount of recombinant protein at the end time t_f.

The third subproblem is to find the maximum of the performance index (9) by repeatedly solving of subproblems 1 and 2, and varying induction time t_{ind}, i.e. the maximal amount of the desired protein at the end of the cultivation.

Solving the first subproblem

The objective function in the first subproblem is:

$$J = x(t_{ind})w(t_{ind}) \rightarrow \max \tag{11}$$

where t_{ind} is a given induction time instant.

Model equations (Eqs. 1, 2, 3, and 4) are solved with initial conditions $x(0)$, $s(0)$, $p_x(0)=0$, $w(0)$ (see Table 1, experiment 4) to determine the feed rate u_{opt}, which maximizes the objective function (11). The solution of this optimization subproblem is trivial: the objective function (11) is maximized if the specific growth rate is controlled at its maximum value. The optimal feed rate, maximizing the specific growth rate, can be calculated from Eq. (2) at $ds/dt=0$:

$$u_{opt} = w\frac{q_s(s_{opt})x}{s_f - s_{opt}} + \frac{s_{opt}F_1}{s_f - s_{opt}} \tag{12}$$

where s_{opt} is the s value satisfying the condition

$$\frac{d\mu(s)}{ds} = 0 \tag{13}$$

Solving the second subproblem

Optimization of the feed rate within the recombinant protein production phase, i.e., the time interval [t_{ind} t_f],

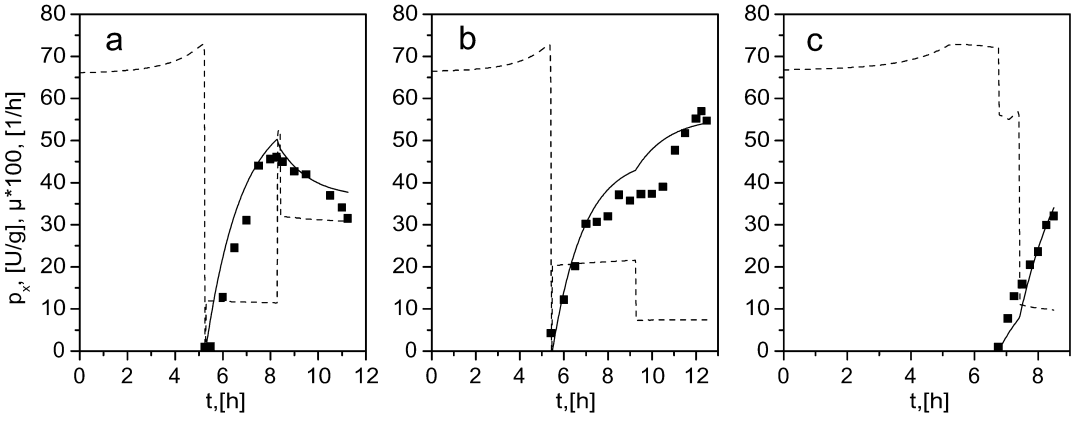

Fig. 2a–c. Specific concentration of viral capsid protein VP1, p_x (*solid squares*) measured during three fed-batch cultivations (subplots **a–c** correspond to experiments 1–3 respectively) together with the corresponding model output (*solid line*), and specific biomass growth rate, μ (*dashed line*) determined from the Eqs. (1), (2), (3), (4), (5), (6), (7), (8), and (9)

Table 3. Values of the parameters in Eq. (16), identified from experiments 1–3

Abbreviation	Value before induction	Value after induction	Dimensions
m_{ox}	0.15	0.015	g/g/h
Y_{ox}	0.5	0.75	g/g

is solved using Pontryagin's maximum principle (see Appendix). Model equations (Eqs. 1, 2, 3, and 4) are solved with initial conditions $x(t_{ind})$, $s(t_{ind})$, $p_x(t_{ind})=0$, $w(t_{ind})$ to determine the feed rate u_{opt}, which maximizes the objective function (Eq. 9):

$$u_{opt} = w\frac{\left(\frac{d\mu}{dt}\right)_{opt}\Big/\frac{\partial\mu}{\partial s} + q_s x}{s_f - s} + \frac{sF_1}{s_f - s} \tag{14}$$

where

$$\left(\frac{d\mu}{dt}\right)_{opt} = \frac{\frac{\partial q_{px}}{\partial\mu}\frac{\partial q_{px}}{\partial p_x} - \frac{\partial^2 q_{px}}{\partial\mu\partial p_x}q_{px}}{\frac{\partial^2 q_{px}}{\partial\mu^2}} \tag{15}$$

The corresponding derivatives can be directly determined from the model equations (5), (7) and (8) already discussed.

Solving the third subproblem

Once the algorithm for determining the optimal profiles (subproblems 1 and 2) corresponding to a predetermined induction time t_{ind} is implemented, one can vary this single parameter by means of a simple search in order to find the t_{ind} corresponding to the maximal value of the performance index (Eq. 9). This can be done by applying any search algorithm, e.g., the golden search method. Alternatively, as the computation time for a single optimization run is below a second on a standard PC, it can be solved by polling across the relevant time interval with a time increment Δt_{ind} corresponding to the time resolution of the experiments. For the experiments discussed, Δt_{ind}=3 min proved to be sufficient for such a polling.

3.3 Optimization results

The optimal feed-rate time profiles were calculated by integrating the model equations (1, 2, 3, 4, 5, 6, 7, 8). The feed rate in the growth phase was determined by relationship (12) while relationships (14) and (15) were used for the calculation during the protein production phase. If the calculated feed-rate values violate restriction (10), they are replaced by the boundary values.

In practice, one additional restriction on the substrate feed rate was introduced in order to avoid oxygen limitation in the reactor. The oxygen uptake rate was modeled by means of

$$OUR = Y_{ox}\mu\, x + m_{ox}x \tag{16}$$

The relationship (16) together with Eqs. (5), (6), and (12) are used to calculate the maximal feed rate u, which does not lead to violation of the oxygen supply restriction. Such a restriction of the feed rate was implemented in the control algorithm in experiment 4 during the cultivation time interval between t=7 h and t=8 h. This is illustrated in Fig. 4, where the oxygen uptake rate approaches the constraint of $OUR \le 6$ g/kg/h (see Fig. 4a), and u (see Fig. 4c) is continuously reduced to maintain the necessary μ (see Fig. 4b) and OUR. The identified values of the parameters Y_{ox} and m_{ox} are presented in Table 3.

The optimal induction time t_{ind}=8 h was determined. The biomass concentration, the specific activity of recombinant protein and the optimized feed-rate time profiles maximizing the amount of recombinant protein at t_f=10 h are depicted in Fig. 3 and 4c. Note, the curves are model predictions and the symbols are measurement values obtained from the subsequently performed experiment.

The optimized feed-rate profile in the protein production phase depicts a specific biomass growth rate μ=0.12 1/h at $t=t_{ind}$=8 h, which is then monotonically reduced to μ=0.105 1/h at the end of cultivation. These values of the specific growth rate are slightly higher than the one that maximizes p_{max} in Eq. (8). The optimal specific growth rate μ depends on the structure of Eqs. (14) and (15).

Until t=7 h , the specific growth rate was slightly smaller than expected, as can be seen in Fig. 3: where only a level of about 95% was obtained when compared to the predicted values. Nevertheless, at the predetermined induction time, the predicted biomass concentration could be achieved.

Specific protein activities in the cell mass were found in the predicted range. The total productivity J of the process also met the predicted values.

The experimental results showed notable improvements in the productivity with respect to the total amount of desired protein (Eq. 9) as compared to the exploratory experiments 1–3 (see Fig. 5).

4 Conclusions

The mechanical mathematical model proposed for prediction and optimization of the operational procedure in recombinant protein production processes reliably predicts the experimental data.

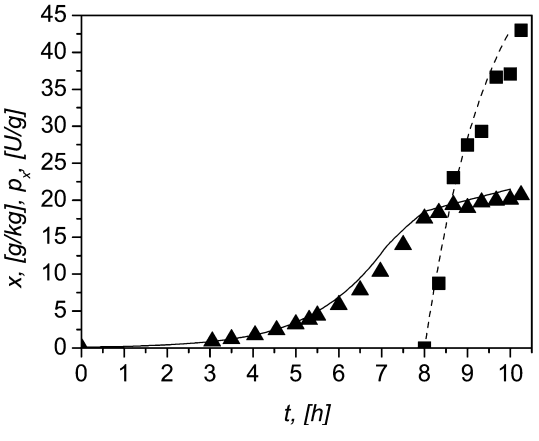

Fig. 3. Predicted trajectories of the state variables: biomass concentration, x (*solid line*), and specific protein concentration, p_x (*dashed line*) in experiment 4, where t_{ind}=8 h. Additionally, the obtained experimental data for biomass (*solid triangles*) and specific protein (*solid squares*) concentrations are depicted

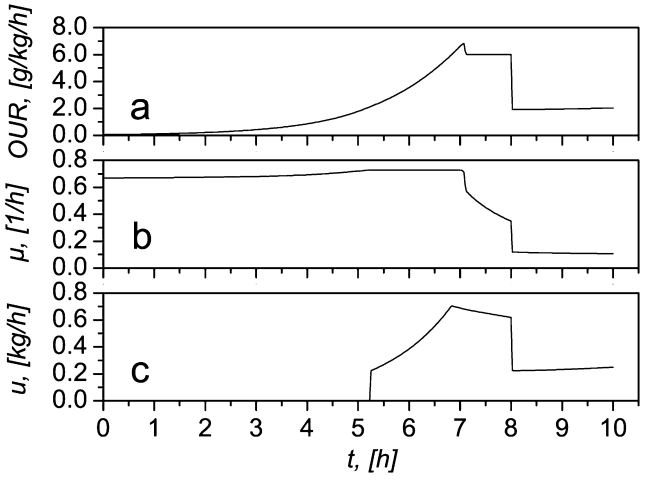

Fig. 4a–c. Profiles of the process variables in experiment 4. **a** predicted oxygen uptake rate, *OUR*; **b** predicted specific biomass growth rate, μ; **c** optimized substrate feed rate, u

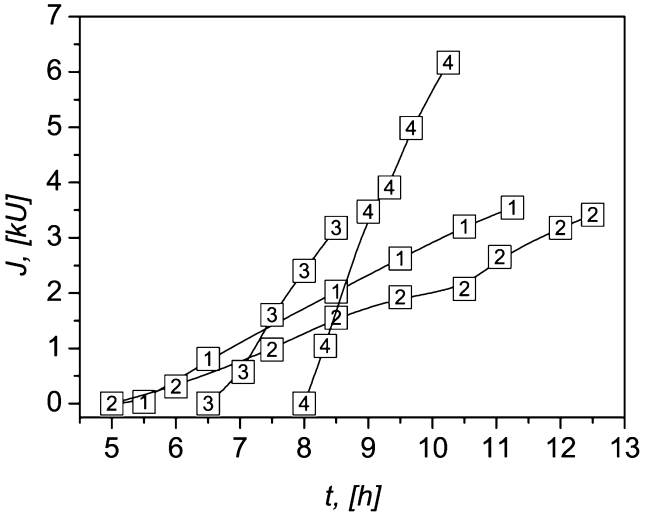

Fig. 5. Process performance index (Eq. 9) achieved in exploratory experiments 1–3 (*boxed 1*, 2, and *3* respectively) and in optimized experiment 4 (*boxed 4*), calculated from the experimental data

Equations (3), (7) and (8) describe the dynamics of the protein formation in a different way from the conventional approaches. The dynamics of the protein formation reflect the experimentally observed fact that the desired protein VP1 is preferentially produced at low specific growth rates of ~0.10–0.12 1/h, indicating that the metabolic load of the cells should be kept lower in order to produce a larger amount of soluble protein within the given cultivation time. The adaptation of the cells to a new physiological protein production state, as characterized by a new specific growth rate, is dynamically described by a first-order dynamical process. The small number of parameters in the process model allows a reliable identification of the parameters from the usually quite small experimental data base.

The final model was used to predict the optimal feeding strategy. With Pontryagin's maximum principle, the optimization problem could be solved directly. The resulting optimal control strategy was tested in a validation experiment. A good agreement between prediction and experimental data was observed, indicating the good prediction properties of the model. As compared to the initially obtained amounts of product, the optimization led to a significant improvement in the process performance.

The computation time required to determine induction time and the corresponding feed-rate profiles is a few seconds only. Hence, the procedure proposed in this report can be applied for the online correction of the optimal induction time and the feed-rate profile. This is of practical advantage wherever deviations in the operating conditions from those assumed during the optimization appear. To perform an online re-optimization, online estimations of the biomass growth and substrate consumption rates together with measurements of these variables are required. Based on the deviations between the modeled and measured quantities, the model parameters in Eqs. (5) and (6) should be re-identified online. Nevertheless, an appropriate minimal number of measurement points is necessary in order to assure the convergence of the identification algorithm.

As measurement values of the specific protein concentrations are not available during the cultivation, the online correction of the parameters in Eqs. (7) and (8) is not possible. Thus, a recalculation of specific protein concentrations must be based on the assumption that the product formation dynamics remains the same as identified from the previously performed experiments.

After model re-identification, optimal feed-rate profiles and the induction time can be recalculated using the procedure described in the optimization section. As initial conditions, the currently estimated process state must be used.

As the proposed model is simple and the proposed algorithm allows a quick calculation of optimal feed-rate profiles, the optimization approach presented is an attractive solution to optimization problems in industrial recombinant protein production processes.

Appendix: Employing Pontryagin's maximum principle for feed-rate optimization

The optimization problem involves maximizing the objective function (9), when process state variables are related by model equations (1), (2), (3) and (4) and the restrictions (10) are imposed on the duration of the process and the control variable.

The original problem given by Eqs. (1), (2), (3), (4), (9), and (10) can be reduced by introducing a new state variable, total biomass $X=xw$, and a new control variable, specific growth rate μ.

With the new variables, the optimization problem can be formulated as follows:

Model equations:

$$\frac{dX}{dt} = \mu X \tag{A1}$$

$$\frac{dp_x}{dt} = q_{px}(\mu, p_x) \tag{A2}$$

Objective function:

$$J = p_x(t_f)X(t_f) = \int_0^{t_f} \frac{d(p_x X)}{dt} dt$$

$$= \int_0^{t_f} \left(p_x \mu X + q_{px}(\mu, p_x)X\right) dt \rightarrow \max \quad (A3)$$

In using the maximum principle for the problem given by Eqs. (A1), (A2), and (A3), the Hamiltonian becomes:

$$H = p_x \mu X + q_{px}(\mu, p_x)X + \lambda_X \mu X + \lambda_p q_{px}(\mu, p_x) \quad (A4)$$

Equations for the adjoint variables are:

$$\frac{d\lambda_X}{dt} = -\frac{\partial H}{\partial X} = -p_x \mu - q_{px}(\mu, p_x) - \lambda_X \mu \quad (A5)$$

$$\frac{d\lambda_p}{dt} = -\frac{\partial H}{\partial p_x} = -\mu X - \frac{\partial q_{px}}{\partial p_x} X - \lambda_p \frac{\partial q_{px}}{\partial p_x} \quad (A6)$$

From the transversality conditions, the final values of the adjoint variables are:

$$\lambda_X(t_f) = 0 \quad (A7)$$

$$\lambda_p(t_f) = 0 \quad (A8)$$

According to the necessary conditions of optimality, the Hamiltonian must remain constant along the optimal trajectory:

$$H = C \quad (A9)$$

where C is some constant related to the fixed duration of the controlled process.

From the maximum principle, the control law must satisfy the following condition:

$$\frac{\partial H}{\partial \mu} = p_x X + \frac{\partial q_{px}}{\partial \mu} X + \lambda_X X + \lambda_p \frac{\partial q_{px}}{\partial \mu} = 0 \quad (A10)$$

As follows from equation (A9), the time derivatives of the Hamiltonian must be identically zero along the optimal trajectory:

$$\frac{d^n H}{dt^n} = 0, \qquad n = 1, 2 \quad (A11)$$

By combining the equations (A1), (A2), (A4), (A5), (A6), (A9), (A10), and (A11) the adjoint variables λ_X and λ_p can be eliminated, and the necessary condition of optimality for the controlled process can be expressed in terms of the control and state variables:

$$\frac{\partial^2 q_{px}}{\partial \mu^2} \frac{d\mu}{dt} + \frac{\partial^2 q_{px}}{\partial \mu \partial p_x} q_{px} - \frac{\partial q_{px}}{\partial \mu} \frac{\partial q_{px}}{\partial p_x} = 0 \quad (A12)$$

From the equation (A12) the optimal change of the specific growth rate can be determined as:

$$\left(\frac{d\mu}{dt}\right)_{opt} = \frac{\frac{\partial q_{px}}{\partial \mu}\frac{\partial q_{px}}{\partial p_x} - \frac{\partial^2 q_{px}}{\partial \mu \partial p_x} q_{px}}{\frac{\partial^2 q_{px}}{\partial \mu^2}} \quad (A13)$$

Returning to the original control variable (feed rate, u), the specific growth rate μ and the feed rate u are related by equation:

$$\frac{d\mu}{dt} = \frac{\partial \mu}{\partial s}\frac{ds}{dt} = \frac{\partial \mu}{\partial s}\left(-q_s x + u\frac{s_f - s}{w} - F_1 \frac{s}{w}\right) \quad (A14)$$

By combining equations (A13) and (A14), the following functional relationship is obtained for the optimal feed-rate calculation:

$$u_{opt} = w \frac{\left(\frac{d\mu}{dt}\right)_{opt} \Big/ \frac{\partial \mu}{\partial s} + q_s x}{s_f - s} + \frac{sF_1}{s_f - s} \quad (A15)$$

References

1. Reid B (2002) US biotech prepares to fight generic biologics. Nature Biotechnol 20:322–322
2. Lee J, Ramirez WF (1994) Optimal fed-batch control of induced foreign protein production by recombinant bacteria. AIChE J 40:899–907
3. Levisauskas D, Galvanauskas V, Simutis R, Lübbert A (1999) Model based calculation of substrate/inducer feed-rate profiles in fed-batch processes for recombinant protein production. Biotechnol Tech 13:37–42
4. Ryu DDY, Kim JY, Lee SB (1991) Bioprocess kinetics and modeling of recombinant fermentation. In: Schügerl K (ed) Biotechnology, vol. 4: Measuring, modeling and control. VCH, Weinheim, Chap. 15
5. Lim HC, Lee KS (1991) Control of bioreactor systems. In: Schügerl K (ed) Biotechnology, vol. 4: Measuring, modeling and control. VCH, Weinheim, Chap. 16
6. Pontryagin LS, Boltyanskii YG, Gamkrelidze RV, Mishchenko EF (1962) The mathematical theory of optimal processes, Wiley, New York
7. Leavitt AD, Roberts TM, Garcea RL (1986) Polyoma virus major capsid protein VP1. Purification after high level expression in *Escherichia coli*. J Biol Chem 260:12803–12809
8. Salunke DM, Caspar DL, Garcea RL (1986) Self-assembly of purified polyomavirus capsid protein VP1. Cell 46: 895–904
9. Forstova J, Krauzewicz N, Sanding V, Elliott J, Palkova Z, Strauss M, Griffin BE (1995) Polyoma virus pseudocapsids as efficient carriers of heterologous DNA into mammalian celss. Hum Gene Ther 6:297–306
10. Ginkel SZ, Dooley TP, Suling WJ, Barrow WW (1997) Identification and cloning of the *Mycobacterium avium fol*A gene, required for dihydrofolate reductase activity. FMS Microbiol Lett 156:69–78
11. Galvanauskas V, Simutis R, Volk N, Lübbert A (1998) Model based design of a biochemical cultivation process. Bioproc Eng 18:227–234
12. Volk N, Hertel T, Lübbert A (1998) Modellgestützte Optimierung der Produktion des Virus-Hüllproteins VP1-DHER mit *E. Coli* BL21. Chem Tech 4:192–197
13. Miao F, Kompala DS (1992) Overexpression of cloned genes using recombinant *Escherichia coli* regulated by a T7 promoter: I. Batch cultures and kinetic modeling. Biotechnol Bioeng 40: 787–796
14. Bellgardt KH (2000) Bioprocess models In: Schügerl K, Bellgardt KH (eds) Bioreaction engineering: modeling and control. Springer, Berlin Heidelberg New York, Chap. 2
15. Deb K, Beyer, HG (1999) Self-adaptive genetic algorithms with simulated binary crossover, Report Cl-61/99, SFB 531, Universität Dortmund