

Efficient feeding profile optimization for recombinant protein production using physiological information

Patrick Wechselberger · Patrick Sagmeister ·
Helge Engelking · Torsten Schmidt ·
Jana Wenger · Christoph Herwig

Received: 15 March 2012 / Accepted: 2 May 2012 / Published online: 28 June 2012
© Springer-Verlag 2012

Abstract A multivariate study was performed aiming at the optimization of a recombinant rhamnose inducible *E. coli* induction system with alkaline phosphatase as target product. The effects of typical factors with impact on post- as well as pre-induction feeding rates were investigated with respect to the space–time yield of the target product. The goal was increased understanding as well as quantitative characterization of these factors with respect to their physiological impact on the model system. The optical density (OD) at which the culture was induced had a strong positive effect on the space–time yield. Pre-induction growth rate (k) had a second-order effect, while induction feed rate drop (J), a factor defining the linear post-induction feed rate, was interacting with (k). However, explanation of the observed effects to acquire more understanding regarding their effect on cell metabolism was not straight forward. Hence, the original process parameters were transformed into physiological more meaningful parameters and served as the basis for a multivariate data analysis. The observed variance with respect to observed volumetric activity was fully explained by the specific substrate uptake rate (q_s) and induction OD, merging the process parameters pre-induction growth rate (k) and feed rate drop (J) into the physiological parameter specific substrate uptake rate (q_s). After transformation of the

response volumetric activity (U/ml) into the biomass specific activity (U/g_{biomass}), the observed variance was fully explained solely by the specific substrate uptake rate (q_s). Due to physiological multivariate data analysis, the interpretation of the results was facilitated and factors were reduced. On the basis of the obtained results, it was concluded that the physiological parameter q_s rather than process parameters (k , J , induction OD) should be used for process optimization with respect to the feeding profile.

Keywords Quality by design · Process analytical technology · Multivariate data analysis · Design of experiments · Bioprocess optimization · Feed profile/feed strategy optimization · Design space development · rhaBAD promoter · Alkaline phosphatase · *E. coli*

List of symbols

t	Time (h)
S	Total amount of substrate in the cultivation broth (C-mol)
\dot{S}	Substrate feed rate (C-mol/h)
r	Conversion rate (C-mol/h)
q	Specific rate (g/g/h)
Y	Yield (C-mol/C-mol)
F	Flow/feed rate (g/h) for liquid
C	Concentration (C-mol/l)
X	Total amount of biomass in the cultivation broth (C-mol)
N	Total amount of ammonium in the cultivation broth (mol)
O_2	Total amount of oxygen in the cultivation broth (mol)
Z_i	Elemental composition of component i in biomass
p_i	Elemental composition of component i in substrate

P. Wechselberger (✉) · P. Sagmeister · C. Herwig (✉)
Vienna University of Technology, Institute of Chemical
Engineering, Research Area Biochemical Engineering,
Gumpendorfer Straße 1a, 1060 Vienna, Austria
e-mail: pwechsel@mail.tuwien.ac.at

C. Herwig
e-mail: christoph.herwig@tuwien.ac.at

H. Engelking · T. Schmidt · J. Wenger
Custom Manufacturing, Development Services,
Lonza Ltd., Visp, Switzerland

V	Volume of the cultivation broth (l)
M	Molecular weight (g/c-mol)
ρ_{feed}	Density of feed (g/l)
S_0	Feed concentration (g/l)
OD	Optical density 600 nm
J	Drop factor in percent of feed rate at end of exp. feed profile (%)
k	Specific growth rate if used for feed rate calculations (h^{-1})
μ	Measured specific growth rate (h^{-1})

Indices

in	Input
out	Output
conv.	Conversion
acc.	Accumulation
s	Substrate
x	Biomass
CO_2	Carbon dioxide
O_2	Oxygen
N	Ammonium
b	Base
O	Oxygen
H	Hydrogen
n	Nitrogen
m	Measured
c	Estimated
i	Item number i
j	Item number j
t	Time point t
0	Initial

Introduction

Quality by design (QbD approach)

Following the recent quality by design (QbD) initiatives, pharmaceutical process development based on sound science to increase process understanding emerged as a key demand from the side of the regulatory bodies [1–3]. There is a clear regulatory drive for more science and risk evaluation-based process development in place of empiric approaches. Besides the possibility for more process flexibility granted by the regulatory authorities, the enhanced process knowledge gained through QbD can pave the way for safer and more efficient processes, reducing product recalls, compliance procedures and post-approval modifications of change. The toolset for gaining process understanding includes design of experiments (DOE), multivariate data analysis (MVDA), quality risk management (QRM) and process analytical technology (PAT), as proposed from the regulatory authorities and exemplified in several excellent recent publications [4–9].

While the initiatives PAT and QbD are there for several years now, few processes are being registered as QbD processes. One of the problems might be that the idea is still very new and involved parties are lacking experience with QbD. This contribution aims at providing a generic methodology on how to apply QbD in recombinant bioprocess development to gain and demonstrate process understanding. We want to show feeding profile optimization for a recombinant process, describing the impact of typical feeding strategy-related process parameters from a cell's perspective. Furthermore, we want to stress that a process should be optimized for real quality attributes such as enzyme activities or product purity rather than plain concentrations. On the basis of this study, a generic science-based QbD methodology for the development of a feeding strategy for a process in red biotechnology is suggested, involving the use of physiological scaling parameters rather than empirically determined process parameters for the benefit of a faster, more cost-efficient process development according to QbD principles.

Quantitative data analysis for physiological state identification

In complex systems such as bioprocesses, experimentation provides the basis for process understanding. Following the path to process understanding [10], information needs to be extracted from the vast amount of measurable data and checked for consistency in order to distinguish between sensor, process and biological variability. The toolset for this task includes, but is not limited to, quantitative, multivariate evaluation of process data, statistical experimental design (DOE) and PAT. Information can refer to variables describing the process such as specific rates and yields [11]. Relevant process information is further analyzed to acquire knowledge on the process [3].

In a bioprocess, the product quality attributes and the process productivity are results of the trajectory of the physiological state of the culture [12, 13], hence we are looking for time-resolved variables describing that state. One way to do this is quantification of stress-related signal molecules such as ppGpp [14] and/or approaches such as metabolomics [15, 16] or transcriptomics [17, 18]. These methods can provide valuable insight into cell physiology, however, they tend to be very time consuming and it is not always possible to do this in industrial bioprocess development. Similar information can be acquired by analysis of primary metabolism, e.g. by quantification and comparison of specific rates and yields [12, 19, 20] which is faster and can be also done in real time using appropriate methods. To extract scale-independent information from bioprocesses and exclude effects related to the size of the cell population and initial conditions, specific rates and yields rather than

concentrations can be used as a valuable alternative for the comparison of experiments [10]. Following the calculation of rates and yields, the system of interest can be subjected to elemental balancing (system consistency check). This constitutes a valuable tool for the detection of gross errors or wrong assumptions, e.g. a sensor miscalibration or the formation of an unknown metabolite. In this contribution, the described rate-based methodology for quantitative bioprocesses analysis is applied on each run of a statistical experimental design, extracting information that is used to decipher the physiological state of the culture.

Optimization of feeding profiles in process development

Currently, empiricism and tedious, time-consuming experimentation efforts dominate the development and optimization of multiple parameters describing feeding profiles for (bio-) processes aiming at recombinant protein production. Next to the non-induced feed profile, an induction-phase feeding strategy (=time trajectory for the feed rate) must be developed that meets technical requirements of the industrial reactor as well as physiological requirements of the investigated host/vector/product system. While the former task is simply dependent on technical limits that need to be considered, the latter is typically overcome with elaborate experimentation in small-scale bioreactors, desirably by DoE approaches. Typically, the goal of optimization of bioprocesses is maximum space–time yield, which is defined as the target product in relation to space and time (in this contribution: active enzyme per culture broth volume (U/ml) in a defined time window, also see Fig. 2). The optimum feeding strategy with regard to optimum space–time product yield is connected to multiple physiological factors. It is typically promoter and/or product related [21], but also related to the primary metabolism including effects such as carbon depletion [22] as well as metabolic load and byproduct formation [23, 24]. Hence, finding the best feeding strategy requires optimization between the poles of production bioreactor capabilities and physiological requirements of the production system. The biomass or the amount of catalyst is an important factor when setting up a feeding strategy, since the target for optimization of the feeding strategy should be product related to biomass (catalyst), which is also a real quality attribute rather than protein concentrations and was shown by many authors [25–27]. It is known that there is a rapid break down of cell division in induced *E. coli* cultures [12, 28, 29], hence the biomass present at the beginning of the induction (induction OD) is an interesting tuning factor, which is also easily accessible. Pre-induction, the feed profile is

typically further defined by the specific growth rate. While some authors showed a clear impact on the product yield and specific productivity [30], other authors have not found such an effect [31, 32]. Another important tuning factor for feed profile optimization is the post-induction feeding rate. The C-flux into the cell typically has significant impact on the cell physiology and the product yield [25, 29, 33]. Summing up, a feed profile for a recombinant bioprocess can be described by induction OD, pre-induction growth rate and post-induction feed rate. In this contribution, the impact of such typical feed strategies-related process parameters on the specific and volumetric activity of a recombinant model system is investigated in lab-scale by means of a multivariate study. Consecutively, the physiological impact of the investigated process parameters on the biological system is discussed using specific rates and yields in combination with multivariate data analysis.

Goals

The goal of this contribution is the quantitative optimization of an expression system (lab-scale bioreactor). We want to increase process understanding and explore of interactions of key process variables using multivariate data analysis. Finally the relevancy of conventional feeding profile factors is discussed regarding optimization. On the basis of the obtained results, we suggest a methodology to replace conventional parameters related to the feed profile with physiologically more meaningful parameters.

Materials and methods

Expression system

Host

A modified K12 *E. coli* strain (kindly provided by Lonza Ltd., Visp, Switzerland) was used as a model system for the project. The strain has a Rhamnose-inducible expression system (rhaBAD promoter). The recombinant protein product was alkaline phosphatase. This is the same as native *E. coli* alkaline phosphatase, hence induced productivity has to be differentiated from native activity. The strain is unable to utilize Rhamnose as a C-source, therefore one time addition of low amounts of inducer was sufficient.

Media

A stoichiometrically defined media from the literature [26] was used in this contribution.

Bioreactor setup and on-line analytics

Bioreactor

Two stainless steel bioreactors Techfors-S with working volumes of 10 and 20 l from Infors (Bottmingen, Switzerland) were used. The system comes with a controller unit, which was used to adjust the process parameters: pH, temperature, aeration, reactor pressure and stirrer speed. Dissolved oxygen (DO_2) was controlled using a step controlled with reactor pressure, stirrer speed and air flow as manipulated variables. The pH was controlled using a Techfors-S integrated digital peristaltic pump and NH_4OH as a base. Air was filtered by a membrane-type filter and dispensed by a ring sparger. The culture vessel was sterilized at 121°C for 20 min by in situ steam sterilization prior to inoculation.

Off-gas analysis

CO_2 and O_2 in the off-gas were quantified by a gas analyzer (Servomex, UK; M. Müller AG, Switzerland), using infrared and paramagnetic principle, respectively. Air flow was quantified by a mass flow controller (Vögtlin, Aesch, Switzerland).

Process management

For recording of process data, the process information management system Lucullus from Biospectra (Schlieren, Switzerland) was used. This system was also used for closed loop control (feed bottle on balance) of the different feed profiles in the project.

Culture mode

A shake flask preculture (100 ml for inoculation of 6 l batch medium, in 1 l shaking flask with baffles) was inoculated from frozen stocks. After 8 h at 35°C , 180 rpm in the shaker, the preculture was used to inoculate the bioreactor. After depletion of glycerol in the batch mode, which was detected by a drastic drop in the CO_2 off-gas signal and an increase in dissolved oxygen (DO_2), an exponential fed-batch with specific growth rates from 0.1 to 0.2 h^{-1} was initiated to increase biomass concentration to OD 50 or up to 100. This was followed by sterile addition (membrane filtration $0.2\ \mu\text{m}$, directly into the reactor using a syringe filter) of low amounts of inducer (Rhamnose). Afterwards, a linear fed-batch phase with varying drop factors (J) was initiated, specified in relation to the feed rate at the end of the exponential growth phase. Such linear feed profiles are commonly used in industrial fed-batch processes to deal with limits for mass transfer of

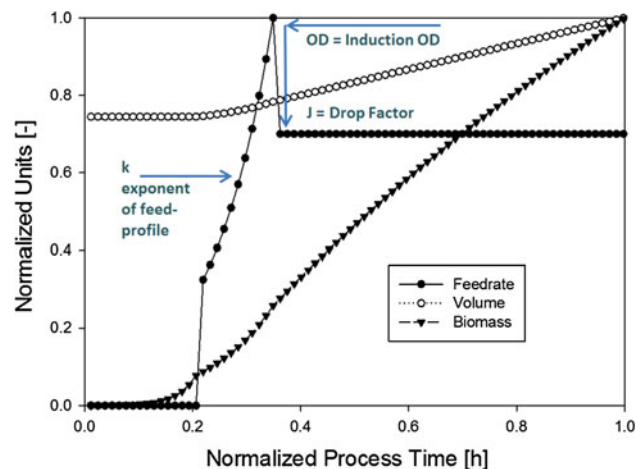


Fig. 1 Definition of factors with impact on the feed profile: OD, k and J . Units were normalized between 0 and 1

the bioreactor system. Oxygen demand is directly connected to the feed rate and the feed profile is often designed in a way to reach near the limit for oxygen mass transfer at the end of the exponential growth phase.

Feed profiles and inoculation OD were varied according to the experimental plan (see “Design of experiments”). An example for one DoE experiment, and how the DoE factors (OD, k , J) impact on the feed profile, is given in Fig. 1. The batch was followed by an exponential fed-batch and a linear induction phase. Equations 1 and 2 were used to calculate the feed profile for the exponential fed-batch. Symbols for these formulas are explained in Table 1. The linear feed in the induction phase was set equal to the last feed rate of the exponential feed profile adjusted by a drop factor according to the DoE experiment.

Feed rate in exponential fed-batch

$$F(t) = F_0 \times e^{k \times t} \quad (1)$$

Initial feed rate in exponential fed-batch

$$F_0 = \frac{k \times X_0 \times M_s \times \rho_{\text{feed}} \times V}{S_0 \times Y_{x/s} \times M_X} \quad (2)$$

Cultivation parameters for both the 10 and 20 l bioreactor were set as shown in Table 2.

Off-line analytics

Biomass

Biomass concentrations were quantified by gravimetric measurement after drying for 72 h at 105°C . Samples were centrifuged (5,000 rpm, 10 min) and the pellet was washed twice with distilled water to remove salts. The initial biomass concentration, which was required for the

Table 1 Constants for Eq. 2

Label	Symbol	Source
Specific growth rate before induction	k	DoE factor
Feed concentration	S_0	Gravimetric determination
Feed density	ρ_{feed}	Gravimetric determination
Initial biomass concentration	X_0	OD600
Molecular weight of substrate	M_S	Literature
Molecular weight of biomass	M_X	Literature or CHON analyzer
Biomass yield	$Y_{x/s}$	Literature or previous experiments

Table 2 Cultivation parameters for 10 and 20 l reactor

Parameter	Symbol
Temperature (°C)	35
Pressure (bar gauge)	0.3–1.6
Air flow rate (l/min)	About 1 vvm
Oxygen flow rate (l/min)	0–2
Stirring speed (rpm)	900–1,495
DO ₂	>20 %
pH	7 ± 0.01

calculation of F_0 (Eq. 2) and also for the induction OD (Fig. 1), was measured by photometric principle (OD 600 nm). Samples were diluted to OD < 0.8 and when applicable converted to a biomass concentration by means of a previously established linear regression.

Substrate and small metabolites

Substrate and small metabolite concentrations were quantified by enzymatic photometric principle in a roboting system (CuBiAn XC; Innovatis, Germany). As photometric principle can interfere with colored samples (supernatant of cultivation broth had a brownish color which increased in intensity with induction time), results were cross checked using an HPLC method (Supelcogel C-610, Sigma Aldrich, flowrate: 0.5 ml/min, eluent: 0.1 % H₃PO₄/NaN₃, 30 °C, RI detector).

The feed density was determined gravimetrically. Feed substrate concentration was determined by a density/feed concentration correlation.

Homogenization procedure

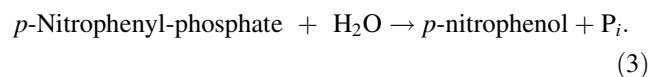
Two milliliters of the cell suspension was washed twice with 0.1 M Tris buffer pH 8.4 and frozen at –20 °C for further cell rupture. Samples were re-suspended in 20 ml of Tris buffer and pre-treated (2 min, setting 6) using a mixer

(UltraTurrax, IKA, Staufen, Germany) to break up any aggregates >50 µm which is the limit for the subsequent high pressure homogenization (Avestin EmulsiFlex, Canada). Cells were finally disrupted at 700 bar for 11 passages, which means more passages with lower pressure compared to the standard settings for *E. coli* recommended by the supplier (about 3 passages at 1,100 bar). A DoE study showed that this is beneficial for the alkaline phosphatase activity (data not shown), probably due to cavitation stress at higher pressures.

AP activity assay

Alkaline phosphatase catalyzes the de-phosphorylation of phosphate groups on a broad range of substrates at alkaline pH (Eq. 3). Product activity was determined by monitoring the absorption change (415 nm) of the substrate pNPP at 37 °C, pH 9.8. The Alkaline Phosphatase Yellow (pNPP) Liquid Substrate System for ELISA (Sigma, P7998) was implemented in an enzymatic analyzer robot (CuBiAn XC, Roche Diagnostics, Rotkreuz, Switzerland). After dilution of the sample, the reaction of the substrate was started by automatic pipetting of 10 µl of the sample to 120 µl of the substrate system. An absorption change rate was calculated from the time window between 20 and 60 s after reaction start. For direct calculation of activities from the reaction rate with the CuBiAn XC analytical robot, a calibration ranging from 0 to 1.6 U/ml with alkaline phosphatase from *E. coli* (Sigma, P5931) as standard was established. Activities are given with respect to this standard. The limit of quantitation (blank + 9 standard deviations) was determined to be 0.007 (U/ml) with a residual standard deviation of 0.0005 (U/ml).

Enzymatic reaction used in AP assay



Gel electrophoresis

SDS-PAGE was used for the qualitative assessment of intra- and extracellular protein according to methods in the literature [34].

Quantitative evaluation of bioprocess data

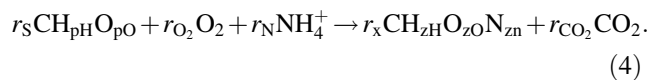
Data pretreatment

Off-line data were smoothed using the cubic smoothing spline function *csaps* included in the Matlab curve fitting toolbox (Curve Fitting Toolbox Software Version 3.2, Matlab R2011b). To get on- and off-line data on a uniform time interval for further data evaluation, the *interp1* Matlab function was used.

Conversion rates

Assuming oxidative metabolism, the bioreaction can be described by the following stoichiometric equation. Although there are many different chemical reactions running in parallel in living cells, the conversion rates in Eq. 4 represent the overall summarized effect of all the different reactions.

Stoichiometric equations for oxidative growth



General material balance

$$\text{Conversion} = \text{accumulation} - \text{input} + \text{output} \quad (5)$$

The conversion rates in Eq. 4 for the species substrate (S), biomass (X), carbon dioxide (CO₂), ammonia (N) as well as oxygen (O₂) can be derived from the general form of the material balance equation 5.

In fed-batch mode, the conversion rates can be calculated as follows (Eqs. 6 and 7):

Conversion rate for substrate uptake

$$r_S = \frac{d(S)}{dt} - \dot{S}_{\text{in}} + \dot{S}_{\text{out}} = -\frac{F(t)}{\rho_{\text{feed}}} S_0 \quad (6)$$

In fed-batch mode, the outflow term \dot{S}_{out} is zero and the accumulation term $\frac{d(S)}{dt}$ can be neglected, as long $\mu < \mu_{\text{max}}$ hence the conversion rate r_s is only dependent on the inflow term \dot{S}_{in} which is calculated from the feed rate.

Conversion rate for biomass production

$$r_x = \frac{d(X)}{dt} - \dot{X}_{\text{in}} + \dot{X}_{\text{out}} = \frac{d(X)}{dt} \quad (7)$$

Since there are no in- and outflow terms, r_x is equal to the accumulation term $\frac{d(X)}{dt}$ (Eq. 7).

Specific rates and yields

Conversion rates are the basis for the computation of yields (Eq. 8). Specific rates are calculated according to Eq. 9.

Calculation of yields

$$Y_i = \frac{r_i}{r_j} \quad (8)$$

Calculation of specific rates

$$q_i = \frac{r_i}{X} \quad (9)$$

Design of experiments

A standard central composite face design (CCF) was applied to model the dependency of enzymatic activity on variables of the feed profile: induction optical density

Table 3 DoE experiment coded factors

Exp no.	Exp name	k exp. feed	OD at induction	Jump at induction
1	Edge 1	-1	-1	-1
2	Edge 2	1	-1	-1
3	Edge 3	-1	1	-1
4	Edge 4	1	1	-1
5	Edge 5	-1	-1	1
6	Edge 6	1	-1	1
7	Edge 7	-1	1	1
8	Edge 8	1	1	1
9	Face 1	-1	0	0
10	Face 2	1	0	0
11	Face 3	0	-1	0
12	Face 4	0	1	0
13	Face 5	0	0	-1
14	Face 6	0	0	1
15	Center 1	0	0	0
16	Center 2	0	0	0
17	Center 3	0	0	0

-1, low; 0, mid; 1, high

(OD), pre-induction specific growth rate (k) and a factor related to the induction feed rate, the drop factor (J), also see Fig. 1. These factors were selected after a literature survey on factors frequently used to define the feed profile (see “[Optimization of feeding profiles in process development](#)”). The factor J was specified in percent of the feed rate at the end of the exponential feed profile. The experimental plan was implemented in a DoE Tool (Modde, Umetrics, Sweden). The CCF design provides information on the curvature of the relation between factors and responses and allows for quadratic type models requiring a low amount number of experimental runs. The design resulted in 17 experiments including 3 center points; ranges for the factors are shown in Table 3. The factors are shown in coded format (3 level setting for the factors: -1 = low, 0 = mid, 1 = high), which is frequently used in design of experiments.

Results

Multilinear regression (MLR) model development: impact of process parameters on the volumetric activity

A design of experiments (DOE) with feed profile parameters (OD, k , J) as factors (see Tables 3, 4) was carried out. The intracellular volumetric activity (U/ml) was chosen as a response, hence as CQA, since this factor represents quality with respect to the space time yield of the active

Table 4 Verification of the volumetric activity models and the specific activity model

Model	OD	k	J	q_s	Predicted activity (U/ml), (U/g)	Lower	Upper	Actual activity (U/ml), (U/g)
1	1	1	1	–	13.09	10.68	15.50	11.62
2	1	–	–	0.80	12.71	10.48	14.94	11.62
3	–	–	–	0.80	0.350	0.286	0.414	0.306

product. Typically maximum activity was reached after about 10 h of induction, afterwards the measured activities started to decline (Fig. 2). For a commercial process the maximum activities are of interest. Consequently, results from samples after 10 h of induction ± 5 h were averaged (=active enzyme per culture broth volume (U/ml) in a defined time window, which corresponds to space–time yield) and used as responses for the first MLR model development. Recombinant expression of alkaline phosphatase was further observed using SDS gel electrophoresis. A timely increasing band at 49 kDa in the supernatant of disrupted cells was identified as alkaline phosphatase (Fig. 3).

The volumetric activity was set in context with the original factors (OD, k , J , see “Design of experiments”). A MLR model was developed using backward selection. All process parameters proved to be significant. Figure 4 shows that the optical density at induction (OD) had a strong positive impact on the volumetric activity. Pre-induction growth rate (k) had a quadratic effect while induction feed rate drop (J) was interacting with (k). Although insignificant, pre-induction growth rate (k) and feed rate drop (J) were included as linear factors due to the obligatory hierarchy of the model. Figure 5 displays the model in the form of a response-surface plot. High volumetric activities can be found at high induction OD values. The quadratic effect of the pre-induction growth rate k results in predicted high activities at low and high k values. The positive interaction effect of the pre-induction growth rate k and the feed rate drop J impacts on the predicted activity in a positive way, therefore J shows a positive effect at high k values and a negative effect at low k values. Statistical evaluation of the model yielded a residual standard deviation of 1.453 and an R^2 of 0.918. ANOVA (F statistic, [35]) showed that the model was statistically significant and had no lack of fit.

Physiological multivariate data analysis and reduction of DoE factors

While the model in “Multilinear regression (MLR) model development: impact of process parameters on the volumetric activity” can be used to predict the volumetric activity, it provides limited physiological insight or process understanding. It is not straight forward to explain the

various local maxima and minima which are shown in Fig. 5 based on the factor OD, k and J . The obtained data was further analyzed using multivariate data analysis and quantitative evaluation by rates and yields. These are basically linear combinations of the original factors; hence they can be directly calculated and contain condensed information. The initial induction feed rate specific to the biomass or initial specific substrate uptake rate (q_s) can be calculated by Eq. 10 and depends on factors k and J . It decreases with induction time since the biomass further increases while the feed rate is linear.

Calculation of initial specific substrate uptake rate q_s (g/g/h)

$$q_{s\text{initial}} = \frac{k \times J}{100 \times Y_{X/S\text{fed-batch}}} \quad (10)$$

Hence, two of the factors in model 1 (k , J) are directly connected to the induction feed rate specific to the biomass and, since no substrate accumulation was detected, to the specific substrate uptake rate (q_s) as well. It was tested, whether the factors (k) and (J) can be condensed to the factor specific substrate uptake (q_s) without the loss of information. A second model was built with the induction OD and the computed specific substrate uptake rate q_s as factors (averaged in the time window of induction time = 10 ± 5 h, as established in “Multilinear regression (MLR) model development: impact of process parameters on the volumetric activity”) and the volumetric activity as response. The MLR model was built in analogy to the process parameter model of “Multilinear regression (MLR) model development: impact of process parameters on the volumetric activity”. Similarly induction OD showed a positive impact on the volumetric activity. The gross of the remaining variation was successfully explained by the specific substrate uptake rate q_s , which was included in the model also as a quadratic and cubic factor (Fig. 6). Figure 7 displays the obtained regression model as response contour plot. Values for q_s were normalized between 0 and 1, while OD is shown as coded factor (low, mid, high = -1 , 0 , 1). Statistical evaluation of the model yielded a residual standard deviation of 1.647 and an R^2 of 0.871. ANOVA (F statistic) showed that the model was statistically significant and had no lack of fit. Summarising, the observed variation in volumetric activity was successfully explained by the induction OD and the feed profile condensed physiological parameter q_s which means the factors were reduced from three to two.

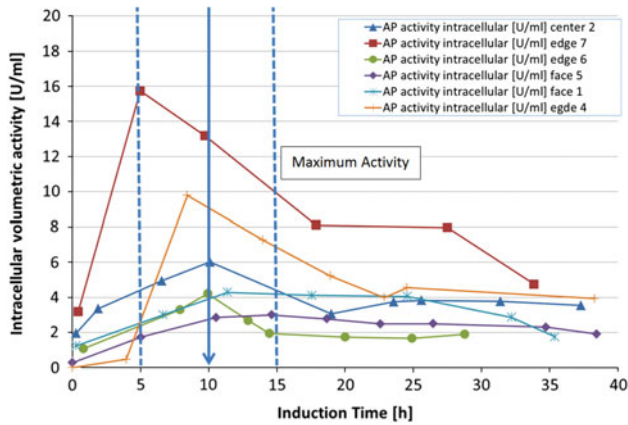


Fig. 2 Typical progression of AP activity after induction

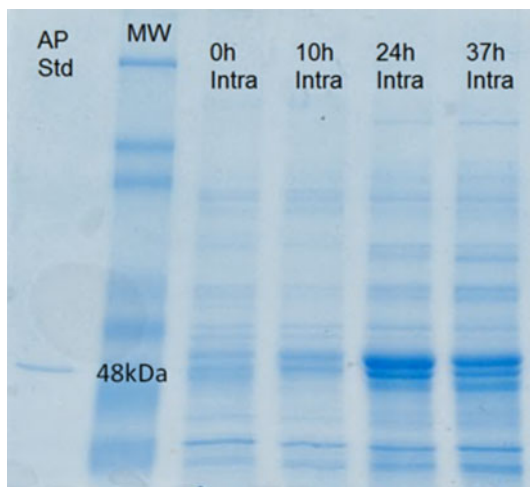


Fig. 3 SDS gel electrophoresis for experiment center 3 (see “Design of experiments”). Alkaline phosphatase standard (AP Std), molecular weights (MW), Intracellular samples (Intra, supernatant of disrupted cells). Dilution 1:10

Specific activities versus volumetric activities

The induction time OD had impact on both the process parameter model of “Multilinear regression (MLR) model development: impact of process parameters on the volumetric activity” and the physiological parameter model developed above, in a positive way. This is most probably due to the size of the cell population, hence the increased amount of catalyst (biomass) present at the time of induction. While the CQA should be the volumetric activity, since it directly relates to the space–time yield of the active enzyme, further knowledge can be acquired by looking at specific activities. For this purpose, it was tested whether the observed variation can solely be explained by the specific substrate rate q_s or also the C-flux per cell, with the specific activity (activity/biomass) as a response

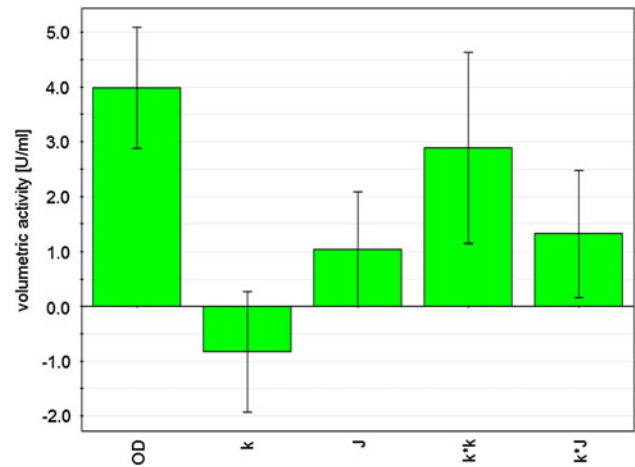


Fig. 4 Scaled and centered coefficients for the model described in “Multilinear regression (MLR) model development: impact of process parameters on the volumetric activity”. Volumetric activity served as response. Induction OD (OD), pre-induction growth rate (k) as well as the feed rate drop (J) were included as linear terms. Pre-induction growth rate (k) was included as quadratic term. The interaction term of J and k was included as well. The interaction term of pre-induction growth rate (k) and the induction OD (OD) showed the highest coefficient

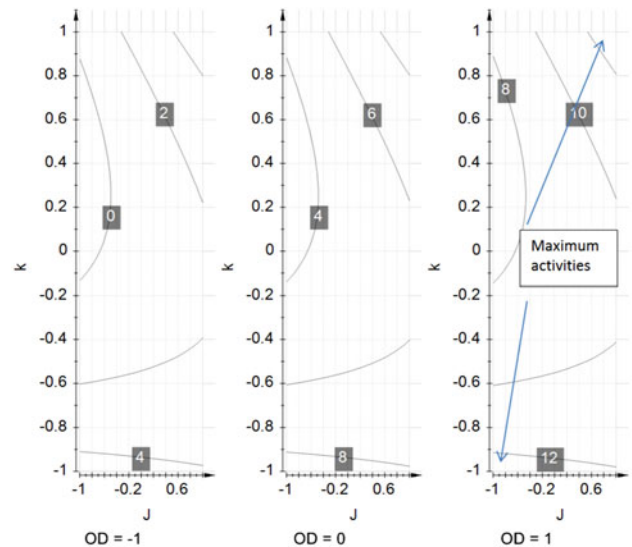


Fig. 5 Response contour plot of the model described in “Multilinear regression (MLR) model development: impact of process parameters on the volumetric activity”. The predicted volumetric activities as a function of the pre-induction growth rate (k) and the feed rate drop (J) are displayed in three plots at fixed OD values. Factors are shown in coded format (low, mid, high = $-1, 0, 1$). Isolines are labeled with the volumetric activity given in U/ml

accordingly. In analogy to regression model development of “Multilinear regression (MLR) model development: impact of process parameters on the volumetric activity” a very simple model with q_s as the only remaining factor

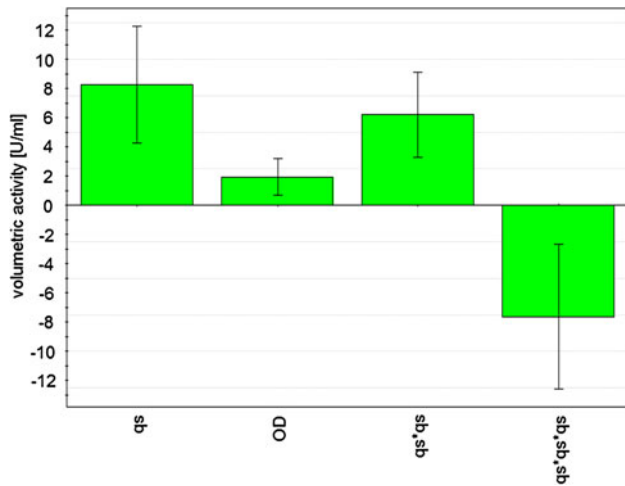


Fig. 6 Scaled and centered coefficients for the model described in “Physiological multivariate data analysis and reduction of DoE factors”. Volumetric activity served as response. Induction OD (OD) and the specific substrate uptake rate (q_s) were included as linear terms. The specific substrate uptake rate (q_s) was furthermore included with a quadratic term and a cubic term

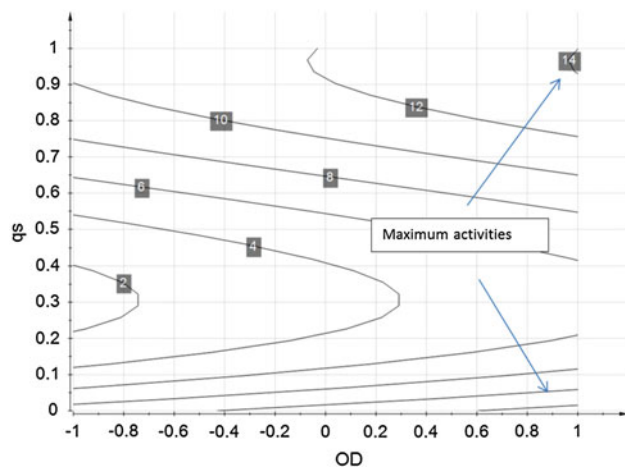


Fig. 7 Response contour plot of the model described in “Physiological multivariate data analysis and reduction of DoE factors”. The predicted volumetric activities as a function of the induction OD (x axis, coded values: low, mid, high) and the specific substrate uptake rate q_s (y axis, normalized between 0 and 1). Isolines are labeled with the volumetric activity given in U/ml

(Figs. 8, 9) was built. Very similar to the physiological parameter volumetric activity model described above, the specific substrate uptake rate (q_s) showed a significant linear, quadratic and cubic effect. Statistical evaluation of the model yielded a residual standard deviation of 0.04875 and an R^2 of 0.796. ANOVA (F statistic) showed that the model was statistically significant and had no lack of fit.

In addition to the analysis of the q_s on the specific and volumetric activity, it was checked whether similar information can be extracted from the specific growth rate μ . Hence, correlation of the specific- or volumetric activity

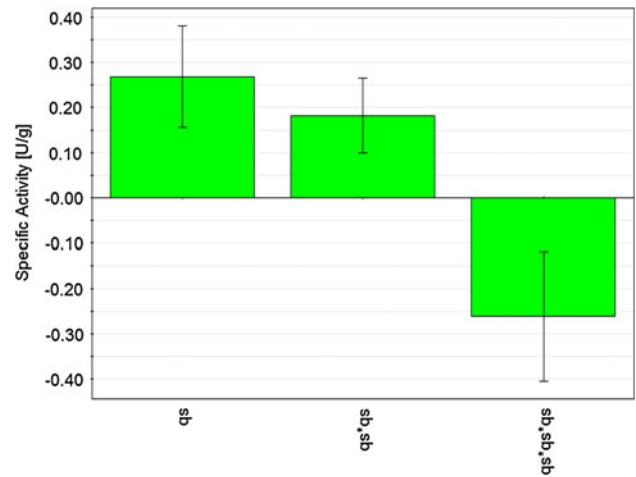


Fig. 8 Scaled and centered coefficients for the specific activity model described in “Physiological multivariate data analysis and reduction of DoE factors”. Specific activity served as response. The specific substrate uptake rate (q_s) was used as a factor in a cubic model

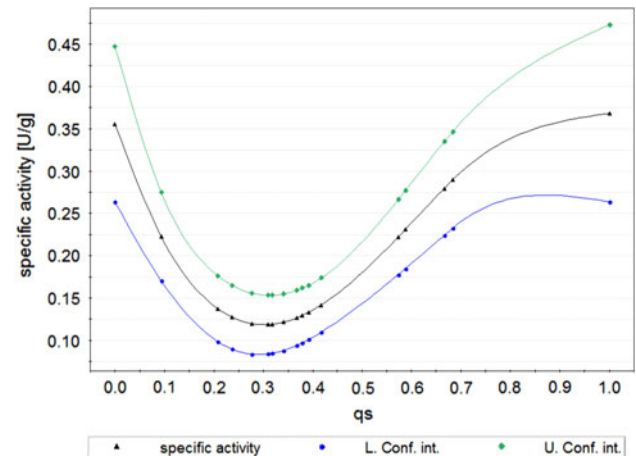


Fig. 9 Response prediction plot for the specific activity model describes in “Physiological multivariate data analysis and reduction of DoE factors”. The predicted specific activities (x axis, triangles) as a function of the specific substrate uptake rate q_s (y axis), which was normalized to values between 0 and 1. Confidence intervals at α 95 (balls, squares) are given

with the specific growth rate μ was evaluated (Fig. 10). No statistically significant correlation was found, ANOVA (F statistic) showed that the model was not statistically significant hence the model coefficients may be due to random noise. R^2 for this model was 0.259.

Verification

To confirm the results of the DoE, the run with the maximum volumetric activity was repeated (OD = 1, k = 1, J = 1). The resulting activities were within the error bars (α = 0.05) of the three models (see Table 4). The measured

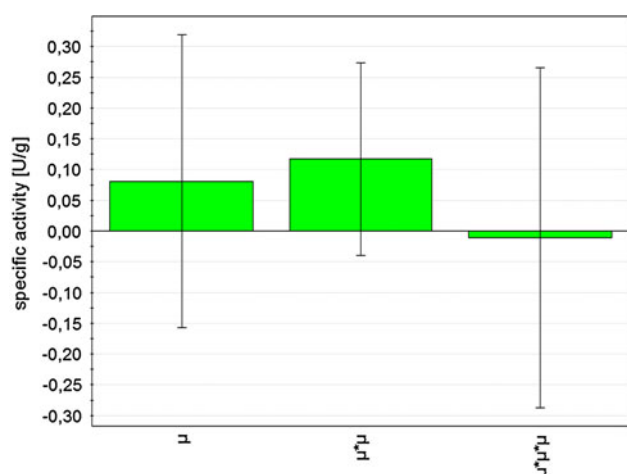


Fig. 10 Scaled and centered coefficients for the specific activity model described in “Physiological multivariate data analysis and reduction of DoE factors”. Specific activity served as response and the specific growth rate (μ), which was normalized to values between 0 and 1, was used as a factor in a cubic model. As error bars indicate the model was not statically significant and the effects may be due to random noise

q_s in the verification run was lower ($q_s = 0.8$, normalized value) than the q_s of the run in the corresponding DoE experiment ($q_s = 1$, normalized value).

Discussion

Impact of the specific substrate rate q_s on the biological system

The multivariate data analysis based on the physiological parameters showed that the observed variance in specific activity can solely be explained by the different specific substrate uptake rates q_s (Fig. 8). The effect is cubic with a minimum at medium to low substrate uptake $q_s = 0.3$ (normalized value, Fig. 9). Third-order terms should be supported by a sufficient number of experiments or resolution along the axis of the factor. Since there were enough experiments with different levels for q_s , the resolution was sufficient. For the limits set in this contribution, the maximum was found at highest substrate take ($q_s = 1$, normalized value), hence higher C-flux into the cell means elevated expression rates. A positive effect of q_s on the specific activity, in an *E. coli* fed-batch process for recombinant protein production, was also found by other authors [36], up to a level of $q_s = 0.85$ (g/g/h). A local maximum was also found at lowest substrate uptake ($q_s = 0$, normalized value). This can probably be attributed to the elevated native expression rates induced by C-starvation stress [37]. From a physiological point of view, a negative third-order term was expected, since it seems

reasonable that higher q_s and also higher protein expression comes with diminishing returns, due to increased metabolic load. Other authors have shown that foreign protein expression will even break down at some point with increased q_s (>0.85 (g/g/h)) [36]. Together with the observed effects at lowest q_s , the resulting kind of relationship has to be cubic.

All DoE models (see “Multilinear regression (MLR) model development: impact of process parameters on the volumetric activity”, “Physiological multivariate data analysis and reduction of DoE factors”, “Specific activities versus volumetric activities”) were able to predict the enzymatic activity of the verification run accurately within the error bars (see “Verification”). The measured q_s in the verification run was lower than in the DoE run (see “Verification”), which seems to be due to the feed forward kind of feed profile (without direct control of q_s , see “Culture mode”) and also due to measurement error. Nevertheless, also model 1 (see Table 4) was able to predict the enzymatic activity within error bars, even though it does not include q_s as factor. This seems to be due the diminishing returns of q_s at higher values as discussed above.

Optimization of feeding profiles using the specific substrate rate q_s according to QbD principles

The goal of optimization of process parameters is typically maximization of the space–time yield. The highest volumetric activities in this contribution were found after about 10 h of induction. Specific activity was not dependent on the induction OD, only on the specific uptake rate (Fig. 8). Consequently the highest space–time yield can be expected at highest specific uptake rates and highest induction OD. Furthermore, the observed variation in activity specific to the induction parameters was successfully explained solely by specific substrate uptake rate q_s , hence the linear positive effect of induction OD (Fig. 6) can be explained straight forward as an increase in biomass (catalyst). This simple relation yielded the same amount of information as the more elaborate 3-factor design with conventional factors. It was also shown that it is not important how biomass is produced before inoculation, which was also found by previous experiments [28]. Though this seems to be strain or expression system dependent, since other authors came up with different results [30], this is a very interesting understanding, since this makes developing feeding strategies much easier. From an experimental design point of view, feed strategy-related process parameters can be condensed into one physiological factor (q_s). Furthermore, the induction OD or also amount of catalyst (biomass) available for the production of alkaline phosphatase had strong positive impact. While this positive dependency on

induction OD does not necessarily hold true for any kind bioprocess, still a fair amount of experimentation efforts can be saved by using induction OD and q_s as DoE factors instead of the original three factors. The information content is the same and the feed profile can be optimized with significantly less experiments. Furthermore, this facilitates interpretation of results for the sake of increased process understanding and speeds up process development.

While the highest space–time yield was found at highest specific substrate uptake rates (=high C-flux into the cell) in this contribution, this is probably also dependent on other items such as strain, product and promoter. More importantly, the effect of different feed profiles can be analyzed by simply applying different levels of starting q_s or even q_s trajectories for the induced culture, also considering prior knowledge if available. Hence, this approach may be especially beneficial when extrapolating from a platform technology to a new recombinant product.

Highest and lowest q_s as shown in Fig. 9 are also practical limits for this process. At highest q_s , it is increasingly challenging to supply sufficient oxygen to the culture. With a q_s near zero, the culture runs into maintenance issues; also it is doubtful whether lowest substrate uptake is beneficial if a foreign recombinant protein is produced contrary to one which is also expressed natively.

An (approximate) control of the q_s is necessary following this approach. The authors suggest that the specific substrate uptake rate q_s is controlled utilizing real-time biomass estimation using available in-line hard or soft sensors. The optimum q_s for the individual process is found by adjusting different constant levels. This is a significant improvement from conventional empiric feed profiles to optimization by a physiological parameter such as q_s . As this requires equipment and/or know-how which is not available everywhere, we propose an alternate strategy which is much easier to execute in any lab. The specific rates can be adapted using OD correlations [25] or initial q_s values can be included in experimental designs. At induction, the biomass should be quantified by OD or any other fast off-line method. Afterwards the induction feed rate (g/h) is set by relation to the biomass (g) resulting in different starting levels of q_s . The disadvantage of this method is that during induction, the biomass will increase resulting in a decrease in specific uptake rate which approaches zero with various slopes, hence results will only be comparable within the first few hours. However, the adaption of the feed profile can be repeated in time to maintain a certain specific rate.

Other authors [38, 39] have also found dependency of the space–time yield on another physiological parameter, the specific growth rate (μ), but in this contribution the effect was not statistically significant, if present. This might be due to higher standard deviation on μ due to more

unfavorable error propagation from off-line samples compared to gravimetric feed rate measurement, which is in favor of the specific uptake rate over specific growth rate. Another problem with the specific growth rate is that it is not always straight forward to quantify and control this variable due to varying biomass stoichiometry.

Physiological downscale models based on q_s

In the course of this study it was shown that the variance in volumetric activity can be solely explained by the induction OD and q_s during the induction phase. Volumetric and also specific activity is independent of how the biomass is produced before induction. This opens new perspectives with respect to physiological downscale models for fast optimization of recombinant industrial processes. Typically, optimization efforts are carried out in fully equipped lab-scale bioreactors, using downscale models based on the volumetric power input, the stirrer tip speed or similar bioreactor-related parameters [40]. Furthermore, bioprocess-related parameters such as feeding profiles, media concentration and biomass concentrations are typically held constant during scale-up. From a physiological point of view, it would be more desirable to keep physiological parameters such as the specific substrate rate (q_s) constant, instead of process parameters detailed above. However, this is not straight forward due to inhomogeneity in the bioreactor mainly as a result of limitations in the power input for mixing [41, 42]. The dependency of space–time yield on substrate uptake, as shown in this contribution, clearly shows once more that this is a very important issue and should be considered during process development. The optimization of the specific substrate rate can be performed in small scale, e.g. multi bioreactor systems, also at low biomass concentrations, since it is not important how the biomass was produced and/or using diluted feeding solutions, which allows for very large range of different levels of specific substrate rates. If the inhomogeneity in a large scale bioreactor can be properly understood using appropriate models [43, 44], it is possible to predict the effect of nutrient inhomogeneity, once the relation between specific substrate uptake rate and space–time yield is identified.

Conclusion

- While Design of Experiments can be a useful tool, DoE factors have to be carefully selected e.g. by risk-based approaches involving a team of experts [3]. This is important, since one fails to do that, results of the DoE are useless. The interpretation of results in this contribution was greatly facilitated by using physiologically meaningful factors such as specific rates in contrast to

the original factors (OD, k , J). A statistically relevant relation between the specific uptake rate (q_s) and the specific activity was detected. No such statistically relevant relation could be shown for the specific growth rate due to increased noise on that rate, hence q_s is superior to μ .

- We want to stress that physiological variables such as the specific rates contain condensed information, which can accelerate process development due to a reduced number of experiments. Looking at specific activities it was shown that one specific rate contains the same information as three conventional factors describing the feed profile, while for volumetric activities the induction OD was required as well.
- Scale-up is potentially easier with condensed factors, since fewer factors have to be scaled.
- If information on the inhomogeneity in a large scale bioreactor is available, it is possible to predict the effect of nutrient inhomogeneity, once the relation between specific substrate rate and space–time yield is identified.

Acknowledgments The authors are very grateful to Lonza Ltd. (Visp, Switzerland) for their support and providing the *E. coli* strain used in this study. The study was financially supported by the Austrian Science Fund (FWF Project P24154-N17).

References

1. FDA (2003) Guidance for industry PAT-A framework for innovative pharmaceutical development, manufacturing, and quality assurance. <http://www.fda.gov>. Accessed March 2012
2. ISPE (2007) Creating QbD/PAT management awareness. <http://www.ispe.org>. Accessed March 2012
3. ICH (2008) Q8, Pharmaceutical development (R1). <http://www.ich.org>. Accessed March 2012
4. Rathore AS (2009) Roadmap for implementation of quality by design (QbD) for biotechnology products. *Trends Biotechnol* 27(9):546–553
5. Harms J, Wang X, Kim T, Yang X, Rathore AS (2008) Defining process design space for biotech products: case study of *Pichia pastoris* fermentation. *Biotechnol Prog* 24(3):655–662
6. Mandenius C-F, Graumann K, Schultz TW et al (2009) Quality-by-design for biotechnology-related pharmaceuticals. *Biotechnol J* 4(5):600–609
7. Rathore AS, Bhambure R, Ghare V (2010) Process analytical technology (PAT) for biopharmaceutical products. *Anal Bioanal Chem* 398(1):137–154
8. CMC-Biotech-Working-Group (2009) A-Mab: a case study in bioprocess development. 1–278. <http://www.ispe.org/pqli/a-mab-case-study-version-2.1>. Accessed March 2012
9. Thomas Garcia, Crahan Cook, Roger N (2008) PQLI key topics—criticality design space and control strategy. *J Pharm Innov* 3:60–68
10. Wechselberger P, Herwig C (2012) Model-based analysis on the relationship of signal quality to real-time extraction of information in bioprocesses. *Biotechnol Prog* 28(1):265–275
11. Wechselberger P, Seifert A, Herwig C (2010) PAT method to gather bioprocess parameters in real time using simple input variables and first principle relationships. *Chem Eng Sci* 65:5734–5746
12. Looser V, Hammes F, Keller M, Berney M, Kovar K, Egli T (2005) Flow-cytometric detection of changes in the physiological state of *E. coli* expressing a heterologous membrane protein during carbon-limited fedbatch cultivation. *Biotechnol Bioeng* 92(1):69–78
13. Christiansen T, Michaelsen S, Wuempelmann M, Nielsen J (2003) Production of savinase and population viability of *Bacillus clausii* during high-cell-density fed-batch cultivations. *Biotechnol Bioeng* 83(3):344–352
14. Striedner G, Cserjan-Puschmann M, Poetschacher F, Bayer K (2003) Tuning the transcription rate of recombinant protein in strong *Escherichia coli* expression systems through repressor titration. *Biotechnol Prog* 19(5):1427–1432
15. Korneli C, Bolten CJ, Godard T, Franco-Lara E, Wittmann C (2012) Debottlenecking recombinant protein production in *Bacillus megaterium* under large-scale conditions-targeted precursor feeding designed from metabolomics. *Biotechnol Bioeng* (Epub ahead of print)
16. Wittmann C, Weber J, Betiku E, Kroemer J, Boehm D, Rinas U (2007) Response of fluxome and metabolome to temperature-induced recombinant protein synthesis in *Escherichia coli*. *J Biotechnol* 132(4):375–384
17. Singh AB, Sharma AK, Mukherjee KJ (2012) Analyzing the metabolic stress response of recombinant *Escherichia coli* cultures expressing human interferon-beta in high cell density fed batch cultures using time course transcriptomic data. *Mol Biosyst* 8(2):615–628
18. Duerrschmid K, Reischer H, Schmidt-Heck W, Hrebicek T, Guthke R, Rizzi A, Bayer K (2008) Monitoring of transcriptome and proteome profiles to investigate the cellular response of *E. coli* towards recombinant protein expression under defined chemostat conditions. *J Biotechnol* 135(1):34–44
19. Herwig C, Marison I, Von Stockar U (2001) On-line stoichiometry and identification of metabolic state under dynamic process conditions. *Biotechnol Bioeng* 75(3):345–354
20. Jazini M, Herwig C (2011) Effect of post-induction substrate oscillation on recombinant alkaline phosphatase production expressed in *Escherichia coli*. *J Biosci Bioeng* 112(6):606–610
21. 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. *Microb Cell Fact* 7
22. Sanden AM, Prytz I, Tubulekas I, Forberg C, Le H, Hektor A, Neubauer P, Pragai Z, Harwood C, Ward A, Picon A, Teixeira de Mattos J, Postma P, Farewell A, Nystrom T, Reeh S, Pedersen S, Larsson G (2003) Limiting factors in *Escherichia coli* fed-batch production of recombinant proteins. *Biotechnol Bioeng* 81(2):158–166
23. Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D (2008) Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. *Biotechnol Bioeng* 100(1):177–183
24. Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D (2004) Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. *Biotechnol Bioeng* 85(4):367–375
25. Dietzsch C, Spadiut O, Herwig C (2011) A dynamic method based on the specific substrate uptake rate to set up a feeding strategy for *Pichia pastoris*. *Microb Cell Fact* 10(1):14
26. Wilms B, Hauck A, Reuss M, Syltatk C, Mattes R, Siemann M, Altenbuchner J (2001) High-cell-density fermentation for production of L-N-carbamoylase using an expression system based

- on the *Escherichia coli* rhaBAD promoter. *Biotechnol Bioeng* 73(2):95–103
27. Lin HY, Neubauer P (2000) Influence of controlled glucose oscillations on a fed-batch process of recombinant *Escherichia coli*. *J Biotechnol* 79(1):27–37
 28. Zabriskie DW, Wareheim DA, Polansky MJ (1987) Effects of fermentation feeding strategies prior to induction of expression of a recombinant malaria antigen in *Escherichia coli*. *J Ind Microbiol* 2(2):87–95
 29. Wong HH, Kim YC, Lee SY, Chang HN (1998) Effect of post-induction nutrient feeding strategies on the production of bioadhesive protein in *Escherichia coli*. *Biotechnol Bioeng* 60(3):271–276
 30. Curless C, Pope J, Tsai L (1990) Effect of preinduction specific growth rate on recombinant alpha consensus interferon synthesis in *Escherichia coli*. *Biotechnol Prog* 6(2):149–152
 31. Shin CS, Hong MS, Lee J, Bae CS (1997) Enhanced production of human mini-proinsulin in fed-batch cultures at high cell density of *Escherichia coli* BL21(DE3)[pET-3aT2M2]. *Biotechnol Prog* 13(3):249–257
 32. 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):1–11
 33. Levisauskas D, Galvanauskas V, Henrich S, Wilhelm K, Volk N, Lubbert A (2003) Model-based optimization of viral capsid protein production in fed-batch culture of recombinant *Escherichia coli*. *Bioprocess Biosyst Eng* 25(4):255–262
 34. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London, UK)* 227(5259):680–685
 35. Funk W, Dammann V, Donnevert G (2005) Qualitätssicherung in der Analytischen Chemie
 36. Ramalingam S, Gautam P, Mukherjee KJ, Jayaraman G (2007) Effects of post-induction feed strategies on secretory production of recombinant streptokinase in *Escherichia coli*. *Biochem Eng J* 33(1):34–41
 37. Ozkanca R (2002) β -Galactosidase activity of *Escherichia coli* under long-term starvation, alterations in temperature, and different nutrient conditions in lake water. *Int Microbiol* 5(3):127–132
 38. Yoon SK, Kwon SH, Park MG, Kang WK, Park TH (1994) Optimization of recombinant *Escherichia coli* fed-batch fermentation for bovine somatotropin. *Biotechnol Lett* 16(11):1119–1124
 39. Kilikian BV, Suarez ID, Liria CW, Gombert AK (2000) Process strategies to improve heterologous protein production in *Escherichia coli* under lactose or IPTG induction. *Process Biochem (Oxford)* 35(9):1019–1025
 40. Santek B, Novak S, Horvat P, Maric V (2004) Scale-up criteria of bioreactor. *Kem Ind* 53(1):7–24
 41. Bylund F, Guillard F, Enfors SO, Tragardh C, Larsson G (1999) Scale down of recombinant protein production: a comparative study of scaling performance. *Bioprocess Eng* 20(5):377–389
 42. Neubauer P, Junne S (2010) Scale-down simulators for metabolic analysis of large-scale bioprocesses. *Curr Opin Biotechnol* 21(1):114–121
 43. Moilanen P, Laakkonen M, Aittamaa J (2006) Modeling aerated fermenters with computational fluid dynamics. *Ind Eng Chem Res* 45(25):8656–8663
 44. Delvigne F, Destain J, Thonart P (2005) Structured mixing model for stirred bioreactors: an extension to the stochastic approach. *Chem Eng J (Amsterdam, The Netherlands)* 113(1):1–12