BIOTECHNOLOGICAL PRODUCTS AND PROCESS ENGINEERING

Selective expression of the soluble product fraction in *Escherichia coli* cultures employed in recombinant protein production processes

Stefan Gnoth · Rimvydas Simutis · Andreas Lübbert

Abstract Recombinant proteins produced in *Escherichia coli* hosts may appear within the cells' cytoplasm in form of insoluble inclusion bodies (IB's) and/or as dissolved functional protein molecules. If no efficient refolding procedure is available, one is interested in obtaining as much product as possible in its soluble form. Here, we present a process engineering approach to maximizing the soluble target protein fraction. For that purpose, a dynamic process model was developed. Its essential kinetic component, the specific soluble product formation rate, if represented as a function of the specific growth rate and the culture temperature, depicts a clear maximum. Based on the dynamic model, optimal specific growth rate and temperature profiles for the fed-batch fermentation were determined. In the course of the study reported, the mass of desired soluble protein was increased by about 25%. At the same time, the formation of inclusion bodies was essentially avoided. As the optimal cultivation procedure is rather susceptible to distortions, control measures are necessary to guarantee that the real process can be kept on its desired path. This was possible with robust closed loop control. Experimental process validation revealed that, in this way, high dissolved product fractions could be obtained at an excellent batch-to-batch reproducibility.

S. Gnoth (✉) · A. Lübbert (✉)
Center for Bioprocess Engineering,
Martin-Luther-University, Halle,
Wittenberg, Germany
e-mail: stefan.gnoth@biochemtech.uni-halle.de
e-mail: andreas.luebbert@biochemtech.uni-halle.de

R. Simutis
Institute of Automation and Control Systems,
Kaunas University of Technology,
Kaunas, Lithuania

Introduction

In *Escherichia coli* cultures, heterologous proteins often appear as a mixture of soluble molecules and aggregates referred to as inclusion bodies. In cases where proper refolding procedures are not available, process operational conditions must be selected that primarily lead to the soluble product fraction.

Literature contains many hints that the relative amount of soluble protein is mainly influenced by the growth conditions and the culture temperature adjusted during the product formation phase. The knowledge about these influences is qualitative:

According to Schein (1989) the easiest way of lowering the fraction expressed as inclusion bodies is to reduce the fermentation temperature, thereby favoring the activity and expression of chaperones and, on the other hand, reducing aggregation reactions (Sørensen and Mortensen 2005). Many authors provided evidence that a reduction in the culture temperature during the product formation phase increases the soluble fraction of the desired product (Piatak et al. 1988; Chalmers et al. 1990; Liao 1991; Vasina and Baneyx 1997). Endogenous proteins aggregated in *E. coli* may resolubilize when cells are subjected to lower temperature or cultivated in presence of chaperones (Kedzierska et al. 1999). Berwal et al. (2008) showed considerable temperature influences on the relative amount of the product forms. They presented an example where *E. coli* expressed its heterologous protein under normal operating conditions of $T=35$–37°C nearly completely in

form of inclusion bodies; but, upon a strong reduction in the culture temperature after induction, the soluble protein fraction appeared increased up to 85%. If cultivation temperatures are decreased, however, biomass growth is significantly lowered and with it the total amount of soluble protein as well. Because of these conflicting effects, high productivity expression of soluble protein appears to be an optimization problem.

In other studies, improved expression strains were employed (Bessette et al. 1999; Steinfels et al. 2002), which focused on co-expression of chaperones (Nishihara et al. 2000; Ikura et al. 2002; Baneyx and Palumbo 2003) or combination with soluble fusion partners (Davis et al. 1999; Kim and Cha 2006). Also, lowering the inducer concentration (Weickert et al. 1996; Heo et al. 2006) was proposed.

The influence of the specific growth rate on the ratio between dissolved product and inclusion body formation is much less investigated (Kopetzki et al. 1989; Shin et al. 1998; Hoffmann et al. 2001). Qualitatively, it is believed that, during expression of correctly folded recombinant proteins, the folding process is rate-limiting, e.g., by a limitation of chaperones (Thomas and Baneyx 1996; Villaverde and Carrió 2003; Rinas et al. 2007; Gasser et al. 2008). If more raw protein is formed than can be folded by the cells, the amount of protein exceeding the maximal folding capacity is shuffled to insoluble aggregates (inclusion bodies). Consequently, by manipulation of the specific growth rate, the expression rate of recombinant protein can be adjusted to this rate-limiting bottleneck of the cell, thus favoring the soluble protein fraction.

In this work, we focus attention on the quantitative influence of the specific growth rate and fermentation temperature on the formation of soluble target protein as well as on the concurring inclusion body formation. From the engineering perspective, soluble product fraction optimization requires a process model that can quantitatively be exploited in order to find optimal operational conditions, i.e., with respect to:

– The ratio soluble/IB product fraction,
– The overall productivity, and, finally,
– The reproducibility of the operational procedure.

Here, we present a robust engineering procedure of solving this optimization problem. Its performance is demonstrated at a concrete example. For that purpose, we took an *E. coli* strain from an industrial partner expressing a recombinant protein that appears partly in soluble form and partly in form of inclusion bodies. We started with two alternative modeling approaches, a conventional model and a data-based hybrid model. Both were parameterized using the same data set. Biomass and product concentration data were predicted by the hybrid model more precisely. Hence, the latter was used to determine an optimal process operational procedure. The optimization result showed that the final process must closely follow the optimal trajectories in order to exclusively express the soluble protein fraction. The reproducibility of the process was finally achieved by controlling the total amount of CO_2 produced by the cells.

Materials and methods

Experimental setup

E. coli BL21(DE3) pET28a (Novagen) was used as the host system in all experiments. The organism expresses a commercially interesting protein, in the following text referred to as HumaX for secrecy reasons. Part of this product appears in its active soluble form, whereas the remaining is formed as inclusion bodies. The protein was expressed under control of the T7 promoter after induction with 1 mM IPTG.

All cultivations were conducted in a B.Braun 10-L-bioreactor Biostat C started with 5 L defined mineral salt medium of the following composition, i.e., Na_2SO_4 2.0 g L^{-1}, $(NH_4)_2SO_4$ 2.46 g L^{-1}, NH_4Cl 0.5 g L^{-1}, K_2HPO_4 14.6 g L^{-1}, $NaH_2PO_4 \times H_2O$ 3.6 g L^{-1}, $(NH_4)_2$-H–citrate 1.0 g L^{-1}, $MgSO_4 \times 7H_2O$ 1.2 g L^{-1}, trace element solution 2 ml L^{-1}, thiamine 0.1 g L^{-1}, and kanamycin 0.1 g L^{-1}. The process was entirely operated in fed-batch mode, i.e., no glucose was present in the medium at inoculation time. The feed, containing glucose as the carbon and energy source and mineral salts in the same composition as the initial mineral medium, was pumped into the reactor directly after inoculation (Jenzsch et al. 2006a). As the starting biomass concentration was only 0.25 g L^{-1} (DCW), the substrate feed rates had to be kept very low in the beginning in order to avoid overfeeding and thus formation of overflow metabolites or even substrate inhibition. Since the feed pumping rates were limited towards low feed rates, the feeding was started with a glucose concentration of 300 g kg^{-1}. When enough biomass had been formed, the feeding solution was switched to a glucose concentration of 600 g kg^{-1}. Substrate feeding addition was measured gravimetrically and recorded on-line using a Sartorius balance. The entire process was operated under the control of a Siemens Simatic PCS7 system (Kuprijanov et al. 2008).

During the biomass growth phase, an exponential feeding led to an exponential growth corresponding to a fixed specific growth rate of $\mu = 0.5$ h^{-1}. This growth rate was chosen in order to run the process in a substrate-limited way just below the specific substrate uptake rate that leads to acetate formation. In the open-loop-controlled phase, the temperature was kept at $T = 35°C$. Within 1 h upon induction at $t = 10$ h, the temperature was linearly ramped down to the specific product formation temperature chosen.

Then, during product formation, the temperature was kept constant. A series of experiments were conducted with different production-phase temperature values in the interval 27°C≤T≤35°C. The experiments were performed at different specific growth rates during the post-induction phase. Again, an open-loop-control procedure was used, i.e., predefined feed rate profiles were applied. At t=9.5 h, the growth rate was linearly ramped down to the desired value which was then kept constant during the product formation phase. In this way, specific growth rates in the interval 0.08 h^{-1}≤μ_{set}≤ 0.16 h^{-1} were adjusted during the product formation time, see Table 1. Additionally, some fermentation runs were performed employing a modified substrate pulse control technique to determine the critical specific substrate uptake rate profile at which an accumulation of overflow metabolites was just avoided. At all times, in the substrate-limited, open-loop controlled fermentations, as well as in fermentation applying feed pulse control, no acetate formation was observed.

Respiration was monitored via exhaust gas analysis in the vent line, i.e., O_2 was measured on-line with a paramagnetic oxygen sensor (Maihak Oxor 610) while CO_2 was detected with an infrared detector (Maihak Unor 610). An Ingold DO probe (Mettler Toledo) was used for monitoring and control of dissolved oxygen to a value of 25%sat applying the gain scheduling approach (Kuprijanov et al. 2009). pH was measured with an Ingold pH probe (Mettler Toledo) and maintained at pH7 using a parameter adaptive proportional integral derivative (PID) controller (Gnoth et al. 2009). Monitoring of the integral base consumption during pH control was accomplished with another balance. Temperature was measured with a PT100 sensor and controlled with a standard PID controller. All on-line data was recorded in a central SQL data-base with a time increment of 0.005 h.

Off-line measurements were performed with time increments of half an hour. Biomass concentration was determined by means of an OD_{600} measurement with a spectral photometer (Shimadzu UV-2102PC). Additionally, some dry weight measurements were performed in order to test the correlation between biomass dry weight and the OD_{600} values. Glucose concentration was measured enzymatically with a YSI 2700 Select Bioanalyzer (Yellow Springs Inc.). Acetate was determined using an enzymatic kit (Boehringer Mannheim) after centrifugation and subsequent heating of the supernatant (15 min, 80°C). Finally, target protein was quantified by gel densitometry of stained sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) after cell disruption, separation of the soluble fraction, and solubilization of inclusion bodies (IB's). All target protein that appeared in the soluble fraction had the desired activity as was shown by our industrial partner.

Model for process optimization

Dynamic process models for process optimization and control purposes are based on mass balances for the main state variables, such as biomass, substrate and product.

$$\frac{dc}{dt} = R + \frac{F}{V} \cdot (c_F - c) \tag{1}$$

$$\frac{dV}{dt} = F \tag{2}$$

where the concentration vector c=[$X\ S\ P_{Sol}\ P_{IB}$] consists of the elements biomass concentration X, substrate concentration S, concentration P_{Sol} of the soluble part of the desired protein, and P_{IB}, the concentration of the inclusion body protein. c_F stands for the vector containing the concentrations of these variables in the feed. F is the feed rate and V the volume of the culture.

As all the protein production processes are operated in the fed-batch mode, we additionally need to consider the dynamic change of the entire culture volume (Eq. 2). The different models are mostly distinguished by the way they are used to formulate the kinetics which makes up the biochemical conversion rate vector R=[$\mu-\sigma\ \pi_{Sol}\ \pi_{IB}$]·$X$. In the following, we first present such a conventional model specifying the specific rates μ, σ, π_{Sol}, and π_{IB}. As this is quite time-consuming, we will further show that data-driven model versions of the same kinetic expressions are easier to develop and are more accurate.

Results

Conventional process model

In a conventional unstructured process model, the kinetics for the biomass and product formation is most often formulated in terms of Monod-type expressions, whereas temperature dependencies are usually modeled by Arrhenius expressions.

Table 1 Experimental setup of the fermentations performed

	Screening experiments	Validation runs
Specific growth rate μ_{Set} [h^{-1}]	0.08, 0.12, 0.16, μ_c	0.12
Fermentation temperature after induction T [°C]	27, 29, 31, 33, 35	31
Type of control	Open-loop, feed-pulse control	tcCPR-control

Cells take up substrate in order to grow and to produce the desired product. Hence, we start with the specific substrate uptake rate as the basic kinetic variable.

$$\sigma = \sigma_{\max} \cdot \frac{S}{K_S + S} \cdot \frac{K_{X-\sigma}}{K_{X-\sigma} + X} \quad (3)$$

The third factor describes the generally observed fact that the specific substrate uptake rate becomes smaller at higher biomass concentrations.

The temperature dependency of the specific biomass growth rate μ is described by an Arrhenius expression (Roels 1983; Esener et al. 1983; Schügerl and Bellgardt 2000). Additionally, by means of a logistic expression, it is considered that biomass growth rates decrease at elevated biomass concentrations.

$$\mu = \sigma \cdot Y_{XS} \cdot \frac{k_1 \cdot \exp(-E_{A1}/(R \cdot T))}{1 + k_2 \cdot \exp(-E_{A2}/(R \cdot T))} \cdot \left(1 - \frac{X}{K_{X-\mu} + X}\right) \quad (4)$$

The specific formation rate of soluble protein, π_{sol}, critically depended on the specific growth rate as well as on the cultivation temperature. The following expression reflects experimental results that the specific product formation rate shows up a distinct maximum with respect to both temperature and product concentration.

$$\pi_{Sol} = \pi_{\max}^{Sol} \cdot \mu \cdot \left(1 - \frac{\mu}{K_{\mu 1}}\right) \cdot \frac{k_3 \cdot \exp(-E_{A3}/(R \cdot T))}{1 + k_4 \cdot \exp(-E_{A4}/(R \cdot T))} \quad (5)$$
$$\cdot \left(1 - \frac{P_{Sol}^2}{K_P}\right)$$

Finally, the specific formation rate of the inclusion body protein, π_{IB} is described by Eq. 6, saying that inclusion body formation is being favored at increasing cultivation temperatures and biomass growth rates.

$$\pi_{IB} = \pi_{\max}^{IB} \cdot \frac{\mu}{K_{\mu 2} + \mu} \cdot \frac{k_5 \cdot \exp(-E_{A5}/(R \cdot T))}{1 + k_6 \cdot \exp(-E_{A6}/(R \cdot T))} \quad (6)$$
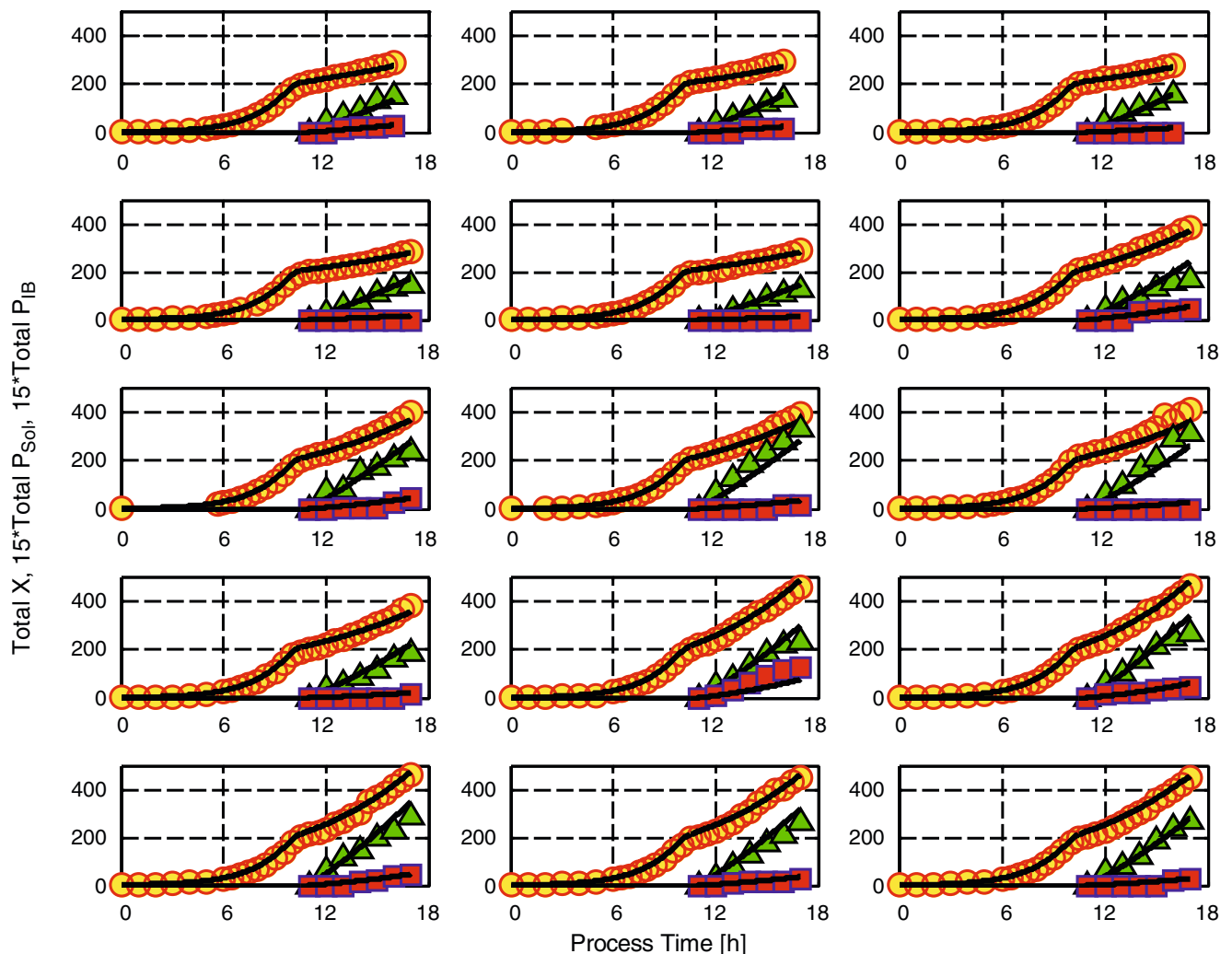

Fig. 1 Results of simulations using the conventional process model as described in Eqs. 1–7 (curves) compared with the experimental data (circles total biomass, triangles total soluble protein, squares total inclusion body protein)

With these kinetic expressions, the system of mass balances Eqs. 1 and 2 was solved numerically in Matlab®. The results are depicted in Fig. 1 together with the experimental data recorded during the open-loop controlled fermentations from which the initial and operating conditions were taken.

Hybrid model

Formulating a conventional model based on complex Monod-type kinetic expressions including temperature influences and determining their parameter values to a sufficient accuracy is time-consuming, and the results are restricted to the particular example investigated. Thus, a more general hybrid model is proposed that is easier to develop and leads to a higher accuracy.

This hybrid model combines feed-forward artificial neural networks (ANNs) describing the kinetic model components, the specific rate expressions with simple mass balances as shown in Fig. 2. Separate ANNs were taken for estimation of the specific growth rate μ and the specific product formation rate π. The single mass balances have not been formulated in terms of the biomass or product concentrations as in Eq. 1, but instead directly in terms of the total mass values x and p (Gnoth et al. 2008). The
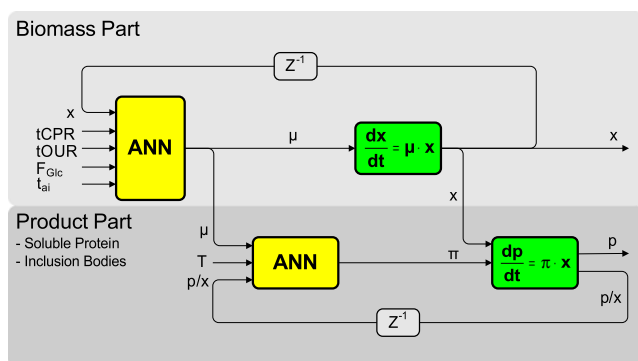

Fig. 2 General structure of the hybrid model describing the specific biomass and product formation rates μ and π together with the biomass x and the respective product mass p. At each on-line sampling time, the substrate addition rate and tCPR as well as tOUR are processed in the first ANN (five inputs, one bias, ten hidden-layer nodes and one output) to estimate the specific biomass growth rate and subsequently, the total biomass. Once the biomass part hybrid net is trained, the outputs together with the cultivation temperature and the specific protein load act as inputs for the following hybrid net predicting the target protein (three inputs, one bias, seven hidden-layer nodes and one output). The same architecture of the product part network is used for estimation of either soluble or inclusion body protein formation. For simplicity reasons, only one product network respective for both protein fractions is shown. Solving the optimization problem was performed by exploitation of the second trained part of the hybrid network system ("*Product Part*") with fixed network weights. Estimation of specific biomass growth rate μ then drops, as it is defined in Eq. 10. In this way, the hybrid model is used to easily simulate various specific growth rate profiles and cultivation temperatures

model parameters (structure and weights of the ANN components) were determined using the sensitivity equation approach (Schubert et al. 1994) and the "Leave-One-Out" cross-validation procedure (Hjorth 1994).

The performance of this hybrid model can be seen in Fig. 3, where the off-line biomass and the product mass profiles recorded in 15 different cultivation runs are compared with the corresponding solutions computed with one and the same hybrid model.

As quantified in Table 2 by means of root mean square deviations (RMSEC values) between that of biomass, soluble protein, and IB-protein data and their corresponding computational results, the hybrid model leads to a considerably better fit to the measurement data than the conventional process model.

From the process engineering point of view, it is instructive to know how soluble and inclusion body product formation rates vary with the specific biomass growth rate μ and the cultivation temperature. The output of the hybrid model for the specific formation rates of soluble and IB-protein fraction at a fixed value of the respective specific target protein load p/x, i.e., at a specified time after induction, is depicted in Fig. 4. While the inclusion body formation rate generally increases with rising temperature within the temperature interval investigated, the soluble fraction depicts a clear maximum at intermediate temperatures of 31°C. At a given temperature, the IB formation rate rises with the specific growth rate as expected, while the product formation rate for the soluble fraction again passes a pronounced maximum at value of about 0.12 h^{-1}.

Finally, it should be noted that the inputs of the hybrid model shown in Fig. 2 are on-line measured variables. The only variable that is not available on-line is the total biomass x, which, however, is used recursively. That means we only need its initial value. The same applies for starting values of p_{Sol} and p_{IB} which are considered to be zero in practical applications. In other words, if the inoculum size is known, the hybrid model depicted in Fig. 2 is not only able to estimate biomass and product formation from historical fermentation data sets, it can also be used as a soft-sensor for the important variables biomass x, product mass p_{Sol}, and p_{IB}, the specific growth rate μ, as well as the specific product formation rates π_{Sol} and π_{IB}. Because it is a purely data-driven approach, it can easily be adapted to different hosts and product formation kinetics by training with the data from the corresponding cultivations. Finally, the hybrid model is not restricted to a fixed fermentation time as measurement data is given to network every time interval.

Optimization

After the specific product formation rate patterns $\pi(\mu, T)$ have been determined, the trained hybrid network can
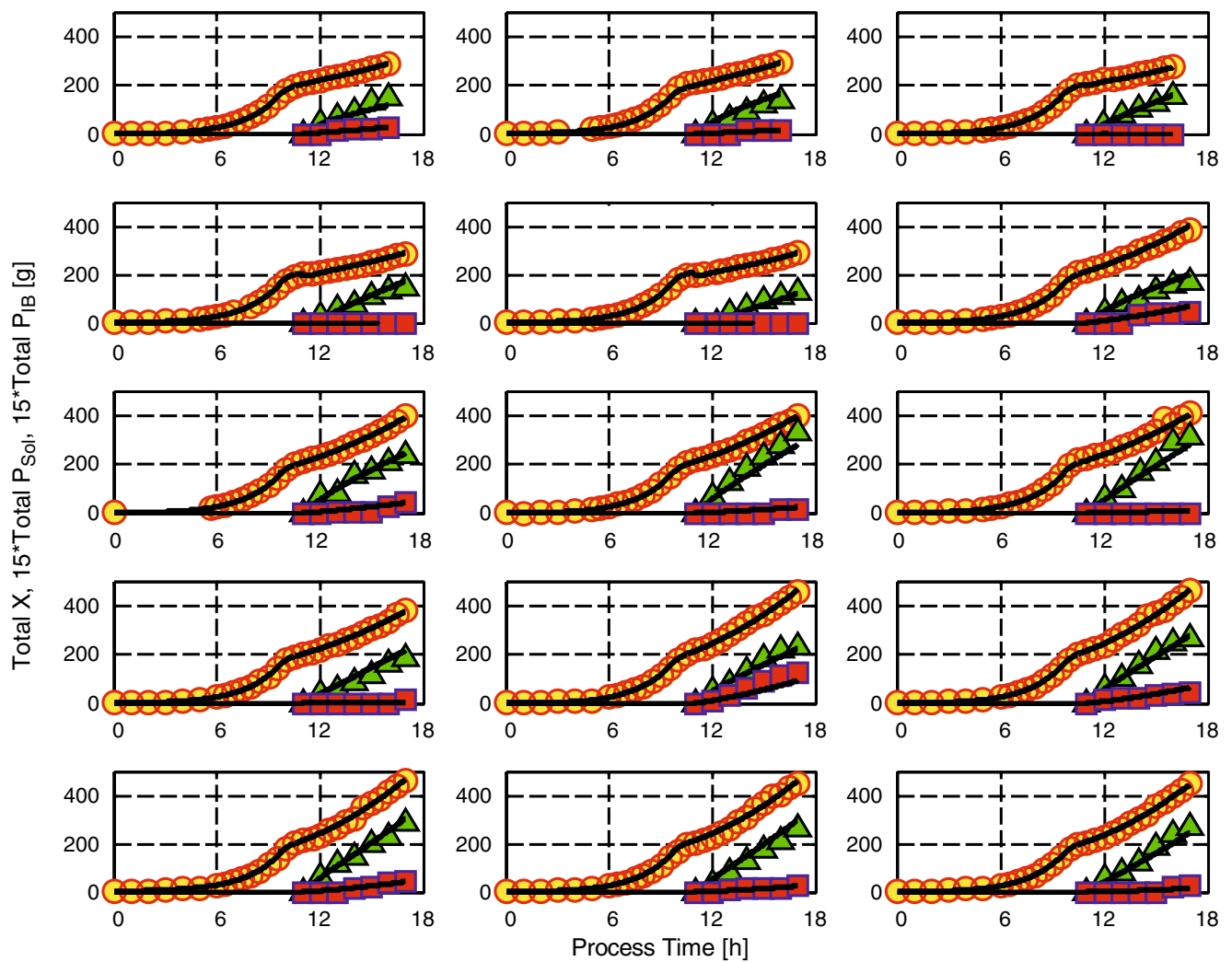
Fig. 3 Comparison between modeled (*lines*) and measured values of biomass and protein mass for the experiments performed (*circles* total biomass, *triangles* total soluble protein, *squares* total inclusion body protein). One and the same set of hybrid models was used in all cases

readily be used to find optimal μ and T profiles with respect to a given objective function. For that purpose, only a part of the full hybrid model is required. As the total biomass is directly accessible from a given μ-profile via the ordinary differential equation,

$$\frac{dx}{dt} = \mu \cdot x \qquad (7)$$

only the "Product Part" (second part) of the hybrid model is required. For any combination of μ and T, the hybrid model is capable to, first, determine the total biomass x and, subsequently, to calculate the specific product formation rate directly yielding the total product mass via balance equation. The weights of the ANN computing π remain the same as in the original hybrid model depicted in Fig. 2.

Table 2 Comparison of prediction quality of the conventional process model and the hybrid ANN-approach, expressed by the root mean square error

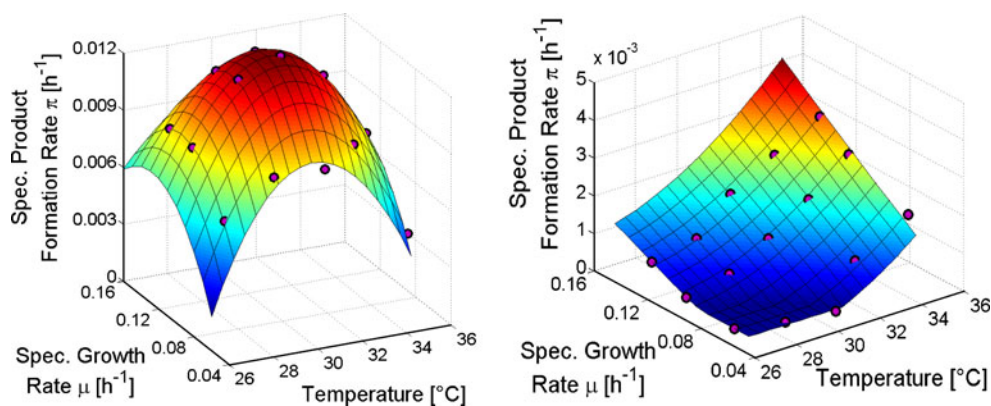|	RMSEC [g]		
	Total biomass	Total soluble protein	Total IB-protein
Conventional model	10.81	1.78	0.87
Hybrid network	4.71	1.28	0.62

Fig. 4 Specific product formation rates of the soluble protein fraction (*left*) and inclusion body fraction (*right*) as a function of the specific growth rate and fermentation temperature in the product formation phase. The marked dots depict fermentation data used for model identification. Compared with the desired soluble protein formation which depicts distinct optimum at lower growth rates and temperatures, IB-protein formation mainly occurs at elevated growth rates and temperatures

Beforehand, a quantitative formulation of the objective function J must be made. In order to keep the procedure simple, we fixed the duration of the post-induction phase and chose

$$J = p_{Sol}(t_f) - \beta_1 \cdot p_{IB}(t_f) - \beta_2 \cdot x(t_f) \qquad (8)$$

where the first term denotes the desired soluble product mass available at the end of the fermentation ($t=t_f$). The second part considers the inclusion body formation, whereas the last term was added to avoid excessive biomass formation. The parameters β_1 and β_2 must be specified from economical points of view. They can be considered penalty factors punishing the formation of inclusion bodies and excessive biomass without contribution to target protein formation.

In reality, the solution of the optimization procedure is constrained by physical as well as physiological factors. The main physical constraint is the limited oxygen transfer rate that can be provided by the bioreactor. It influences the maximal amount of biomass that can be employed for product formation. Since the oxygen uptake rate in the current cultivation system drops roughly half an hour after induction, the aeration capacity of the reactor mainly determines the biomass at which induction has to be performed.

From physiological point of view, the limiting constraint is often the critical specific substrate uptake rate σ_c of the cells. In the current context, this is defined as the maximal specific substrate uptake rate of the cells just avoiding the formation of overflow products. In case of *E. coli*, acetate represents the main overflow metabolite, and it has been shown to inhibit growth and product formation and thus the overall process productivity (Luli and Strohl 1990; Jensen and Carlsen 1990; Roe et al. 2002). Because the critical substrate uptake rate considerably drops with time after

induction, knowledge of σ_c is required to avoid overfeeding and thus the formation acetate. Mechanistic models about the dynamical change of σ_c are not available. Thus, σ_c was determined experimentally using a modified substrate pulse-response technique based on the approach proposed by Åkesson (Åkesson et al. 1999). For that purpose, cultivation experiments were performed where the substrate feed rate was modulated, and the corresponding response in the dissolved oxygen concentration (DO) was examined. The modulated feed rate was increased as long as there was a significant reaction in the DO. When a response was no longer observed, the critical substrate uptake rate was reached. Then, the mean feed rate was slightly reduced not to run into a state of acetate formation. By periodically approaching σ_c, its time profile was determined.

To ensure validity of estimation near the constraints, the hybrid model was additionally trained with on-line data of a feed pulse-controlled fermentation. In this way, the specific formation rate of the soluble fraction of the desired protein was determined as a hyper-surface on the plane spanned by the specific growth rate μ and the specific product concentration p_{Sol}/x. This is depicted in Fig. 5. Using the data of the feed modulation experiments just described, we can get the corresponding path on this hyper-plane. The periodic approach to the critical substrate uptake rate can be seen in the wavy nature of the red curve. As the p_{Sol}/x-axis can be viewed as a slightly nonlinear axis of the cultivation time after induction, one can see that the specific growth rate drops sharply with time after induction; however, this drop starts after some time delay, an effect that is qualitatively mentioned in literature (Bentley et al. 1990; Bhattacharya and Dubey 1995; Hoffmann and Rinas 2001). The essential information from Fig. 5 with respect to process optimization is that the process must be kept below
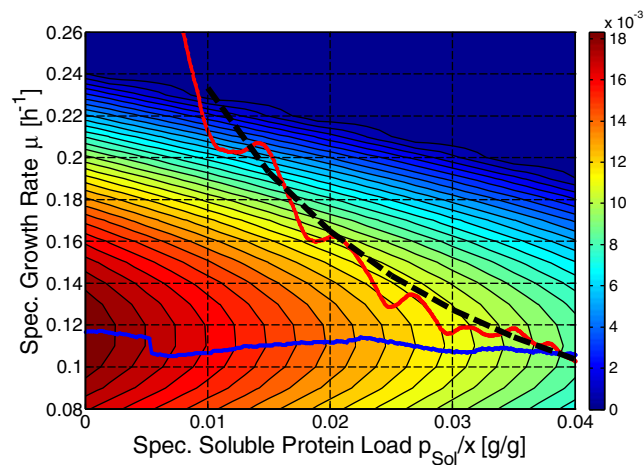
Fig. 5 Specific soluble product formation rate hyperspace on plane spanned by the specific growth rate and the specific soluble protein load of the cells at a fixed fermentation temperature. The *red curve* depicts the result of the probing control experiment, saying that maximal biomass growth capacity is rapidly decreasing after induction. The *dashed black curve* is a smoothed version of the result defined as a constraint for optimization. The *blue curve* is a typical trajectory of an experiment guided along a fixed specific growth rate under growth limited conditions

the red curve if substrate overflow is to be avoided. Thus, this curve can be considered a constraint to the optimal process control strategy.

In order to get an expression that can conveniently be used in optimization studies, the red curve was approximated by

$$\mu < \mu_c, \ \mu_c = 0.4 \cdot \left(1 - \frac{p_{Sol}/x}{0.014 + p_{Sol}/x}\right) \qquad (9)$$

This is represented by the dashed black line in Fig. 5.

Keeping the numerical optimization feasible from the practical point of view, some general simplifications were made. The first was to keep the temperature constant after ramping down to the desired value. The temperatures adjusted during the production phase were constrained to the interval $27°C \leq T \leq 35°C$.

The second simplification regards the specific biomass growth rate. In literature, it was most often assumed to keep the specific growth rate as far as possible constant at its optimal value (e.g., Jenzsch et al. 2006b). Here, we allowed for a slight linear shift in μ and assume a linear time function for the specific growth rate profile.

$$\mu(t) = a_0 + a_1 \cdot t \qquad (10)$$

In order to optimize the $\mu(t)$ profile with respect to the objective function J, the parameters a_0, a_1, and the fermentation temperature T were varied. The results of the numerical optimization performed with Matlab® depended on the choice of the parameters in the objective function:

For the case $\beta_1 = \beta_2 = 0$, the optimal values of the free model parameters were $T = 31°C$, $a_0 = 0.2$, and $a_1 = -0.01$. The solutions for the biomass x and product masses p_{Sol}, p_{IB}, as well as for the specific growth μ and formation rates π_{Sol} and π_{IB} are presented in Fig. 6a. Here, the amount of desired protein was calculated to be 21.36 g. At the same time, the undesired IB-protein was being formed at an amount of 3.91 g. The corresponding biomass was $x = 537.6$ g.

Alternatively, we considered the situation of a trade-off between soluble protein and inclusion body formation concurrently optimizing substrate to soluble product yield. The parameters here were chosen to be $\beta_1 = 1.0$ and $\beta_2 = 0.025$, saying that IB formation should be kept on a lower level and biomass should not be too excessive. Then, the estimates of
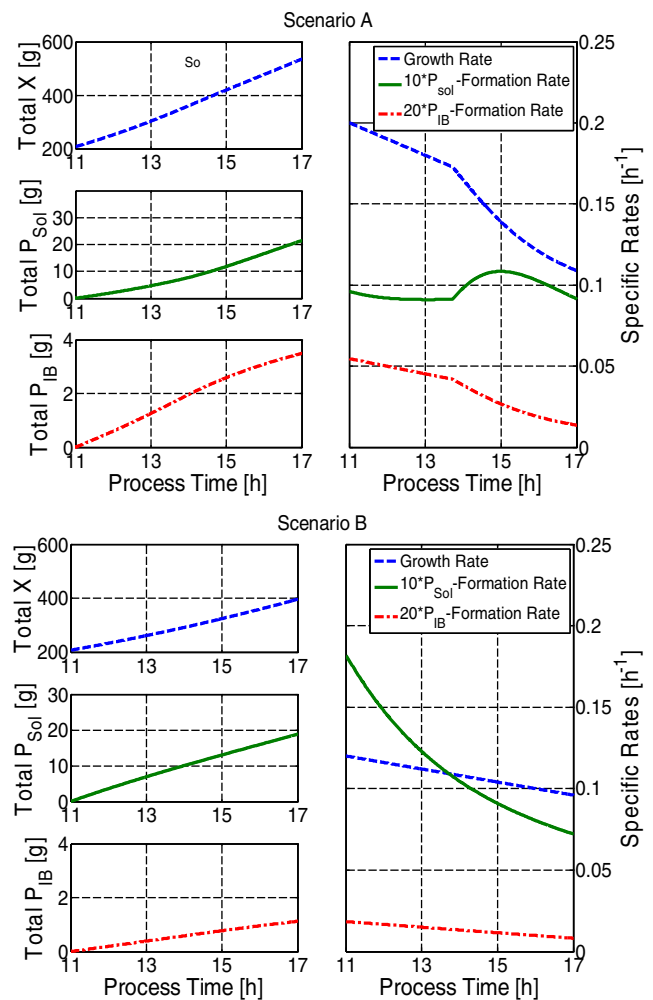

Fig. 6 Simulated profiles of the desired soluble protein mass (*solid green line, left side*), the IB-protein mass (*dash-dot red line, left side*) and biomass (*dashed blue line, left side*) as a function of process time together with the corresponding specific growth (μ *dashed blue line, right side*) and product formation (π_{Sol} *solid green line, right side*, π_{IB} *dash-dot red line, right side*) rates for optimization *scenario A* penalty factors $\beta_1 = 0.0$, $\beta_2 = 0.0$ and optimization *scenario B* penalty factors $\beta_1 = 1.0$, $\beta_2 = 0.025$. The outcome of both optimization runs for the cultivation temperature was 31°C

the optimal parameters were $T=31°C$, $a_0=0.12$, and $a_1=-0.004$. The corresponding results are depicted in Fig. 6b.

Here, the amount of the desired product was 19 g, but the total mass of IB-protein was only 1.4 g. Biomass ended up at 396 g.

The second optimization scenario was considered more related to practice. Hence, its results were further investigated. An experiment performed with the optimized parameter values resulted in the blue curve in Fig. 5. It only approached the constraint with respect to the critical substrate uptake rate at quite high specific protein loads of the cells, i.e., at the end of the cultivation. This only required a small reduction in the specific growth rate, i.e., a small feed reduction.

Figure 7 compares the results with parameters from the second optimization scenario with those obtained when a standard fermentation strategy for recombinant proteins was applied. The latter strategy was proposed at the beginning of the investigation reported here, when the strain was delivered. At that time, a temperature of 35°C and a constant specific growth rate of $\mu_{set}=0.16$ h^{-1} during the protein formation phase were chosen.

It becomes immediately clear from Fig. 7 that the improvements with respect to the optimization task to be solved are considerable. The soluble product fraction became larger and the inclusion body fraction was nearly suppressed. The improved results were obtained at decreased cultivation temperatures after induction and at lower total biomass values indicating a significant increase of substrate to soluble protein yield.
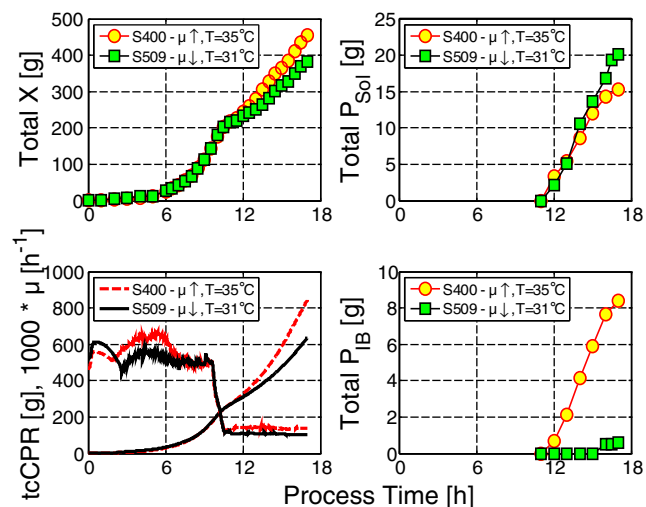
Figure 7 reveals product mass profiles of the two solutions being much different while the biomass profiles are quite close to each other. Thus, the optimal trajectories are sensitive to small distortions in the biomass growth profile. The question therefore arises how to assure that the good results can be produced with a high reproducibility.

Assuring reproducibility

Process variability is a major problem in fermentation plants. The batch-to-batch variability in the biomass profiles typically found in biopharmaceutical industry is known to be so high that the two biomass profiles, depicted in Fig. 7, are completely within the normal scatter. Hence, feedback process control was employed in order to improve reproducibility of the process operation to a grade that the process consistently produces the dissolved product at its optimized fraction.

For that purpose, the process was controlled along the time profile of the total mass of CO_2 generated by the cells. As was demonstrated by Jenzsch et al. (2007), this approach is a reliable indirect control of a corresponding biomass and specific growth rate profile, as well, circumventing sophisticated state estimation methods of these variables. The profile of the integral CO_2 produced was taken from cultivation with optimal cultivation parameter settings (as shown in Fig. 7). As the manipulated variable the feed rate F was used. The batch-to-batch reproducibility obtained with this control procedure can easily be judged from the five validation fermentations depicted in Fig. 8.
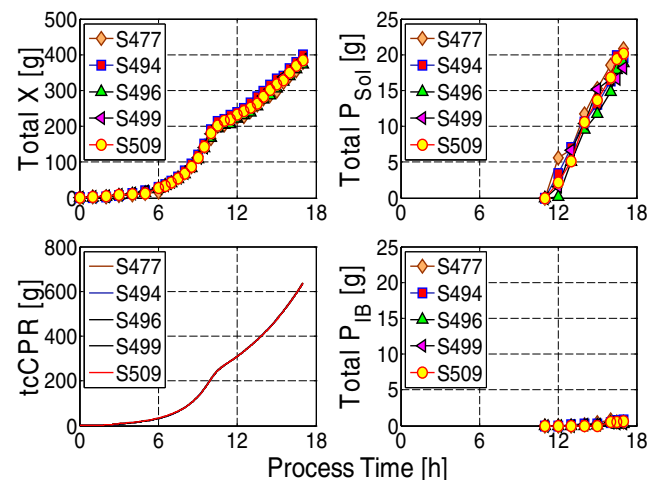

Fig. 7 Comparison of the trajectories of a standard fermentation strategy applied at the early beginning of our studies (*yellow circles*, *red line*) and after model supported process optimization (*green squares*, *black line*). The improved strategy leads to higher amounts of soluble protein and, simultaneously, to a reduction of the inclusion body fraction

Fig. 8 Results of the five experiments performed to validate the resulting operational procedure that aims in maximizing the formation of soluble product fraction (*upper right*) with simultaneous reduction of the inclusion body fraction (*lower right*). The high degree of batch-to-batch reproducibility was assured by application of feedback control of the integral mass of carbon dioxide generated by the cells (*tcCPR; lower left*)

It can be seen that the total CO_2-mass profiles measured in the different cultivations (depicted in lower left part of Fig. 8) almost completely overlap. Deviations from the set point profile are less than 1%. Thus, controller works well. As the process is a deterministic system, the low variability in the total CO_2 leads to small variations in biomass and protein masses, as well. Practically all target protein appears in its soluble form while the inclusion body fraction is almost completely suppressed. The larger scatter in the protein data is due to the fact that the precision of the protein concentration measurements by gel densitometry of SDS-PAGE is considerably smaller than the one in the measurements of biomass and CO_2.

Discussion

Here, we present an engineering approach to finding a process operational procedure that channels practically all the desired product formed by the *E. coli* bacteria into its soluble form. As such, the method makes use of numerically exploitable models quantitatively describing the process dynamics of product formation. We first used a conventional model where the kinetics is represented by Monod-like expressions. This model required considerable development time because different hosts and types of target proteins demand for a careful quantitative, knowledge-based description of process kinetics. In a next step, an approach was chosen which, first of all, is universally applicable and, moreover, depicts a higher accuracy. It is based on general valid mass balances and purely data-driven expressions for the kinetics. The latter allow for an easy model adaptation to various strains as well as to differing product formation kinetics and were formulated by means of artificial neural networks.

It was pinpointed that the employed hybrid process model could not only be used for off-line data-driven estimation of biomass and product formation kinetics, but also, since all input variable values are available on-line, as a soft-sensor to monitor the important process variables μ, π_{Sol}, π_{IB}, x, as well as both fractions of the target protein p_{Sol} and p_{IB}. Moreover, the hybrid model was shown to be exploitable in numerical optimization studies, where the aim was to find process operational protocols maximizing the soluble fraction of the desired product formed by the cells. For that purpose, we formulated a concrete objective function and determined the process constraints.

The results showed that the process can be run in such a way that the inclusion body formation is practically negligible. However, the process performance was sensitively affected by deviations from the optimal path of the process. Hence, in order to stabilize the process, advanced feedback control adjusting biomass growth was employed.

This guaranteed a high reproducibility of all important state variables, particularly a nearly exclusive production of the desired soluble target protein fraction. As the controlled process variable the time profile of the overall mass of CO_2 generated by the cells was used. This has the advantage that rather noiseless on-line measurement signals can be supplied to the controller instead of indirectly measured variables such as the specific biomass growth rate μ. The total integral CO_2 profile in turn can directly be related to a profile of the specific biomass growth rate which is most often used in literature as the controlled variable and which appears as the result from the optimization procedure described in the previous section.

The developments reported address the central requirements of FDA's PAT initiative with respect to comprehensive use of knowledge-based process design and operation. First, the identification of the variables that critically affect product formation was performed. Subsequently, quantitative process models were formulated that can be and were used for constraint process optimization. Finally, advanced feedback control was implemented to assure reproducibility. So far, the methodology provides innovative tools and methods for a knowledge-based process design and operation.

Symbol	Description	Value	Dimension
c	Concentration vector		[g/L]
c_F	Concentration vector of species in feeding		[g/L]
DO	Dissolved oxygen		[% sat]
E_{A1}	Activation enthalpy for biomass growth	12.4321	[kJ/mol]
E_{A2}	Inactivation enthalpy for biomass growth	298.5476	[kJ/mol]
E_{A3}	Activation enthalpy for soluble protein expression	94.5376	[kJ/mol]
E_{A4}	Inactivation enthalpy for soluble protein expression	290.2615	[kJ/mol]
E_{A5}	Activation enthalpy for inclusion body protein expression	92.0230	[kJ/mol]
E_{A6}	Inactivation enthalpy for inclusion body protein expression	280.3431	[kJ/mol]
F	Feeding rate		[g/h]
F_{Glc}	Feeding rate of pure glucose		[g/h]
J	Objective function of optimization		[g]
k_1	Constant for activation of biomass growth	130.0307	[−]
k_2	Constant for inactivation of biomass growth	3.8343e48	[−]
k_3	Constant for activation of soluble protein expression	2.4099e16	[−]
k_4	Constant for inactivation of soluble protein expression	3.0156e49	[−]

k_5	Constant for activation of inclusion body protein expression	7.1563e14	[–]
k_6	Constant for inactivation of inclusion body protein expression	1.3982e46	[–]
K_P	Constant of inhibition of soluble protein expression caused by increasing protein load	40.1263	[g^2/L^2]
K_S	Monod constant	0.08	[g/L]
$K_{X\text{-}\sigma}$	Constant for inhibition of substrate uptake caused by higher cell densities	36.8123	[g/L]
$K_{X\text{-}\mu}$	Constant for inhibition of biomass growth substrate caused by higher cell densities	121.8669	[g/L]
$K_{\mu 1}$	Constant for inaction of soluble protein expression	0.25	[–]
$K_{\mu 2}$	Constant for inaction of inclusion body protein expression at lower growth rates	0.082	[h^{-1}]
p	General annotation of total mass of target protein		[g]
p_{IB}	Total mass of target inclusion body protein fraction		[g]
p_{Sol}	Total mass of soluble target protein fraction		[g]
P_{Sol}	Concentration of soluble target protein fraction		[g/L]
P_{IB}	Concentration of target inclusion body protein fraction		[g/L]
R	Universal gas constant	0.0083145	[kJ/ (K mol)]
R	Vector of conversion rates		[h^{-1}]
$RMSE$	Root mean square error		[g]
S	Substrate concentration		[g/L]
t	Process time		[h]
t_F	Duration of fermentation		[h]
t_{ai}	Time after induction		[h]
$tCPR$	Total carbon dioxide production rate		[g/h]
$tOUR$	Total oxygen uptake rate		[g/h]
T	Cultivation temperature		[°C],[K]
x	Total biomass dry weight		[g]
X	Concentration biomass dry weight		[g/L]
Y_{XS}	Yield coefficient biomass–substrate	0.535	[g/g]
V	Culture volume		[L]
μ	Specific biomass growth rate		[h^{-1}]
μ_c	Maximal specific biomass growth rate where no overflow metabolism occurs		[h^{-1}]
π_{Sol}	Specific product formation rate of soluble protein		[h^{-1}]
π_{IB}	Specific product formation rate of inclusion body protein		[h^{-1}]
σ	Specific substrate uptake rate		[h^{-1}]
σ_c	Maximal specific substrate uptake rate where no overflow metabolism occurs		[h^{-1}]
σ_{max}	Maximal specific substrate uptake rate	1.5008	[h^{-1}]
π_{max}^{Sol}	Maximal specific product formation rate of soluble protein	0.2160	[h^{-1}]
π_{max}^{IB}	Maximal specific product formation rate of inclusion body protein	0.024	[h^{-1}]

References

Åkesson M, Nordberg E, Karlsson K, Axelsson JP, Hagander P, Tocaj A (1999) On-line detection of acetate formation in *Escherichia coli* cultures using dissolved oxygen responses to feed transients. Biotechnol Bioeng 64(5):590–598

Baneyx F, Palumbo JL (2003) Improving heterologous protein folding via molecular chaperone and foldase co-expression. Meth Mol Biol 205:171–197

Bentley WE, Noushin M, Andersen DC, Davis RH, Kompala DS (1990) Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. Biotechnol Bioeng 35(7):668–681

Berwal R, Gopalan N, Chandel K, Prasad GBKS, Prakash S (2008) *Plasmodium falciparum*: enhanced soluble expression, purification and biochemical characterization of lactate dehydrogenase. Exp Parasitol 120(2):135–141

Bessette PH, Aslund F, Beckwith J, Georgiou G (1999) Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. Proc Natl Acad Sci USA 96:13703–13708

Bhattacharya SK, Dubey AK (1995) Metabolic burden as reflected by maintenance coefficient of recombinant *Escherichia coli* overexpressing target gene. Biotechnol Lett 17(11):1155–1160

Chalmers JJ, Kim E, Telford JN, Wong EY, Tacon WC, Shuler ML, Wilson DB (1990) Effects of temperature on *Escherichia coli* overproducing β-lactamase or epidermal growth factor. Appl Environ Microbiol 56:104–111

Davis GD, Elisee C, Newham DM, Harrison RG (1999) New fusion protein systems designed to give soluble expression in *E. coli*. Biotechnol Bioeng 65:382–388

Esener AA, Roels JA, Kossen NWF (1983) Theory and applications of unstructured growth models: kinetic and energetic aspects. Biotechnol Bioeng 15:2803–2841

Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodriguez-Carmona E, Baumann K, Giuliani M, Parrilli E, Branduardi P, Lang C, Porro D, Ferrer P, Tutino ML, Mattanovich D, Villaverde A (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microbial Cell Factories 7:11

Gnoth S, Jenzsch M, Simutis R, Lübbert A (2008) Product formation kinetics in genetically modified *E. coli* bacteria: inclusion body formation. Bioprocess Biosyst Eng 31:41–44

Gnoth S, Kuprijanov A, Simutis R, Lübbert A (2009) Advanced control of dissolved oxygen concentration in fed-batch cultures during recombinant protein production. Appl Micriobiol Biotechnol 82:221–229

Heo MA, Kim SH, Kim SY, Kim YJ, Chung J, Oh MK, Lee SG (2006) Functional expression of single-chain variable fragment antibody against c-Met in the cytoplasm of Escherichia coli. Prot Expr Purif 47:203–209

Hjorth JSU (1994) Computer intensive statistical methods: validation model selection and bootstrap. Chapman and Hall, London

Hoffmann F, Rinas U (2001) On-line estimation of the metabolic burden resulting from synthesis of plasmid-encoded and heat-shock proteins by monitoring respiratory energy generation. Biotechnol Bioeng 76(4):333–340

Hoffmann F, Posten C, Rinas U (2001) Kinetic model of in vivo folding and inclusion body formation in recombinant Escherichia coli. Biotechnol Bioeng 72(3):315–322

Ikura K, Kokubu T, Natsuka S, Ichikawa A, Adachi M, Nishihara K, Yanagi H, Utsumi S (2002) Co-overexpression of folding modulators improves the solubility of the recombinant guinea pig liver transglutaminase expressed in Escherichia coli. Prep Biochem Biotechnol 32:189–205

Jensen EB, Carlsen S (1990) Production of recombinant human growth hormone in Escherichia coli: expression of different precursors and physiological effects of glucose, acetate and salts. Biotechnol Bioeng 36:1–11

Jenzsch M, Simutis R, Lübbert A (2006a) Generic model control of the specific growth rate in Escherichia coli cultivations. J Biotechnol 122:483–493

Jenzsch M, Gnoth S, Beck M, Kleinschmidt M, Simutis R, Lübbert A (2006b) Open-loop control of the biomass concentration within the growth phase of recombinant protein production processes. J Biotechnol 127:84–94

Jenzsch M, Gnoth S, Kleinschmidt M, Simutis R, Lübbert A (2007) Improving the batch-to-batch reproducibility of microbial cultures during recombinant protein production by regulation of the total carbon dioxide production. J Biotechnol 128:858–867

Kedzierska S, Staniszewska M, Wegrzyn A, Taylor A (1999) The role of DnaK/DnaJ and GroEL/GroEL systems in the removal of endogenous proteins aggregated by heat-shock from Escherichia coli cells. FEBS Lett 446:331–337

Kim YS, Cha HJ (2006) Solubility dependency of coexpression effects of stress-induced protein Dps on foreign protein expression in Escherichia coli. Enzyme Microb Technol 39:399–406

Kopetzki E, Schumacher G, Buckel P (1989) Control of formation of active soluble or inactive insoluble baker's yeast–glucosidase PI in Escherichia coli by induction and growth conditions. Mol Gen Genet 216:149–155

Kuprijanov A, Schaepe S, Sieblist C, Gnoth S, Simutis R, Lübbert A (2008) Variability control in fermentations—meeting the challenges raised by FDA's PAT initiative. Bioforum Europe 12:38–41

Kuprijanov A, Gnoth S, Simutis R, Lübbert A (2009) Advanced control of dissolved oxygen concentration in fed batch cultures during recombinant protein production. Appl Microbiol Biotechnol 82:221–229

Liao HH (1991) Effect of temperature on the expression of wild-type and thermostable mutants of kanamycin nucleotidyl transferase in Escherichia coli. Prot Expr Purif 2:43–50

Luli GW, Strohl WR (1990) Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations. Appl Environ Microbiol 56 (4):1004–1011

Nishihara K, Kanemori M, Yanagi H, Yura T (2000) Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. Appl Environ Microbiol 66:884–889

Piatak M, Lane JA, Laird W, Bjorn MJ, Wang A, Williams M (1988) Expression of soluble and fully functional ricin A chain in Escherichia coli is temperature-sensitive. J Biol Chem 263:4837–4843

Rinas U, Hoffmann F, Eriola B, Estapé D, Marten S (2007) Inclusion body anatomy and functioning of chaperone-mediated in vivo inclusion body disassembly during high-level recombinant protein production in Escherichia coli. J Biotechnol 127:244–257

Roe AJ, O'Byrne C, McLaggan D, Booth IR (2002) Inhibition of Escherichia coli growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. Microbiology 148:2215–2222

Roels JA (1983) Energetics and kinetics in biotechnology. Elsevier Biomedical Press, Amsterdam

Schein CH (1989) Production of soluble recombinant proteins in bacteria. Bio/Technology 7:1141–1148

Schügerl K, Bellgardt KH (2000) Bioreaction engineering. Springer, Heidelberg

Schubert J, Simutis R, Dors M, Havlik I, Lübbert A (1994) Hybrid modeling of yeast production processes-combination of a priori knowledge on different levels of sophistication. Chem Eng Technol 17(1):10–20

Shin CS, Hong MS, Kim DY, Shin HC, Lee J (1998) Growth-associated synthesis of recombinant human glucagon and human growth hormone in high-cell-density cultures of Escherichia coli. Appl Microbiol Biotechnol 49:364–370

Sørensen HP, Mortensen KK (2005) Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. Microbial Cell Factories 4:1

Steinfels E, Orelle C, Dalmas O, Penin F, Miroux B, Di Pietro A, Jault JM (2002) Highly efficient over-production in E. coli of YvcC, a multidrug-like ATP-binding cassette transporter from Bacillus subtilis. Biochim Biophys Acta 1565:1–5

Thomas JG, Baneyx F (1996) Protein misfolding and inclusion body formation in recombinant Escherichia coli cells overexpressing heat-shock proteins. J Biol Chem 271(19):11141–11147

Vasina JA, Baneyx F (1997) Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the Escherichia coli cspA and tac promoter systems. Prot Expr Purif 9:211–218

Weickert MJ, Doherty DH, Best EA, Olins PO (1996) Optimization of heterologous protein production in Escherichia coli. Curr Opin Biotechnol 7:494–499

Villaverde A, Carrió M (2003) Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnol Lett 25:1385–1395