

IMPLEMENTATION OF
PROCESS ANALYTICAL TECHNOLOGY
TOOLS IN
MICROBIAL FERMENTATION PROCESSES
FOR RECOMBINANT THERAPEUTIC
PROTEIN PRODUCTION

Dissertation

zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

genehmigt durch das

- Zentrum für Ingenieurwissenschaften -
der Martin-Luther-Universität Halle-Wittenberg

von

Herrn Dipl.-Ing. Stefan Gnoth
geb. am 22. März 1981 in Neubrandenburg

Geschäftsführender Direktor: Prof. Dr.-Ing. habil. Holm Altenbach

Gutachter: Prof. em. Dr. rer. nat. habil. Andreas Lübbert
Prof. Dr.-Ing. habil. Jochen Büchs

Tag der öffentlichen Verteidigung: 28.02.2011

Halle (Saale), den 07. September 2010

CONTENTS

CHAPTER		PAGE
<i>1</i>	Introduction	1 – 22
<i>2</i>	Simple Adaptive pH Control in Bioreactors Using Gain-Scheduling Methods	23 – 42
<i>3</i>	Advanced Control of Dissolved Oxygen Concentration in Fed Batch Cultures during Recombinant Protein Production	43 – 58
<i>4</i>	Product Formation Kinetics in Genetically Modified <i>E. coli</i> Bacteria: Inclusion Body Formation	59 – 68
<i>5</i>	Selective Expression of the Soluble Product Fraction in <i>Escherichia coli</i> Cultures Employed in Recombinant Protein Production Processes	69 – 88
<i>6</i>	Fermentation Process Supervision and Strategies for Fail-Safe Operation: A Practical Approach	89 – 113
<i>SUMMARY</i>		114 – 115
<i>ZUSAMMENFASSUNG</i>		116 – 117
<i>PUBLICATION LIST</i>		118 – 119
<i>CURRICULUM VITAE</i>		120
<i>DANKSAGUNG</i>		121
<i>ERKLÄRUNG</i>		122

INTRODUCTION

Parts of this chapter have been published in:

Bioprocess Biosystems Engineering

Gnoth, S., Jenzsch, M., Simutis, R., Lübbert, A., 2008. Control of cultivation processes for recombinant protein production: a review. *Bioproc. Biosyst. Eng.* 31: 21–39.

Journal of Biotechnology

Gnoth, S., Jenzsch, M., Simutis, R., Lübbert, A., 2007. Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control. *J. Biotechnol.* 132: 180–186.

PROCESS ANALYTICAL TECHNOLOGY

In the last decades biotechnology has been evolved into a promising key technology with application fields in medical care, food technology, agriculture and fine chemicals. Generating annual sales of approximately 33 billion dollars in 2004 and representing about 20-25 % share of the total number of new medical entities in the United States (US) and Europe (Walsh 2006), the production of biopharmaceuticals for the health care sector is an important and profitable market. Among the top selling products the majority are oncology and oncology supporting products (monoclonal antibodies and growth factors), followed by hormones (predominantly insulin), fusion proteins and cytokines (Lawrence 2007). The future seems promising, as the yearly growth rates of sales of biologics in the United States are in the double digit range of 10 %, a development clearly outperforming US pharmaceuticals which lagged behind with 3-4 % in 2007 (Aggarwal 2008).

However, contrary to the big milestones made in the fields of biochemistry, molecular biology and microbiology for the development of active pharmaceutical ingredients (API) as well as the identification of new application fields, less effort is spent in bioprocess engineering and the cost effective large scale production of the biologics. It is widely accepted that, compared to other sectors of production industry, e.g., automobile and petrochemistry, these processes generally have deficits in terms of rates of technological process innovation as well as detailed, quantitative process understanding (Dünnebier and Tups 2007). These facts seem to be not surprising, since the selling products exhibit a long duration of patents and generate high revenues in almost exclusive application fields. Moreover, after years of intensive research, subsequent testing of promising drug candidates in capital-intensive clinical studies and the final drug approval, it is generally difficult from the manufacturer's point of view to perform significant changes in the production process. In addition, rigid requirements of regulatory authorities in the past have generally lowered the drive of continuous process optimization.

Recently, current pricing systems for pharmaceuticals in many industrialized countries are up for debate, because of fast-growing health care costs as a consequence of the demographic change. Both, political pressure as well as the fact of the standstill in technical innovation in biopharmaceutical production, have brought the US American Food and Drug Administration (FDA) to raise their Process Analytical Technology (PAT) initiative (FDA 2004). Moreover, with the advent of approval and production of biogenerics (biosimilars) the established companies become increasingly pressurized, since they have to compete with comparable but cheaper APIs. This development is now explicitly supported by the European Medical Agency (EMA) with their biosimilar guidance documents (EMA 2005, Walsh 2006, Rathore 2009a).

The intent of the PAT framework is to catalyze improvements in process understanding and control of the manufacturing process while at the same time being consistent with the FDA's maxim: "quality cannot be tested into products; it should be built-in or should be by design". As it might be misinterpreted by the term "Analytical", the PAT guidance is not just about implementing modern measurement devices. It further aims at using modern multivariate techniques to gain quantitative understanding, especially about the complex cause-effect interrelationships between the various process variables and their detailed influence on the process quality. Concomitantly, reproducibility is assured by utilizing signals of on-line available process data and, subsequently, by using this information for feedback control of critical process variables within tight specification limits. Finally, the FDA's focus is on the quality by design approach, which means the consequent utilization of gained process knowledge and its quantitative conversion into models, ending up in a continuous process improvement. PAT has been discussed in industry (DePalma 2004) as well as in academia much more intensively than any other regulatory topic before (Hinz 2006, Rathore 2009b). It is recognized as an important paradigm change in the agencies' way to inspect and approve production processes for

pharmaceuticals. The initiative was immediately taken up by the EMEA and the Japanese Ministry of Health, Labour and Welfare (MHLW) and continues within the common steady International Conference on Harmonization (ICH) of the leading agencies. In particular, the ICH issued a couple of guidelines (Q7–Q10) defining the framework for the new approach to process approval (ICH 2000-2008). PAT is embedded into the new concept of the agencies, which is going to leave the more or less formal fixed set of standard operational procedures that do not take the specialties of the concrete production processes into account. It will be replaced by a risk-based procedure, where the importance of the individual operations with respect to their impact on the quality of the final product and, hence, on the safety for the patients is examined.

From the viewpoint of biopharmaceutical production, changes at the production process always bear a certain risk (Chirino and Mire-Slius 2004). Possible violation of specification limits or a loss of a production lot must be traded off against desired outcomes of enhanced yields, improved technologies and economical benefits. Process alterations still involve regulatory adaptation and have to be evaluated with respect to possible negative effects on the safety and efficacy of the product. This can only be quantified if the influence of the changes on the process or product quality is sufficiently well understood. Hence, central to the new approach is the knowledge about the essential details of process equipment and its dynamical operation during up- and downstream processing. The intent is to directly use the gained knowledge for a continuous improvement of process quality.

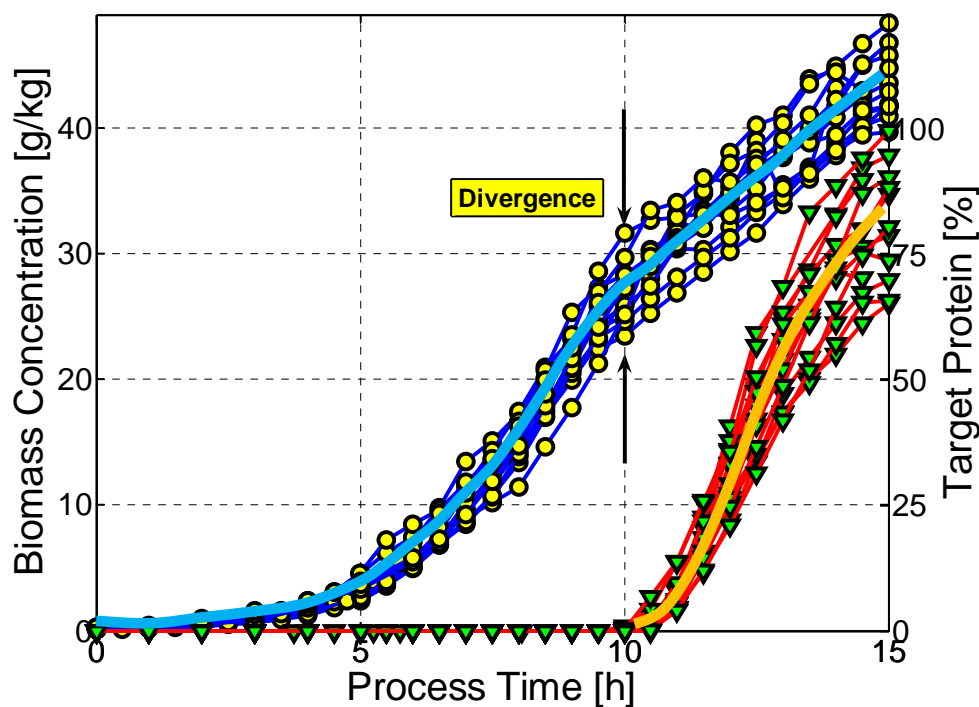


Figure 1. Off-line measurements of product (*triangles*) and biomass concentration (*circles*), typically found in recombinant protein manufacturing for a standard cultivation process without implementation of advanced feedback and supervisory control systems (Jenzsch 2006a, Gnoth *et al.* 2007). As can be seen, the deviation from the desired profiles (*blue* - biomass, *orange* - product) is high and therefore the batch-to-batch reproducibility of both state variables is rather low.

In particular, the quality of recombinant therapeutic proteins, which cannot be measured on-line, is closely related to the quality of the production process. Process quality itself is directly related to its reproducibility. Thus, in practice, the straightforward way of assuring protein quality is guaranteeing reproducibility of the processes wherein the product is formed. A typical example, that represents the actual situation of the batch-to-batch reproducibility in biopharmaceutical

production, is given in Figure 1. It shows profiles of biomass and product concentration values of 12 cultivation processes operated with the same control strategy. Here, the batch-to-batch reproducibility can be characterized by the standard deviation of the corresponding concentration trajectories of successive production runs from a desired ideal profile (blue and orange line in Figure 1). It is immediately clear from Figure 1 that reproducibility, expressed by a significant scatter of biomass and protein values, and therefore process robustness against disturbances is rather low.

On the way to solve that general problem in biopharmaceutical production processes PAT sets up the framework for purposeful and defined process improvements. According to the guidance, a process is considered well understood when, (I) all critical sources of variability are identified and explained, (II) variability is managed by the process, (III) product quality properties can be accurately and reliably predicted over the design space established, for materials used, process parameters, manufacturing, environmental, and other conditions.

In the thesis presented here, several specific examples are given how to implement the general PAT idea into practical operation of fermentations for recombinant therapeutic protein production. The approaches developed apply to different aspects of fermentation technology and basically encompass the improvement of the measurement data quality and the use of process models. Subsequently, these models are exploited for optimization, prediction of process behavior as well as for the application of feedback control and process supervisory systems to counteract random errors and systematic failures. The highlighted problems are of practical value and the developed solutions meet the demands of application in research as well as in industry.

CONCEPTUAL FRAMEWORK

Quality of Measurement Data

The quality of fermentation processes crucially depends on the precision of which central process parameters are maintained with respect to the desired values. Prior to building complex advanced monitoring, feedback control and supervision techniques, it is worthwhile to focus attention on the improvement of measurement data quality, i.e., the signal-to-noise ratio of process variables has to be enhanced. The reason is that systematic or random errors made during measurement and control of essential process variables, basically affect all other recorded process variables. Since these errors increasingly propagate in the direction of more complex models, the best way of increasing the overall prediction quality of models is to enhance the information content of measurement data itself.

In this context, the parameters of fermentation temperature, dissolved oxygen concentration pO_2 as well as the pH value have to be singled out. These parameters tremendously influence the physico-chemical observed in the liquid culture medium and the gas-liquid inferential area. For instance, even small fluctuations in the pH value cause variations of the carbon dioxide solubility in water and the equilibrium with bicarbonate ions leading to either additional CO_2 accumulation or dissolution effects in the culture medium (Royce 1992, Spérandio and Paul 1997). The result of those fluctuations is monitored by the off-gas analytics, where the variance in the CO_2 volume fraction might be misinterpreted as changes in the metabolic activity of the cells. Consequently, after the effects between pH changes and CO_2 solubility had been identified, process models were applied to quantitatively understand the fundamental cause-effect relationship followed by appropriate controller design on the basis of model-supported simulation studies. In practical validation experiments, the designed parameter adaptive controller, which used the gain-scheduling approach, showed to be highly capable to control pH within narrow tolerance limits. Thereby, associated fluctuations in the CO_2 off-gas signals were eliminated.

Another important challenge of the process operator is to strictly control the pO_2 , i.e., the partial pressure of oxygen dissolved in the culture medium, to a given setpoint. With approximately 8 mg L^{-1} , oxygen has a low solubility in water and especially in aerobic fermentations in large scale bioreactors oxygen transfer into the liquid phase might become a limiting problem. Moreover, the fast dynamics in oxygen transfer and uptake requires a special design of the pO_2 controllers. Standard proportional-integral-derivative (PID) controllers with fixed parameters are not able to cope with this problem, since a good performance is only assured for a certain working point. Instead, process analysis showed that parameter adaptation by a simple gain-scheduling approach significantly improved control quality (Kuprijanov *et al.* 2009). In doing so, the PID parameters are continuously adjusted to the changing process dynamics, which are primarily caused by the metabolic activity of the cells. As a metabolic indicator for the process dynamics the CO_2 production rate was selected to be the scheduling variable. The improvement of pO_2 control by the gain-scheduling approach is illustrated in Figure 2.

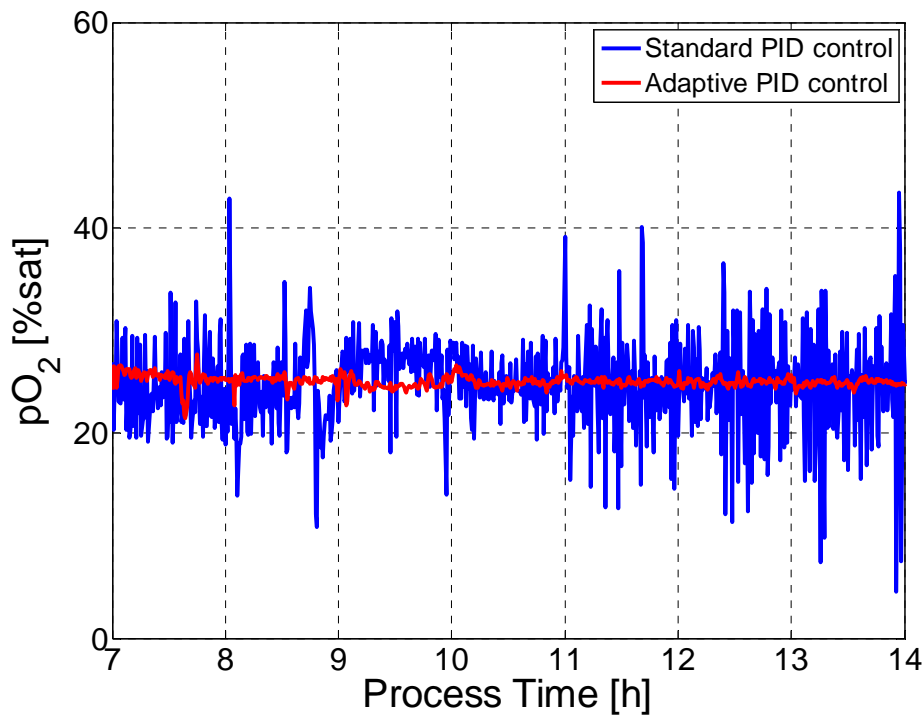


Figure 2. Comparison of two fermentation runs with respect to the measurement data of pO_2 . The *blue* curve depicts the typical oscillating behavior of pO_2 control obtained for a standard PID controller with fixed parameters. The application of gain-scheduling on the controller parameters results in a significant reduction of measurement noise (*red line*).

The benefits obtained by an improved pO_2 control are primarily a more stable stirrer speed profile and, which might be more important for model development, a significant increase in the signal-to-noise ratio of off-gas values O_2 and CO_2 . These on-line measurements are central variables for superior monitoring and control models later on used, because they provide information about the physiological state of the culture as well as about the amount of biomass formed.

Quality by Design

A central point often claimed in the course of the PAT discussion is about building quality into products by design. Process improvements should be performed in a directed manner on the basis of detailed quantitative knowledge about the underlying functional interrelationships. As an indispensable prerequisite, systematic experiments allow for the identification of influencing factors and for the development of a suitable simplified model idea. In this context, the design space is defined as the combinational range of the investigated parameters where the designated model maintains its validity. In terms of fermentation development, it is desired to extract the functional relationship between growth and product formation with a minimum of effort. Conventional batch and fed-batch control strategies enable the exploration of growth and product formation kinetics either under substrate conditions exceeding the decisive K_s -value or under substrate limitation. However, in the substrate limited fed-batch mode investigation of growth and product formation is restricted toward maximum growth rates. On the other hand, in excess of substrate significant amounts of overflow metabolites (e.g., acetate) are formed, which were shown to inhibit product formation (Luli and Strohl 1990, Jensen and Carlsen 1990, Roe *et al.* 2002) and therefore distort the estimation of the pure π - μ -relationship.

Feed Pulse Control

The probing control technique, first described by Åkesson *et al.* (1999), proved to be suitable to investigate product formation under near-maximal growth conditions without accumulation of substrate and inhibiting by-products. However, in the course of the presented work, the original concept was further extended by performing quasi continuous feed pulses and immediate analysis of the corresponding respiratory responses in terms of pO_2 . By switching periodically between substrate limited and slightly unlimited states, the trajectories of substrate uptake and the corresponding biomass growth rate, at the same time avoiding overflow metabolism, can be identified.

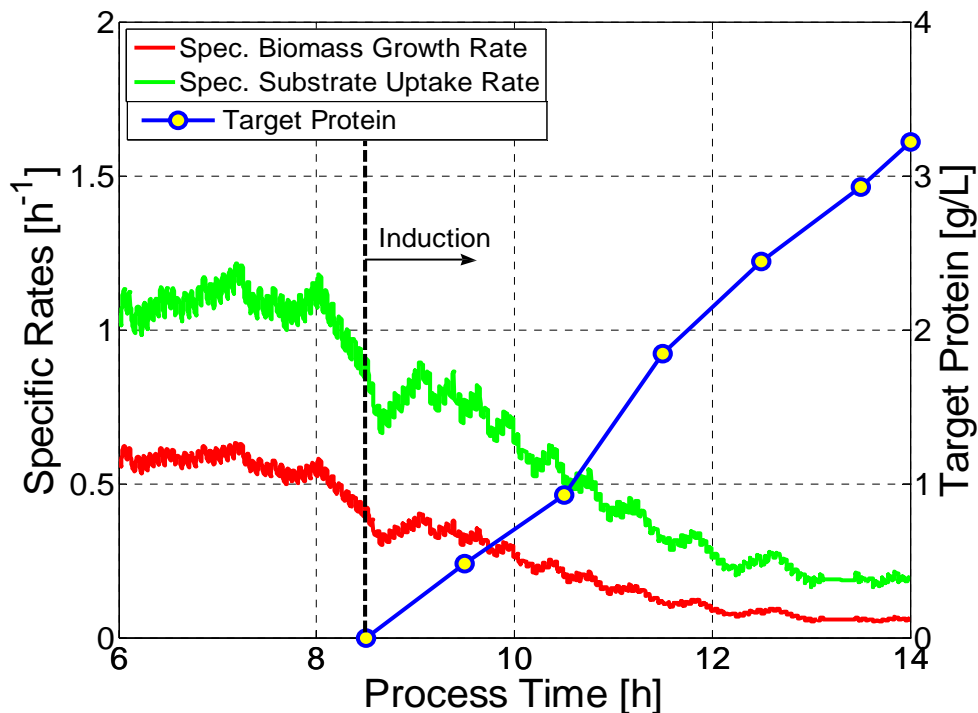


Figure 3. Typical behavior of the critical substrate uptake rate (*green line*) and the corresponding biomass growth rate (*red line*) after induction with 1 mM IPTG, here at the example of a heterologous recombinant protein expression of the reference protein HumaX (*circles*). As can be seen, the respective rates significantly drop with evolving target protein. The results were obtained by using the probing control technique and allow for the investigation of product formation kinetics under quasi maximal growth conditions without accumulation of inhibiting overflow metabolites, e.g., acetate.

It is known, that upon induction of heterologous protein expression in high cell density cultures the biomass growth rate tends to dynamically change, primarily caused by the metabolic burden (Bentley *et al.* 1990, Bhattacharya and Dubey 1995, Glick 1995, Neubauer *et al.* 2003, Rozkov *et al.* 2004). The simple technique paves the way for an easy identification of such fast changing and complex growth trajectories. As an example, experimentally determined substrate uptake and biomass growth rate profiles for an *Escherichia coli* (*E. coli*) fermentation process after induction of the recombinant protein expression, are shown in Figure 3. Moreover, as the functional relationship between substrate uptake and respiratory response is used, the proposed controller is not restricted to a specific host, but instead is universally applicable in a wide range of aerobic fermentations. The feed pulse controller therefore provides a simple means to obtain the practical achievable design space, which plays a prominent role in the early stage of process development. It proved to be a valuable tool in fermentation processes for the identification of product formation kinetics of different recombinant proteins (NGIPR, HumaX).

Process Models

A functional description of process know-how is a key requirement for a continuous improvement of process quality. A major approach in bioprocess engineering is to represent this knowledge by means of quantitatively exploitable process models. The models developed herein can be seen as a simplified representation of complex relationships observed during fermentation. They are results of detailed process analysis, which finally leads to an improved process understanding. The established models have been directly utilized for simulation studies by explaining the process variability in a knowledge-based cause-effect analysis. Moreover, quantitative process models, either mechanistic or data-driven, have been used for process optimization (Levisauskas *et al.* 2003), feedback control (Jenzsch *et al.* 2006b) as well as for model-based process supervision (Jenzsch *et al.* 2006a). Obviously, such models are of practical use, since they enable reliable prediction of the process behavior as well as the occurring kinetic process dynamics. And this prediction capability is a clear evidence of profound process knowledge.

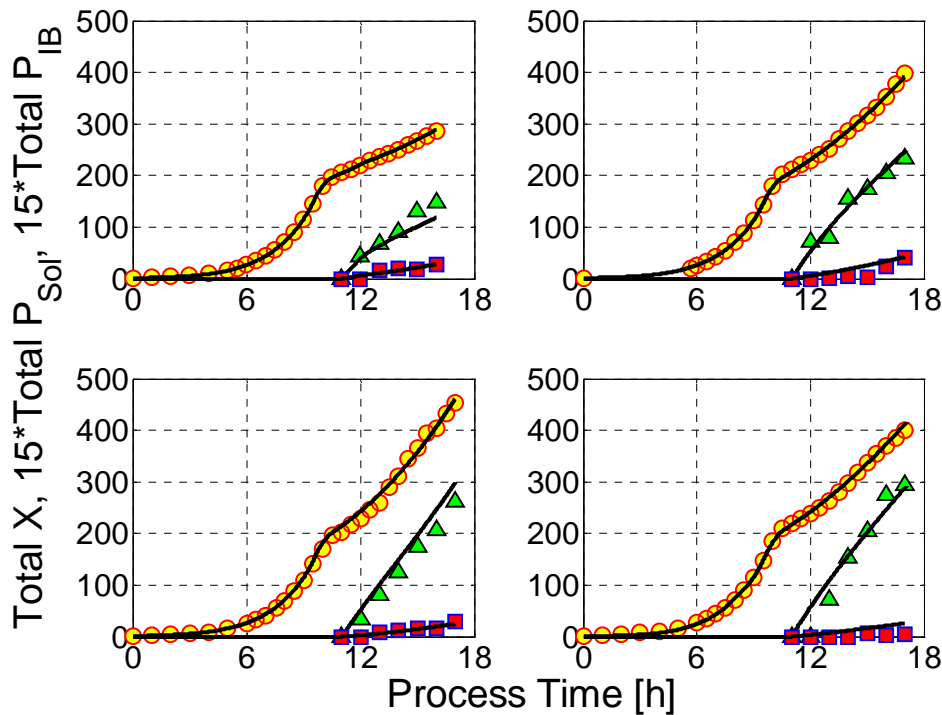


Figure 4. Model estimations (*black solid line*) using the proposed hybrid network approach together with the corresponding off-line measured values of biomass (*circles*) and both fractions of target protein HumaX (*soluble – triangles, inclusion bodies – squares*). The hybrid network is capable to accurately predict the state variables and their kinetics for various cultivation conditions, i.e., different growth rates and temperatures during fermentation.

The description of the observed process characteristics in form of quantitatively exploitable, conventional models is a major problem, since it often turns out to be a time consuming procedure not always leading to satisfying results. It requires detailed knowledge about the cultivation process as well as a high degree of modeling experience to extract the complex underlying relationships among process variables and further to express the corresponding quantitative mechanistic terms. Thus, it was a main objective during this work to develop a universal approach for the identification and the quantitative prediction of state variables as well as their dynamical behavior. In the presented hybrid modeling approach, the specific rate kinetics is predicted by artificial neural networks (ANN) and is combined with the general valid mass balance equations. The implemented artificial neural networks were trained with on-line

measurable process data. By using nonlinear response functions the inherently nonlinear process dynamics were identified in a completely data-driven way. It resulted in an excellent prediction quality compared to conventional process models, and, moreover, circumvented the formulation of complex mathematical terms. Results obtained by the hybrid network approach are presented in Figure 4 at the example of the HumaX fermentation process.

Since the developed approach works completely data-driven, there is no need for process specific model assumptions, and it is therefore not restricted to a certain cultivation process or host strain. This was demonstrated by the identification of completely different product formation kinetics (Gnoth *et al.* 2008, Gnoth *et al.* 2010a). Moreover, the system is perfectly suited as an on-line state estimation and process supervision tool, because all required inputs for the hybrid network are real-time accessible.

Process Optimization

Incorporating process know-how into quantitatively utilizable models is essentially the basis for directed process optimization procedures. With the presented approach as well as with conventional models process design studies are easy to be performed. Once the network is trained, optimal strategies for the used network inputs are determined by common optimization search routines.

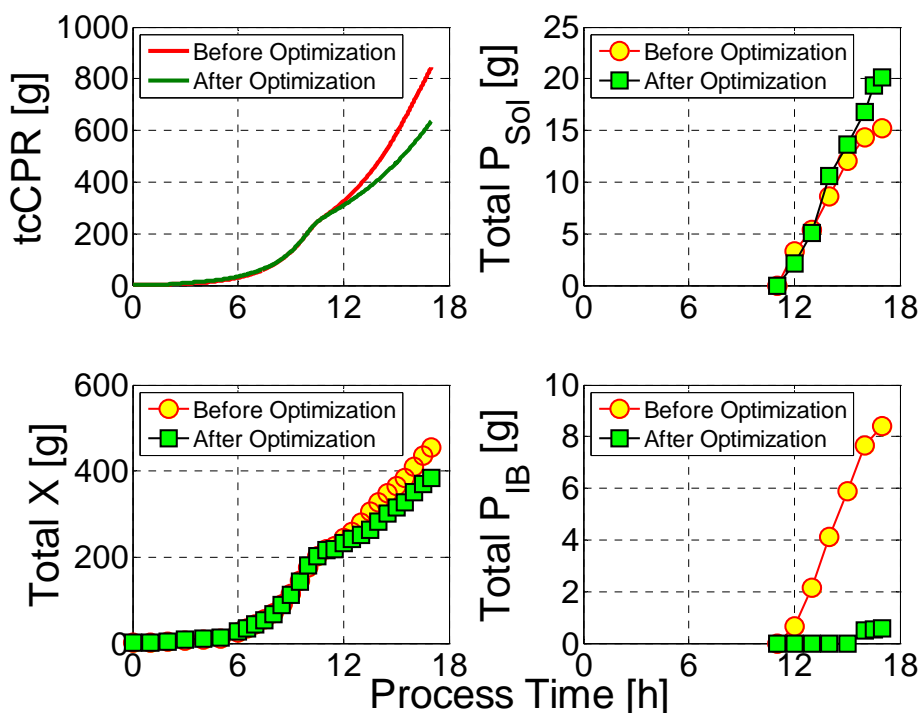


Figure 5. Comparison of central state variables characterizing the quality and the performance of the fermentation process for production of the recombinant therapeutic protein HumaX. After model-based process optimization (*green*), the mass of desired soluble protein was increased (*upper right*), whereas the unwanted inclusion body fraction (*lower right*) was suppressed concomitantly. The specific biomass growth rate as well as the fermentation temperature after induction were identified to strongly affect soluble protein formation.

Results of practical validation experiments after numerical optimization are exemplarily shown in Figure 5. Here, the aim was to maximize the expression of the soluble protein fraction of the heterologous protein HumaX simultaneously suppressing the undesired inclusion body fraction. In contrast to the situation of just evaluating single biomass and product values, the knowledge about the underlying dynamics enables the design of an optimal process considering the

dynamical changes. In this special case, the specific growth rate as well as the fermentation temperature after induction were manipulated to meet the optimization goal. It can immediately be seen that the proposed optimized process control led to an increase of the desired soluble protein fraction and additionally to a suppression of unwanted inclusion bodies.

Figure 5 reveals the general idea behind the development of process models. With the quantitative knowledge of the process dynamics it is possible to identify critical parameters affecting the process (Franco-Lara and Weuster-Botz 2005, Zheng *et al.* 2005, Kavanagh and Barton 2008). These parameters are available to be subsequently optimized in a knowledge based way and/or used for advanced feedback control to increase process performance.

Feedback Control

In real industrial environment disturbances may appear which cannot simply be compensated for by a robust process operation. In order to guarantee that even then the process will follow the desired paths, the advice in PAT is to use control in the engineering sense (Lee *et al.* 1999). Control strategies are intended to actively manipulate the actual process in such a way that it maintains a desired state (setpoint). The overall aim is to ensure a high batch-to-batch reproducibility of the process, which is tantamount to a consistent quality of output materials and the final product.

Generally, the major objective is to maximize the amount of the desired protein. This mass m_p can be represented as a time integral of the specific product formation rate π and the target protein expressing biomass x within a certain time interval of cultivation. The more cells are formed, the more protein m_p can be produced. Thus, manufactures are interested in high cell density cultivations (Lee 1996, Riesenberg and Guthke 1999). However, a high number of cells is not sufficient to get high amounts of product. Essentially, these cells have to perform well and this performance property is characterized by the specific product formation rate π . Both variables must, in theory, simultaneously reach values as large as possible to maximize the mass of target protein m_p . In this regard, the primary influence variable for both is the specific biomass growth rate μ . This correlation is trivial for x , but in most industrially relevant fermentations it was found that it is also valid for π (e.g., Neidhardt *et al.* 1990, Pirt 1994). The reason is that cell replication needs a well running and efficient protein formation machinery, which is obviously the case for foreign protein production as well.

In practice, operators have to keep in mind that there is a trade-off between the specific growth rate μ and the product formation performance of the cells. So it is not always advisable to maintain μ as high as possible in order to achieve maximal final target protein values. The optimal control strategy is dependent on the quantitative $\pi(\mu)$ -relationship, which is a function of several influencing factors, e.g., host system, type of heterologous protein, nutritional conditions. A typical example of such a $\pi(\mu)$ -relationship for a recombinant, soluble protein expressed in the cells' cytoplasm is presented in Figure 6. In this particular case, the curve depicts a maximum at about 0.12 h^{-1} . Hence, in order to keep the performance with respect to recombinant protein production as high as possible, the best solution is to keep cells' specific growth rate at the specified value.

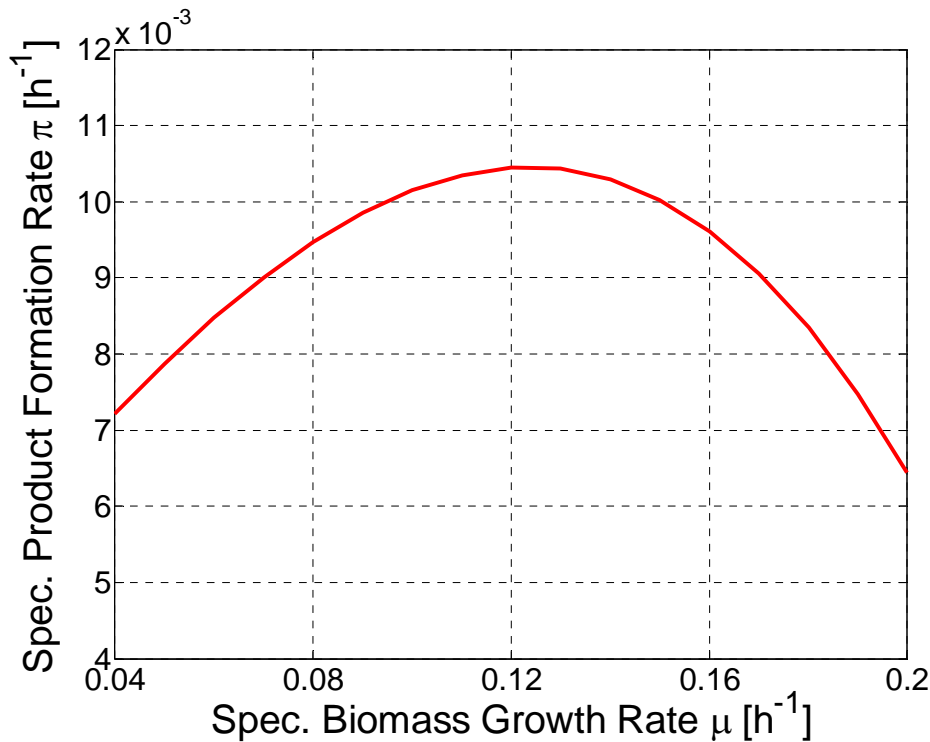


Figure 6. Specific product formation rate π for the HumaX production as a function of the specific biomass growth rate μ .

From the physiological point of view, μ is obviously the preferred control variable. Many bioprocess engineers considered μ control as the method of choice (e.g., Shioya *et al.* 1985, Lee *et al.* 1989, Paalme *et al.* 1990, Riesenber *et al.* 1991, Yoon *et al.* 1994, Levisauskas *et al.* 1996, Soons *et al.* 2006). In this context, “Generic Model Control” proved to be a simple model-based technique well suited for μ control (Preuß *et al.* 2000, DeLisa *et al.* 2001, Repsyte and Simutis 2006). However, specific growth rate control has a fundamental problem. As shown by Jenzsch *et al.* (2006c), variations in the biomass concentration, for instance, caused by different amounts of inoculum, inevitably lead to different trajectories of biomass during fermentation.

A further step toward a robust and precise control technique, which, by the way, avoids indirect state estimation, was done by controlling the total cumulative carbon dioxide production rate tcCPR. It turned out, that this variable was perfectly suited to match all requirements for bioprocess operation. The tcCPR mirrors the respiration of the cells, which in turn directly correlates with the actual total amount of biomass and thereby with the specific biomass growth rate, as well. Jenzsch *et al.* (2007) demonstrated the applicability of this approach for the production of recombinant proteins in *E. coli*. In the course of this work tcCPR control was improved with respect to the relative deviation from the setpoint through integration of the control algorithm into the novel process control system SIMATIC PCS 7 (Kuprijanov *et al.* 2008). It is shown, that the control technique is a ready-to-use application, mainly due to its simple integration into industrial biopharmaceutical production environment. Besides that, an outstanding advantage is the exclusive use of a global variable measured by common off-gas analyzers. This leads to a high degree of batch-to-batch reproducibility in biomass and target protein. The most important advantage from such tight control is taken in downstream processing. Because of the lower variability of the product concentrations, it can be run much easier with smaller losses of product and maybe with a lower number of processing steps.

Fault Detection and Process Supervision

Feedback control algorithms were shown to enable a compensation of randomly appearing process disturbances and therewith to increase process reproducibility. However, in feedback control, the controller assumes measurement devices, which provide correct information about the changes of the controlled variable. Now, imagine a situation where process analytics depicts systematic errors, e.g., due to sensor drifts, sampling and instrumentation faults, or even worse, completely fails. In those cases, feedback control is not capable to work properly. As the controller is not able to differentiate between normal process data and process errors, faulty measurement signals are misinterpreted by the controller causing completely wrong control actions. In the worst case, the actual fermentation deviates from its specification limits, if the abnormal process behavior is not detected as early as possible by the process operator. Moreover, industrial experience shows that manual supervision of process data and control actions by the process operator is time consuming and requires a certain degree of fermentation experience.

Consequently, there is a strong need for automated fault detection and supervisory systems. The intention of these algorithms is to check process data for consistency simultaneously as feedback control actions takes place. In doing so, supervisory and feedback control complement each other, since, compared to feedback control, the data is treated on a much higher integral level. Obviously, the main task is to detect failures and to disclose inconsistencies in the process behavior. In cases, where process errors occur, alarms should be sent automatically to the operator. In a further step, automated fail-safe actions should be activated concomitantly enforcing the engineering control to enter a different program, for instance, switching over to a safe open-loop control mode. In this way, process supervision is understood as a means of continuous process verification, which is in the sense of ICH's guideline Q8 (ICH 2005) an alternative approach to process validation where the manufacturing process performance is continuously monitored and evaluated.

An easy and straightforward way to check process data for consistency is to apply simple balances. In this context, mass, carbon and degree of reduction balances are of first choice. Elemental balances were often described (e.g., Chattaway and Stephanopoulos 1989, Nielsen and Villadsen 1994) but seldom used on-line during fermentation. These techniques are generally used in a first step for process supervision of measurement values, that are involved in mass and energy flow rates.

A logical step is to combine balancing techniques with the accumulated quantitative process understanding to yield advanced process models. By this means, it is possible to check the consistency of measurement data as well as to reliably predict central variables and kinetics signaling process performance. More sophisticated model-based methods involve state estimators, e.g., Kalman filters, fuzzy and neural data-driven techniques or hybrid models (Dochain and Bastin 1990, Claes and van Impe 1999, Grewal and Andrews 2001, James *et al.* 2002, Galvanauskas *et al.* 2004).

In recent years, an often proposed method is statistical process control. This is not a control in the engineering sense, but a method of multivariate statistical data analysis. It does not execute a particular action on the process, but aims at process monitoring. With statistical process control techniques the process is supervised primarily in order to detect deviations beyond common cause variation. It operates independently from feedback controllers and detects problems caused by disturbances that cannot be addressed by the controllers.

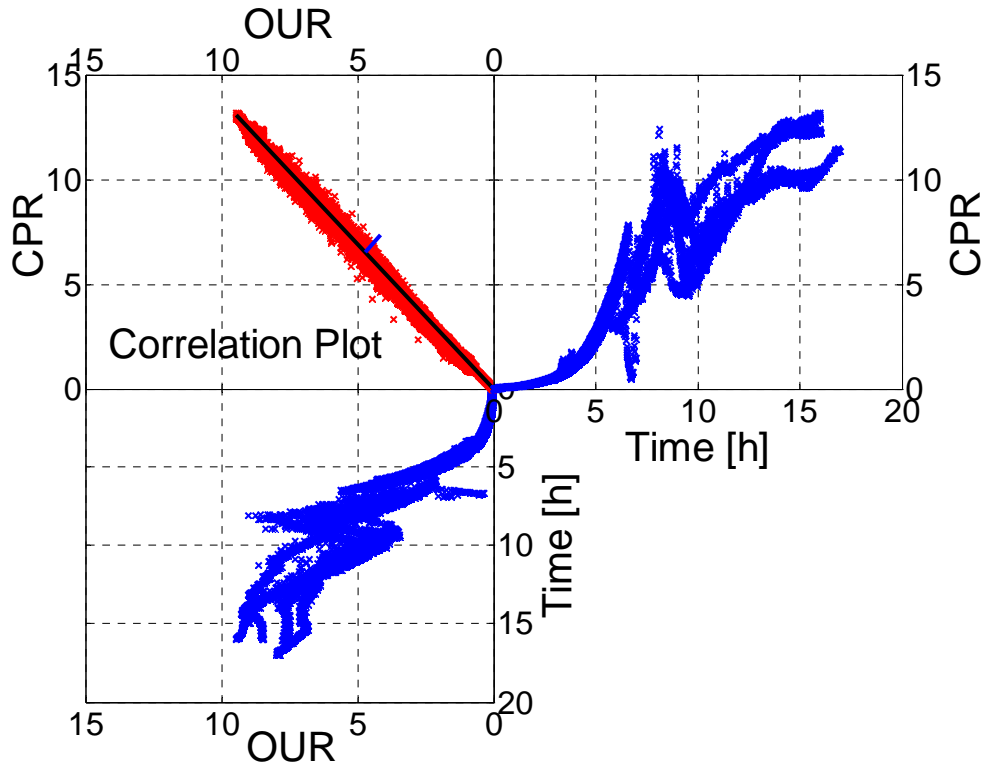


Figure 7. A simple to understand example of the principal component analysis principle with real process data from recombinant protein production processes. Here, the oxygen uptake rate OUR and the carbon dioxide production rate CPR measurement data of about 30 fermentation runs were plotted to visualize the correlation between these two variables. It can be simply described by a linear regression line, which is equivalent to the first principal component corresponding to the original data matrix.

In statistical process control, principal component analysis (PCA) is most often employed to get a quick overview and to detect deviations from the desired process behavior. PCA is a linear theory that tries to extract the static interrelationship between process variables. The idea behind PCA is to simply reduce the dimension of the space of related variables toward a lower-dimensional representation space. Herein, the remaining variables appear to be independent of each other, i.e., the redundancy in the measurements is removed. In Figure 7 a plausible example is shown, constructed from on-line measured data of about 30 experiments performed of varying conditions with *E. coli* BL21 DE3 cells. The so-called correlation plot depicts a high degree of correlation between oxygen uptake rate OUR and carbon dioxide production rate CPR so that, if no measurement errors occur, the data can be reduced to a one-dimensional representation, a straight line. Deviations from this line could be used to characterize the inconsistency of a particular set of OUR and CPR data.

PCA is not restricted to only two dimensions or to continuous processes operated at a single operating point. Extensions toward batch processes were proposed by several groups (Nomikos and MacGregor 1995, Wold *et al.* 1998). These techniques have also been applied to fermentation processes, where such multivariate statistical process control (MSPC) tools were useful to detect sensor faults and process deviations in seed, batch and fed-batch cultivations (Gregersen and Jørgensen 1999, Biccato *et al.* 2002, Cimander and Mandenius 2002, Glassey *et al.* 2000, Lennox *et al.* 2002).

As recombinant protein production processes are intrinsically nonlinear, the use of principal components, which is a linear theory, might be problematic (Shimizu *et al.* 1998). This restriction does not apply to nonlinear alternatives to PCA. The most prominent among them is

the autoassociative artificial neural network (aANN) technique developed by Kramer (Kramer 1991) for chemical engineering applications. The aANN operates for the most part analogous to PCA. The first step is to train a simple feedforward network, that maps n inputs onto m outputs ($m < n$) being practically independent of each other. The reduction of dimensionality of the input variables into the feature space is performed in the bottleneck layer. In order to make sure that this mapping describes the process behavior well enough, the reverse mapping of the bottleneck layer variables onto the input variables by a second feedforward network should be an overall result of the input layer onto itself. Training of the aANN is performed off-line on the basis of reference fermentations run under normal operating conditions. The result is a unique correlation structure of the process, which is expressed by a certain parameter set of the aANN weights. Subsequently, the trained network system is used on-line to supervise an unknown running process in a way that the measurement data is compared with the correlation structure of the reference model. Deviations from these normal operating trajectories indicate process faults caused by abnormal variations in the monitored variables. The aANNs have been used in case studies of fault diagnosis in fermentation processes (Glassey *et al.* 1994, Ignova *et al.* 1995), but also at (even industrial) fermentations (Simutis and Lübbert 1998, Hiden *et al.* 1999).

CONCLUSIONS

While being top in the development of new biologics, biopharmaceutical industry lags far behind other sectors of industry in terms of manufacturing technology. This led to the situation, that US Food and Drug Administration (FDA 2004) as well as the European Medicines Agency (EMA 2006) found themselves, in a reversal of their traditional role, pushing the industry they regulate to adopt more progressive approaches in quality and manufacturing systems (McKenzie *et al.* 2006). They encourage manufacturers to utilize “quality by design” techniques and to develop manufacturing processes on the basis of mechanistic understanding - measures that have been introduced in other industries long years ago. By using modern optimization, supervision and control methods, recombinant therapeutic protein production processes can be operated with a very good batch-to-batch reproducibility. This property can be exploited in order to run the processes closer to the design-space boundary, which in turn results in higher product yields. The benefit is possibly enhanced even more in the subsequently performed downstream processing steps, where the small variability in the raw product stream results in reliable and reproducible purification steps.

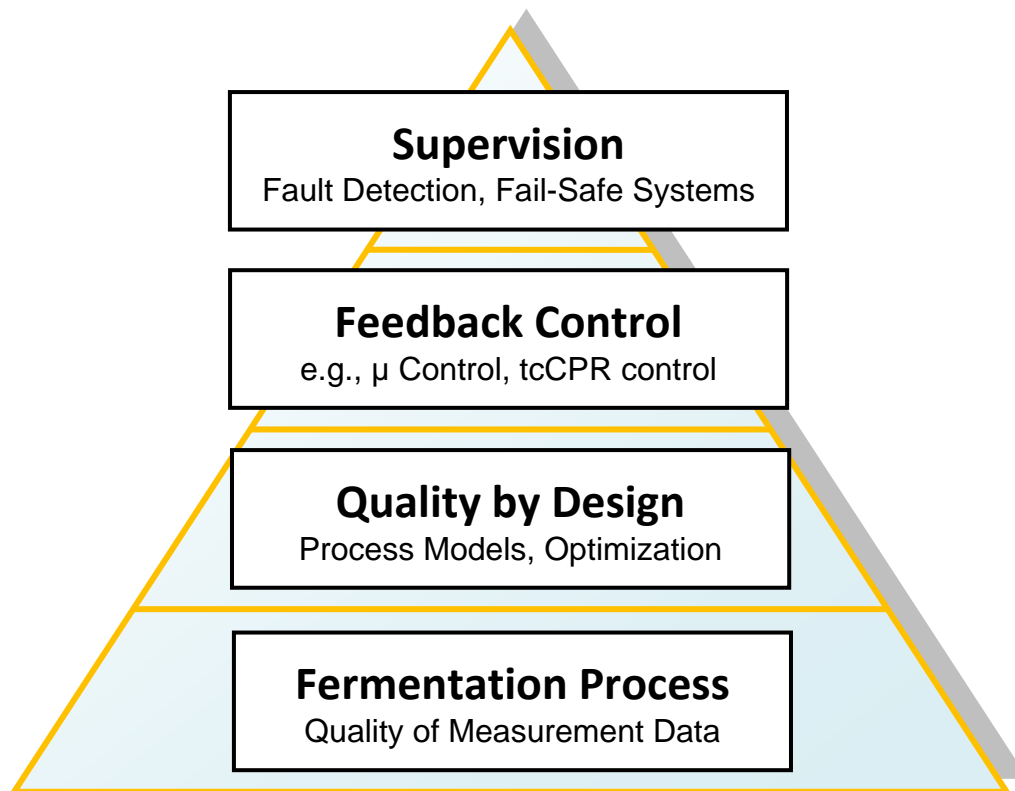


Figure 8. Hierarchical structure of the fundamental concept, which is intended to be practically implemented. First, special emphasis is laid on the improvement of the quality of measurement data, a point that is often underestimated or even completely skipped by process operators. Increasing the signal-to-noise ratio directly results in an improved process understanding, which basically assists in developing fundamental process interrelations into quantitative models. These models offer opportunity to optimize the process performance by manipulating central influence variables in a knowledge based way. In a next level, advanced feedback control techniques are implemented to combat process variations and to increase the batch-to-batch reproducibility. Finally, the intention of process supervision is the continuous check for consistency of measurement data, the on-line detection of critical process errors, and, in a consequent step, the activation of backup systems for automated fail-safe operation within the specification limits.

Quality is not the only aspect calling for the techniques described here. Extremely high and still increasing costs for the new drugs will no longer be tolerable by the authorities. That is why the development and manufacturing costs must be reduced in future. Hence, there is a strong focus on productivity and manufacturing efficiency as well as on rapid process development. As with quality assurance, it is mandatory to initiate the development of these topics on the basis of more comprehensive mechanistic process understanding and with manufacturing feasibility in mind. It should be emphasized, that the process operational strategy can only be as good as the knowledge about the process behavior and dynamics that can be made quantitatively available to process optimization studies. Therefore, quantitative models describing the process aspects relevant in keeping the process objectives under control, may they be mechanistic or data-driven ones, are not simply a play with computers, but tools indispensable for optimizing the benefit/cost ratio.

The broad PAT framework embedded in the latest initiatives of FDA and EMEA will serve to support the conceptual idea of this dissertation. Its practical application in fermentation processes for the production of recombinant therapeutic proteins will be discussed. The techniques are designed to achieve a high process performance and a robust process operation. In this context, process robustness in terms of a high degree of batch-to-batch reproducibility, is obtained by the consequent use of tools as suggested within PAT (Gnoth *et al.* 2007). Figure 8 summarizes the specific topics as well as the order to be treated in the present thesis. One of the major objectives of this thesis was the ease of implementation as well as the simple transferability of the methods proposed. These methods were tested at different recombinant proteins. The work provides insofar tools for the application in industrial as well as in the research process environment.

Basically, the work starts with the detailed evaluation of measurement data, both on-line and off-line available, and the subsequent identification of critical process parameters. These parameters, which dramatically affect the measurement noise, are explicitly analyzed and the resulting functional relationships are formulated in terms of quantitative utilizable models. Consequently, a model-based elimination of systematic measurement errors primarily results in a significant improvement of the overall measurement data quality. At the examples of the critical to control parameters pH and pO_2 the gain-scheduling approach demonstrates to be absolutely beneficial for the measurement quality of fermentation data. These results were presented in two publications (Kuprijanov *et al.* 2008, Gnoth *et al.* 2010).

Following the general conceptual idea, it will be shown, that the increased information content of measurement data is beneficial for the prediction quality of advanced process models as well as for advanced process control. First, process models are developed that quantitatively describe the observed process behavior. Especially the knowledge about the current product formation kinetics and its functional interrelationship with the specific growth rate is of primary importance. An approach that proved to be suitable for a quick and reliable identification of product formation as a function of different parameters was the hybrid artificial neural network. In this way, product formation kinetics of different types of target protein expression were identified. The use of this approach enabled the investigation and quantitative estimation of target protein formation as inactive insoluble inclusion bodies. The corresponding findings were published in Gnoth *et al.* 2007.

In another case study the hybrid modeling approach was further demonstrated to be a valuable tool for the identification of formation kinetics of a soluble expressed heterologous protein as well as its competing formation as insoluble inclusion bodies. The trained model herein was exploited for a constrained process optimization of central process variables, where the formation of the soluble protein fraction was favored while almost completely suppressing the insoluble part. As the batch-to-batch reproducibility of the optimized results was sensitively affected by even small disturbances, advanced feedback control was applied. This ensured

process variability to fall within tight specification limits as well as a high degree of batch-to-batch reproducibility. Gnoth *et al.* 2010a summarized the investigations regarding this topic.

Finally, as feedback control cannot handle process faults, which inevitably appear in fermentation processes, a system of superior fault detection and process supervision was implemented. It is capable of checking process data for consistency, identifies the responsible variables in cases when process faults occur and, if necessary, activates back-up systems that guarantee a secure process operation within the specification limits. A practical approach concerning the use of a fault diagnosis system and an automated fail-safe operation is presented in Gnoth *et al.* 2010b.

REFERENCES

- Aggarwal, S., 2008. What's fueling the biotech engine – 2007. *Nat. Biotechnol.* 26(11): 1227–1233.
- Åkesson, M., Karlsson, E.N., Hagander, P., Axelsson, J.P., Tocaj, A., 1999. On-line detection of acetate formation in *Escherichia coli* cultures using dissolved oxygen responses to feed transients. *Biotechnol. Bioeng.* 64: 590–598.
- Bentley, W.E., Mirjalili, N., Andersen, D.C., Davis, R.H., Kampala, D.S., 1990. Plasmid-encoded protein: the principal factor in the “metabolic burden” associated with recombinant bacteria. *Biotechnol Bioeng* 35: 668–681.
- Bhattacharya, S.K., Dubey, A.K., 1995. Metabolic burden as reflected by maintenance coefficient of recombinant *Escherichia coli* overexpressing target gene. *Biotechnol. Lett.* 17: 1155–1160.
- Bicciato, S., Bagno, A., Solda, M., Manfredini, R., di Bello, C., 2002. Fermentation Diagnosis by multivariate statistical analysis. *Appl. Biochem. Biotechnol.* 102/103: 49–62.
- Chattaway, T., Stephanopoulos, G.N., 1989. An adaptive state estimator for detecting contaminants in bioreactors. *Biotechnol. Bioeng.* 34: 647–659.
- Chirino, A.J., Mire-Slius, A., 2004. Characterizing biological products and assessing comparability following manufacturing changes. *Nat. Biotechnol.* 22: 1383–1391.
- Cimander, C., Mandenius, C.F., 2002. On-line monitoring of a bioprocess based on the use of a multi-analyser system with multivariate statistical process modelling. *J. Chem. Technol. Biotechnol.* 77: 1157–1168.
- Claes, J.E., van Impe, J.F., 1999. On-line estimation of the specific growth rate based on viable biomass measurements: experimental validation. *Bioproc. Biosyst. Eng.* 21: 389–395.
- DeLisa, M.P., Chae, H.J., Rao, G., Weigand, W.A., Valdes, J.J., Bentley, W.E., 2001. Generic model control of induced protein expression in high cell density cultivation of *Escherichia coli* using on-line GFP-fusion monitoring. *Bioproc. Biosyst. Eng.* 24: 83–91.
- DePalma, A., 2004. PAT: taking process monitoring to next level. *Gen. Eng. News* 24: 46–47.
- Dünnebier, G., Tups, H., 2007. FDA PAT Initiative – Eine Anwendersicht zu technischen Möglichkeiten und aktueller industrieller Umsetzung. *Chem. Ing. Tech.* 79: 2019–2028.
- Dochain, G., Bastin, D., 1990. On-line estimation and adaptive control of bioreactors, Elsevier Amsterdam.
- European Medicines Agency (EMA), 2005. Guideline on similar biological medicinal products. London.
- European Medicines Agency (EMA), 2006. Reflection paper: chemical, pharmaceutical and biological information to be included in dossiers when Process Analytical Technology (PAT) is employed. London.
- U.S. Food and Drug Administration (FDA), 2004. Guidance for Industry: PAT-a framework for innovative pharmaceutical Manufacturing & Quality Assurance. Rockville.
- Franco-Lara, E., Weuster-Botz, D., 2005. Estimation of optimal feeding strategies for fed-batch bioprocesses. *Bioproc. Biosyst. Eng.* 27: 255–262.
- Galvanauskas, V., Simutis, R., Lübbert, A., 2004. Hybrid process models for process optimisation, monitoring and control. *Bioproc. Biosyst. Eng.* 26: 393–400.
-

-
- Glasse, J., Ignova, M., Montague, G.A., Morris, A.J., 1994. Autoassociative neural networks in bioprocess condition monitoring. Proceedings of the IFAC symposium on advanced control of chemical processes: 447–451. *Elsevier*, Oxford.
- Glasse, J., Ignova, M., Montague, G.A., Morris, A.J., 2000. Issues in the development of an industrial bioprocess advisory system. *Trends Biotechnol.* 18: 136–141.
- Glick, B.R., 1995. Metabolic load and heterologous gene expression. *Biotechnol Adv.* 13: 247–261.
- Gnoth, S., Jenzsch, M., Simutis, R., Lübbert, A., 2007. Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control. *J. Biotechnol.* 132: 180–186.
- Gnoth, S., Jenzsch, M., Simutis, R., Lübbert, A., 2008. Product formation kinetics in genetically modified *E.coli* Bacteria: inclusion body formation. *Bioproc. Biosyst. Eng.* 31: 41–46.
- Gnoth, S., Kuprijanov, A., Simutis, R., Lübbert, A., 2010. Simple adaptive pH control in bioreactors using gain-scheduling methods. *Appl. Microbiol. Biotechnol.* 85: 955–964.
- Gnoth, S., Simutis, R., Lübbert, A., 2010a. Selective expression of the soluble product fraction in *Escherichia coli* cultures employed in recombinant protein production processes. *Appl. Microbiol. Biotechnol.* 87: 2047–2058.
- Gnoth, S., Simutis, R., Lübbert, A., 2010b. Fermentation process supervision and strategies for fail-safe operation: a practical approach. *Eng. Life Sci.*, accepted for publication.
- Gregersen, L., Jørgensen, S.B., 1999. Supervision of fed-batch fermentations. *Chem. Eng. J.* 75: 69–76.
- Grewal, M.S., Andrews, A.P., 2001. Kalman Filtering Theory and Practice using MATLAB, 2nd ed., *Wiley*, New York.
- Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, 2nd ed., *Prentice Hall*, Upper Saddle River, New Jersey.
- Hiden, H.G., Willis, M.J., Tham, M.T., Montague, G.A., 1999. Nonlinear principal components analysis using genetic programming. *Comp. Chem. Eng.* 23:413–425.
- Hinz, D.C., 2006. Process analytical technologies in the pharmaceutical industry: the FDA's PAT initiative. *Anal. Bioanal. Chem.* 384: 1036–1042.
- ICH, 2000-2008. Q7 - good manufacturing practice guide for active pharmaceutical ingredients, Q8 - pharmaceutical development, Q9 - quality risk management, Q10 - pharmaceutical quality system, <http://www.ich.org>.
- Ignova, M., Glasse, J., Ward, A.C., Montague, G.A., Irvine, T.S., 1995. Seed data analysis for production fermentor performance estimation. Proceedings of the 6th international conference on computer application in biotechnology: 53–58. *Elsevier*, Oxford.
- James, S., Legge, R., Budman, H., 2002. Comparative study of black-box and hybrid estimation methods in fed-batch fermentation. *J. Process. Contr.* 12: 113–121.
- Jensen, E.B., 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., 2006a. Advanced monitoring and control in microbial cultivation processes for recombinant protein production. *PhD thesis*, University of Halle-Wittenberg, Germany.
-

- Jenzsch, M., Simutis, R., Lübbert, A., 2006b. Generic model control of the specific growth rate in *E. coli* cultivations. *J. Biotechnol.* 122: 483–493.
- Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., Lübbert, A., 2006c. Improving the batch-to-batch reproducibility in microbial cultures during recombinant protein production by guiding the process along a predefined total biomass profile. *Bioproc. Biosyst. Eng.* 29: 315–321.
- 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.
- Kavanagh, J.M., Barton, G.W., 2008. Productivity improvement of recombinant *Escherichia coli* fermentation via robust optimization. *Bioproc. Biosyst. Eng.* 31: 137–143.
- Kramer, M.A., 1991. Nonlinear principal component analysis using autoassociative neural networks. *AIChE. J.* 37: 233–243.
- 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.
- Lawrence, S., 2007. Billion dollar babies – biotech drugs as blockbusters. *Nat. Biotechnol.* 25: 380–382.
- Lee, J., Youn-Hee, C., Shin-Kwon, K., Hyung-Hwan, P., Ik-Boo, K., 1989. Production of human leukocyte interferon in *Escherichia coli* by control of growth rate in fed-batch fermentation. *Biotechnol. Lett.* 11: 695–698.
- Lee, S.Y., 1996. High cell-density culture of *Escherichia coli*. *Trends Biotechnol.* 14: 98–105.
- Lee, J., Lee, S.Y., Park, S., Middelberg, A.P.J., 1999. Control of fed-batch fermentations. *Biotechnol. Adv.* 17: 29–48.
- Lennox, B., Kipling, K., Glassey, J., Montague, G., Willis, M., Hiden, H., 2002. Automated production support for the bioprocess industry. *Biotechnol. Prog.* 18: 269–275.
- Levisauskas, D., Simutis, R., Borvitz, D., Lübbert, A., 1996. Automatic control of the specific growth rate in fed-batch cultivations processes based on exhaust gas analysis. *Bioproc. Eng.* 15: 145–150.
- Levisauskas, D., Galvanauskas, V., Henrich, S., Wilhelm, K., Volk, N., Lübbert, A., 2003. Model-based optimization of viral capsid protein production in fed-batch culture of recombinant *Escherichia coli*. *Bioproc. Biosyst. Eng.* 25: 255–262.
- Luli, G.W., Strohl, W.R., 1990. Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Appl. Environ. Microb.* 56: 1004–1011.
- McKenzie, P., Kiang, S., Tom, J., Rubin, A.E., Futran, M., 2006. Can pharmaceutical process development become high tech?, *AIChE J.* 52:3990–3994.
- Neidhardt, F.C., Ingraham, J.L., Schaechter, M., 1990. Physiology of the Bacterial Cell, a Molecular Approach. *Sinauer Associates Inc.*, Sunderland, Massachusetts.

Neubauer, P., Lin, H.Y., Mathiszik, B., 2003. Metabolic load of recombinant protein production: inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in *Escherichia coli*. *Biotechnol. Bioeng.* 83: 53–64.

Nielsen, J., Villadsen, J., 1994. Bioreaction engineering principles. *Plenum press*, New York

Nomikos, P., MacGregor, J.F., 1995. Multivariate SPC charts for monitoring batch processes. *Technometrics* 37: 41–59.

Paalme, T., Tiisma, K., Kahru, A., Vanatalu, K., Vilu, R., 1990. Glucose-limited fed-batch cultivation of *Escherichia coli* with computer-controlled fixed growth rate. *Biotechnol. Bioeng.* 35: 312–319.

Pirt, S.J., 1994. Product Formation in Cultures of Microbes. *Pirtferm Papers*, Pirtferm Ltd., London.

Preuß, K., Le Lann, M., Proth, J., Pingaud, H., 2000. Modelling and predictive control of fed-batch yeast growth on industrial pilot plant scale. *Chem. Eng. J.* 78: 53–59.

Rathore, A.S., 2009a. Follow-on protein products: scientific issues, developments and challenges. *Trends Biotechnol.* 27: 698–705.

Rathore, A.S., 2009b. Roadmap for implementation of quality by design (QbD) for biotechnology products. *Trends Biotechnol.* 27: 546–553.

Repsyte, J., Simutis, R., 2006. Application of Generic Model Control for autotrophic biomass specific growth rate control. Computer Aided Methods in Optimal Design and Operations: 217–266, *Workshop on Optimal Process Design*, Vilnius, Lithuania.

Riesenberg, D., Schulz, V., Knorre, W.A., Pohl, H.D., Korz, D., Sanders, E.A., Ross, A., Deckwer, W.D., 1991. High cell density cultivation of *Escherichia coli* at controlled specific growth rate. *J. Biotechnol.* 20: 17–28.

Riesenberg, D., Guthke, R., 1999. High-cell-density cultivation of microorganisms. *Appl. Microbiol. Biotechnol.* 51: 422–430.

Roe, A.J., O’Byrne, C., McLaggan, D., Booth, I.R., 2002. Inhibition of *Escherichia coli* growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. *Microbiology* 148: 2215–2222.

Royce, P.N., 1992. Effect of changes in the pH and carbon dioxide evolution rate on the measured respiratory quotient of fermentations. *Biotechnol. Bioeng.* 40: 1129–1138.

Rozkov, A., Avignone-Rossa, C.A., Ertl, P.F., Jones, P., O’Kennedy, R.D., Smith, J.J., Dale, J.W., Bushell, M.E., 2004. Characterization of the metabolic burden on *Escherichia coli* DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. *Biotechnol. Bioeng.* 88: 909–915.

Rumelhard, D.E., McClelland, J.L., 1986. Parallel Distributed Processing, vol. 1. *MIT Press*, Cambridge, MA.

Shimizu, H., Yasuoka, K., Uchiyama, K., Shioya, S., 1998. Bioprocess fault detection by non-linear multivariate analysis: application of an artificial autoassociative neural network and wavelet filter bank. *Biotechnol. Progr.* 14:79–87.

Shioya, S., Shimizu, H., Ogata, M., Takamatsu, T., 1985. Simulation and experimental studies of the profile control of the specific growth rate in a fed-batch culture. IFAC Modelling and Control of Biotechnological Processes, vol. 10. *Pergamon Press*, Oxford, Noordwijkerhout, 79–84.

Simutis, R., Lübbert, A., 1998. Advances in modeling for bioprocess supervision and control. *Bioseparation and Bioprocessing: Biochromatography, Membrane Separations, Modeling, Validation*: 411–461, *Wiley-VCH*, New York.

Soons, Z.I.T.A., Voogt, J.A., van Straten, G., van Boxtel, A.J.B., 2006. Constant specific growth rate in fed-batch cultivation of *Bordetella pertussis* using adaptive control. *J. Biotechnol.* 125: 252–268.

Spérandio, M., Paul, E., 1997. Determination of carbon dioxide evolution rate using on-line gas analysis during dynamic biodegradation experiments. *Biotechnol. Bioeng.* 53: 243–252.

Walsh, G., 2006. Biopharmaceutical benchmarks 2006. *Nat. Biotechnol.* 24: 769–776.

Wold, S., Kettaneh, N., Fridén, H., Holmberg, A., 1998. Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemom. Intell. Lab. Syst.* 44: 331–340.

Yoon, S.K., Kang, W.K., Park, T.H., 1994. Fed-batch operation of recombinant *Escherichia coli* containing Trp promoter with controlled specific growth rate. *Biotechnol. Bioeng.* 43: 995–999.

Zheng, Z.Y., Yao, S.J., Lin, D.Q., 2005. Using a kinetic model that considers cell segregation to optimize hEGF expression in fed-batch cultures of recombinant *Escherichia coli*. *Bioproc. Biosyst. Eng.* 27: 143–152.

CHAPTER 2

Simple Adaptive pH Control in Bioreactors Using Gain-Scheduling Methods

ABSTRACT

A simple well-performing adaptive control technique for pH control in fermentations of recombinant protein production processes is described and its design procedure is explained. First, the entire control algorithm was simulated and parameterized. Afterwards it was tested in real cultivation processes. The results show that this simple technique leads to significant reductions in the fluctuations of the pH values in microbial cultures at a minimum of expenditures. The signal-to-noise ratio and thus the information captured by the pH signal were increased by about an order of magnitude. This leads to a substantial improvement in the noise of many other process signals that are used to monitor and control the process. For instance, respiratory off-gas data of CO₂ and its derived carbon dioxide production rate signals from the cultures carry much less noise as compared to those values obtained with conventional pH control. Detailed process analysis revealed that even very small pH jumps of 0.03 values during the fermentation were shown to result in pronounced deflections in CO₂-volume fraction of 8 % (peak to peak). The proposed controller, maintaining the pH within the interval of 0.01 around the setpoint, reduces the noise considerably.

This paper has been published in *Applied Microbiology and Biotechnology*:

Gnoth, S., Kuprijanov, A., Simutis, R., Lübbert, A., 2010. Simple adaptive pH control in bioreactors using gain-scheduling methods. *Appl. Microbiol. Biotechnol.* 85: 955–964.

INTRODUCTION

Temperature T and pH belong to the first choice of variables that are to be controlled in industrial fermentation processes. This is due to the fact that the specific growth (μ) and product formation (π) rates sensitively depend on T and pH (Nielsen and Villadsen 1994, Johnston *et al.* 2002). As the $\mu(\text{pH})$ curve often depicts a broader and flatter maximum as compared with the $\mu(T)$ relationship, pH control was not given much attention. Hence, in practice, simple conventional proportional integral derivative (PID) controllers are installed at most bioreactors.

In the context of the current process analytical technology (PAT) initiative, process reproducibility is a lively discussed topic. The batch-to-batch reproducibility can significantly be improved with advanced control techniques that tightly keep the process' key variables, such as biomass x or the specific growth rate μ , on predefined trajectories. Improved controllers need on-line measurement data reflecting the metabolic activity of the cells. First choice are measurements of the volume fractions of O_2 and CO_2 in the fermenters' vent lines and the oxygen uptake as well as the carbon dioxide production rates (CPR), which are determined from them. Examples of such advanced control techniques are RQ control (Zeng *et al.* 1994, Franzen *et al.* 1996), the control of the specific growth rate μ (Shioya 1992, Levisauskas *et al.* 1996, Soons *et al.* 2006, Wang *et al.* 2006), the total biomass control (Jenzsch *et al.* 2006a), or the total cumulative carbon dioxide production rate control (Jenzsch *et al.* 2007).

If the on-line signals from the off-gas analyzers are used to improve the batch-to-batch reproducibility by means of advanced control, the success depends on the noise in the measurement signals. Carbon dioxide production rates derived from exhaust gas analyzer data are only accurate if pseudo steady-state conditions can be assumed in the liquid media. For instationary systems, e.g., in cases of fluctuating pH values, the true CPR may be different to that estimated from the measured CO_2 volume fractions. (Royce 1992, Spérandio and Paul 1997). The reason is that the reaction of dissolved CO_2 with water and further the dissociation to bicarbonate sensitively depend on the pH value, leading to additional CO_2 accumulation or dissolution effects in the liquid. Hence, it is straightforward to avoid pH fluctuations as far as possible and thus to look for improved controllers in order to increase precision by which pH is controlled. This will, consequently, lead to an increase in the signal-to-noise ratio in the key on-line signals of fermentations such as the carbon dioxide production rate.

Here we present a simple adaptive pH controller that significantly reduces the fluctuations in the pH and the corresponding base consumption signal during the entire fermentation process by dynamically adapting the controller parameters to changes in the process' dynamics. For that purpose, we use the gain-scheduling technique. We start the discussion with a sensitivity analysis and a simulation of the controller dynamics. Subsequently, the controller's parameters are determined followed by the implementation of the controller in microbial cultivation processes producing recombinant proteins.

MATERIALS AND METHODS

Experimental Setup

All cultivations were performed in a B. Braun 10-Lbioreactor Biostat-C, entirely operated in fed-batch mode, i.e., no initial glucose was present and substrate was added directly after inoculation. The experiments were fully documented in a central database and are identified by their consecutive 'study numbers', e.g., S486, which appear in some of the plots in the following text. The feeding profile was designed according to Jenzsch *et al.* (2006b). This assures substrate limited conditions direct after inoculation and a high degree of batch-to-batch reproducibility.

The feeding solution consisted of glucose with a concentration of 400 g L^{-1} . It contained the same composition with respect to mineral salts as the initial cultivation medium, i.e., Na_2SO_4 2 g L^{-1} , $(\text{NH}_4)_2\text{SO}_4$ 2.46 g L^{-1} , NH_4Cl 0.5 g L^{-1} , K_2HPO_4 14.6 g L^{-1} , $\text{NaH}_2\text{PO}_4 \times \text{H}_2\text{O}$ 3.6 g L^{-1} , ammonium-hydrogencitrate 1.0 g L^{-1} , $\text{MgSO}_4 \times 7 \text{ H}_2\text{O}$ 1.2 g L^{-1} , trace element solution 2 ml L^{-1} .

Two BL21(DE3)-strains with antibiotic resistance against kanamycin/ampicillin were used in the cultivation studies. In all cases the expression of the recombinant protein took place under the control of the T7/lac promoter upon induction with isopropylthiogalactopyranoside. The first host-system *Escherichia coli* BL21, harboring the plasmid pET11a GFP, releases the green fluorescent protein (GFP) in its active form into its cytoplasm, whereas the second one expresses a commercially important recombinant protein (here referred to as HumaX) in its soluble form and as inclusion bodies.

Off-gas analysis was performed on-line with a paramagnetic sensor (Maihak Oxor 610) for O_2 and an infrared detector (Maihak Unor 610) for CO_2 . The dissolved oxygen concentration was monitored with an Ingold pO_2 probe (Mettler Toledo InPro 6800) and maintained around 25 %sat by increasing the airflow rate and subsequently the stirrer speed. Additionally, a fluorescence pO_2 probe (Presens) was applied as a redundant measurement device. pH was measured with another standard Ingold-probe (Mettler Toledo 405-DPAS-SC-K85-120) connected to a 16-bit transmitter (Knick) and controlled by appropriately adding NH_4OH solution (25 % w/w) to the culture. Note that no special pH probe is needed. In order to quantify precision and random error part of the pH measurement, a long-term experiment (15 h) under defined conditions was conducted where measurement values were recorded every 9 s. For that purpose pH-buffer 7 was used at a controlled temperature of $25 \text{ }^\circ\text{C}$. Signal values with a difference of pH 0.0004 could be distinguished and the nominal standard deviation of the measurement noise was measured to be $8\text{e-}4$, saying that the predominant part of the later observed pH variations in fermentations accounts for systematic errors. The base addition was measured on-line by means of a balance. Substrate feeding was conducted with a gravimetric feed rate controller (Sartorius) connected with a balance (Sartorius).

Off-line measurements were performed with time increments of half an hour: final biomass concentrations around 60 g (DW)/L were usually obtained in the cultures reported about here. The values were estimated from OD_{600} measurements performed with a spectral photometer (Shimadzu UV-2102PC). Additionally, some dry weight measurements were performed in order to validate the available correlation between biomass dry weight and the OD_{600} values. Glucose concentration was quantitatively determined by YSI-Glucose-Analyzer (Yellow Springs).

Gain-Scheduling Controller

In cases where pH fluctuations have to be kept very small, the conventional PID controllers do not work properly as the process dynamics in batch and fed-batch procedures, most often applied in industry, cannot be considered constant. With changes in the dynamics, the optimum controller parameters also change so that the controller performance may become insufficient. A straightforward measure to cope with this problem has been automatic adaptation of the controller parameters to the changes in the dynamic behavior of the culture. In literature, model-aided adaptation procedures are proposed for this purpose. In industry, however, these are considered difficult and not simple to use. Hence, there is a need for simple transparent techniques of controller parameter adaptation. Gain-scheduling is an advanced PID control technique well suited for nonlinear processes where the process dynamics are varying with the operating conditions (Cardello and San 1988, Aström and Hägglund 2006) such as the dynamics of biotechnical production processes that change drastically during the cultivations (Franzen *et al.* 1996, Shuler and Kargi 2001).

Gain-scheduling control can follow rapid changes in the process dynamics, if values or estimates of quantities signaling the corresponding changes in dynamics, can be measured with a sufficiently small delay time. Possible applications of the gain-scheduling technique for control of biotechnological processes have already been discussed in literature (Levisauskas 1995, 2001), but to our knowledge, the technique has not yet been practically used in bioprocess engineering, because the most important variable that reflects changes in bioprocess dynamics - the absolute and the specific biomass growth rate - normally cannot be measured on-line. However, in cases where the biomass growth rate can be indirectly estimated from the carbon dioxide production rate or other biomass growth rate estimators, gain-scheduling algorithms can be easily implemented and used for control of biotechnological processes. A block diagram of the presented control system working with gain-scheduling is shown in Figure 1. The system has an inner loop composed of the process and the controller, and an outer loop which adjusts the controller parameters upon changes in the process dynamics.

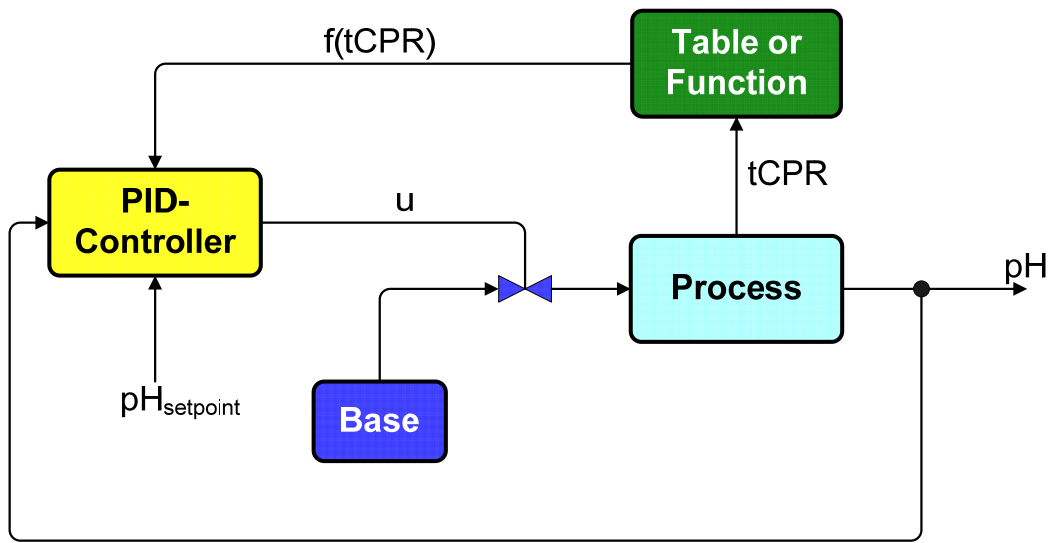


Figure 1. Block diagram of a PID control system for adaptive pH control where gain-scheduling is used to adapt the controller to changes in the process dynamics.

RESULTS

pH Dependency of the Off-gas Volume Fraction y_{CO_2}

Microorganisms transfer CO_2 , a main metabolic product, into their environment in dissolved form. Their activity is characterized by the carbon dioxide production rate CPR. In aerated bioreactors, main part of the dissolved CO_2 is transferred into the dispersed gas phase. This transfer rate is usually described by the corresponding volumetric rate, the carbon dioxide transfer rate CTR. With the gas throughput, the carbon dioxide finally leaves the reactor through the reactor's vent line, where the CO_2 volume fraction y_{CO_2} [L_{CO_2} / L_{Gas}] is measured by means of infrared sensors or mass spectrometers.

The CO_2 concentration CO_2^{out} in the gas leaving the culture broth is primarily determined by the CTR. Its dynamics can be described by

$$\frac{dCO_2^{out}}{dt} = \frac{1}{V_L \cdot \varepsilon} \cdot \left(\frac{V_L \cdot CTR}{M_{CO_2}} + Q^{in} \cdot CO_2^{in} - Q^{out} \cdot CO_2^{out} \right) \quad (1)$$

where ε is the gas holdup, M_{CO_2} the molecular weight of CO_2 , Q^{in} and Q^{out} are the volumetric gas flow rates in the inlet and outlet gas and CO_2^{in} the concentration of CO_2 in the gas supply.

The response in the concentration CO_2^{Sensor} reaching the exhaust gas sensor is delayed in the gaseous headspace volume and the analyzer. Assuming a time constant T_N we can describe this time delay by

$$\frac{dCO_2^{Sensor}}{dt} = \frac{1}{T_N} \cdot (CO_2^{out} - CO_2^{Sensor}) \quad (2)$$

The corresponding volume fraction recorded by the off-gas analyzer is

$$y_{CO_2}^{Sensor} = CO_2^{Sensor} \cdot \frac{1000 \cdot R \cdot T}{p^{out}} \quad (3)$$

Compared to oxygen, CO_2 has a much higher solubility in water and, additionally, dissolved CO_2^L reacts with water to form bicarbonate. The conversion of dissolved carbon dioxide into bicarbonate leads to a dynamic change in the dissolved CO_2 concentration. This is considered by the CO_2 balance:

$$\frac{dCO_2^L}{dt} = \frac{1}{M_{CO_2}} \cdot (CPR - CTR) - R_{CO_2} \quad (4)$$

Royce (1992) showed that in real cultivations the rate can simply be described by

$$R_{CO_2} = k_1 \cdot CO_2^L - k_2 \cdot \frac{10^{-pH} \cdot HCO_3^-}{K_a} \quad (5)$$

depicting the pH-dependency and the influence of the bicarbonate concentration HCO_3^- on the rate by which CO_2 is removed from the liquid phase by chemical reaction. k_1 and k_2 are reaction constants whereas K_a represents the carbonic acid dissociation constant.

$$CTR = M_{CO_2} \cdot k_L^{CO_2} \cdot a \cdot (CO_2^L - CO_2^*) \quad (6)$$

Upon changes in the pH of the culture, significant changes in CTR appear even if CPR remains constant. Decreasing pH values lead to temporary increases in CTR and decreasing CTRs appear upon jumps to higher pH values.

Figure 2 shows an experimental validation of this model with data from the *E. coli* cultivation system used in this work. The experiment was performed during the product formation phase of the *E. coli* cultivation. During the measurements, the airflow rate was kept constant at 20 sLpm and the feed rate remained at 100 g h^{-1} . The mean carbon dioxide volume fraction in the vent line could therefore be considered roughly constant in that time interval at about $y_{CO_2} = 0.02 \text{ L}_{CO_2} / \text{L}_{Gas}$. Changes in the pH were induced by step changes in the base addition through the NH_4OH -solution supply line used for pH control.

A positive pH shift leads to an immediate drop of y_{CO_2} from which the culture recovers with a time constant of a few minutes. As can be seen from the figure, even very small pH changes of 0.05 units lead to heavy effects in CTR. Moreover, pH changes in the order of 0.03 lead to measurable responses in the CO_2 volume fraction in the vent line.

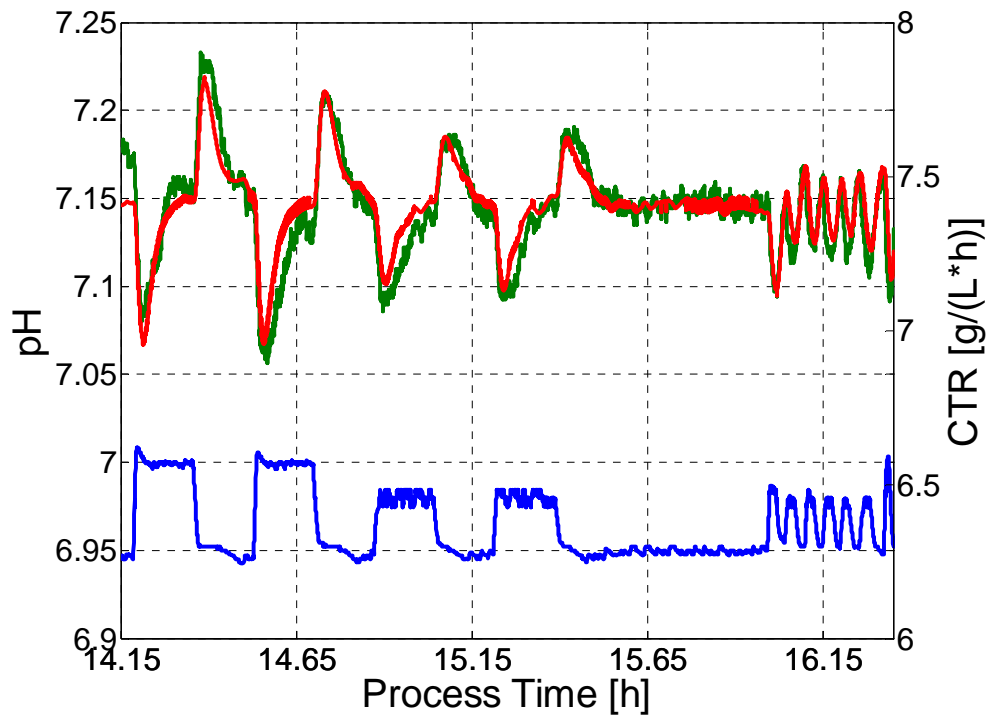


Figure 2. Systematic step changes of pH (step size 0.05 and 0.03) induced during the *E. coli* culture and corresponding answers in CTR as measured by the off-gas analyzer. The lower curve shows the pH profile (scale on the left axis); the upper two curves show the measured CTR (scale on the right axis) together with the profile computed with the model equations (Eqs. 1-6).

As the pH variations in most real cultivation experiments are larger than 0.05, much of the CO_2 fluctuations in the off-gas of real cultures can be traced back to pH fluctuations. Consequently, if the off-gas CO_2 values are to be used in advanced controllers for RQ-, biomass-, or tcCPR control, the pH value in the culture must be controlled very precisely in order maintain stationary conditions (see Eqs. 4 and 5), to avoid strong fluctuations and thus information losses from the off-gas measurement signals.

RESULTS

Modeling the pH Dynamics

In most fermenters, pH control is performed with separately operating PID controllers, which respond to deviations in the pH values, measured with conventional electrochemical electrodes, from the desired set point profiles by pumping base or acid into the culture medium. The PID controller parameters are most often kept constant and are usually chosen in such a way that the controller can be used across the entire cultivation.

A closer look at the performance of the pH control with a standard PID controller reveals that the change in the process dynamics of the culture is reflected by the changing fluctuations in the controlled pH value. Figure 3 shows a typical example of a culture in a laboratory-scale reactor, where pH is controlled with B. Braun's DCU, using standard controller settings. The important point is that the control performance is not constant across the entire cultivation and the envelope of the deviations in the controlled pH signal resembles the variations in the feeding profile and the total carbon dioxide production rate (tCPR). The tCPR signal is known to reflect the metabolic activity of the cells and is dependent on both, the total biomass and the specific growth rate.

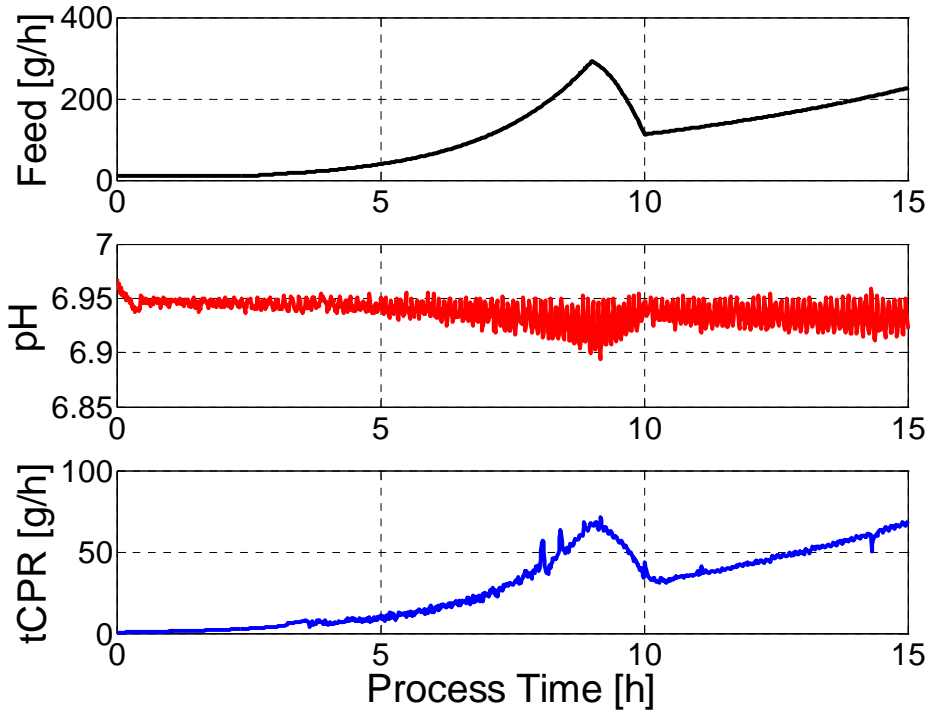


Figure 3. Illustration of pH control quality in a 10 L B. Braun Biostat reactor by means of measured trajectories of the substrate feed rate F , the controlled pH and the total carbon dioxide production rate (tCPR). The measurements were performed during a GFP production process.

This observation can serve as the basis of a simple model. We used a Luedeking–Piret type of equation describing the carbon dioxide production rate as a function of the biomass formation. Accordingly, we also need an equation for biomass production. In our accompanying example of a recombinant protein production process, biomass growth is divided into two phases operated at different specific growth rates, i.e., the biomass and the product formation phase. The following simple equations are sufficiently accurate for controller design purposes:

$$\frac{dx}{dt} = \mu \cdot x \quad \mu(t) = \begin{cases} 0.5 \text{ h}^{-1} & t \leq 9 \text{ h} \\ 0.5 \text{ h}^{-1} - 0.36 \text{ h}^{-2} \cdot (t - 9 \text{ h}) & 9 \text{ h} < t < 10 \text{ h} \\ 0.14 \text{ h}^{-1} & t \geq 10 \text{ h} \end{cases} \quad (7)$$

$$tCPR = Y_{cx} \cdot \mu \cdot x + m_c \cdot x \quad (8)$$

Where x [g] is the total biomass in the reactor, μ [h^{-1}] the specific growth rate, tCPR [g/h] the total carbon dioxide production rate in the reactor and Y_{cx} and m_c are parameters of the Luedeking–Piret equation. During the biomass growth phase, cells were fed with an exponential substrate feed rate profile that adjusts growth to a relatively high specific growth rate $\mu=0.50 \text{ h}^{-1}$. In the product formation phase, the specific growth rate was kept at its optimal value for product formation, which was $\mu=0.14 \text{ h}^{-1}$ in this particular case. Between both process phases, i.e., in the time interval ($9 \text{ h} \leq t \leq 10 \text{ h}$), μ was ramped down linearly. For the Luedeking–Piret type expression in Eq. 8 the parameter values were determined to be $Y_{cx}=0.935 \text{ g}_{\text{CO}_2} / \text{g}_x$ and $m_c=0.065 \text{ g}_{\text{CO}_2} / \text{g}_x/\text{h}$.

The change of the pH value in the culture is a function of the metabolic conversion rate. It is dependent on the carbon dioxide production rate CPR [g/kg/h] and the rate of NH_4OH

addition [mL/h] to the culture. As already mentioned, part of CO_2 being produced during the cultivation process appears in the dissolved form and reacts with water primarily to form bicarbonate. Ammonium hydroxide is taken up by the cells as nitrogen source. To a practical approach, CO_2 production and ammonium consumption can be considered correlated (Siano 1995). This relationship is changing during the cultivation process with the process dynamics. Thus, the pH change was assumed to be a nonlinear function of the key variables determining the dynamics of the process, for instance CPR (CPR and tCPR can be converted vice versa via the culture weight) and the specific gravimetric rate of base addition R_{Base}/W . Here, an artificial neural network (ANN) is used to quantitatively describe pH changes in the reactor upon changes in the specific rate of base addition R_{Base}/W , CPR and the overall amount of CO_2 produced (cCPR). With this rate expression the dynamics of the pH signal is described by:

$$\frac{dpH}{dt} = \text{ANN}\left(\frac{R_{\text{Base}}}{W}, \text{CPR}, \text{cCPR}, \mathbf{P}\right) \quad (9)$$

where \mathbf{P} represents the set of parameters (weights) of the artificial neural network. Numerical experiments showed that a simple feedforward ANN with a single hidden layer containing three nodes responding with hyperbolic tangent functions is sufficient for this purpose. For network training, the experimental data of pH, R_{Base}/W , CPR, cCPR were taken from several cultivation runs. The training was performed using the cross-validation technique with MATLAB's optimization subroutine *lsqcurvefit* (Mathworks). Typical simulation results together with the experimental observation are presented in Figure 4.

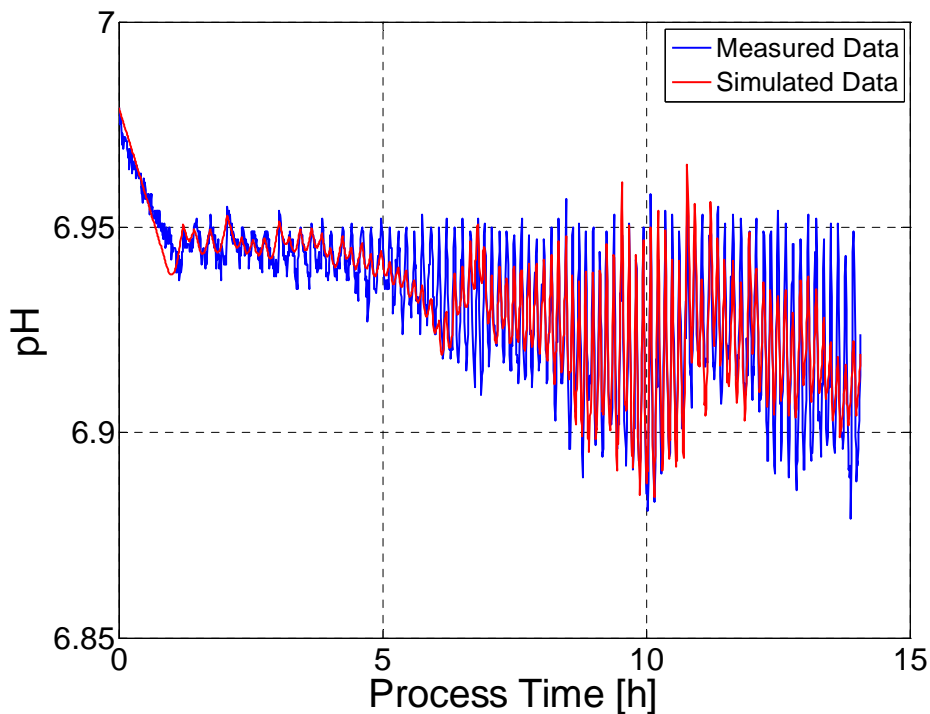


Figure 4. Comparison of modeled (*red*) and measured (*blue*) pH signals.

For controller design it must be taken into account that the experimentally determined pH values (measured values) appear with some time delay τ with respect to the base addition. For that purpose, experiments were made with 5 kg mineral salt medium and a base addition rate of $R_{\text{Base}}=0.17 \text{ mL}_{\text{Base}}/\text{s}$ in order to identify the delay time τ . The estimate of time delay was $\tau=5 \text{ s}$. The pH sensor signal was computed as a simple delay function in the following way:

$$pH_{\text{Sensor}}(t) = pH(t - \tau) \quad (10)$$

Simulation of PID Controller

For pH control we used the positional form of digital PID controllers. In this form the PID algorithm is expressed by:

$$u(t) = K_c \left(e(t) + \frac{T_c}{T_i} \cdot \sum_{t=0}^t e(t) + \frac{T_d}{T_c} \cdot (e(t) - e(t-1)) \right) + u_0 \quad (11)$$

where T_i and T_d are the parameters of the integral and derivative terms, T_c is the control cycle, K_c the controller gain, and u_0 the base level control signal.

The advantage of the positional form is that it actually determines the full value of the action variable, which can more easily be supervised than its alternative, the velocity or incremental representation. This control expression is first used for simulation and then applied in experimental tests. Because of process nonlinearities (different slopes at increasing and decreasing pH values) it is difficult to apply the classical PID controller tuning rules, such as Ziegler and Nichols (1942) or Cohen and Coon (1953) for controller parameter tuning (Seborg *et al.* 2004). Hence, in such situations, PID controller parameters for bioreactors are most often tuned experimentally. So we did in a first approach.

In the cultivation experiments we initially used a set of PID controller parameters, which were obtained after using Ziegler–Nichols rules (Aström and Hägglund 2006) and additional experimental tuning during some test cultivations. Despite that, the quality of the pH control was not sufficiently good in some intervals of the cultivation process. The simulated and the experimentally obtained results are compared in Figure 5 (left part). Since the simulated pH values reflect the experimentally observed structure in the pH fluctuations fairly well, the assumption in the ANN model (Eq. 9) that the pH changes are strongly correlated with CPR gets experimental support. Other possible contributions do not seem to make a big qualitative change in the profile. Hence, the experiments also suggest considering the CO_2 formation rate a key influence variable on pH changes in the culture.

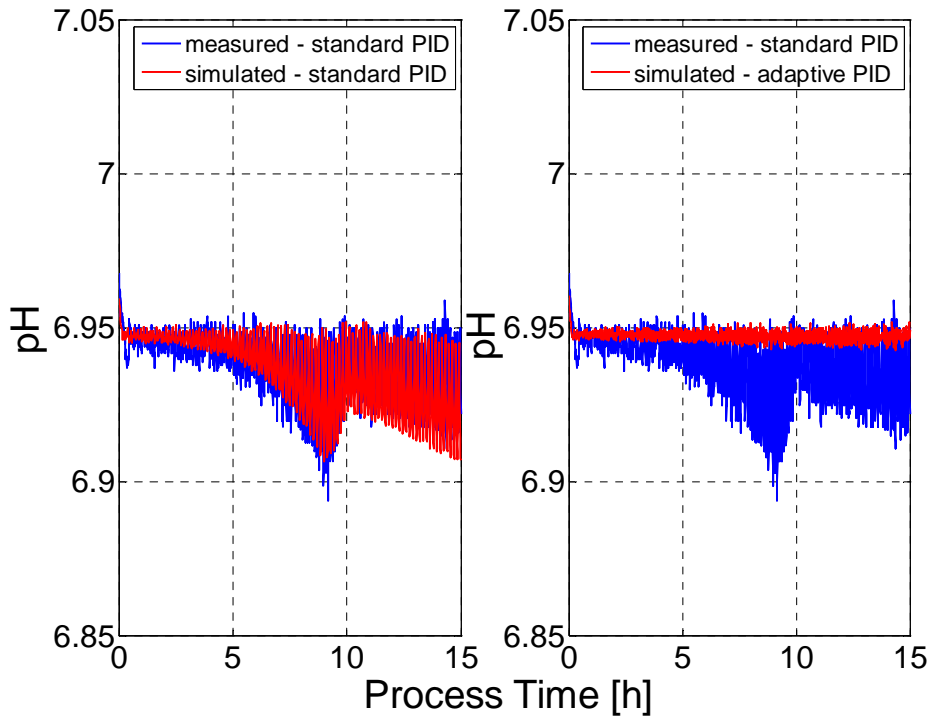


Figure 5. Comparison of simulated (*red*) and experimental (*blue*) results with the standard PID controller installed in a B. Braun DCU (*left part*). The same standard PID controller is compared with simulation results obtained with the gain-scheduling approach (*right part*).

In order to improve the pH controller, the information contained in tCPR is processed using the gain-scheduling technique (see Figure 1). The straightforward way is simply to adapt the PID controller gain in the following linear way:

$$K_c(t) = K_c + K_{tcpr} \cdot tCPR(t) \quad (12)$$

$$K_c(t) = 45 + 5 \cdot tCPR(t)$$

The parameter K_{tcpr} was obtained in a linear search procedure by simulation runs using process model and pH controller. The integrated absolute error (IAE) from the pH setpoint value was used as the criterion for quality of the controller (Aström and Hägglund 2006, Seborg *et al.* 2004).

$$IAE = \int_{t_1}^{t_2} |pH_{Setpoint} - pH_{real}| \cdot dt \quad (13)$$

As further shown in Figure 5 (right part), with the simple adaptation of the gain, the controller performance can be improved significantly as compared to the normal PID controller and there is no longer a structure in the fluctuations that reflects the tCPR profile. The IAE for the controlled pH per analyzed time interval 2-15 h was: IAE=0.0236 pH-unit·h (adaptive PID), and IAE=0.1367 pH-unit·h (normal PID).

Experimental Test of the pH Control System

The gain-scheduled controller designed according to the procedure described in the previous sections was first applied to *E. coli* cultures producing GFP. A typical result is depicted in Figure 6, which experimentally confirms the outcome of the simulation studies. The improvement as compared to the standard PID controller is evident. The integrated absolute error was $IAE=0.0189$ pH-unit·h for adaptive PID, and $IAE=0.1887$ pH-unit·h for normal PID (pH control starts at time=2 h). Hence, the adaptive controller improved the control performance by about one order of magnitude.

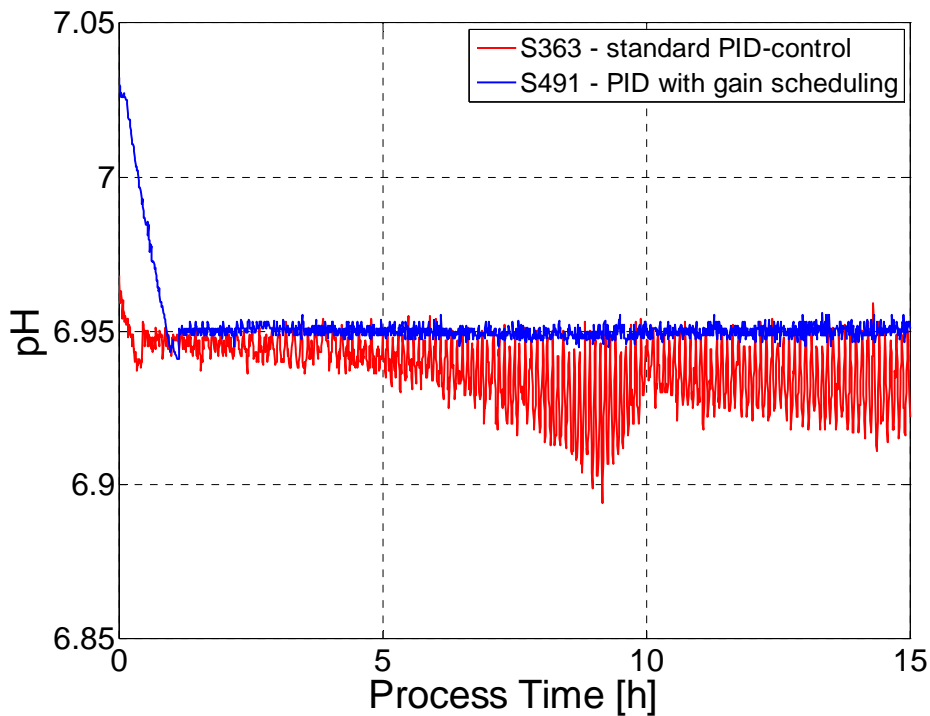


Figure 6. Experimental result of the pH control with a gain-scheduled controller (*blue*) compared with a corresponding profile obtained with a standard PID controller (*red*) installed in commercially available controllers (B. Braun DCU).

In order to show that the results are not restricted to a particular bioreaction system, fermentation runs with the same controller have also been conducted with another *E. coli* strain that expresses its product partly in form of inclusion bodies and partly in a soluble active form. Results are presented in Figure 7 giving an impression of the controller performance over the entire fermentation time.

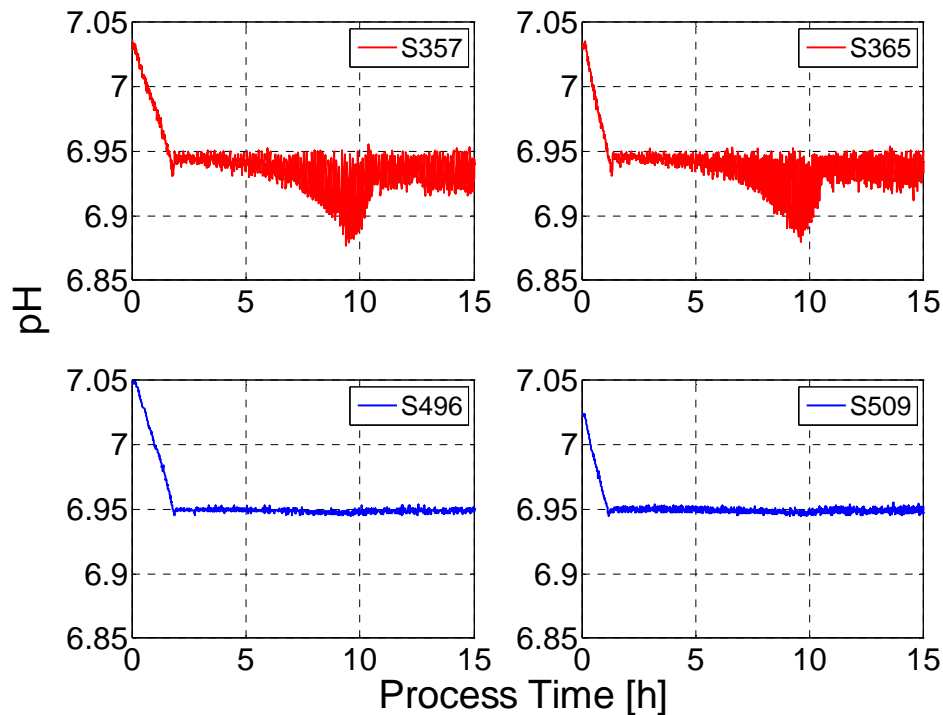


Figure 7. Experimental results of pH controlled experiments. The upper two plots (*red lines*) are from standard controlled experiments (HumaX studies S357, S365), the lower two (*blue lines*) were controlled with a gain-scheduling controller (HumaX studies S496, S509).

Simplification

If the cultivation process is run under substrate limited conditions throughout, as it is usual in processes operated in the fed-batch mode, gain-scheduling control of pH can be simplified even more. By taking the actual feed rate $F(t)$ instead of the total $tCPR(t)$ profile, it is also possible to adapt the controller parameters. This possibility becomes obvious from Figure 3, which shows that the course of the feed rate profile F is very similar to that of the $tCPR$ profile.

In Figure 8, an experimental validation of this fact is shown. A pH profile is depicted that was measured during an experiment where the PID controller gain was adapted using the actual feed rate F . It can be seen that the controller performance is as good as in the cases discussed above, where the total carbon dioxide production rate $tCPR$ was used for this purpose. For a comparison a pH profile of a culture where $tCPR$ measurements were used for scheduling is also plotted in the figure.

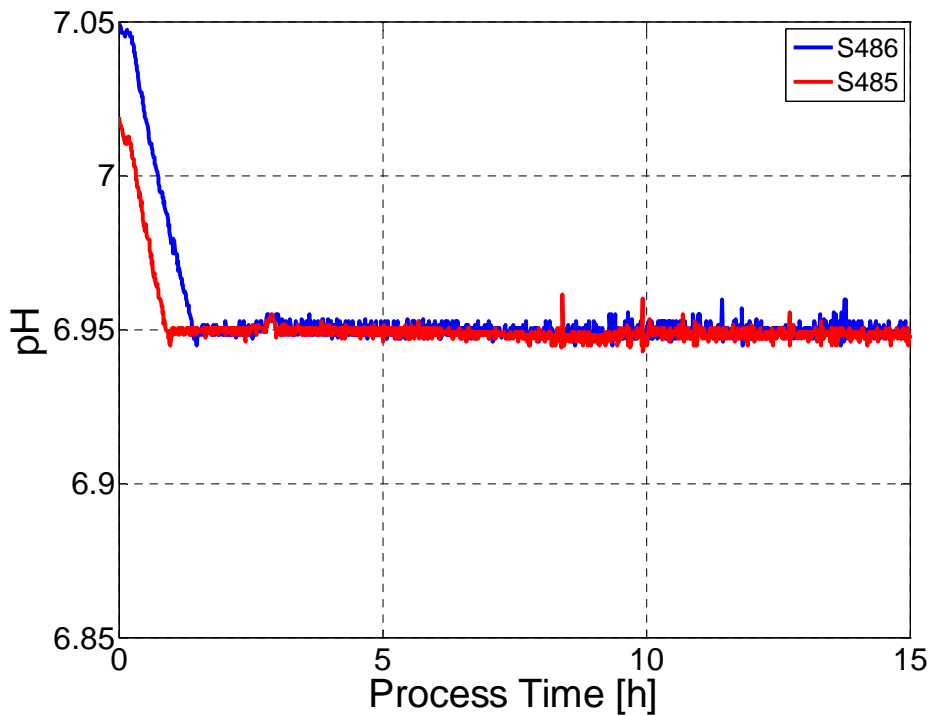


Figure 8. Measured pH profile in a cultivation performed with a gain-scheduled PID controller using the feed rate $F(t)$ as the scheduling variable (Study S486 - red line). For comparison, the corresponding profile of a study in which parameter adaptation is made with the $tCPR(t)$ measurement (Study S485 - blue line) is shown.

If gain-scheduling is the issue, we do not need an exact estimation of the particular variable used to monitor the changes in the process dynamics. Here it is sufficient to grossly quantify major changes. Thus, in cases of fed-batch cultures that entirely run in a substrate limited way, the equipment required to install this adaptive controller is much simpler, as the off-gas analyzer required for calculating $tCPR$ is no longer necessary. Then it suffices to use the feed rate profile to adapt the controller gain.

DISCUSSION

Benefits of Improved pH Control

Besides the primary aim of reduction of pH variation, the controller is of advantage in cultivations where the base consumption during pH control is used to estimate important state variables. The total amount of base, $Base(t)$ [mL], that has already been added to the fermenter during pH control as well as its time derivative $dBase/dt$ can be used to predict biomass and specific growth rates on-line (Iversen *et al.* 1994, Vicente *et al.* 1998, Jenzsch *et al.* 2006c, Sundström and Enfors 2008). The generally assumed approximation of the relationship between $Base(t)$ and the state variables biomass x and the specific biomass growth rate μ is similar to the Luedeking–Piret approach:

$$\frac{dBase}{dt} = Y_{bx}(t) \cdot \mu \cdot x \quad (14)$$

where $Y_{bx}(t)$ is the yield coefficient of base-biomass as a function of process time, i.e., changing process dynamics. By combining Eqs. 7 and 14, the base consumption rate is a valuable signal for estimation of biomass growth and the specific biomass growth rate on-line.

$$\frac{dx}{dt} = \frac{1}{Y_{bx}(t)} \cdot \frac{dBase}{dt} \quad (15)$$

$$\mu = \frac{1}{x} \cdot \frac{dx}{dt}$$

The estimation quality crucially depends on the quality of the base consumption rate signal and is significantly improved with the proposed pH control. The influence of the improved pH control on the estimates can be judged from the corresponding data shown in Figure 9. The comparison between conventionally and adaptively controlled experiments demonstrates that the base consumption signal in the adaptively controlled fermentation is significantly improved and therefore facilitates state estimation.

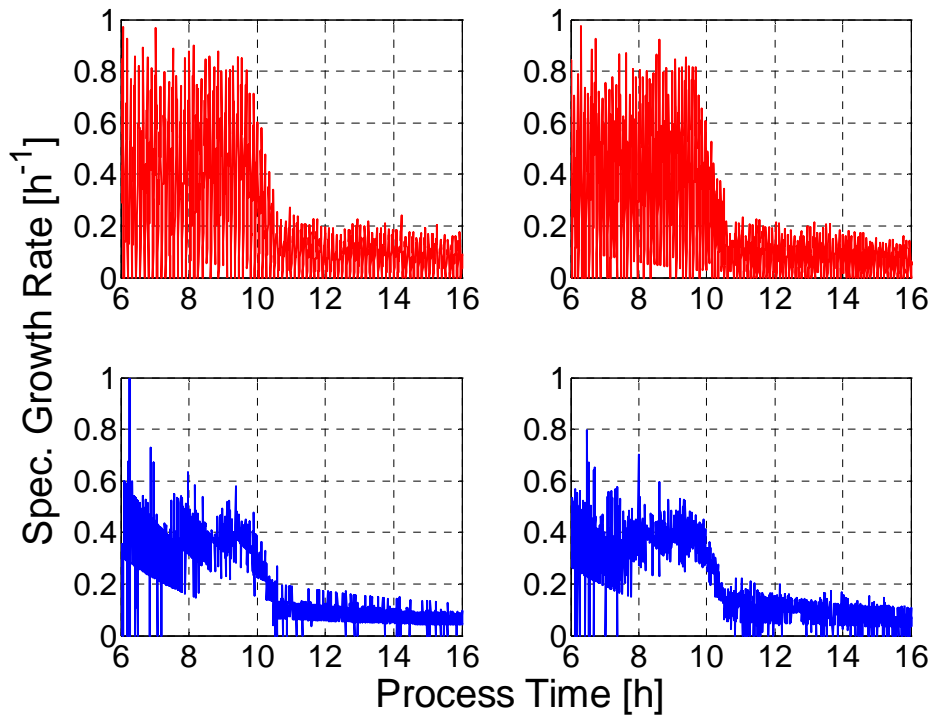


Figure 9. On-line estimates of specific growth rates from base consumption. With gain-scheduling control (*blue lines*, HumaX studies S496, S509) the signal-to-noise ratio of the specific growth rate estimation becomes significantly better compared to that obtained from standard PID control (*red lines*, HumaX studies S357, S365). The improved signals are now more suitable and valuable for state estimation and monitoring techniques.

Moreover, as already pointed out, pH is a key influencing factor for the concentration of CO_2 dissolved in water. Even small variations in pH cause significant changes in the CO_2 volume fractions measured by the off-gas analyzer in the vent line (Royce 1992, Spérandio and Paul 1997). This can easily be seen by comparing data from cultivation runs performed with standard PID and gain-scheduling control, as depicted in Figure 10. Consequently, the measurement of the respiratory quotient, a central monitoring variable of metabolic state, becomes less noisy when pH control is improved.

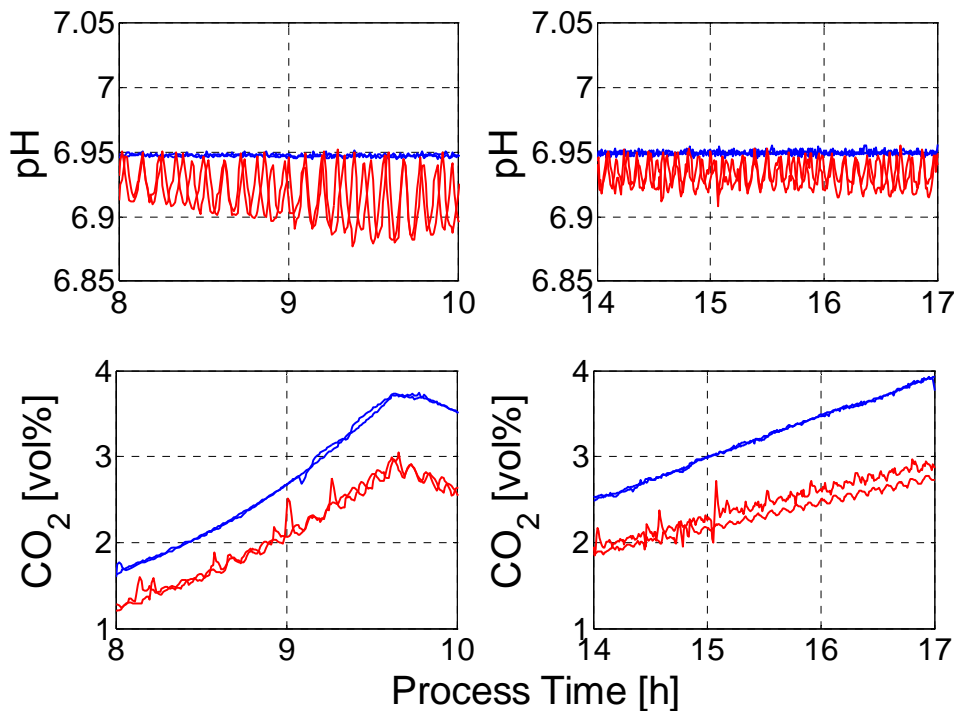


Figure 10. Effect of improved pH on carbon dioxide volume fraction observed by the off-gas analyzer in fermentations with standard PID control (*red lines*, HumaX studies S357, S365) and gain-scheduling control (*blue lines*, HumaX studies S496, S509) operated at a constant airflow rate $Q^{\text{in}}=1,200 \text{ L h}^{-1}$. A closer look at two different fermentation time intervals shows that the measurement noise of CO_2 is decreased due to a reduction in the pH fluctuations.

Finally, an improved $d\text{Base}/dt$ signal can also be used for process control purposes. Control of the total cumulative CPR was shown to lead to a perfect batch-to-batch reproducibility in microbial cultivations (Jenzsch *et al.* 2007). The detectors of CO_2 in off-gas are most often based on infrared sensors. In cases where the $t\text{CPR}$ signal fails, the faulty CO_2 measurement signal can be replaced by an estimate based on the total base consumption $\text{Base}(t)$. This allows keeping the controller running in a safe operational mode, provided the signal-to-noise ratio of this signal is high enough. The adaptive pH controller presented leads to $\text{Base}(t)$ data that allow very good $t\text{CPR}$ signal estimates as shown in the experimental example depicted in Figure 11. In this example, this indirectly measured $t\text{CPR}$ signal was used to control the cultivation process. The difference to the reference process is very low.

In conclusion, here we showed for the first time at real cultivation systems that gain-scheduled PID controllers are easy to apply for pH control in fermentation processes for recombinant protein production. As the key scheduling variable the total carbon dioxide formation rate is the natural choice. It allows continuously adapting the controller gain to changes in the dynamics of the cultivation process. If the entire cultivation is running under substrate limited conditions, the gain adaptation can be simplified even more. Then, the substrate feed rate profile can be used directly as the scheduling variable. The feed rate is available as a noise-free signal and does not require a costly off-gas analyzer. With both approaches the final pH fluctuations in the culture could be reduced significantly. This was demonstrated at several cultivation systems. Improved pH control also led to a largely increased signal-to-noise ratio of other important signals, e.g., the base consumption rate or the rate of CO_2 appearance in off-gas. Both signals are often used in various process state estimation, monitoring, control and fault analysis tasks (e.g., Jenzsch *et al.* 2007).

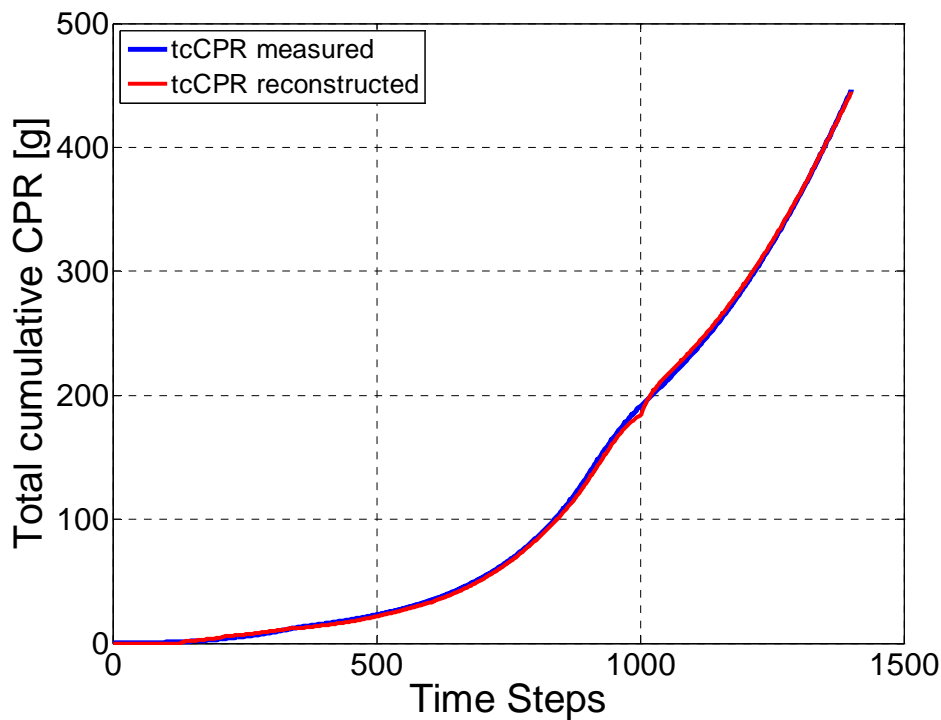


Figure 11. Fermentation controlled along a predefined total cumulative CPR profile. In the experiment, the tcCPR signal was estimated (*red line*) from the base consumption measured (*blue line*) during a fermentation where the pH was controlled with the gain-scheduled pH controller described above.

Thus, with the proposed pH controller the central requirement of FDA/ EMEA (PAT) with respect to data quality and batch-to-batch reproducibility can be met in a simple way. The benefit of the improved controller is that it drastically reduces the variability of decisive process variables and therewith increases the performance of advanced controllers.

Acknowledgments

This work has been performed within the “Excellence Cluster Biotechnology” of the State of Sachsen-Anhalt and was financially supported by the Ministry of Science and Technology. We gratefully thank for that support.

NOMENCLATURE

B	Base addition	[g]
CO_2^{in}	Concentration of CO_2 in the inlet gas	[mol/L _{Gas}]
CO_2^{out}	Concentration of CO_2 entering reactor's headspace	[mol/L _{Gas}]
$\text{CO}_2^{\text{Sensor}}$	Concentration of CO_2 reaching off-gas sensor	[mol/L _{Gas}]
CO_2^{L}	Concentration of CO_2 in the liquid phase	[mol/L _{Liquid}]
CO_2^*	Saturation concentration of CO_2 in the liquid phase	[mol/L _{Liquid}]
CPR	Specific carbon dioxide production rate	[g _{CO2} /(L _{Liquid} ·h)]
CTR	Carbon dioxide transfer rate	[g _{CO2} /(L _{Liquid} ·h)]
tCPR	Total carbon dioxide production rate	[g _{CO2} /h]
cCPR	Cumulative specific carbon dioxide production rate	[g _{CO2} /kg]
ctCPR	Cumulative total carbon dioxide production rate	[g _{CO2}]
e	Control error between setpoint and actual value	[-]
HCO_3^-	Concentration of bicarbonate in the liquid phase	[mol/L _{Liquid}]
IAE	Integral absolute error	[pH-unit·h]
k_1	Forward reaction rate constant of CO_2 hydration	[h ⁻¹]
k_2	Reverse reaction rate constant of CO_2 hydration	[h ⁻¹]
$k_L^{\text{CO}_2 a}$	Volumetric mass transfer coefficient for CO_2	[h ⁻¹]
K_a	Dissociation constant for carbonic acid	[mol/L _{Liquid}]
K_c	Basic controller gain	[-]
K_{ICPR}	Adaptive controller gain from tCPR measurements	[h/g _{CO2}]
m_c	Maintenance contribution to tCPR	[g _{CO2} /g _x /h]
M_{CO_2}	Molecular weight of CO_2	[g/mol]
p^{out}	Pressure	[Pa]
Q^{in}	Inlet volumetric aeration rate	[L _{Gas} /h]
Q^{out}	Outlet volumetric aeration rate	[L _{Gas} /h]
R	Universal gas constant	[J/(K·mol)]
R_{Base}	Rate of base addition	[mL/h]
R_{CO_2}	Rate of CO_2 hydration in water	[mol/(L _{Liquid} ·h)]
T	Temperature	[K]
T_c	Control cycle	[s]
T_d	Differential time constant in PID algorithm	[s]
T_i	Integral time constant in PID algorithm	[s]

T_N	Delay time	[h]
u	Control signal	[-]
V_L	Liquid volume in reactor	[L]
W	Weight of liquid volume	[kg]
x	Total biomass	[g]
$y_{CO_2}^{Sensor}$	Volume fraction of outlet CO ₂ measured by off-gas sensor	[L _{CO2} /L _{Gas}]
Y_{bx}	Yield coefficient base – biomass	[g _{Base} /g _x]
Y_{cx}	Yield coefficient carbon dioxide – biomass	[g _{CO2} /g _x]
ε	Gas holdup	[L _{Gas} /L _{Liquid}]
μ	Specific biomass growth rate	[h ⁻¹]
τ	Time constant of pH electrode	[s]

REFERENCES

- Aström, K., Hägglund, T., 2006. Advanced PID control. *ISA-Instrumentation, Systems, and Automation Society*, Research Triangle Park. North Carolina.
- Cardello, R.J., San, K.Y., 1988. The design of controllers for batch bioreactors. *Biotechnol. Bioeng.* 32: 519–526.
- Cohen, G.H., Coon, G.A., 1953. Theoretical considerations of retarded control. *Trans ASME* 75: 827–834.
- Franzen, C.J., Albers, E., Niklasson, C., 1996. Use of the inlet gas composition to control the respiratory quotient in microaerobic bioprocesses. *Chem. Eng. Sci.* 51: 3391–3402.
- Iversen, J.J.L., Thomsen, J.K., Cox, R.P., 1994. On-line growth measurements in bioreactors by titrating metabolic proton exchange. *Appl. Microbiol. Biotechnol.* 42: 256–262.
- Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., Lübbert, A., 2006a. Improving the batch-to-batch reproducibility in microbial cultures during recombinant protein production by guiding the process along a predefined total biomass profile. *Bioproc. Biosyst. Eng.* 29: 315–321.
- 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., Simutis, R., Eisbrenner, G., Stückerath, I., Lübbert, A., 2006c. Estimation of biomass concentrations in fermentation processes for recombinant protein production. *Bioproc. Biosyst. Eng.* 29: 19–27.
- 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.
- Johnston, W., Cord-Ruwisch, R., Cooney, M.J., 2002. Industrial control of recombinant *E. coli* fed-batch culture: new perspectives on traditional controlled variables. *Bioproc. Biosyst. Eng.* 25: 111–120.
- Levisauskas, D., 1995. An algorithm for adaptive control of dissolved oxygen concentration in batch culture. *Biotechnol. Techn.* 9: 85–90.
- Levisauskas, D., 2001. Inferential control of the specific growth rate in fed-batch cultivation processes. *Biotechnol. Lett.* 23: 1189–1195.
- Levisauskas, D., Simutis, R., Borvitz, D., Lübbert, A., 1996. Automated control of the specific growth rate in fed-batch cultivation processes based exhaust gas analysis. *Bioproc. Eng.* 15: 145–150.
- Nielsen, J., Villadsen, J., 1994. Bioreaction engineering principles. *Plenum Press*, New York.
- Royce, P.N., 1992. Effect of changes in the pH and carbon dioxide evolution rate on the measured respiratory quotient of fermentations. *Biotechnol. Bioeng.* 40: 1129–1138.
- Seborg, D.E., Edgar, T.F., Mellichamp, D.A., 2004. Process dynamics and control. *Wiley*, New York.
- Shioya, S., 1992. Optimization and control in fed-batch bioreactors. *Adv. Biochem. Eng. Biotechnol.* 46: 111–142.
- Shuler, M., Kargi, F., 2001. Bioprocess engineering: basic concepts. *Prentice Hall International*, New York.
-

- Siano, S.A., 1995. On the use of the pH-control reagent addition rate for fermentation monitoring. *Biotechnol. Bioeng.* 47: 651–665.
- Soons, ZITA, Voogt, J.A., van Straten, G., van Boxtel, A.J.B., 2006. Constant specific growth rate in fed-batch cultivation of *Bordetella pertussis* using adaptive control. *J. Biotechnol.* 125: 252–268.
- Spérandio, M., Paul, E., 1997. Determination of carbon dioxide evolution rate using on-line gas analysis during dynamic biodegradation experiments. *Biotechnol. Bioeng.* 53: 243–252.
- Sundström, H., Enfors, S.O., 2008. Software sensors for fermentation processes. *Bioproc. Biosyst. Eng.* 31: 145–152.
- Vicente, A., Castrillo, J.I., Teixeira, J.A., Ugalde, U., 1998. On-line estimation of biomass through pH control analysis in aerobic yeast fermentation systems. *Biotechnol. Bioeng.* 58: 445–450.
- Wang, F., Du, D., Li, Y., Chen, J., 2006. Regulation of CCR in the CGTase production from *Bacillus macorous* by the specific cell growth rate control. *Enzyme. Microb. Technol.* 39: 1279–1285.
- Zeng, A.P., Buyn, T.G., Posten, C., Deckwer, W.D., 1994. Use of respiratory quotient as a control parameter for optimum oxygen supply and scale-up of 2,3-butanediol production under microaerobic conditions. *Biotechnol. Bioeng.* 44: 1107–1114.
- Ziegler, J.G., Nichols, N.B., 1942. Optimum settings for automatic controllers. *Trans. ASME* 64: 759–768.

CHAPTER 3

Advanced Control of Dissolved Oxygen Concentration in Fed Batch Cultures during Recombinant Protein Production

ABSTRACT

Design and experimental validation of advanced pO_2 controllers for fermentation processes operated in the fed-batch mode are described. In most situations, the presented controllers are able to keep the pO_2 in fermentations for recombinant protein productions exactly on the desired value. The controllers are based on the gain-scheduling approach to parameter-adaptive proportional integral controllers. In order to cope with the most often appearing distortions, the basic gain-scheduling feedback controller was complemented with a feedforward control component. This feedforward/feedback controller significantly improved pO_2 control. By means of numerical simulations, the controller behavior was tested and its parameters were determined. Validation runs were performed with three *Escherichia coli* strains producing different recombinant proteins. It is finally shown that the new controller leads to significant improvements in the signal-to-noise ratio of other key process variables and, thus, to a higher process quality.

This paper has been published in *Applied Microbiology and Biotechnology* (Kuprijanov and Gnoth contributed equally to that publication):

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.

INTRODUCTION

Controlling the dissolved oxygen concentration in bioreactors that are used for recombinant protein production is commonplace. Usually, the objective is not to allow the pO_2 dropping below a critical limit where the product quality is affected. It is, however, less known that a tighter pO_2 control in the bioreactor significantly reduces product variability and considerably increases the batch-to-batch reproducibility. Both issues, tight feedback control and high batch-to-batch reproducibility, are crucial in quality control in the course of the production of biopharmaceuticals and main goals of the Food and Drug Administration's (FDA) process analytical technology (PAT) initiative. Besides this, many fermentation processes are finally limited by the realizable oxygen transfer rate (OTR) of the bioreactor system. These processes critically depend on the proximity the pO_2 is maintained at the lowest possible value without negatively affecting the product formation rate. A well performing pO_2 control allows optimization of bioprocesses with respect to the maximization of OTR, as the actual working point of the controller can be approached closer to the desired minimum controllable pO_2 setpoint. In this way, the amount of total biomass per amount of supplied air can be further increased. Additionally, as will be shown later, bad tuned controllers exhibit large variations in pO_2 during fermentation, followed by significant fluctuations in pO_2 -related variables, i.e., off-gas data of CO_2 and O_2 . Consequently, the implementation of process models, state estimation or algorithms for biomass, product or growth rate estimation based on respiratory off-gas data is hindered. A strictly controlled pO_2 leads to significantly lower noise levels on signals of all key process variables and, therefore, to an increased reliability of advanced models and monitoring systems (Kuprijanov *et al.* 2008).

Controlling the dissolved oxygen concentration exactly to a given setpoint is difficult since various effects influence its actual value. As the dynamics of the fermentation processes are usually rapidly changing, classical proportional-integral-derivative (PID) controllers cannot be used for high-performance dissolved oxygen concentration control in bioreactors. Instead, parameter adaptive or model-supported controllers must be employed. Here, an improved gain-scheduling controller complemented with a feedforward component is designed and tested in laboratory bioreactors operated with different *Escherichia coli* strains expressing recombinant proteins. It is shown that it performs much better in compensating for the most probable process distortions with influence on pO_2 than the conventional pO_2 controllers. For industrial environments, only very simple and robust controllers can be proposed. In this paper, we present a simple gain-scheduling approach which, combined with feedforward disturbance elimination, leads to very stable and tight pO_2 control performance over the entire fermentation process. The approach is demonstrated at examples of concrete *E. coli* cultures but can easily be adapted to other systems.

Advanced pO_2 Control Algorithm

Gain-scheduling is an advanced PID control technique well suited for nonlinear processes in which the dynamics are varying with time and the operating conditions (Åström and Hägglund 1988, 2006, Åström and Wittenmark 1989). In bioprocess engineering, we are generally confronted with such processes. The dynamics of fermentation processes change drastically during the cultivation (Bastin and Dochain 1990, Shuler and Kargi 2001). Making use of gain-scheduling techniques, it is preliminarily necessary to find suitable measurable variables or quantities that can accurately be estimated and carry enough information about changes in process dynamics. Gain-scheduling methods use the information about the process dynamics to adapt the parameters of PI or PID controllers. For historical reasons, the phrase gain-scheduling is used even if PID controller parameters other than the gain, such as the derivative time or integral time, are changed in the course of the process. A block diagram of a control system

working with gain-scheduling is shown in Figure 1a. The system has an inner loop composed of the process and the controller and an outer loop, which adjusts the controller parameters upon changes in the process dynamics. In fermentation processes, formation or consumption rates of metabolites and substrates appear to be a good choice of scheduling variables depicting changes in the process dynamics. They carry more information about the process dynamics than the state variables (concentrations of biomass, substrate, and product). The possibility of applying the gain-scheduling technique for bioprocess control has already been discussed in literature (Cardello and San 1988, Levisauskas 1995, Lee *et al.* 1990), but the technique has not yet been applied in bioprocess engineering because of the pretended complexity of the algorithms or since the most important variables, which reflect changes in bioprocess dynamics, e.g., the absolute and the specific biomass growth rate, could not be measured on-line. In this paper, we show that in cases where the biomass growth rate can be related to the oxygen consumption rate, the gain-scheduling technique can easily be implemented and used for controller setting adjustments. We particularly demonstrate the performance of gain-scheduled PID controllers for dissolved oxygen concentration control in bioreactors. Most well-performing fed batch processes are kept under control by varying the substrate feed rate F . In this case, pO_2 control can be improved exploiting the relationship between pO_2 and its main influence variable, the substrate feed rate in a feedforward mode. This results in an improved control system that combines a simple gain-scheduling feedback controller with a feedforward controller (Figure 1b).

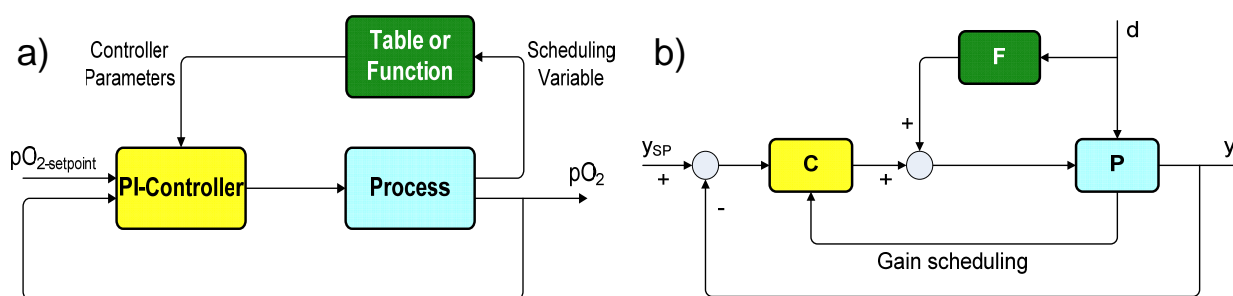


Figure 1. Block diagram of a PI-control system with gain-scheduling (a) and feedforward part, compensating for influences of substrate feed rate changes on the controller performance (b).

MATERIALS AND METHODS

Experimental Setup

All cultivations were performed in a B. Braun 15-L bioreactor Biostat-C operated in the fed-batch mode, i.e., no initial glucose and starting substrate addition directly after inoculation. The feeding profile was designed according to Jenzsch *et al.* (2006). This assured substrate-limited conditions directly after inoculation and a high degree of batch-to-batch reproducibility. The feeding solution consisted of glucose with a concentration of 400 g L^{-1} . It contained the same composition with respect to mineral salts as the initial cultivation medium, i.e., Na_2SO_4 2 g L^{-1} , $(\text{NH}_4)_2\text{SO}_4$ 2.46 g L^{-1} , NH_4Cl 0.5 g L^{-1} , K_2HPO_4 14.6 g L^{-1} , $\text{NaH}_2\text{PO}_4 \times \text{H}_2\text{O}$ 3.6 g L^{-1} , ammonium hydrogen citrate 1.0 g L^{-1} , $\text{MgSO}_4 \times 7 \text{ H}_2\text{O}$ 1.2 g L^{-1} and trace element solution 2 mL L^{-1} . Three BL21(DE3) strains with antibiotic resistance against kanamycin/ampicillin were used in the cultivation experiments. In all cases, the expression of the recombinant protein took place under the control of the T7/lac promoter (Sambrook *et al.* 1989) after induction with isopropylthiogalactopyranoside (IPTG). The first host system *E. coli* BL21 pET11a EGFP releases the green fluorescent protein (GFP) in its active form into its cytoplasm, whereas the second one expresses a commercially important recombinant protein (here referred to as

HumaX), which is encoded in pET28a vector, in two fractions, i.e., dissolved form and as an inclusion body. The third strain expresses a commercially interesting protein (here named GIP) nearly exclusively in form of inclusion bodies. Off-gas analysis was performed on-line with a paramagnetic oxygen sensor (Maihak Oxor 610) for O₂ and an infrared detector (Maihak Unor 610) for CO₂. The dissolved oxygen concentration was monitored with an Ingold pO₂ probe (Mettler Toledo) and maintained around 25 % saturation (%sat) by increasing the airflow rate and, subsequently, the stirrer speed. Additionally, a fluorescence pO₂ probe (Presens) was applied as an additional redundant measurement device. pH was measured by another Ingold probe (Mettler Toledo) and controlled by appropriately adding NH₄OH solution (25 % w/w) to the culture. The base addition was recorded on-line by means of a balance. Substrate addition was conducted with a gravimetric feed rate controller (Sartorius), which was connected with a balance (Sartorius). Off-line measurements were performed with time increments of half an hour: Biomass concentration was estimated from OD₆₀₀ measurements performed with a spectral photometer (Shimadzu UV-2102PC). Additionally, some dry weight measurements were performed in order to validate the available correlation between biomass dry weight and the OD₆₀₀ values. Glucose concentration was quantitatively determined by a YSI glucose analyzer (Yellow Springs Instrument), whereas acetate was measured by gas chromatography (Perkin Elmer) according to Vairavamurthy and Mopper (1990).

Process Model

Characteristic to fed-batch fermentation processes is that their dynamics significantly change with time. Hence, classical PID controllers are not expected to work accurately. PID controllers for the dissolved oxygen concentration lead to significant variation in some phases of the cultivation, i.e., the quality of the pO₂ controller may be good in one process phase but bad in others, as shown in Figure 2.

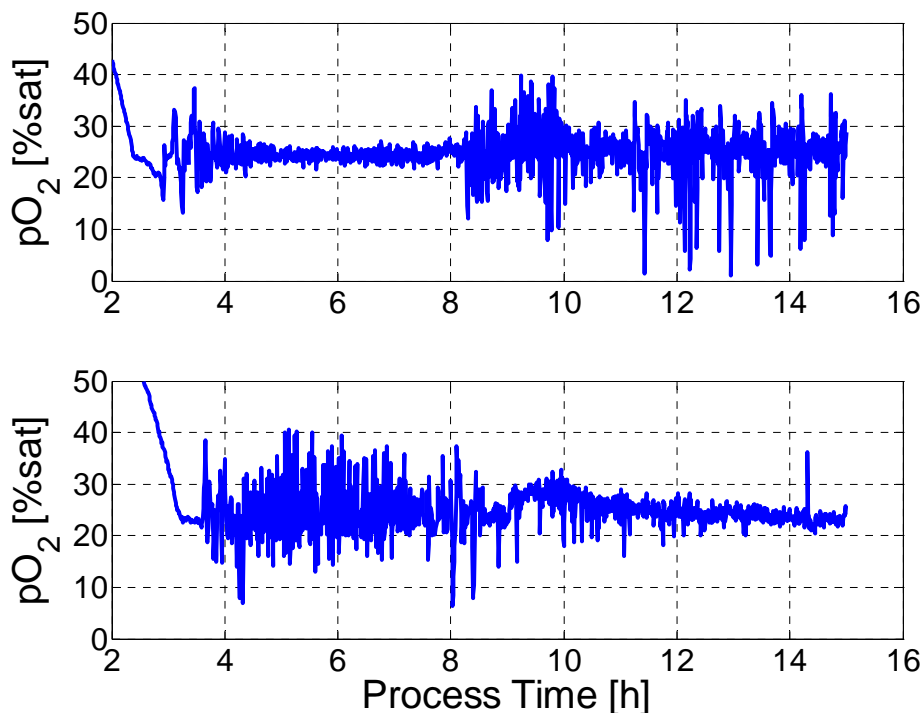


Figure 2. Typical experimental records of pO₂ controllers based on the industrial standard, the PID controller with constant parameters. The controller performs well in some process phases, but bad in others.

Various processes affect pO_2 in a real fermentation process for recombinant protein production. The most important influence variables are: airflow, stirrer speed, oxygen consumption rate (OUR) and more indirectly the substrate feed rate $Feed(t)$. Airflow rate Q_g and stirrer speed $N_{Stirrer}$ are both taken as action variables for pO_2 control in a sequential fashion. In the first phase of cultivation, pO_2 concentration is usually controlled by manipulating the airflow rate Q_g . During this short phase, pO_2 control is uncomplicated and can generally be guaranteed using classical PI and PID algorithms. Problems with pO_2 control usually appear later in the cultivation, when pO_2 is controlled by varying the stirrer speed. In order to estimate the parameters of the gain-scheduling controller, we used a mathematical model describing the dynamics of the dissolved oxygen in the bioreactor.

The concentration of dissolved oxygen O depends on the oxygen transfer rate OTR, i.e., the supply of oxygen from the aeration system of the reactor, and the oxygen consumption rate OUR of the cellular system. The balance equation for the dissolved oxygen O can be written as

$$\begin{aligned} \frac{dO}{dt} &= k_L a (O^* - O) - (Y_{xo} \cdot \mu \cdot X + Y_{mo} \cdot X) = k_L a \cdot (O^* - O) - OUR \\ pO_2 &= O \cdot H_{O_2} \\ \frac{dpO_2^{Sensor}}{dt} &= \frac{1}{T_{sensor}} \cdot (pO_2(t - \tau) - pO_2^{Sensor}) \\ k_L a &= f(N_{Stirrer}) \end{aligned} \quad (1)$$

where O^* is the saturation concentration of dissolved oxygen. pO_2 is used in bar, while in laboratory environments, this variable is usually scaled in percent of its saturation value. OUR [$g\ kg^{-1}\ h^{-1}$] represents the oxygen uptake rate, X [$g\ kg^{-1}$] the biomass concentration, μ [h^{-1}] the specific growth rate and $N_{stirrer}$ [rpm] the stirrer rotational speed. Because the mixing time constant in the laboratory reactor used is small as compared to the time constant of the pO_2 sensor as well as to the dynamics of the process, it was neglected in the model. The signal of the dissolved oxygen sensor pO_2^{Sensor} was modeled assuming a first order response system (Dang *et al.* 1977) with a time constant T_{sensor} and a time delay, τ .

Following the usual mass transfer theory, the $k_L a$ value was assumed to depend on the stirrer speed only using the following equation:

$$k_L a = a_0 + a_1 \cdot (N_{stirrer})^{a_2} \quad (2)$$

The parameters a_0 , a_1 , a_2 , T_{sensor} , and τ were identified using experimental data and Matlab's optimization procedure *lsqcurvefit*.

Active Experiments

Prior to building an adequate model for an improved pO_2 control, the relationships among central process variables have to be identified in fermentation runs switching off pO_2 control and varying the stirrer speed manually. Investigating the dynamic response in pO_2 upon step changes in the stirrer speed in a 10-L bioreactor, systematic experiments were carried out to generate informative data for that relationship. It is important to note that getting informative data, the measurements must be made available with sampling rates of at least $1\ s^{-1}$, since the time constants of the relevant transfer processes are in the order of a few seconds only. Figure 3 displays the experimental values of $N_{stirrer}$, pO_2 and OUR, carbon dioxide production rate (CPR) across the entire cultivation process. Interesting to note is the influence of the stirrer speed on the respiratory data of OUR and CPR. Changes in stirrer speed lead to changes in oxygen and carbon dioxide transfer and are observed as peaks in the respective off-gas data. Therefore, optimal

tuning of pO_2 controllers without great variations in the stirrer speed benefits signal quality of related off-gas measurements, which are central variables in advanced state estimation and monitoring systems.

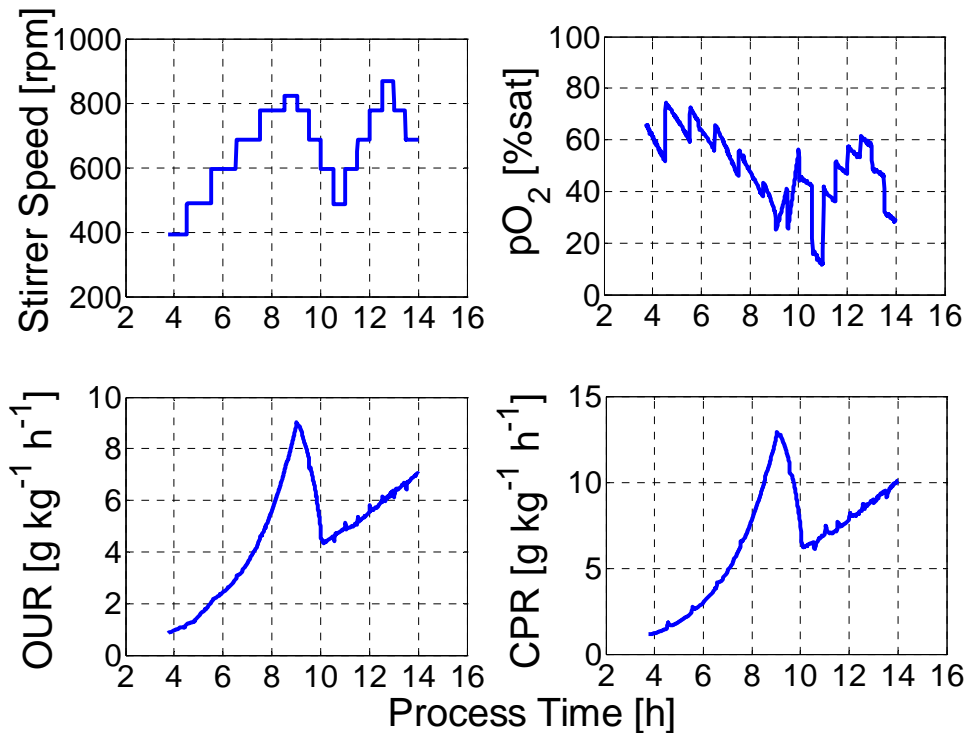


Figure 3. Results of active experiments for model identification. Response of the pO_2 signal to a change of the stirrer speed of the bioreactor. *Upper Part:* pO_2 control was deactivated and the stirrer was manually changed stepwise during the cultivation (working interval 350-950 rpm). *Lower part:* OUR and CPR measured during the same cultivation. The immediate step change of the stirrer speed leads to distortions in the off-gas signals of OUR and CPR.

Model Identification

Model identification, based on data from these measurements, resulted in the following parameter values for the k_1a and pO_2 sensor model:

$$a_0 = -357 \quad a_1 = 1.06 \quad a_2 = 1.07 \quad \tau = 9 \text{ s} \quad T_{\text{sensor}} = 17 \text{ s}$$

Figure 4 shows a comparison between the experimental results and simulated values using the identified parameters.

Despite the fact that quasi-steady-state model and experimental data differ somewhat, the model describes the dynamics of the pO_2 upon changes in the stirrer speed accurately enough for our purpose. In this respect, one must take into account that it is not necessary for the model to predict the pO_2 values with high accuracy in quasi-steady states; it only must be able to follow the gross changes in the dynamic response characteristics. It is the controller, which must force the process to tightly follow the setpoint profile. From this point of view, the model can perfectly be used for controller design. This model was applied to estimate the parameters for gain-scheduling of PI controller.

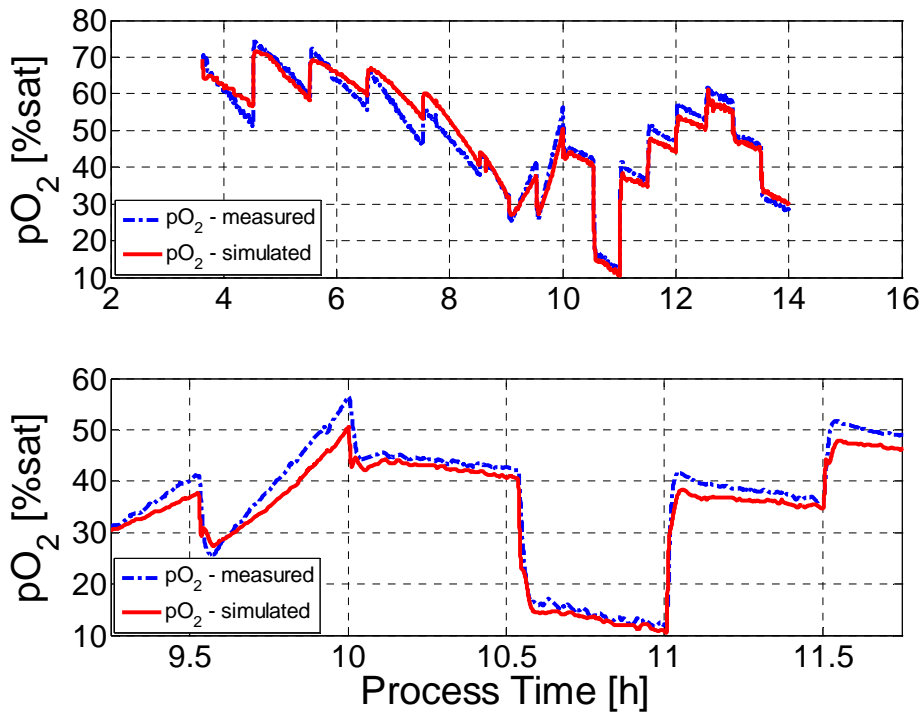


Figure 4. Measured and simulated signals of pO_2 in a cultivation process: *Upper graph*: entire cultivation process; *Lower graph*: detailed representation.

Estimation of Gain-Scheduling Control Parameters

For the gain-scheduling pO_2 control, we used the following representation of digital PI controllers:

$$u(k) = K_c \left(e(k) + \frac{T_c}{T_i} \cdot \sum_{k=0}^n e(k) \right) + u_0 \quad (3)$$

This equation can easily be transformed into the velocity form of PI controllers, which is more conveniently used in practical applications:

$$u(k) = u(k-1) + K_c \cdot [e(k) - e(k-1)] + \frac{K_c \cdot T_c}{T_i} \cdot e(k) \quad (4)$$

$e(k)$ is the control error, T_i the integral time, T_c the control cycle, K_c the controller gain, and u_0 is the initial value of the controller action.

For controller parameter tuning, we used the well known frequency response method described in Åström and Hägglund (2006). This method is based on knowledge about the point on the Nyquist representation of the process transfer function, where the Nyquist curve intersects the negative part of the real axis. For historical reasons, the point has been referred to as the ultimate point. The controller gain at this point is called ultimate gain, and the corresponding oscillation period is the ultimate period.

In the frequency response test, the gain, K_p , of the P part of the controller is increased starting from zero until the system begins to oscillate. When K_p is set such that a constant oscillation (neither increasing nor decreasing amplitude) is reached, that value is called the ultimate gain and is denoted K_u . The oscillation will generally be periodic with some period T_u .

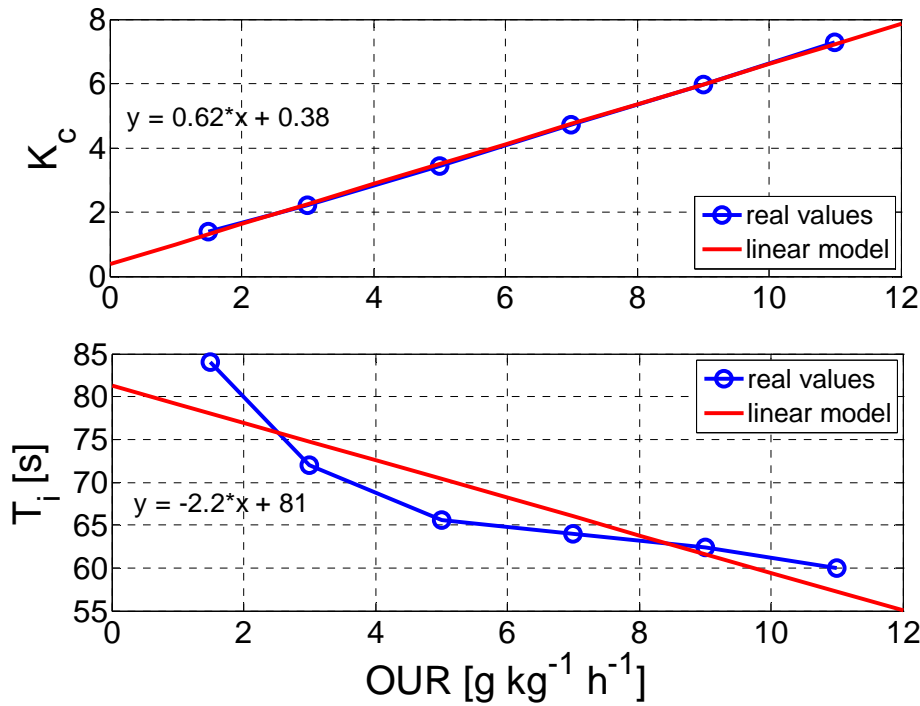


Figure 5. Measured and approximated values of the controller parameter K_c (upper graph) and for T_i (lower graph) for different values of OUR.

The process model described above was used to determine the parameters K_u and T_u . These parameter values are later used for controller design. In this investigation, the setpoint of the dissolved oxygen pO_{2sp} , is assumed to be 25 %sat, and the oxygen uptake rates, OUR, are allowed to vary in the interval ($1.5 \leq \text{OUR} \leq 11 \text{ g kg}^{-1} \text{ h}^{-1}$). The resulting values of the ultimate gain, K_u , and the period, T_u , are processed in a straightforward way to determine the controller parameters K_c and T_i making use of the well-known method of Ziegler and Nichols (1942, Seborg *et al.* 2004). For that purpose, the following relations were used:

$$K_c = 0.4 \cdot K_u \quad T_i = 0.8 \cdot T_u \quad (5)$$

In Figure 5, the measured values of K_c and T_i are compared with those approximated by means of a linear regression with the oxygen uptake rate, OUR, which resulted in:

$$K_c = 0.38 + 0.62 \cdot \text{OUR} \quad (6)$$

$$T_i = 81 - 2.2 \cdot \text{OUR} \quad (7)$$

The simple result of this discussion is that the PI controller parameters K_c and T_i can linearly be related to the actual oxygen uptake rate, OUR, which is signaling changes in the process dynamics. The latter is available on-line from off-gas analyzers, i.e., O_2 and airflow values measured on-line during the fermentation process. In other words, the simple linear relationships (Eqs. 6 and 7) between K_c and T_i and OUR can be used for an automated on-line adaptation of the controller parameters, which are normally kept constant in classical PI controllers.

Improving the Controller by a Feedforward Component

In substrate-limited processes, the substrate feed rate is the most important variable influencing pO_2 . Consequently, if the substrate feed rate $Feed(t)$, which is either known beforehand or measured on-line, changes rapidly during the cultivation process, the information can be used in a feedforward fashion to improve the quality of the pO_2 control. The simplest way is using a static approximation: A simple linear relationship between substrate feed rate and the stirrer speed can be derived from the experimental data depicted in Figure 6. This feedforward control component leads to a significant improvement of the pO_2 control performance, as the results will show.

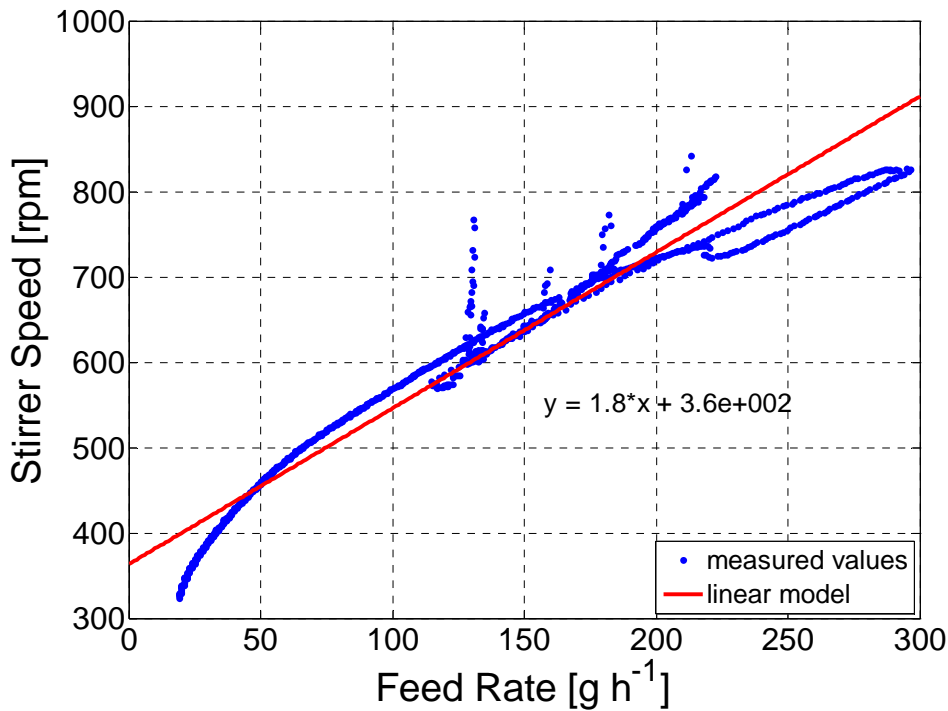


Figure 6. Experimental determination of a static relation between substrate feed rate and stirrer speed rate, when pO_2 is controlled at $pO_{2,sp}=25\%$ sat.

The linear fit can directly be taken as the feedforward part of the controller. It sufficiently approximates the stirrer speed rate, which is necessary to keep pO_2 around 25 %sat for a given substrate feed rate:

$$N = b_0 + b_1 \cdot Feed \quad (8)$$

In this particular case, the regression parameters are $b_0=360$ rpm and $b_1=1.8$ rpm $g^{-1} h^{-1}$. The equation is also useful for pO_2 control when the feedback component in the pO_2 controller is not activated. In this case, the equation can be used in the sense of a simple feedforward control. As we are using the velocity form of the controller, the change in the action variable u due to the feedforward component is:

$$\Delta u_f = \Delta N = b_1 \cdot (Feed(k) - Feed(k-1)) \quad (9)$$

Finally, the feedforward and the feedback algorithm must be combined. The final change in the action variable then simply becomes the sum of the individual actions:

$$\Delta u_a = \Delta u + \Delta u_f$$

$$\Delta u_a(k) = K_c \cdot [e(k) - e(k-1)] + \frac{K_c \cdot T_c}{T_i} \cdot e(k) + b_1 \cdot [Feed(k) - Feed(k-1)] \quad (10)$$

and the control signal is iteratively computed from the previous value by adding the actual total change Δu_a :

$$u_a(k) = u_a(k-1) + \Delta u_a(k) \quad (11)$$

The feedforward part of the control algorithm is only important where the substrate feed rate is changing rapidly. In other cases, good controller performance can be guaranteed with the simple gain-scheduled PI feedback controller.

RESULTS

Gain-Scheduling/Feedforward Controller Simulations

Tests of the gain-scheduling/feedforward controller and comparison of its performance with that of traditional PI controllers can initially be performed numerically using the process model described above.

Representing the noisy behavior of pO_2 in real cultivation, the output of the process model was corrupted with random measurement noise (zero mean, STD=0.5 %).

Additionally, a k_La disturbance was considered in order to simulate the controller's reaction upon an addition of a pulse of an antifoam agent to the culture. For such a situation, a reduction in the k_La value of 30 % was assumed. Such a large jump in k_La upon antifoam pulses is typical for real cultivation processes. In the simulation described here, this distortion was made at the fermentation time $t=8$ h for 6 min.

Moreover, a distortion in substrate feed rate was simulated at fermentation time $t=12$ h. At that time, the feed pump was assumed to be distorted for 6 min. During this time interval, the feed rate was reduced by 35 % from its setpoint profile. In order to simulate this distortion, an OUR step was calculated from the substrate feed rate, using a linear relationship.

A further problem with pO_2 control appears upon induction, which was performed experimentally by addition of IPTG. In the simulations, induction was assumed at the fermentation time $t=9$ h. After induction, the specific substrate uptake rate suddenly drops. Hence, the feed rate must be suddenly be decreased as well. This leads to a decrease in the OUR and, thus, at constant mass transfer characteristics, to an increase in pO_2 when the pO_2 controller cannot cope adequately with the rapidly changing oxygen demand.

The simulations of the pO_2 signal depicted in Figure 7 were computed for the three controller alternatives: the conventional PI controller (*upper graph*), the gain-scheduled controller (*middle*), and the gain-scheduled/feedforward controller (*lower graph*).

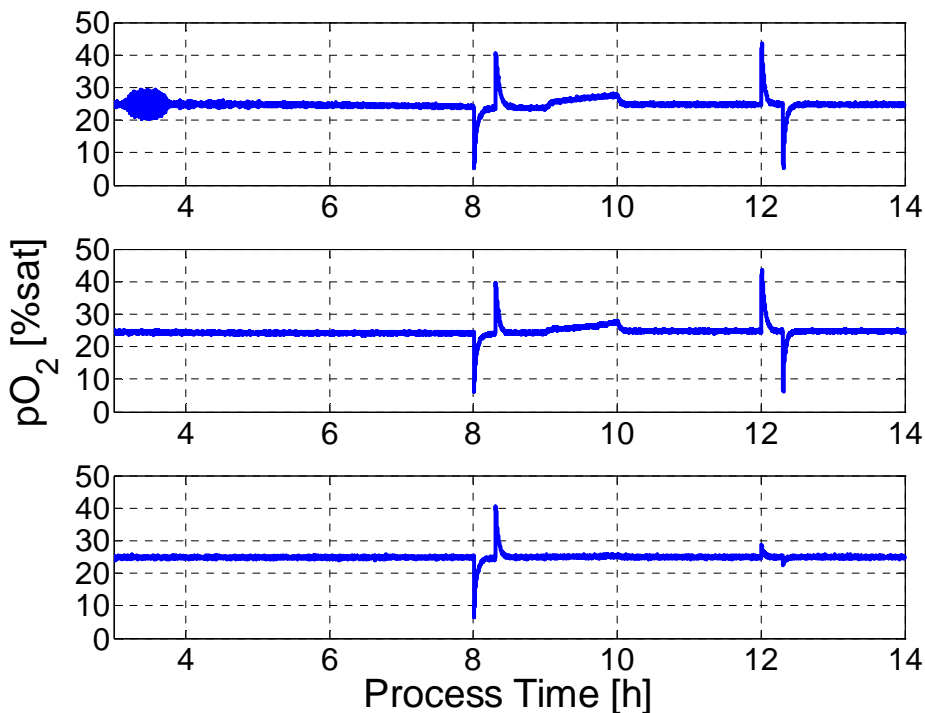


Figure 7. *Upper graph:* pO_2 profiles simulated for a PI-controller (tuned for the working point $OUR=5 \text{ g kg}^{-1} \text{ h}^{-1}$). *Middle:* gain-scheduled PI-controller. *Lower graph:* gain-scheduled/feedforward PI-controller. The typical disturbance of $k_L a$ is assumed to be 30 %. Also, a substrate feed distortion of 35 % was simulated.

The conventional PI controller, tuned to the working point $OUR=5 \text{ g kg}^{-1} \text{ h}^{-1}$, shows considerable fluctuations when the system is operated apart from these conditions. In the upper graph, this can be seen in the time interval ($3 \leq t \leq 4 \text{ h}$) where the controller performance is insufficient. At $t=8 \text{ h}$, the antifoam pulse leads to a severe decrease in pO_2 , which the controller tries to compensate for by increasing the stirrer speed. Typically, an overshoot of pO_2 in the opposite direction is observed when the distortion is removed. Upon induction at $t=9 \text{ h}$, pO_2 increases, but the controller cannot follow fast enough. It takes about 1 h until the controller is able to bring the pO_2 back on track. For the feed pump distortion at $t=12 \text{ h}$, the effect is similar to the antifoam distortion. However, the distortion goes the other way around. When the pump fails, the oxygen concentration rises, and the controller reacts. Then, after removing the distortions, the controller must quickly adapt to the normal situation, leading to an overshoot into the other direction.

The gain-scheduling controller, as displayed in the second row of the figure, is able to keep the pO_2 close to track over the entire cultivation process (Figure 7). Only upon the drastic distortions it cannot react fast enough. Hence, upon antifoam addition, feed pump failure and induction there are still problems in keeping the process on track. Another problem appearing with the gain-scheduled controller is that it leads to a small systematic deviation from the desired setpoint profile. This is due to its finite reaction time on changes in the process dynamics.

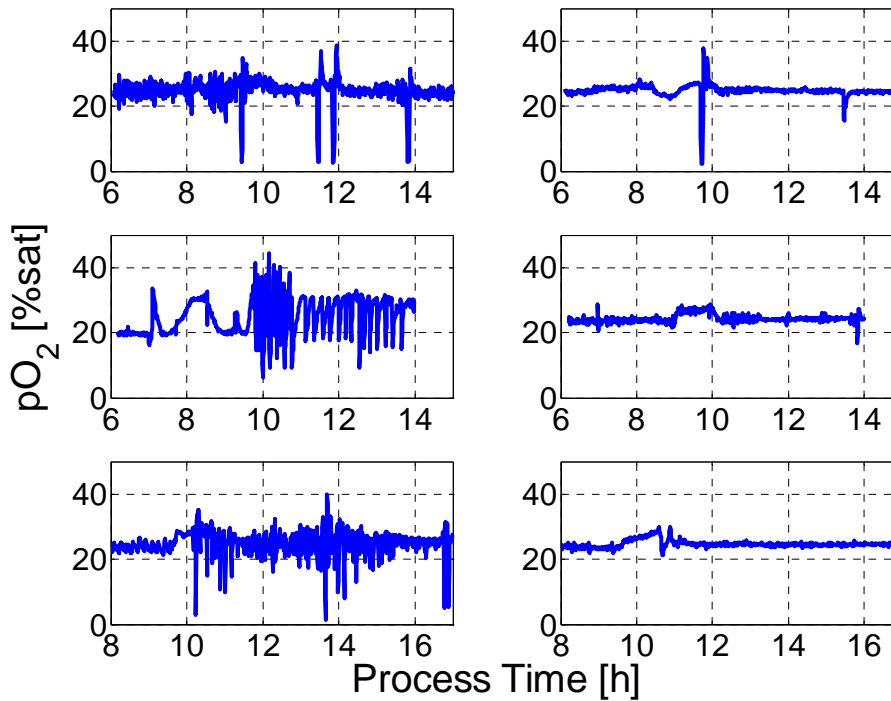


Figure 8. pO_2 profiles measured in experiments with 3 *E. coli* strains expressing the three recombinant proteins: GFP, GIP, HumaX. *Left:* PI control (Cultivations S486, S327, S429). *Right:* gain-scheduled PI control (Cultivations S489, S498, S496).

The feedforward/feedback variant of the controller eliminates the problems with these small systematic deviations, as they can be forecasted. This is shown in the upper part of Figure 9. The changes in the relevant disturbance variable (substrate feed ratio) are known or are measurable and can be compensated by the feedforward component of the controller; this eliminates the small systematic deviations from the setpoint profile. In particular, the drop of the feed rate upon induction is known beforehand or is measurable. Hence, the controller problem at induction can also be removed with the feedforward controller component.

As can be seen from Figure 7, the only distortions that cannot be coped with are the reactions on the randomly appearing antifoam pulses. From the distortions around the induction time and failures at the feed pump, only small reactions remain.

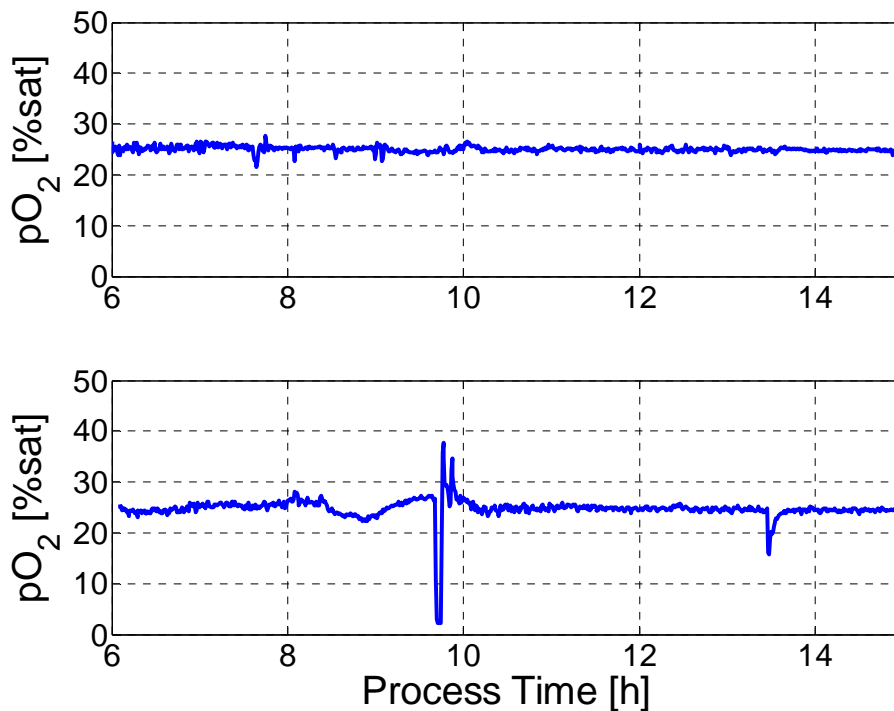


Figure 9. Experimental tests of the pO₂ control performance using feedforward/gain-scheduling PI control (*upper graph*: cultivation S526) and the simple gain-scheduling PI control (*lower graph*: cultivation S489) for a recombinant protein production process (GIP protein)

Experimental Validation of Gain-Scheduling Control

The gain-scheduling controller was applied to cultivations of genetically modified *E. coli* bacteria. Experiments were performed with three different strains. Each expressed a different protein. These are referred to as GFP, GIP, and HumaX. Typical pO₂ profiles obtained with a traditional PI and the simple gain-scheduled PI controller are depicted in Figure 8.

Figure 9 depicts the measured pO₂ signal in cultivations where pO₂ was controlled with the feedforward/gain-scheduled feedback controller. In this case, the controller performed well over the entire time interval it was switched on, i.e., it was not affected by quick changes in substrate feed rate at the induction time where the simple gain-scheduled PI controller still had some problems, as the lower part in the figure shows. The feedforward part of the controller eliminated these problems and worked perfectly. Hence, the proposed feedforward/gain-scheduled feedback control algorithm allows increasing the dissolved oxygen performance significantly. As well known in practice, cultures of genetically modified *E. coli* bacteria may produce foam particularly in the second half of the cultivation process. This is most often destroyed by adding chemical antifoam agents. Additions of antifoam agents immediately increase bubble coalescence, and consequently, the k_{La} suddenly drops and with it associated the oxygen transfer rate. The process proceeds so rapidly that one cannot react fast enough with increasing the stirrer speed or the airflow rate. Thus, unfortunately, the randomly appearing antifoam pulses lead to so drastic distortions that the controller cannot remove the corresponding distortions in pO₂.

Finally, an additional payoff of the improved pO₂ control must be discussed. The significantly reduced pO₂ noise level is expected to lead to an equivalent reduction in the noise level of other culture variables, a goal for process quality improvement stressed by FDA's and European Medicines Agency's PAT initiative. As shown by Jenzsch *et al.* (2007), the batch-to-batch reproducibility of fermentation processes, performed during the production of recombinant

therapeutic proteins, can be significantly improved with total cumulative carbon dioxide production rate (tcCPR) feedback control. Tight control requires accurate and precise measurements of the CPR throughout the entire time the process is running under this controller. It can immediately be seen from Figure 10 that the CPR signal in this fermentation depicts a much lower noise level, if the fermentations are additionally run with a feedforward/feedback (gain-scheduled) controller, as described above.

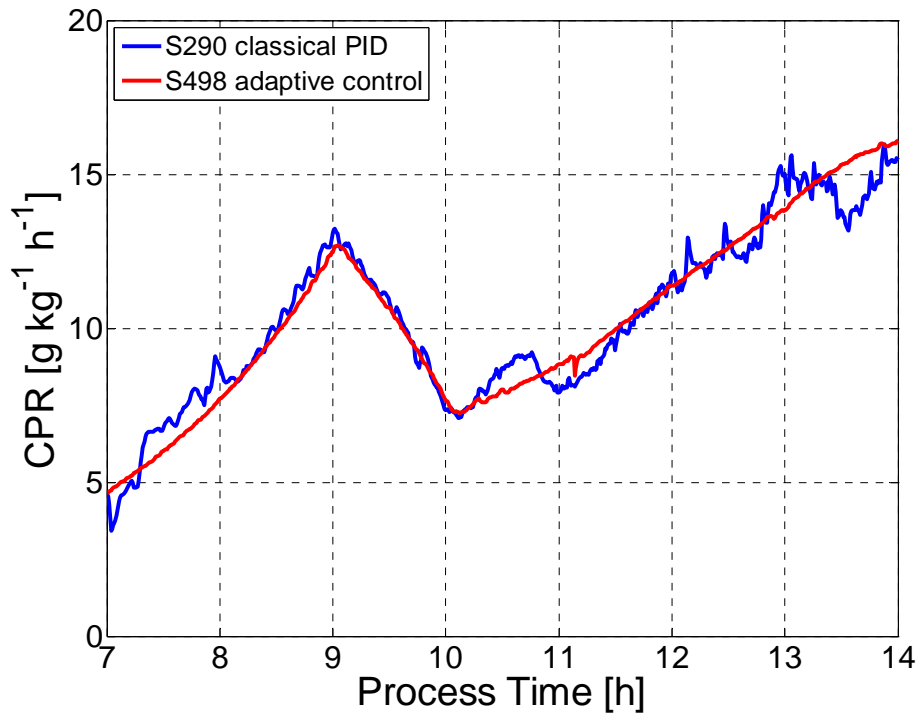


Figure 10. Measurements of the CPR during tcCPR controlled fermentations of *E. coli* producing GIP. In run S290 pO_2 was controlled with a conventional PI controller, in run S498 the gain-scheduled pO_2 controller was used.

The signal-to-noise ratio improvement in this key measurement signal clearly justifies the investment into the improved adaptive pO_2 controller, as it significantly improves the CPR signal quality and, with it, the accuracy by which the tcCPR controller can keep the overall cultivation under tight control.

DISCUSSION

pO_2 control is required in order to keep this quantity at its desired level despite strongly changing oxygen consumption rates. Conventionally used standard PID controllers cannot do this sufficiently well during the entire fermentation, as changes in the oxygen uptake dynamics require adjustments of the controller parameters. The consequences are strong fluctuations in the dissolved oxygen tension pO_2 . In this paper, we proposed a simple and robust adaptation technique that allows keeping pO_2 tightly on its desired level despite the random and systematic distortions on the culture.

Fluctuating oxygen tensions, resulting from inadequately pO_2 controlled cultures, lead to fluctuating OUR and CPR values. Hence, process state estimation and supervisory process control techniques, which make use of these off-gas signals (Jenzsch *et al.* 2007) perform significantly better when pO_2 fluctuations are kept smaller.

Moreover, in order to guarantee a high oxygen transfer rate, the pO_2 has to be kept as low as possible from physiological reasons. Tighter control allows a closer approach of the pO_2 setpoints to the critical oxygen concentration, and this immediately leads to a higher oxygen transfer rate OTR.

It was shown that extending the simple and robust gain-scheduling controller by a feedforward component enables the controller to quickly respond to fast changes in the substrate feed rate or in the volumetric mass transfer coefficient $k_L a$. Such changes generally appear upon antifoam agent pulses or substrate feed pump distortions. Moreover, most pO_2 controllers lead to problems after induction of recombinant protein expression where the substrate uptake rate drops significantly.

All the pO_2 changes resulting from these distortions can be removed by the simple adaptive feedforward/feedback controller proposed, with the exception of the pO_2 peaks appearing upon antifoam pulses. The latter distortions can significantly be reduced, but not completely eliminated. Currently, there is only one way to avoid the strong response of the pO_2 controller on antifoam agent additions to the culture: As these are signaled early enough by the foam sensor, the pO_2 controller can be switched off for about 5 min, just holding the current stirrer speed and airflow constant in that time interval.

Generally, the pO_2 controller helps reducing the process noise and thus improves the batch-to-batch reproducibility of the fermentation in accordance to the actual requirements of FDA.

Acknowledgments

This work was performed within the “Excellence Cluster Biotechnology” of the State of Sachsen-Anhalt, Germany. We gratefully thank the state’s government for the comprehensive support.

REFERENCES

- Åström, K.J., Hägglund, T., 1988. Automatic tuning of PID controllers. *Instrument Society of America*, New York.
- Åström, K.J., Hägglund, T., 2006. Advanced PID control. *ISA Instrumentation, Systems and Automation Society*, Research Triangle Park.
- Åström, K.J., Wittenmark, B., 1989. Adaptive control. *Addison-Wesley*, Massachusetts.
- Bastin, G., Dochain, D., 1990. On-line estimation and adaptive control of bioreactors. *Elsevier*, The Netherlands.
- Cardello, R.J., San, K.Y., 1988. The design of controllers for batch bioreactors. *Biotechnol. Bioeng.* 32: 519–526.
- Dang, N.D.P., Karrer, D.A., Dunn, I.J., 1977. Oxygen transfer coefficients by dynamic model moment analysis. *Biotechnol. Bioeng.* 19: 853–865.
- Jenzsch, M., Gnoth, S., Beck, M., Kleinschmidt, M., Simutis, R., Lübbert, A., 2006. 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.
- Kuprijanov, A., Schaepe, S., 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.
- Lee, S.C., Hwang, Y., Chang, H.N., Chang, Y.K., 1990. Adaptive control of dissolved oxygen concentration in a bioreactor. *Biotechnol. Bioeng.* 37: 597–607.
- Levisauskas, D., 1995. An algorithm for adaptive control of dissolved oxygen concentration in batch culture. *Biotechnol. Tech.* 9: 85–90.
- Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning - a laboratory manual. *Cold Spring Harbour Laboratory Press*, New York.
- Seborg, D.E., Edgar, T.F., Mellichamp, D.A., 2004. Process dynamics and control. *Wiley*, New York.
- Shuler, M., Kargi, F., 2001. Bioprocess engineering: basic concepts. *Prentice Hall International*, New York.
- Vairavamurthy, A., Mopper, K., 1990. Determination of low-molecular-weight carboxylic acids in aqueous samples by gas chromatography and nitrogen-selective detection of 2-nitrophenylhydrazides. *Anal. Chim. Acta* 237: 215–221.
- Ziegler, J.G., Nichols, N.B., 1942. Optimum settings for automatic controllers. *Trans. ASME* 64: 759–768.

CHAPTER 4

Product Formation Kinetics in Genetically Modified *E. coli* Bacteria: Inclusion Body Formation

ABSTRACT

A data-driven model is presented that can serve two important purposes. First, the specific growth rate and the specific product formation rate are determined as a function of time and thus the dependency of the specific product formation rate from the specific biomass growth rate. The results appear in form of trained artificial neural networks from which concrete values can easily be computed. The second purpose is using these results for on-line estimation of current values for the most important state variables of the fermentation process. One only needs on-line data of the total carbon dioxide production rate (tCPR) produced and an initial value x of the biomass, i.e., the size of the inoculum, for model evaluation. Hence, having the inoculum size and on-line values of tCPR, the model can directly be employed as a soft-sensor for the actual value of the biomass, the product mass as well as the specific biomass growth rate and the specific product formation rate. In this paper the method is applied to fermentation experiments on the laboratory scale with an *E. coli* strain producing a recombinant protein that appears in form of inclusion bodies within the cells' cytoplasm.

This paper has been published in *Bioprocess Biosystems Engineering*:

Gnoth, S., Jenzsch, M., Simutis, R., Lübbert, A., 2008. Product formation kinetics in genetically modified *E. coli* bacteria: inclusion body formation. *Bioproc. Biosyst. Eng.* 31: 41–46.

INTRODUCTION

E. coli is the most important host cell system for recombinant protein production systems, if the desired products do not need posttranslational modifications to obtain efficacy (Walsh 2006). In many practical cases, the heterologous products appear in form of inclusion bodies within these bacterial cells. Then, several downstream processing steps including cell disruption followed by solubilization and refolding are necessary before clinical efficacy of the protein is achieved.

In order to obtain a high product titer in the fermenter, the process operational procedure must be optimized for high cell density, i.e., high biomass concentration X , and, at the same time high specific product formation rates π . The latter can only be adjusted to their optimal values if the relationship between π and the variables that can be adjusted during the cultivation process is known. Such relationships are not well investigated, and thus, only very rough estimates can be found in literature. Usually it is simply assumed that the specific product formation rate is in a fixed stoichiometric relationship to the specific biomass growth rate $\pi=Y_{PX}\cdot\mu$; or the specific substrate consumption rate. Only a few groups developed more complex kinetic expressions for product formation (e.g., Levisauskas *et al.* 2003).

Strict mechanistic approaches are extremely difficult to quantify, as the anabolic metabolism of the cells is rather complex and not yet completely understood. Anyway, for process control purposes, it is straightforward to look for correlations, preferably of variables that are on-line accessible during industrial fermentation runs. Such data-driven approaches can be very well performing, provided that the right variables are chosen and the relations are identified using many data sets from the single process under consideration. The approach used here is based on artificial neural networks that are trained on an extended set of data records. These networks are known to depict very good mapping properties for complicated nonlinear relationships (e.g., Haykin 1999).

Data-Driven Approach to Product Formation Kinetics

A schematic view of the data-driven approach used here is shown in Figure 1 (Gnoth *et al.* 2006). It is based on two simple feedforward artificial neural networks (ANNs). The first one determines the specific biomass growth rate μ from on-line measured carbon dioxide production rate (tCPR) data as well as t_{ai} , the time after induction. The time signal t_{ai} is zero before induction and increases continuously thereafter. Acting as a switch, this information first signals the induction point to the network model and then delivers the current process time axis after induction. The specific growth rate μ determined by the identified inputs is used in an ordinary differential equation, i.e., a simple process model, to determine the biomass x , which is then fed back onto the input layer of the ANN. Thus, the model is a hybrid one where the kinetics, represented by ANNs, is combined with dynamic mass balances (Schubert *et al.* 1994).

The estimate of μ from this ANN is then used as an input of the second ANN computing the specific product formation rate π . A further input to this second ANN is again the time after induction, t_{ai} . Additionally, the third input is the specific protein concentration $p_x=p/x$ (where x is the total biomass, and p the total product mass). p is obtained from x and π by solving another simple mass balance equation shown in the Figure 1.

As there are no direct on-line measurements available for μ and π , the networks must be trained using off-line measured biomass x and product mass p data. This can be done with the sensitivity equation technique discussed by Simutis and Lübbert (1997) and Gnoth *et al.* (2006).

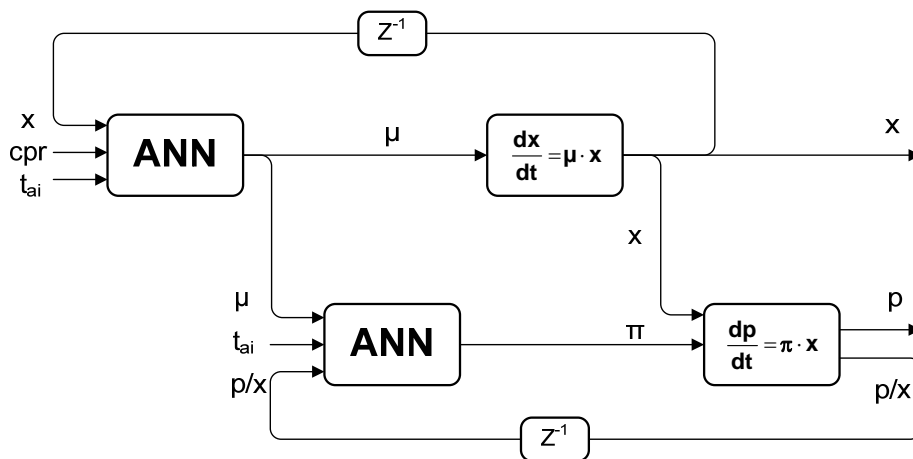


Figure 1. Scheme of the ANN-based approach of deriving the $\pi(\mu)$ -relationship as well as biomass x and product mass p .

EXPERIMENTAL

Experiments were performed with genetically modified *E. coli* bacteria that are able to produce the commercially interesting gastric inhibitory polypeptide GIP (Jenzsch *et al.* 2006). The desired product in the process reported about here is in form of inclusion bodies. All experiments used *E. coli* BL21(DE3) as the host cell. The target protein was coded on the plasmid pET28a and expressed under the control of the T7 promoter after induction with isopropyl-thiogalactopyranoside (1 mM IPTG). The strain was resistant against kanamycin. The product appears as inclusion bodies within the cytoplasm. The particular strain used did not produce notable amounts of acetate (data not shown) under the cultivation conditions adjusted in the experiments reported.

All the experiments were performed within Biostat-C 15-L bioreactor (BBI Sartorius) operated at maximum 8-L volume. The fermenter was equipped with three standard six-blade Rushton turbines that could be run up to 1,400 rpm. The aeration rate could be increased up to 24 sLpm. Aeration rate and then stirrer speed were increased one after the other in order to keep the dissolved oxygen concentration at 25 % saturation.

The fermentations were operated in the fed-batch mode immediately after inoculation. The initial volume was 5 L. Temperature and pH were adjusted to 35 °C and 7, respectively. The main carbon and energy source, glucose, was fed at concentration of 300 and 600 g/kg. For more details about the medium the reader is referred to Jenzsch *et al.* (2006).

All fermentations were started during night by automatic transfer of the inoculum from a refrigerator into the reactor. Substrate feeding was started in an open-loop fashion with predefined exponential profiles. When, after some cultivation time, the signal-to-noise ratio of the off-gas data reached a predefined level, closed-loop control was started with the total biomass or the total carbon dioxide produced (tcCPR) as control variables. Additionally, experiments were performed under unlimited conditions.

CO₂ in the vent line was measured with MAIHAK's Unor 610, O₂ with MAIHAK's Oxor 610. The total ammonia consumption during pH control was recorded by means of a balance beneath the base reservoir. These three quantities were measured on-line.

Biomass concentrations were measured off-line via optical density at 600 nm with a Shimadzu photo-spectrometer (UV-2102PC). Glucose was determined enzymatically with a YSI 2700

Select Bioanalyzer (Yellow Springs Instrument). The product was measured with SDS-PAGE after separation of the inclusion bodies and their solubilization.

RESULTS

Forty-nine data sets from the *E. coli* fermentations described above were used for training and validation (cross-validation procedure) of the hybrid model depicted in Figure 1. These fermentations were performed under very different conditions. Some of them were controlled to small specific growth rates in the order $\mu=0.1 \text{ h}^{-1}$. Others were run in an unlimited way with respect to the substrate concentration S . Furthermore, some runs were controlled to fairly high specific growth rates in the beginning of the product formation phase. Figure 2 depicts the result of the training of the network system illustrated in Figure 1 with respect to the simple $\pi(\mu)$ -relationship.

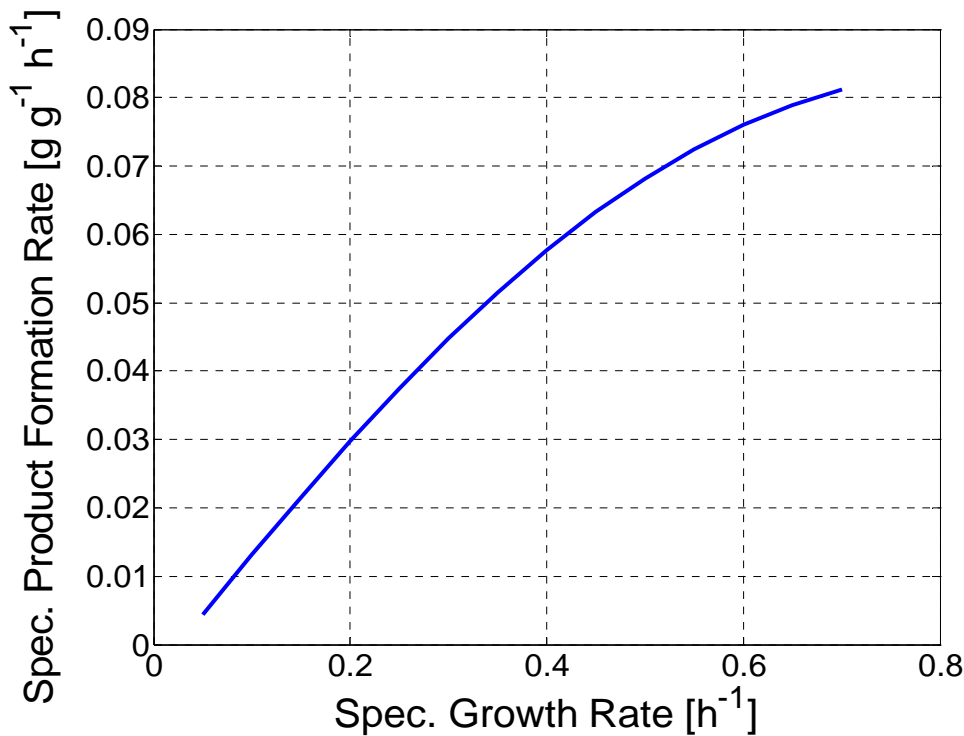


Figure 2. $\pi(\mu)$ -relationship derived from the network system depicted in Figure 1 for fixed $t_{ai}=1 \text{ h}$ and $p_x=0.01 \text{ g/g}$.

In order to assure that the data-driven model is truly mapping the biomass growth and product formation kinetics, the model solutions were compared to the corresponding experimental data. For this purpose, the cross validation technique was employed. The typical results shown in Figure 3 are from experiments, i.e., the data, which were not used during the network training.

Note, that the model depicted in Figure 1 only needs the on-line available tCPR signal, the initial biomass x as well as the time t_{ai} at which the culture is induced. The actual values of total biomass x and total product concentration p then appear as model outputs together with the specific growth and product formation rates at each time where a new tCPR value becomes available. Hence, the method can be used as a soft-sensor for x , p , μ and π . In Figure 3 the lines depict the outputs of this soft-sensor for x and p , the symbols show the corresponding off-line measurement values, which, in the case of the protein data, are available only days after the fermentation had been finished.

It is worthwhile to mention, that the on-line measured variable tCPR and the information t_{ai} about the induction state are sufficient for an on-line adaptation of the ANN-based model to the current state of the process, particularly in the product formation phase. With this on-line information, biomass and product mass as well as the corresponding specific formation rates μ and π can quite accurately be estimated. The full lines in Figure 3, corresponding to the biomass and the product mass shown, confirm this.

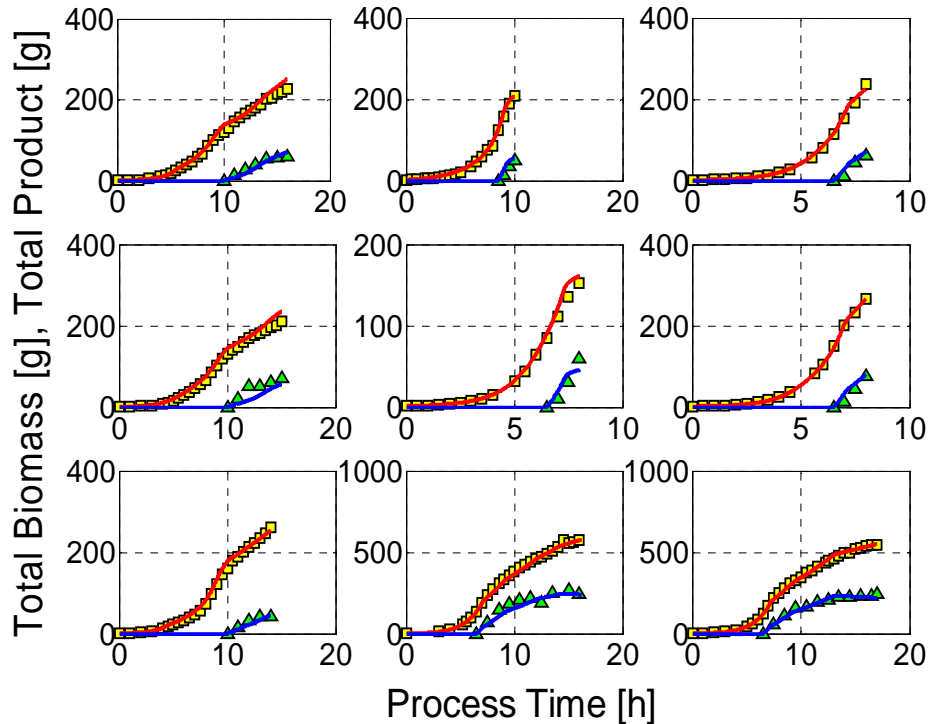


Figure 3. Typical example for simulations using the data-based kinetics. The biomass and product mass profiles (*lines*) are shown together with the corresponding off-line measured data (*symbols*). These data were not used during network training, thus, the comparison in the plot can be considered as a model's validation procedure.

Heterologous protein formation is usually accompanied by a metabolic load of the cells. In order to quantify this, the influence of the specific protein concentration p_x , i.e., the protein load of the cells on μ was examined, as well. The resulting three-dimensional graph is depicted in Figure 4. The specific product concentration p_x is seen not influencing the $\pi(\mu)$ -relationship significantly. At a given μ , π is only slightly decreasing with the accumulation of the inclusion body protein within the cytoplasm of the cells. However, there is a significant influence of p_x , the protein load of the cells, on the specific growth rate μ , which is referred to in literature as a metabolic burden.

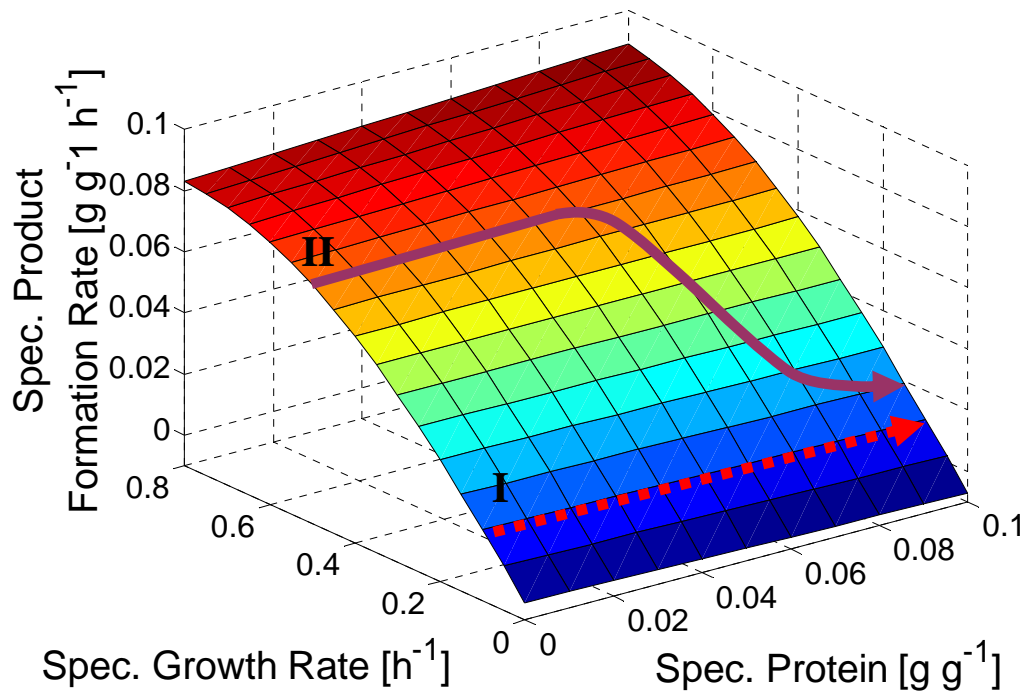


Figure 4. Specific product formation rate as a function of μ , the specific biomass growth rate, and p_x , the specific protein concentration. In general, two types of μ - π - p_x dependencies were obtained. Type I (*dashed line*) shows constant relationship, if the fermentations run under limited conditions below the critical specific growth-rate. Type II (*solid line*) depicts the relationship under maximum growth conditions, i.e., the increasing protein load on the cells reduces the achievable maximum growth rate after induction. Arrows indicate the process evolution with increasing process time.

To further clarify this, the π values, obtained from experimental data were plotted into the $\pi(\mu, p_x)$ -surface depicted in Figure 4. Figure 5 shows the data points for all 49 experiments. The values practically remain on the surface. They clearly show which part of the (μ, p_x) -space has been explored during the fermentations performed. With higher p_x values smaller and smaller μ values were obtained, even when the culture does not run in a substrate-limited way. Thus, only a part of the surface depicted in the model (Figure 4) is accessible during the process.

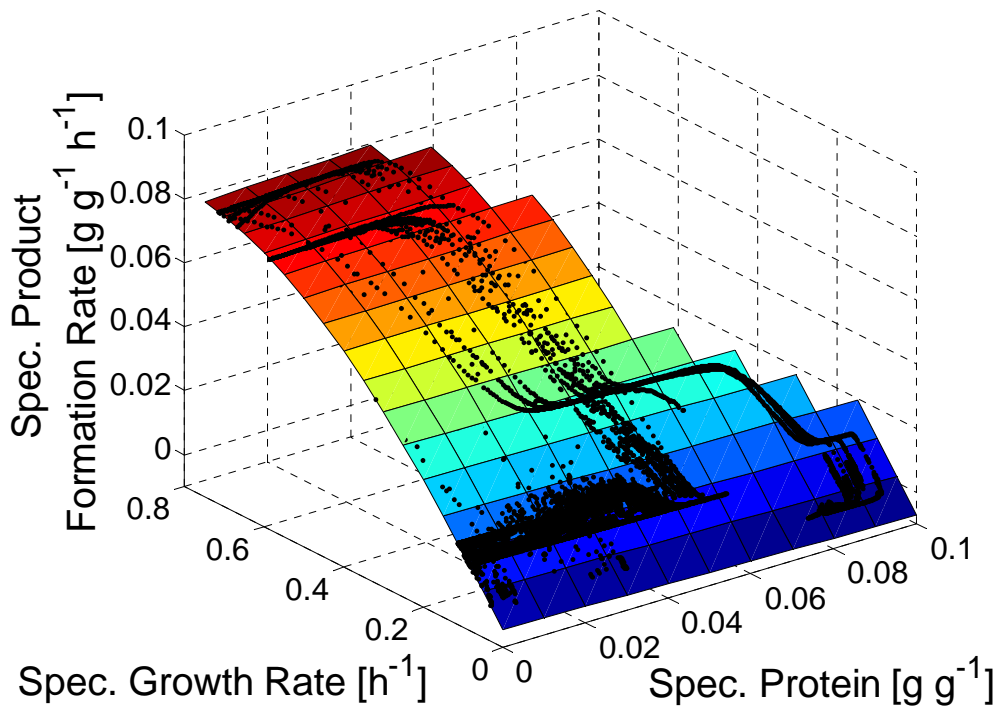


Figure 5. Comparison of the $\pi(\mu, p_x)$ -relationship derived from the ANN-based model with π values obtained from experimental data (*symbols*). The data on the plane shows the size of the design space, i.e., the approximate range of values that is possible at all in the $[\mu, p_x]$ -space.

The monotonic increase of the specific product formation rate with the specific biomass growth rate μ leads to the consequence that the biomass growth rate must be kept as high as possible, in order to obtain a maximum product titer at the end of the cultivation.

These results were successfully tested in some spot checks. For this purpose, additional validation experiments were performed. In the first one, the culture was grown at its maximum specific growth rate after induction, in the second one, a small specific growth rate which was suboptimal for the expression of inclusion body proteins was chosen. Both resulting protein formation patterns are depicted in Figure 6. The product mass profiles p are very different and show the behavior expected from the monotonically increasing π with increasing μ .

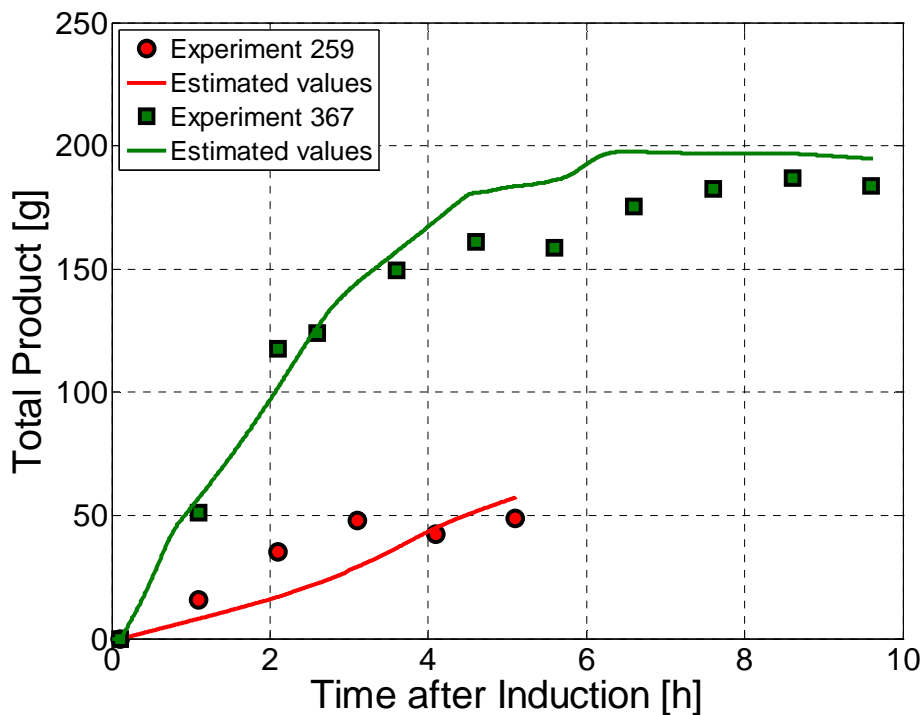


Figure 6. Validation tests of the results depicted in Figure 2 for two cases: one cultivation operated with substrate concentration well above values where substrate limitation occurs and one operated under substrate limitation. Total biomass was similar in both experiments at induction time. The curves present the total product obtained with the data-driven model and the points are experimental values.

DISCUSSION

The most important point to note for the protein investigated here, which is packed in form of inclusion bodies within the cells, is that the $\pi(\mu)$ -relationship is a simple monotonic function of essentially the specific biomass growth rate μ only. The metabolic load of the cell, resulting from the product accumulation within its cytoplasm and characterized by the specific protein concentration p_x , influences the maximum specific biomass growth μ_{\max} , but not directly the specific product formation rate π . In other words, the specific product formation rate π is only dependent on the specific growth rate μ as assumed by many researches in bioprocess engineering (e.g., Shioya 1992, Yoon *et al.* 1994, Soons *et al.* 2006). However, μ cannot be freely adjusted, as its maximum attainable value μ_{\max} decreases with the specific product concentration p_x . Thus, in this particular system, the influence of the cell's internal protein concentration or accumulation on the cell's protein formation performance π is an indirect one.

The $\pi(\mu)$ -relationship of the strain investigated here is qualitatively different from the one of other *E. coli* systems, e.g., one where the product appears in a soluble active form. For a strain expressing the soluble green fluorescence protein (GFP), Gnoth *et al.* (2006) found that the $\pi(\mu)$ -relationship depicts a maximum at a rather low specific biomass growth rate of about $\mu=0.14 \text{ h}^{-1}$. Other strains possibly show further forms of the $\pi(\mu)$ -relationship. Hence, it is straightforward to look for a well-performing technique that allows determining the product kinetics without the assumption of unproven models or constraints on the cell metabolism. Such a method is presented in this paper. It is not restricted to a special product and works without any assumption about kinetic parameters. Insofar, it is suitable for any kind of expressing protein (e.g., soluble/insoluble).

As the approach proposed here is a purely data-driven approach, relatively many data records are required to train the networks and to validate the results. In the beginning of the developments with a new biological system, having a few data records only, the prediction might not be sufficiently good. This could be considered a disadvantage of the proposed method. However, this approach has the advantage of excellent learning abilities. After each cultivation the networks can automatically be retrained using the extended database. In this way, the software learns without much additional efforts of the plant personnel.

Experiments performed in much different ways as compared to the records used for network training, will also not be predicted sufficiently well. However, after adding the data records to the database, the automatic learning will quickly lead to a better model performance. Thus, in the following runs the model will be sufficiently accurate for state estimation.

Providing many data records might be a problem in small laboratories, but it is definitely not a problem in industrial production environments. The number of data records necessary to obtain reliable results is dependent on the quality of the data, particularly on the accuracy of the product concentration values. Typically, data from about ten experiments are needed. Again, this is not so much a problem in industrial environments, where the laboratories are usually very experienced in measuring the concentrations of their particular product.

One of the main advantages of the method summarized in Figure 1 is that the evaluation of the identified model is extremely quick. It can thus perfectly be used for on-line model-supported process monitoring and control purposes. One example is its use as a software sensor for x and p as well as for μ and π . For these quantities no physical sensors are available.

Acknowledgements

The financial support from the “Ministry of Cultural Affairs” of Sachsen-Anhalt, Germany, is gratefully acknowledged.

REFERENCES

- Gnoth, S., Jenzsch, M., Simutis, R., Lübbert, A., 2006. Product formation kinetics in a recombinant protein production process. CAB-2007, IFAC, *Elsevier*, Amsterdam (accepted).
- Haykin, S., 1999. Neural networks: a comprehensive foundation. 2nd edn., *Prentice Hall*, Upper Saddle River.
- Jenzsch, M., Gnoth, S., Beck, M., Kleinschmidt, M., Simutis, R., Lübbert, A., 2006. Open-loop control of the biomass concentration within the growth phase of recombinant protein production processes. *J. Biotechnol.* 127: 84–94.
- Levisauskas, D., Galvanauskas, V., Henrich, S., Wilhelm, K., Volk, N., Lübbert, A., 2003. Model-based optimization of viral capsid protein production in fed-batch culture of recombinant *Escherichia coli*. *Bioproc. Biosyst. Eng.* 25: 255–262.
- Schubert, J., Simutis, R., Dors, M., Havlik, I., Lübbert, A., 1994. Bioprocess optimization and control: application of hybrid modelling. *J. Biotechnol.* 35: 51–68.
- Shioya, S., 1992. Optimization and control in fed-batch bioreactors. *Adv. Biochem. Eng. Biotechnol.* 46:1.
- Simutis, R., Lübbert, A., 1997. Exploratory analysis of bioprocesses using artificial neural network-based methods. *Biotechnol. Progr.* 13: 479–487.
- Soons, Z.I.T.A., Voogt, J.A., van Straten, G., van Boxtel, A.J.B., 2006. Constant specific growth rate in fed-batch cultivation of *Bordetella pertussis* using adaptive control. *J. Biotechnol.* 125: 252–268.
- Walsh, G., 2006. Biopharmaceutical benchmarks 2006. *Nat. Biotechnol.* 24: 769–776.
- Yoon, S.K., Kang, W.K., Park, T.H., 1994. Fed-batch operation of recombinant *Escherichia coli* containing Trp promoter with controlled specific growth rate. *Biotechnol. Bioeng.* 43: 995–999.

CHAPTER 5

Selective Expression of the Soluble Product Fraction in *Escherichia coli* Cultures Employed in Recombinant Protein Production Processes

ABSTRACT

Recombinant proteins produced in *Escherichia coli* hosts may appear within the cells' cytoplasm in form of insoluble inclusion bodies (IBs) 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.

This paper has been published in *Applied Microbiology and Biotechnology*:

Gnoth, S., Simutis, R., Lübbert, A., 2010. Selective expression of the soluble product fraction in *Escherichia coli* cultures employed in recombinant protein production processes. *Appl. Microbiol. Biotechnol.* 87: 2047–2058.

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 $T=35\text{--}37\text{ }^{\circ}\text{C}$ 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 CO₂ 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 part 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., Na₂SO₄ 2.0 g L⁻¹, (NH₄)₂SO₄ 2.46 g L⁻¹, NH₄Cl 0.5 g L⁻¹, K₂HPO₄ 14.6 g L⁻¹, NaH₂PO₄×H₂O 3.6 g L⁻¹, (NH₄)²-H-citrate 1.0 g L⁻¹, MgSO₄×7 H₂O 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 toward 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 $\mu=0.5\text{ h}^{-1}$. 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 fermentations applying feed pulse control, no acetate formation was observed.

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_{crit}	0.12
Fermentation Temperature after induction T [$^{\circ}C$]	27, 29, 31, 33, 35	31
Type of Control	Open-loop, Feed-Pulse Control	tcCPR-control

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 pH=7 using a parameter adaptive proportional integral derivative (PID) controller (Gnoth *et al.* 2010). 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 database 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 $^{\circ}C$). 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 (IBs). 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.

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

$$\frac{dV}{dt} = F \quad (2)$$

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 $R=[\mu - \sigma \ \pi_{Sol} \ \pi_{IB}] \cdot X$. 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.

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.

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

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.

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

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

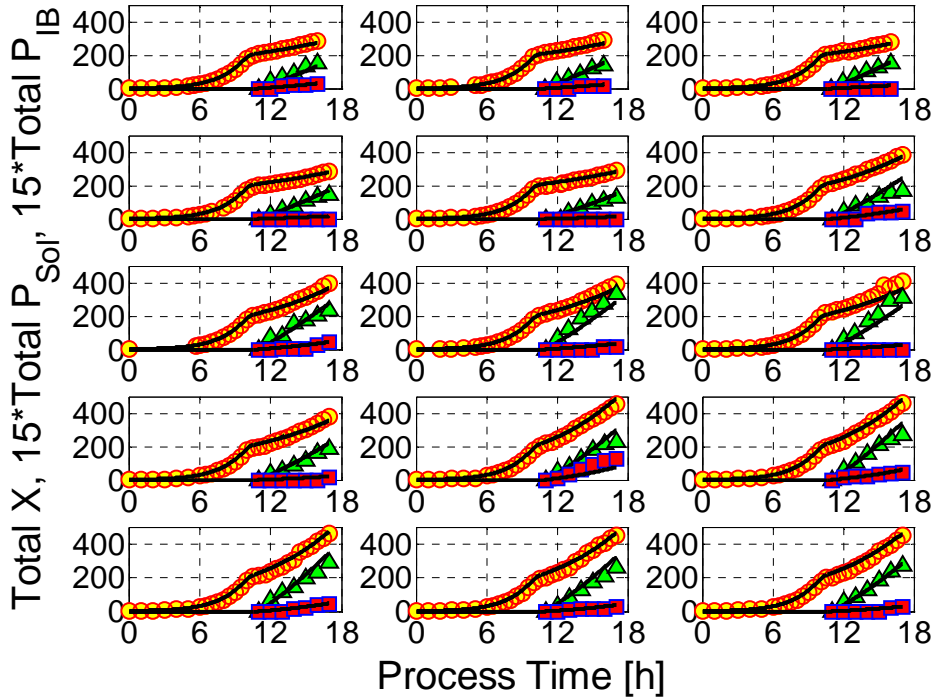


Figure 1. Results of simulations using the conventional process model as described in Eqs. 1–6 (*curves*) compared with the experimental data (*circles*: total biomass, *triangles*: total soluble protein, *squares*: total inclusion body protein).

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 Figure 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 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 Figure 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.

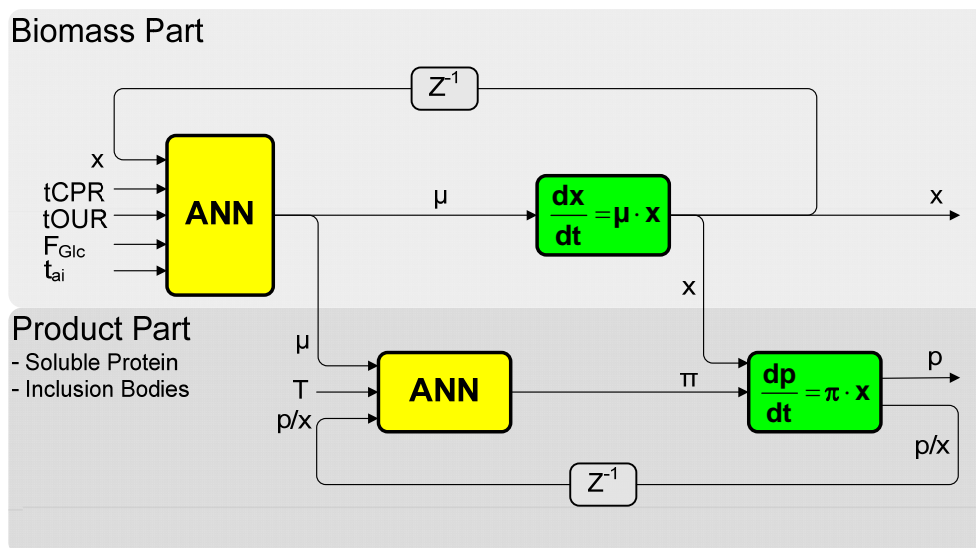


Figure 2. General structure of the hybrid model describing the specific biomass and product formation rates μ and π together with the biomass x and the respective product mass p . At each on-line sampling time, the substrate addition rate and tCPR as well as tOUR are processed in the first ANN (five 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.

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.

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

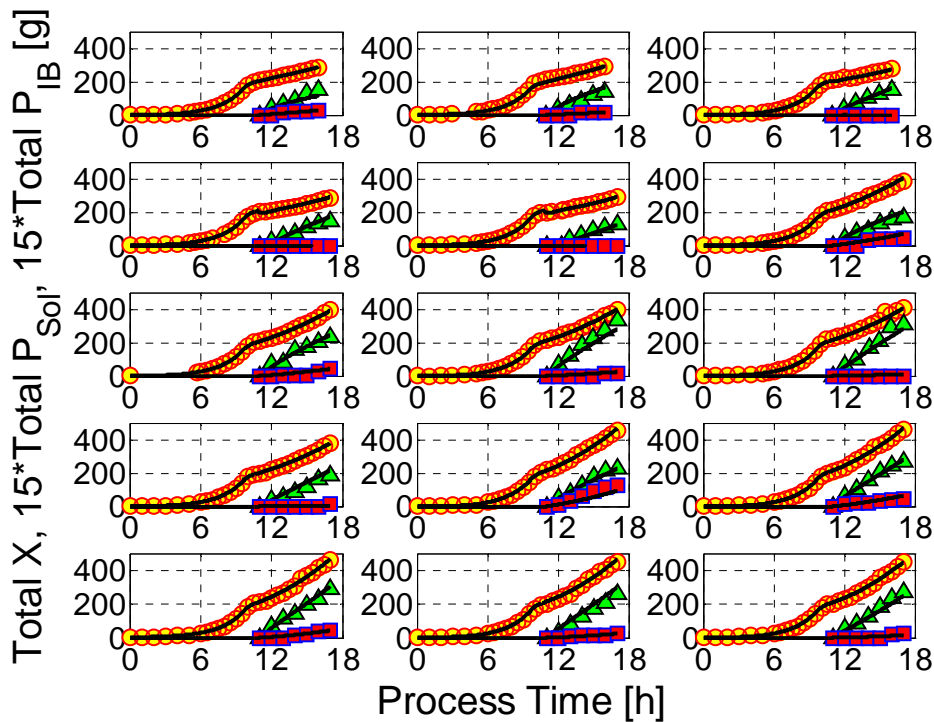


Figure 3. Comparison between modeled (*lines*) and measured values of biomass and protein mass for the experiments performed (*circles*: total biomass, *triangles*: total soluble protein, *squares*: total inclusion body protein). One and the same set of hybrid models was used in all cases.

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 Figure 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 a value of about 0.12 h⁻¹.

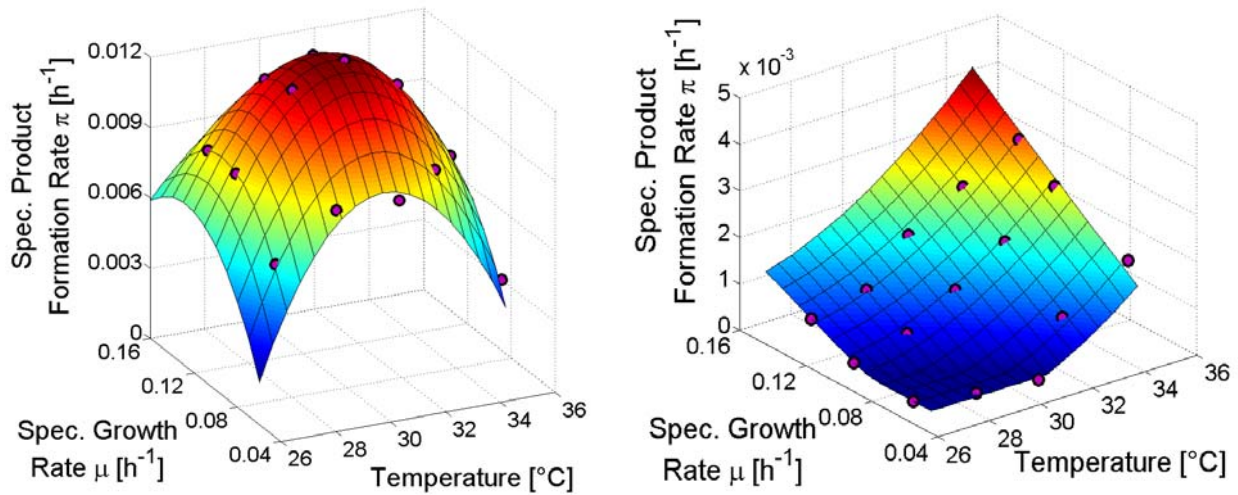


Figure 4. Specific product formation rates of the soluble protein fraction (*left*) and inclusion body fraction (*right*) as a function of the specific growth rate and fermentation temperature in the product formation phase. The marked dots depict fermentation data used for model identification. Compared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

Finally, it should be noted that the inputs of the hybrid model shown in Figure 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 the 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 Figure 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 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,

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

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 Figure 2.

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

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

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 of 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 Figure 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 Figure 5 with respect to process optimization is that the process must be kept below 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

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

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

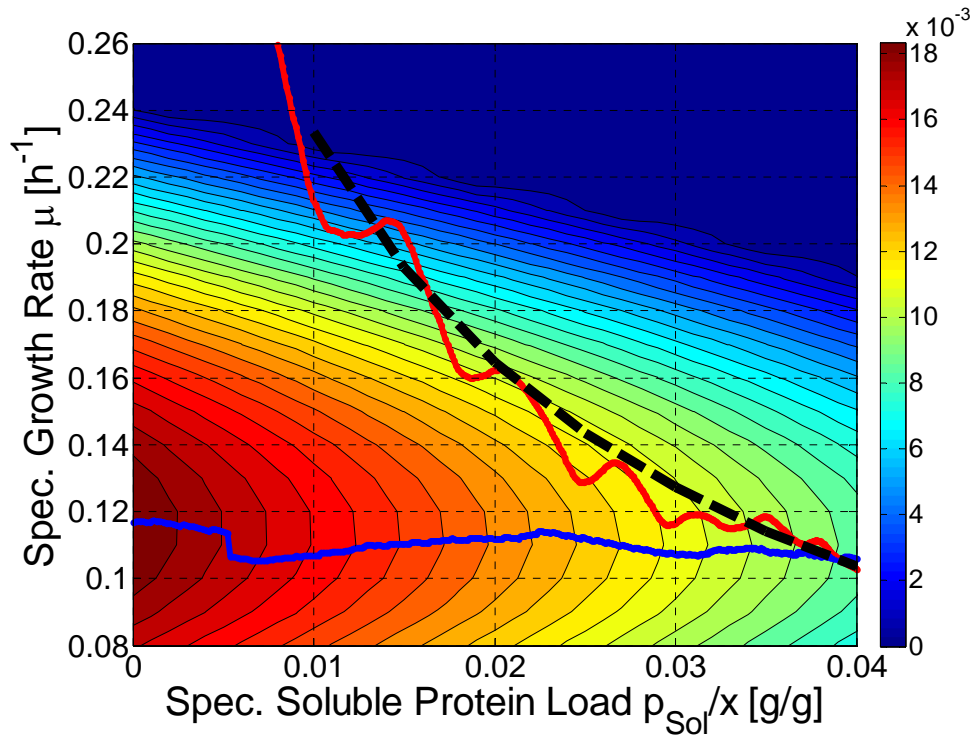


Figure 5. Specific soluble product formation rate hyperspace on plane spanned by the specific growth rate and the specific soluble protein load of the cells at a fixed fermentation temperature. The *red* curve depicts the result of the probing control experiment, saying that maximal biomass growth capacity 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.

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 $27\text{ °C} \leq T \leq 35\text{ °C}$.

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 assumed a linear time function for the specific growth rate profile.

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

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 $T = 31\text{ °C}$, $a_0 = 0.2$, 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 Figure 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 $x = 537.6\text{ g}$.

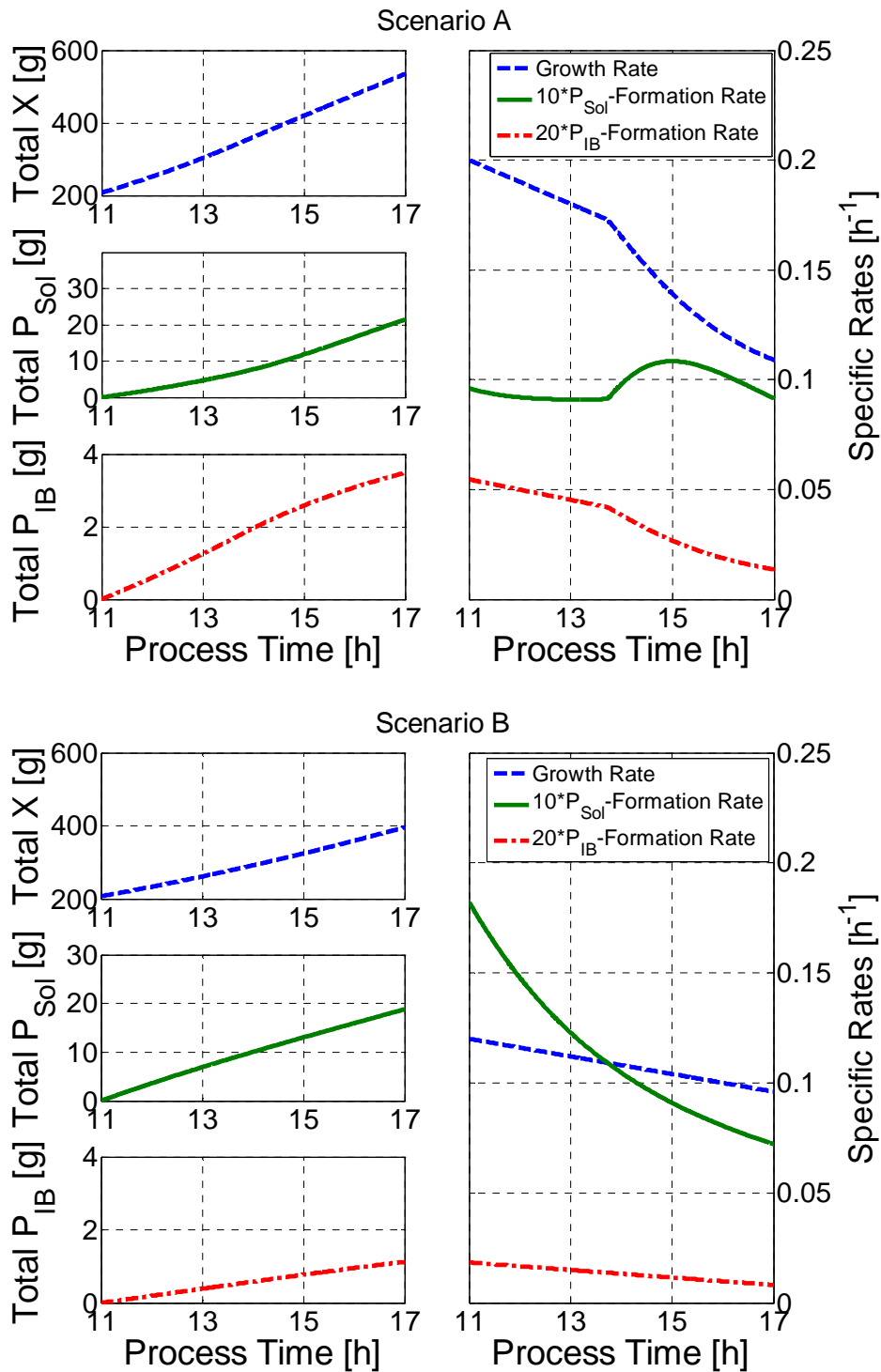


Figure 6. Simulated profiles of the desired soluble protein mass (*solid green line, left side*), the IB-protein mass (*dash-dot red line, left side*) and biomass (*dashed blue line, left side*) as a function of process time together with the corresponding specific growth (μ : *dashed blue line, right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

Alternatively, we considered the situation of a trade-off between soluble protein and inclusion body formation concurrently optimizing the 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 the optimal parameters

were $T=31\text{ }^{\circ}\text{C}$, $a_0=0.12$, and $a_1=-0.004$. The corresponding results are depicted in Figure 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 Figure 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.

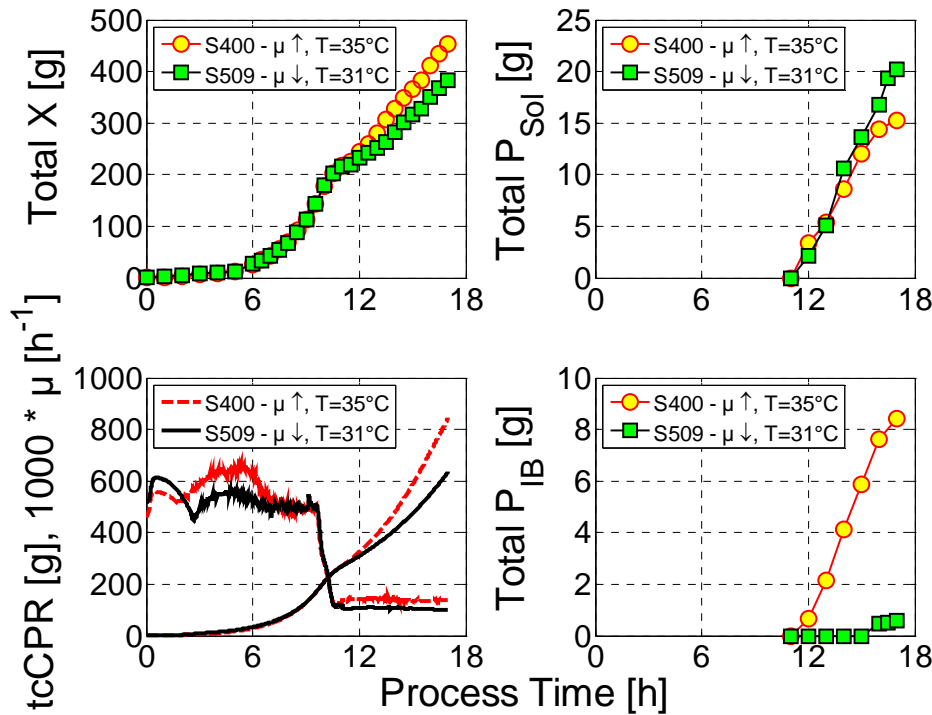


Figure 7. Comparison of the trajectories of a standard fermentation strategy applied at the early beginning of our studies (yellow circles, red line) and after model-supported process optimization (green squares, black line). The improved strategy leads to higher amounts of soluble protein and, simultaneously,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

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 $35\text{ }^{\circ}\text{C}$ and a constant specific growth rate of $\mu_{\text{set}}=0.16\text{ h}^{-1}$ during the protein formation phase were chosen.

It becomes immediately clear from Figure 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 Figure 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.

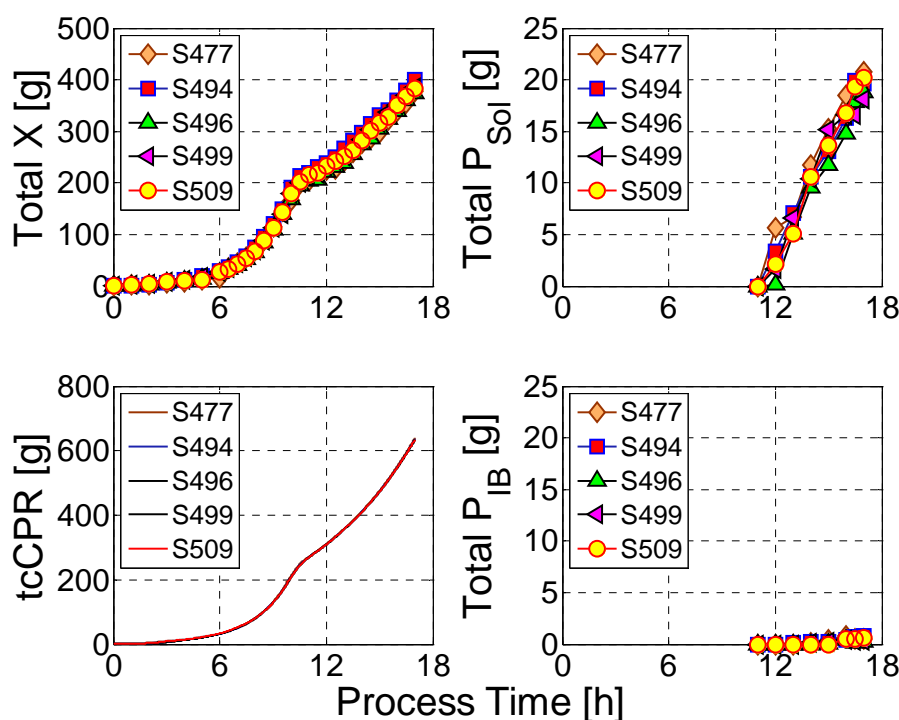


Figure 8. Results of the five experiments performed to validate the resulting operational procedure that aims in maximizing the formation of soluble product fraction (*upper right*) with simultaneous reduction of the inclusion body fraction (*lower right*). The high degree of batch-to-batch reproducibility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

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 the cultivation with optimal cultivation parameter settings (as shown in Figure 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 Figure 8.

It can be seen that the total CO_2 -mass profiles measured in the different cultivations (depicted in lower left part of Figure 8) almost completely overlap. Deviations from the set point profile are less than 1 %. Thus, the 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.

Acknowledgement

The work was performed as a project within the “Excellence Initiative” of the German State Sachsen-Anhalt. We gratefully thank for the financial support. The *E. coli* strain used in the experiments was provided by Probiodrug, Halle, Germany.

NOMENCLATURE

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	3.8343e ⁴⁸	[-]
k ₃	Constant for activation of soluble protein expression	2.4099e ¹⁶	[-]
k ₄	Constant for inactivation of soluble protein expression	3.0156e ⁴⁹	[-]
k ₅	Constant for activation of inclusion body protein expression	7.1563e ¹⁴	[-]
k ₆	Constant for inactivation of inclusion body protein expression	1.3982e ⁴⁶	[-]
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-σ}	Constant for inhibition of substrate uptake caused by higher cell densities	36.8123	[g/L]
K _{X-μ}	Constant for inhibition of biomass growth caused by higher cell densities	121.8669	[g/L]
K _{μ1}	Constant for inactivation of soluble protein expression at higher growth rates	0.25	[-]
K _{μ2}	Constant for inactivation 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 of 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, J.P., 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: 590–598.
- Baneyx, F., Palumbo, J.L., 2003. Improving heterologous protein folding via molecular chaperone and foldase co-expression. *Meth. Mol. Biol.* 205: 171–197.
- Bentley, W.E., Noushin, M., Andersen, D.C., Davis, R.H., Kompala, D.S., 1990. Plasmid-encoded protein: the principal factor in the “metabolic burden” associated with recombinant bacteria. *Biotechnol. Bioeng.* 35: 668–681.
- Berwal, R., Gopalan, N., Chandel, K., Prasad, G.B.K.S., Prakash, S., 2008. *Plasmodium falciparum*: enhanced soluble expression, purification and biochemical characterization of lactate dehydrogenase. *Exp. Parasitol.* 120: 135–141.
- Bessette, P.H., 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, S.K., Dubey, A.K., 1995. Metabolic burden as reflected by maintenance coefficient of recombinant *Escherichia coli* overexpressing target gene. *Biotechnol. Lett.* 17: 1155–1160.
- Chalmers, J.J., Kim, E., Telford, J.N., Wong, E.Y., Tacon, W.C., Shuler, M.L., Wilson, D.B., 1990. Effects of temperature on *Escherichia coli* overproducing β -lactamase or epidermal growth factor. *Appl. Environ. Microbiol.* 56: 104–111.
- Davis, G.D., Elisee, C., Newham, D.M., Harrison, R.G., 1999. New fusion protein systems designed to give soluble expression in *E. coli*. *Biotechnol. Bioeng.* 65: 382–388.
- Esener, A.A., Roels, J.A., Kossen, N.W.F., 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, M.L., 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: 41–46.
- Gnoth, S., Kuprijanov, A., Simutis, R., Lübbert, A., 2010. Simple adaptive pH control in bioreactors using gain-scheduling methods. *Appl. Microbiol. Biotechnol.* 85: 955–964.
- Heo, M.A., Kim, S.H., Kim, S.Y., Kim, Y.J., Chung, J., Oh, M.K., Lee, S.G., 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, J.S.U., 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: 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: 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, E.B., 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, Y.S., Cha, H.J., 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 α -glucosidase PI in *Escherichia coli* by induction and growth conditions. *Mol. Gen. Genet.* 216: 149–155.
- Kuprijanov, A., Schaeppe, 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, H.H., 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, G.W., Strohl, W.R., 1990. Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Appl. Environ. Microbiol.* 56: 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, J.A., Laird, W., Bjorn, M.J., 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, A.J., O'Byrne, C., McLaggan, D., Booth, I.R., 2002. Inhibition of *Escherichia coli* growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. *Microbiology* 148: 2215–2222.

Roels, J.A., 1983. Energetics and kinetics in biotechnology. *Elsevier Biomedical Press*, Amsterdam.

Schein, C.H., 1989. Production of soluble recombinant proteins in bacteria. *Bio/Technology* 7: 1141–1148.

Schügerl, K., Bellgardt, K.H., 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: 10–20.

Shin, C.S., Hong, M.S., Kim, D.Y., Shin, H.C., 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, H.P., Mortensen, K.K., 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, J.M., 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, J.G., Baneyx, F., 1996. Protein misfolding and inclusion body formation in recombinant *Escherichia coli* cells overexpressing heat-shock proteins. *J. Biol. Chem.* 271: 11141–11147.

Vasina, J.A., 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, M.J., Doherty, D.H., Best, E.A., Olins, P.O., 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.

CHAPTER 6

Fermentation Process Supervision and Strategies for Fail-Safe Operation: A Practical Approach

ABSTRACT

In industrial production environment, cultivation processes forming recombinant proteins run along predefined trajectories. Herein, feedback control is the best way to keep the processes on track. However, feedback controllers require accurate on-line values of the controlled variables. In order to find out whether the corresponding measurement signals are correct, process supervision is required. Here we compare some realistic approaches to such supervision at one and the same fermentation process from a practical point of view. As the accompanying example, a fermentation process expressing a pharmaceutically relevant recombinant protein was taken. Classical univariate techniques were found to be still extremely valuable tools, as most deviations from the desired trajectories can first be noticed within these charts. Simple balancing techniques form an important basis for multivariate process supervision, as they characterize the measurement data with respect to consistency. Since the number of measured variables is limited in fermentation technology, univariate fault detection procedures can be employed throughout, thus, if process failures occur, errors can easily be identified by analyzing each measurement variable separately. Thus, data reduction, the main advantage of linear principal component analysis does not lead to a significant benefit with respect to fault detection. Autoassociative artificial neural networks (aANN), the nonlinear extension to PCA, are powerful estimators for values of variables when the corresponding sensors fail. Optimal for process supervision are full dynamic process models that can be used to link data from different sensors and that can be further applied in model aided controllers. After a process failure has been appeared and the corresponding responsible variables have been identified, automated fail-safe techniques are proposed to finish the process properly without violation of its specification limits.

This paper has been accepted for publication in *Engineering in Life Sciences*:

Gnoth, S., Simutis, R., Lübbert, A., 2010. Fermentation process supervision and strategies for fail-safe operation: a practical approach. *Eng. Life Sci.*, accepted for publication.

INTRODUCTION

Highly reproducible operation in fermentation processes is of paramount importance for the production of recombinant therapeutic proteins and thus a major demand expressed by the regulatory authorities, e.g., FDA, EMEA, in order to make sure that products stay within their specification limits (FDA 2004, De Palma 2004, EMEA 2006). Reproducibility can effectively be supported by closed-loop process control (Jenzsch *et al.* 2006). A major prerequisite for correct process control is that the measurement devices employed are working correctly.

In real fermentation plants, however, several failures may appear. Hence, it is straightforward to supervise the fermentation process in such a way that errors are detected as early as possible, in order to have time enough to start corrective measures preventing the process from leaving its specification domain. For that purpose, process monitoring, supervision and control systems are installed at bioreactors to register the on-line measurement signals, display them properly and analyze the on-line information. The errors found during the process are classified into different categories (Hocalar *et al.* 2006).

- Random measurement noise with zero mean and in agreement with measurement accuracy
- Disturbances/abnormalities – sudden gross errors in the measurement data including the complete failure of the sensor
- Distortions/drifts – systematic measurement errors due to, e.g., wrong calibrated sensors, or systematic movements of the sensors signal during the process
- Decreased measurement accuracy

In situations where measured signals start deviating systematically from their specified trajectories, the first task is to find out whether or not measurement signals are correct. This question must be solved by checking them for consistency with simultaneously measured signals. Models interrelating the signal values are then required. The different supervision methods available are essentially distinguished by the level of sophistication of the models employed. When a measurement does not deliver a correct value, it should be replaced by a sufficiently correct estimate. Again, it is straightforward to determine the estimate by means of a model from other simultaneously measured values.

In cases where the measurement variables are consistent but the process is progressing atypical, e.g., off-line values of metabolites are out of their normal ranges, the problem is to identify the reason of the fault in the process. This is not considered in this paper. Here, we focus on supervising the on-line measurement signals in fermentation processes that are used to keep the process under control.

In the project we report about, we investigated several techniques that were proposed for and are in use in other sectors of production industry, mainly in the chemical and petrochemical industries, with respect to their applicability in fermentation processes. Methods that are of high importance in these processes might not be as valuable as in fermentation plants, since conditions are much different in many respects. An important difference to chemical or petrochemical production reactors is the number of on-line measurement devices being restricted as far as possible in bioreactors to reduce the contamination risk and to avoid expensive device validation procedures (Royce 1993, Junker *et al.* 2006).

In order to get a realistic picture, we compared all techniques investigated at one and the same fermentation process for production of a recombinant therapeutic protein. For that purpose, we took an example where the desired product appears in its active form as well as in form of inclusion bodies within genetically modified *E. coli* bacteria.

MATERIAL AND METHODS

Escherichia coli BL21(DE3) pET28a 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 another one in form of 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 with defined mineral salt medium, i.e., Na_2SO_4 2 g L⁻¹, $(\text{NH}_4)_2\text{SO}_4$ 2.46 g L⁻¹, NH_4Cl 0.5 g L⁻¹, K_2HPO_4 14.6 g L⁻¹, $\text{NaH}_2\text{PO}_4 \times \text{H}_2\text{O}$ 3.6 g L⁻¹, $(\text{NH}_4)_2\text{-H-Citrate}$ 1.0 g L⁻¹, $\text{MgSO}_4 \times 7 \text{H}_2\text{O}$ 1.2 g L⁻¹, trace element solution 2 ml L⁻¹, thiamine 0.1 g L⁻¹, and kanamycin 0.05 g L⁻¹. All cultivation experiments were entirely run in the fed-batch mode, i.e., the medium did not contain glucose at inoculation time. At the early stage of fermentation an exponential feeding was performed by open-loop control. After seven hours of fermentation, the process operation was switched to control of total mass of generated carbon dioxide tcCPR (Jenzsch *et al.* 2007). In this way, a specific growth rate of $\mu=0.5 \text{ h}^{-1}$ in the biomass growth phase and an optimal growth rate for soluble recombinant protein production of $\mu=0.12 \text{ h}^{-1}$ after induction was achieved. Within all fermentations the temperature was kept at $T=35 \text{ }^\circ\text{C}$ and subsequently lowered to $31 \text{ }^\circ\text{C}$ prior to induction. Both parameters, the specific growth rate as well as the fermentation temperature, were controlled to optimal values after induction as was found by Gnoth *et al.* (2010a).

Because the initial 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 toward 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.

Respiratory data 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 sensor (Maihak Unor 610). An Ingold pO_2 probe (Mettler Toledo) was used for monitoring and control of dissolved oxygen to a value of 25 %sat. pH was measured with an Ingold pH probe (Mettler Toledo) and maintained at pH 7. The integral base consumption during pH control was monitored with another balance. Control of pO_2 as well as pH was performed using a parameter adaptive gain-scheduling approach (Kuprijanov *et al.* 2009, Gnoth *et al.* 2010b). Temperature was measured with a PT-100 sensor and controlled with a standard PID controller. In addition, PT-100 sensors to measure the temperature of the coolant inflow and outflow were installed at the entry and exit of the cooling jacket. The entire process was operated under the control of a Siemens SIMATIC PCS7 control system (Kuprijanov *et al.* 2008, 2010).

Off-line measurements were performed with time increments of half an hour. Biomass concentration was estimated from an OD_{600} measurement performed with a spectral photometer (Shimadzu UV-2102PC). Additionally, some dry weight measurements were conducted, in order to test the correlation between biomass dry weight and the OD_{600} values. Glucose concentration was determined enzymatically with a YSI 2700 Select Bioanalyzer (Yellow Springs Inc.). Finally, target protein was quantified by gel densitometry after cell disruption, separation of the soluble fraction, solubilization of IB's and subsequent SDS-PAGE.

In developing a reference profile for the investigations of the various process supervision techniques, a couple of experiments were operated under the same controlled conditions similar to production runs in industry. From these fermentations a mean profile of the total mass of

carbon dioxide formed was determined that was used as the setpoint profile for the feedback control. To give an impression of the data quality, the data sets are depicted in Figure 1.

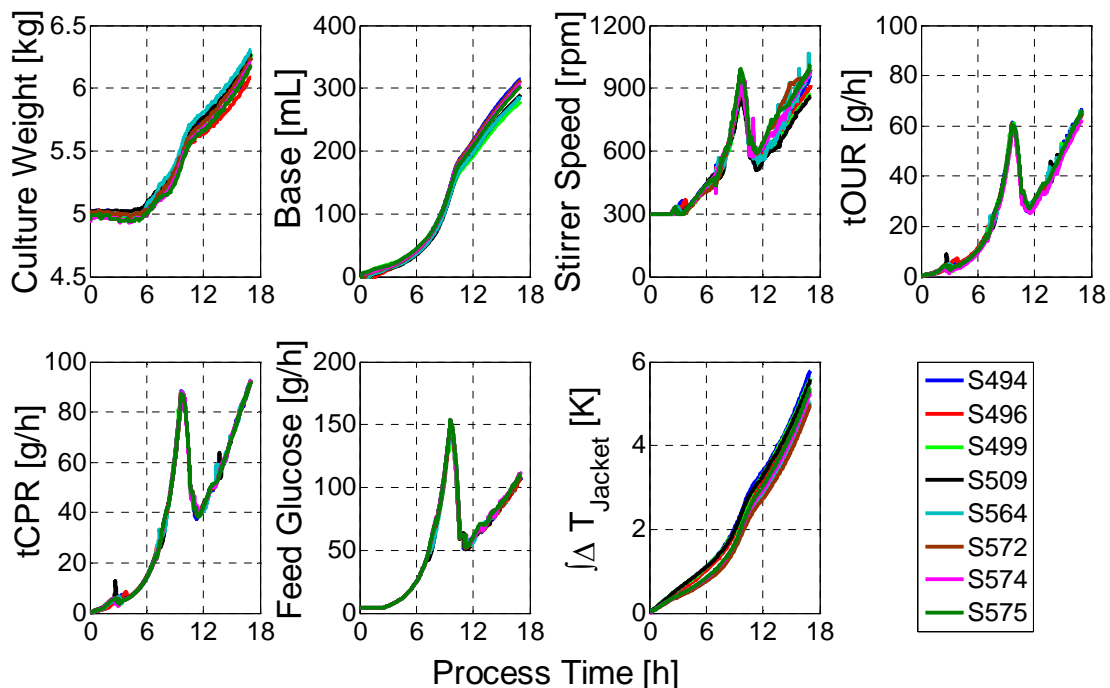


Figure 1. Data set originating from a fermentation system used in recombinant protein production with *E. coli* bacteria. All fermentation runs were successfully controlled to a predefined profile of the total carbon dioxide mass [g] produced by the cells (Jenzsch *et al.* 2007). The data set was used to identify all models discussed in the subsequent parts of this paper. The numbers in the legend are study identifiers of the fermentations in our laboratory database.

In order to test the various monitoring techniques described later, additional experiments were performed where failures were induced during the fermentation.

RESULTS AND DISCUSSION

Univariate Supervision Methods

Control Charts

In continuous processes it is straightforward to make use of control charts where the measurement values of single variables are plotted against time (e.g., Stoumbos *et al.* 2003). If the values of variables have to be kept constant, for instance the pH-value, deviations from the desired values signal can simply be analyzed. In batch or fed-batch processes that are predominating in fermentation plants, things are usually more complicated as the desired values of most state variables, e.g., the biomass, vary with time. A straightforward approach is to set two-sided limits around the desired value within which the fluctuations must stay in an undisturbed process. When unusual deviations appear, operators will be alarmed, and in situations of stronger variations, corrective actions become indispensable. The identification of an unusual deviation is performed after statistical analysis of the data from previous successfully finished cultivations. Generally, two limits are used, the alarm and the action limit. The first only signals a significant deviation from the desired profile, the second says that an immediate response by the operator is required.

In practice, the first problem is to determine the time profiles of these limits. A first guess is derived from the paths of successfully performed fermentations, such as those shown in Figure 1, by computing the alarm limits from the standard deviation of fluctuations around the mean observed in fermentations under normal operating conditions. Since those profiles, e.g., 2-sigma curves, will probably be noisy, they must be smoothed appropriately. In this context, we applied a normal running mean filter, but generally several alternative techniques exist, for instance Savitzky-Golay smoothing (Savitzky and Golay 1964). The alarm and action limits must be separately tuned for each variable under consideration based on experiences with the current process or with similar ones. For simplicity, we took smoothed 2-sigma distances for the alarm and 3-sigma distances for the action limits (Shewart 1931, Fortuna *et al.* 2007). An example is given in Figure 2.

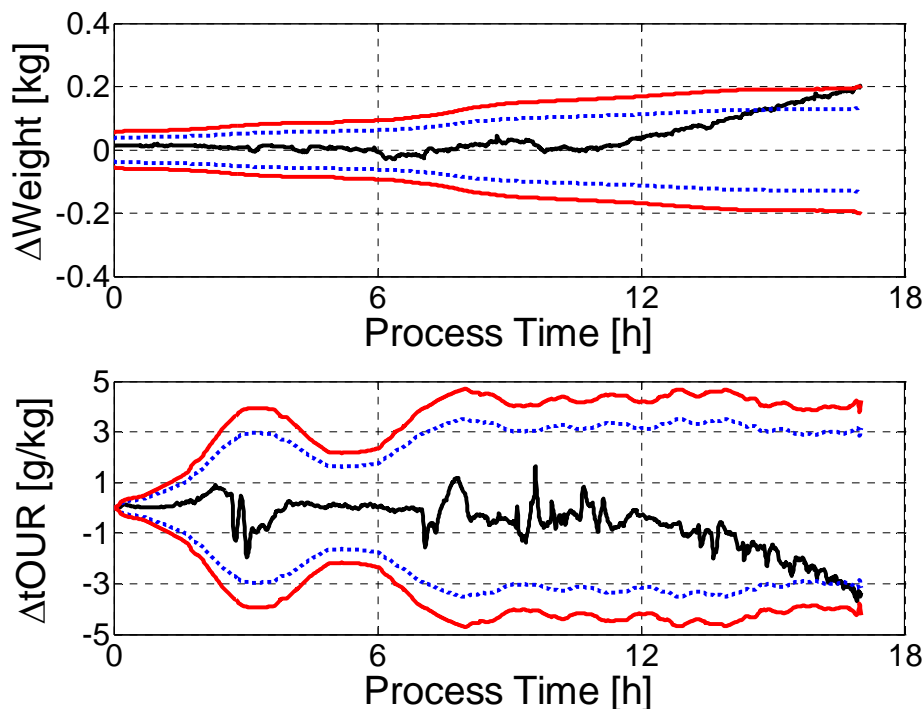


Figure 2. Typical charts used to supervise on-line measured variables in a process control system. They depict deviations from the desired reference value. *Upper part:* for the culture weight, *lower part:* for the total oxygen uptake rate. The *solid black* line is the deviation from the currently measured signal, the *blue dashed* lines are the alarm limits and the *solid red* lines are the action limits.

During the cultivation experiments the on-line measurement values were supervised by a SIMATIC PCS7 control system that alarmed the operator whenever the respective limit was reached, or started an appropriate action procedure in cases where the action limit was violated.

An important point is the selection of the variables to be monitored. Measurements significantly influencing the cell's behavior with respect to biomass growth and product formation are most important. The main physiological state variable of the cells, the specific growth rate, sensitively depends on quantities such as pH, $p\text{O}_2$, temperature, etc.. These are measured on-line, but they are separately controlled to predefined, constant values. If the controllers work correctly, the signals of these variables themselves do not deliver significant information about process faults. However, the action variables are varied and therefore clearly depict the expenses necessary to keep the controlled quantities on track. For pH control this is the amount of base (and, if necessary, acid) needed to hold the desired pH value. The dynamics of temperature control are mirrored by the energy required for cooling, or simply by the integral of the temperature difference between coolant in- and outflow at the cooling jacket, provided that the coolant flow

rate is constant. In this way, a small number of - at the first glance unconventional - on-line measurement variables are used for process supervision together with others that are more directly related to the specific growth rate, e.g., the oxygen uptake rate OUR and the specific carbon dioxide production rate CPR, which both are derived from off-gas analytics.

Trend and Noise Analysis

Control charts can be further exploited by looking whether the signal noise shows an unexpected behavior. For instance, when the noise level on the signal changes unexpectedly, it must be argued that there is something wrong.

The signals are usually allowed to freely move within the alarm limits around the setpoint. During normal process operation, these distortions should randomly fluctuate around the setpoint. Otherwise, a systematic error must be assumed. Such trends or drifts are best recognized by integrating the deviations between measured and reference values. When the integral significantly deviates from zero, a trend must be assumed. In this way, trends can often be detected long before the process reaches the alarm limits. A typical example is shown in Figure 2, here for weight or oxygen uptake rate at times larger than 12 h a trend was detected. Besides this, methods to detect and analyze cyclical trends in the measurement data are summarized in the general topic of time series analysis, which encompass, e.g., auto- and cross correlation or Fourier transformation procedures (Brereton 2005).

Multivariate Process Supervision Procedures

When a key process variable tends to drift away from its desired reference trajectory, the objective is to find out whether or not all process components work correctly. This question must be answered in steps. Looking for consistency of process data is a first measure. It requires a model that incorporates the different variables. The various methods are distinguished by the models employed.

Balances

Mass Balance

A mass balance around the entire culture is a simple and straightforward basic model used to check measurement data for consistency. The mass balance sums up changes in the culture mass W due to incoming and outgoing flow rates. The biggest flow rates increasing W are (i) the substrate feed rate F_S , (ii) the rate F_B of base addition during pH-control, (iii) the total oxygen transfer rate $tOTR$ and (iv) the rate F_{AF} at which the antifoam agent is added to the broth. W is diminished mainly by (i) the rate $tCTR$ at which CO_2 is leaving the culture through the vent line, but also (ii) by the rate of water evaporation F_{Evap} and (iii) F_{Sample} the rate at which culture is removed during the sampling procedure for off-line measurements. The sum of all these contributions should be equal to the measured culture weight W_m .

$$W_m(t) = W_0 + \int_0^t (F_S + F_B + F_{AF} + 0.001 \cdot (tOTR - tCTR) - F_{Evap} - F_{Sample}) \cdot dt \quad (1)$$

Of course, all the feeds F_i are functions of time t . Typical plots of the measured mass W_m together with the right hand side of the balance equation for normal operating fermentations are shown in Figure 3.

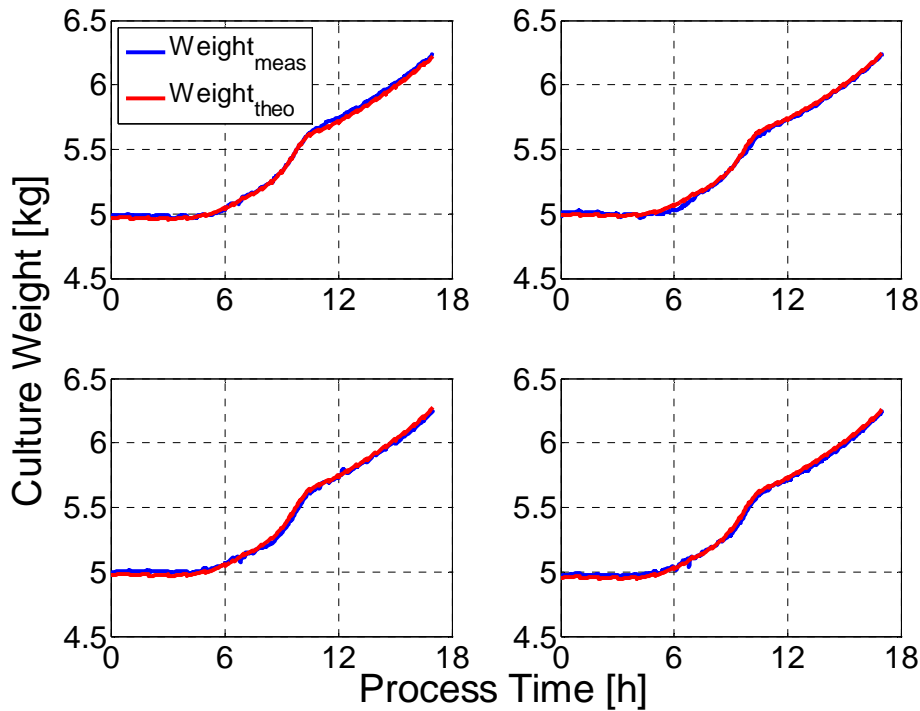
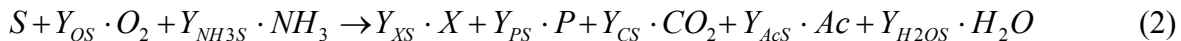


Figure 3. Examples of mass balances computed on-line for 4 different fermentation runs (S481, S572, S598, S600). As can be seen, measured (*blue line*) and balanced (*red line*) culture weight fall together and the corresponding mass balance is considered closed.

This representation of the mass balance can be used on-line during the running process to supervise the process. I.e., the sum of the contributions can directly be displayed with the culture weight W_m measured by the balance. Upon statistically significant deviations between both signals an alarm is triggered.

Carbon Balance

Simultaneously, the C-balance can be used for on-line process supervision (Han *et al.* 2003, Jobé *et al.* 2003). For an aerobic *E. coli* process, it is derived from the gross reaction equation



where S is the substrate, here glucose, NH_3 the ammonia-solution used for pH control, X the biomass, P the target protein, Ac the overflow metabolites, H_2O , O_2 and CO_2 water, oxygen and carbon dioxide, respectively.

Here, the stoichiometric coefficients are written in terms of yield coefficients with respect to C-moles of substrate. In the C-Balance the yield coefficients for ammonia, oxygen and water do not make a contribution. With the elemental composition of the species involved, the change n_i in the amount of C-atoms in species i during the reaction per unit time can be related to each other. It is convenient to resolve this relation for the biomass term. Then the C-balance reads

$$n_{CX} = n_{CF} + n_{CS_0} + n_{CX_0} - (n_{CCO_2} + n_{CP} + n_{CAc} + n_{CSample} + n_{CS}) \quad (3)$$

It is known that carbon dioxide dissolves to a considerable extent in the liquid medium and further is in equilibrium with bicarbonate. However, if the pH is strictly controlled to a constant value (as it is normally done in fermentations), the CO_2 -bicarbonate system is then in steady state and can be neglected (Royce 1992, Spérandio 1997, Gnoth *et al.* 2010b).

Often, the impact of product formation on the overall carbon balance is negligible. The equation is further simplified, if the process is assumed to be operated in substrate limited fed-batch mode. Then, the amount of substrate accumulating in the culture broth as well as the formation of overflow metabolites can be neglected. The amount of carbon being added by the feed as well as the overall amount of carbon from CO_2 production are derived from on-line measured variables, i.e., the substrate feeding rate and the CO_2 fraction in the exhaust gas. With the measured culture mass W_m and the molecular weights of the species, the relationship can be represented in terms of concentrations, which are more conveniently compared with the off-line measured biomass concentrations as shown in Figure 4. The off-line biomass values were obtained from OD_{600} values via correlation curve with time delays of 15 min after sampling.

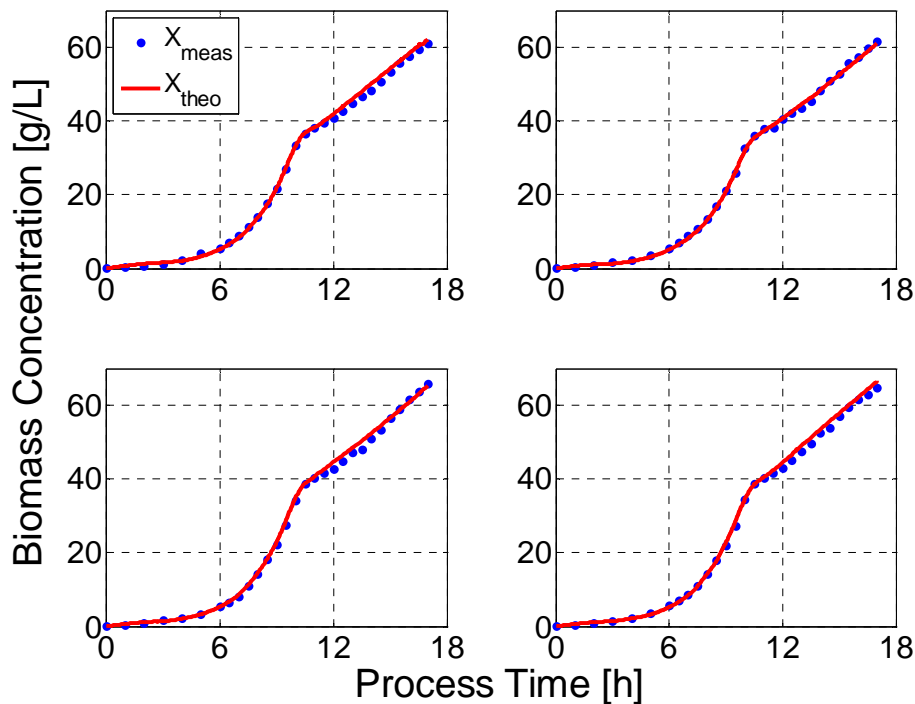


Figure 4. Results of carbon balances of 4 fermentation runs. As explained in the text, the C-containing compounds of the culture are balanced. In this way, biomass concentration is estimated on-line (red line) and is compared with those measured off-line (symbols).

Degree of Reduction Balance

Further, the degree of reduction balance (e.g., Shuler and Kargi 2002, Chattaway *et al.* 1992) can be exploited on-line. This balance is also based on the gross reaction equation just as the C-balance. It makes use of the O_2 -contribution instead of the CO_2 -contribution to the overall reaction. Again, the result can be displayed in terms of biomass concentrations to more conveniently show plots to which the process operators are familiar.

All three balances were evaluated on-line during the fermentation process and were incorporated into our Siemens PCS7 process control system. They are known to be the first consistency check to be made in order to make sure that the measurement equipment works correctly (e.g., Nielsen and Villadsen 1994). Since these balances make use of different measurement variables, best results for identification of process errors are obtained if all balances are combined. Then, in case of an error in the substrate feed rate, for instance, all three balances are not closed, while a faulty CO_2 measurement is signaled by a non-closed carbon balance but probably not in the degree of reduction balance. The other way around, a closed carbon balance and an open degree of

reduction balance reveal a problem with the O_2 measurement in the vent line. More complex balancing methods, i.e., linear constraint and data reconciliation methods, have been developed by Wang and Stephanopoulos (1983) as well as Heijden (1994a,b,c). At the example of industrial fed-batch yeast fermentations Hocalar *et al.* (2006) applied elemental balances and an additional heat balance. The redundant information was utilized for error diagnosis, fault detection and state estimation. In principal, the first action is to set up the elemental component matrix and the conversion rate vector on the basis of the stoichiometry of the current fermentation process and the number of elemental balances, i.e., the constraints of the system. Then, the degree of redundancy is calculated. If the degree of redundancy is > 0 , the system can be used to i) detect gross measurement errors during fermentation, ii) estimate nonmeasured conversion rates and further iii) improve the accuracy of the measured conversion rates by reconciliation methods (Jobé *et al.* 2003). Measurement errors are detected after the elemental component matrix and the conversion rate vector have been divided into a measurable and a non-measurable part. Subsequently, the redundancy matrix, and with it the reduced form specifying the number of independent equations, can be calculated. In further steps the residual vector and its variance-covariance matrix are determined and finally a statistical test introduced by Reilly and Carpani (1963) is made. The test function h is chi-square distributed when the raw measurements are uncorrelated. If h exceeds the given confidence level, the serial elimination method by Ripps (1965) is a tool to identify the source of an error.

Black-box Models

Where there is not enough physically based process knowledge, more or less simple black-box models are practical alternatives. There are two groups of such models, statistical and nonlinear parametric ones. Both are data-driven *per se*, so that little process knowledge is required with the exception of the choice of dependent and independent variables to be considered.

Statistical Models

One of the most well known statistical modeling and analyzing method for fault detection is principal component analysis (PCA), including their multiway versions (Wold *et al.* 1987, Nomikos and MacGregor 1994), which essentially is a multidimensional linear mapping between process variables. The measured data from J variables forms a cloud in an J -dimensional data space. PCA performs a coordinate transformation in such a way that in the new orthogonal coordinate system the first coordinate points to the direction of the largest extension of the cloud. The second coordinate indicates the direction where the variance again is next highest, and so on with the third and further up to the J -th coordinate. In the new linear vector space the coordinates (transformed variables) unfortunately do no longer have a physical meaning, but on the other hand have the advantage that they are independent of each other and are classified with respect to the variability, their signal amplitudes depict during the cultivation.

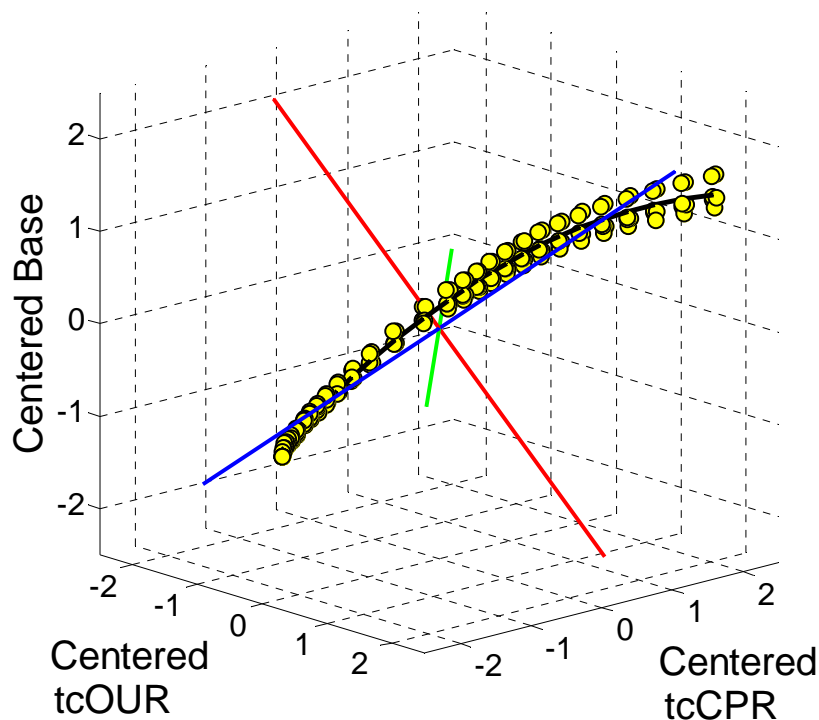


Figure 5. Classical representation of fermentation data (*symbols*) in a three dimensional data space together with the 3 principal components (*lines*). While the $tcOUR=f(tcCPR)$ is strictly linear, the $Base=f(tcOUR, tcCTR)$ is nonlinear. Because of that, data reduction using linear PCA is problematic, since significant information is lost.

A simple basic example with real fermentation data is given in Figure 5 where the centered base consumption is plotted above the plane spanned by the centered cumulative oxygen uptake rate and the centered cumulative carbon dioxide production rate. As the name suggests, PCA deals with principal components and neglects further ones that do not influence the process significantly. This data space reduction property is essential to the success of PCA.

Moreover, the classification capability is the very advantage of the method, as feature variables displaying no significant change during the cultivation do not necessarily need to be considered. The property is particularly important, if very many variables are to be measured at the process under consideration. In fermentation processes for recombinant protein production this feature is not really used, because the number of on-line measured variables is usually very small so that a data reduction cannot bring a very big advantage. Additionally, as shown in Figure 5, in nonlinear systems a data reduction often cannot be performed, since all three components are required in the transformed system to describe the "data cloud".

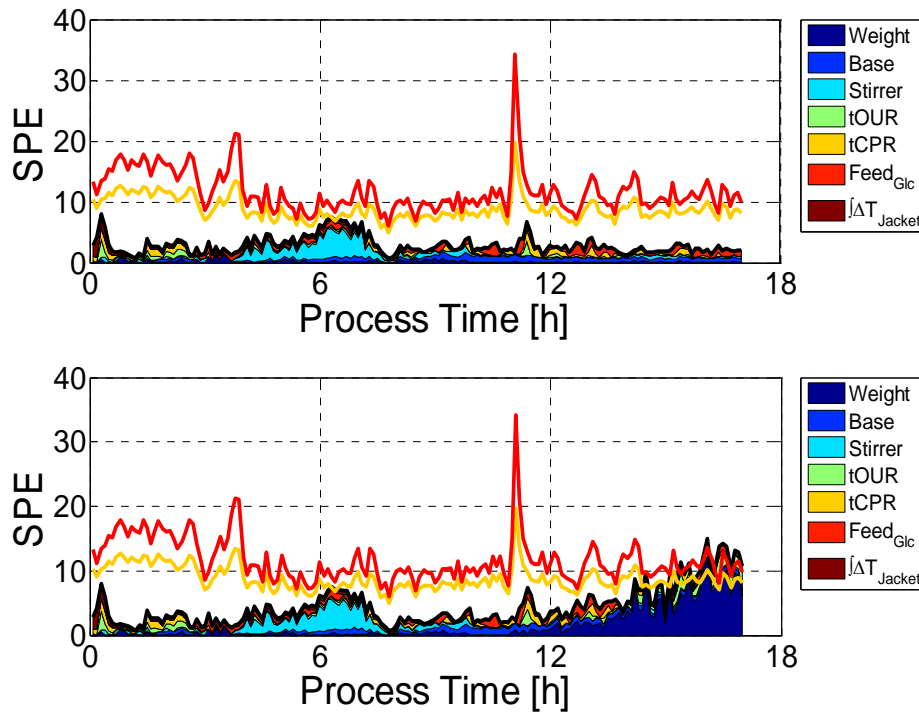


Figure 6. Monitoring of a well performing (*upper graph*) and a faulty (*lower graph*) cultivation process based on the multiway principal component analysis of seven variables. For comparability, the same process faults as in Figure 2 were applied, i.e., error in culture weight and tOUR. In order to establish the PCA model, 16 reference batches performed under normal operating conditions were used. The calculated squared prediction error SPE serves as an indicator for the deviation between actual fermentation and the established process model. Since PCA performs a linear transformation, the SPE can be divided into the single variable contributions. SPE values exceeding the alarm (*solid orange line*) and action limits (*solid red line*) indicate an abnormal process behavior.

The second feature of PCA is that it can easily detect faults, if the correlation structure among process variables is disturbed. In that case, the value appears as an outlier with respect to the data cloud made up by the correct values in the J-dimensional data space.

For on-line fault detection of a fermentation process, the measurement data has to be unfolded in a special way. First of all, data from I reference fermentations with J variables at K time steps is unfolded so that the new two-way matrix has the dimension $X = I \times (J \cdot K)$. Each column in X corresponds now to one variable at a certain time instant. Obviously, the dimension of X is rather high and the columns, i.e., the variables, are not independent of each other. The application of PCA reduces the dimension of X, eliminates redundant information and extracts the principal components (PCs). Often only three to four PCs are sufficient to describe the major part of the system's variability. After PCA has been applied to the reference data set, a unique correlation structure of the measurement data is identified and PCA model is set up. The measurement matrix X can then be divided into the loadings P, the scores T and a residual error E of the model in the following way.

$$X = TP^T + E \quad (4)$$

The established PCA model then can be used to monitor an unknown fermentation process. An advantage of the linear PCA method as compared to more complex nonlinear methods is its ease of identification of variables being responsible for the outlier point. For that purpose, one can compute the squared prediction error (SPE), which can be further divided into its components (Gregersen and Jørgensen 1999, Westerhuis *et al.* 2000), as shown in Figure 6.

$$SPE_k = \sum_{r=(k-1)J+1}^{k \cdot J} e_r^T e_r \quad (5)$$

Where e_r is the r^{th} column of the error matrix $E = X_{\text{new}} - t_{\text{new}} P^T$. For a single batch it is a scalar and the SPE_k becomes the sum of quadratic residues between the values of the batch to be monitored and the values predicted by the PCA model at time point k . The SPE can be approximated by a χ^2 distribution, where m_k and v_k are the mean and the variance of the SPE of the reference batches used for building the PCA model. Based on the confidence levels (in our case $\alpha=0.05$ and $\alpha=0.01$), the alarm and action limits were calculated, respectively.

$$SPE_k \sim \left(\frac{v_k}{2 \cdot m_k} \right) \cdot \chi_{2 \cdot m_k / v_k, \alpha}^2 \quad (6)$$

Our experience showed that no significant time savings were observed with PCA compared to the classical univariate supervisory system, i.e., errors were detected simultaneously. Additionally, reliable process supervision by PCA demands a careful selection of the monitoring variables and especially a high number of model batches to have statistically safe results.

Currently the "Kernel PCA" technique is discussed in literature (Liu and Zhao 2006, Zhang and Qin 2007) as an extension of PCA for dimension reduction and fault analysis of applications depicting a nonlinear behavior. Kernel PCA is very well suited to extract nonlinear structures of interest in the data (Schölkopf *et al.* 1998). It works well for classification purposes; however, the problem is that a reconstruction of original values from the feature space is not generally possible in a unique way. Hence, its use in fermentation fault recovery is questionable.

Parametric Black Box Models

For PCA it is extremely important to choose appropriate input variables, i.e., those having a significant influence on the process dynamics. The same applies for alternatives to PCA, such as autoassociative artificial neural networks (aANNs). These are known to be nonlinear extensions of the PCA techniques (Kramer 1992). Their main advantage is the ability to cope with the intrinsic nonlinearity of fermentation processes. Autoassociative neural networks are black box models where the parameters are determined by means of a nonlinear fit between the network outputs to given process data. They consist of two feedforward artificial neural networks. The first maps the input data to a set of feature variables of such a low dimension that the new variables are almost independent of each other (i.e., they do no longer contain redundant information), and the second feedforward network converts these variables back into the original physical data space. A trained aANN thus transforms the real process signals onto themselves (Figure 7). As in case of PCA, the difference between original and reconstructed measurements is used for calculation of the squared prediction error SPE. Whenever one input value is distorted, the aANN does not map the input signals onto themselves and the SPE rises, indicating that an error in the measurements has occurred. Because of the missing linear properties, the responsible physical quantity cannot be directly identified with the aANN in cases of observed faults in the measurement data.

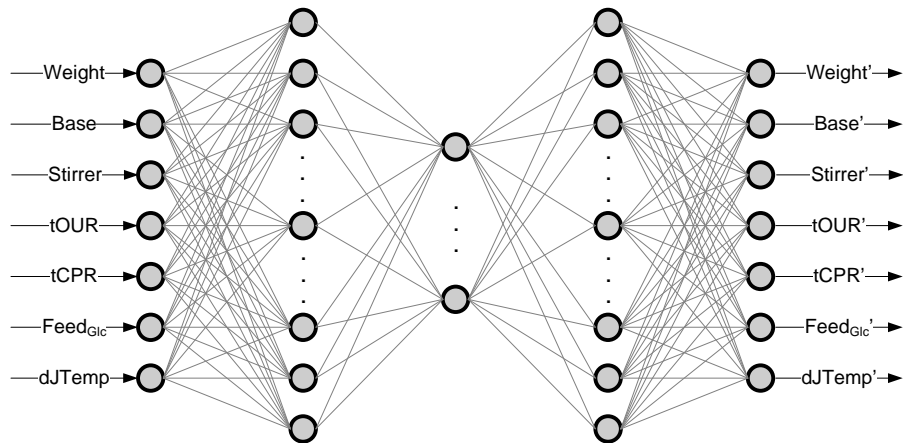


Figure 7. Structure of the autoassociative artificial neural network (aANN) used here for process supervision and signal reconstruction. Measurement values (*left*) are fed into the aANN and, subsequently, the trained network is capable to calculate reconstructed values (*right*). Only if the correlation structure among the variables is correct, original and reconstructed values fall together. In this special case, the number of nodes in the outer hidden layers was 10 and in the bottleneck layer 4 nodes were used.

However, if a single faulty measurement is identified in the univariate signal analysis, the aANN can easily be used to recover the faulty signal. A straightforward way is to simply vary the corresponding input value of the aANN until the SPE becomes minimal. In this way, the signal value, which fits best into the correlation structure with the other input signal values, is obtained.

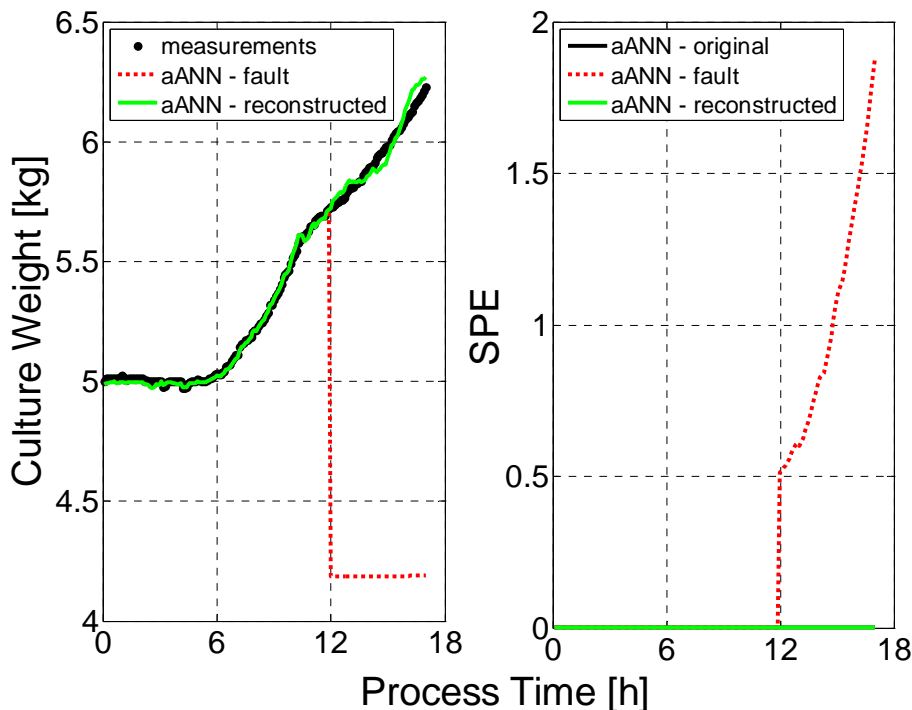


Figure 8. Recovery of a measurement signal by an autoassociative artificial neural network after a complete failure of a single input variable, here the culture weight at $t=12$ h (*left*). First, the responsible variable is identified, e.g., in the univariate monitoring system. By means of a “Golden Search” algorithm this input variable is then varied in a way that the squared prediction error SPE (*right*) as the objective function is minimized, saying that the correlation structure between the measured variable is correct.

The variation is done automatically with a simple search algorithm, e.g., with the "Golden Search" procedure (Galvanauskas *et al.*, 2004). An illustration of this procedure is given in Figure 8, where the measurement of the culture weight failed at $t=12$ h, which was immediately reconstructed by an aANN.

Dynamic Process Models

A highly valuable technique to identify errors is to run a validated dynamic process model in parallel to the real process (Assisa and Filho 2000, Boudreau *et al.* 2006). This directly exploits the mechanistic knowledge of the process engineer about the current state of the process. Examples of dynamic models used for prediction of biomass, substrate or target protein values were presented by (Levisauskas *et al.* 2003, Zheng *et al.* 2005, Nadri *et al.* 2006, Gnoth *et al.* 2010a). Dynamic models are usually based on mass balances around the entire culture and are formulated in terms of systems of ordinary differential equations. The comparison of the modeling results with the current process measurement values directly reveals deviations from the desired process path, a fact that can be utilized for process supervision purposes. More or less complex dynamic process models are essentially the basis of many model aided process variable estimators such as observers and Kalman filters (Bastin and Dochain 1990, Neeleman and van Boxtel 2001, Alvarez and Simutis 2004). In a further step, dynamic process models and black box techniques can be linked to yield hybrid models. These models were successfully applied for monitoring and control of fermentation processes (Oliveira 1998, Roubos *et al.* 2000, Silva *et al.* 2000, Galvanauskas *et al.* 2004). Today, advanced modeling, control and supervisory algorithms are easily implementable into the process control software. The fast progress of these systems has been open the opportunity to simulate the complete bioreactor system in real-time (e.g., kinetics, control action, delay times etc.) simultaneously as the real fermentation takes places. This is referred to as the virtual plant concept. With it, the possibility arises to test process control system's interaction and performance. The virtual plant offers opportunity to reduce costs (e.g., for the development of advanced controllers, process testing), to train personnel in real fermentation environment, and further to accelerate the benefits of the PAT-initiative (McMillan 2007). This concept probably has a great potential, but needs a thorough analysis of the processes involved (McMillan 2004, Boyes *et al.* 2005, Kuprijanov *et al.* 2008).

Recommendation

The backbone of process supervision is and remains univariate analysis of the various on-line signals from the fermenter. This must be installed at all production fermenters integrated in the automatic process control systems (PCS). If variables leave the specified trajectories, alarms are automatically sent to the operator and corrective actions can be performed.

Disturbances are most easily detected by means of trend analysis. Moreover, simple multivariate analyzing techniques are recommended. We propose to perform a mass balance as well as a C-balance on-line during the cultivations. Additionally a degree of reduction balance should be calculated. In this study these balances were utilized to estimate the culture weight (mass balance) and biomass values (C-balance, degree of reduction balance) during the process. The estimations were then compared with on-line and off-line measured values. Analysis of the species involved in the different balances was a simple means to detect the responsible variable in cases of process errors. Methods to combine the different balances and further to automatically detect gross measurement errors have been proposed by (Wang and Stephanopoulos 1983, Heijden 1994a, b, c.). Its practical applicability in industrial fed-batch processes showed Hocalar *et al.* (2006). Balancing techniques require a thorough examination of the species and variables based on the specific conditions of the process. However, these methods lead to a better understanding of the process and the variables included. Compared to statistical process supervision, the methodology does not need extensive fermentation data to

obtain reliable results. In addition, the measurement data involved does not need to be pretreated in a special way.

Abnormalities were first recognized in the original process variables. Operators easier recognize an unusual process behavior when observing original physical signals, e.g., biomass concentrations. A data reduction at the cost of losing physically interpretable measurement values seems to be disputable in bioprocess supervision and control.

Case Studies of Failure Compensating Measures

Process monitoring and fault detection have frequently been discussed in literature (e.g., Glassey *et al.* 1997, Bicciato *et al.* 2000, Albert and Kinley 2001, Gunther *et al.* 2007), however, less information is available about strategies of keeping the process safely within its specification limits upon the identification of an error. It is important to correct faults/disturbances whenever possible. In the following case-studies some practical solutions are proposed.

Error in the pH-Measurement

In a first study we assume that the pH sensor fails. When the sensor electrode breaks, the pH signal jumps beyond the action limit. This is immediately detected by process control system PCS7. If no redundant measurement pH probe is installed, the simplest way of keeping the process running is to directly deactivate the pH controller and switch to an open-loop control mode where the base feed is controlled along its reference profile.

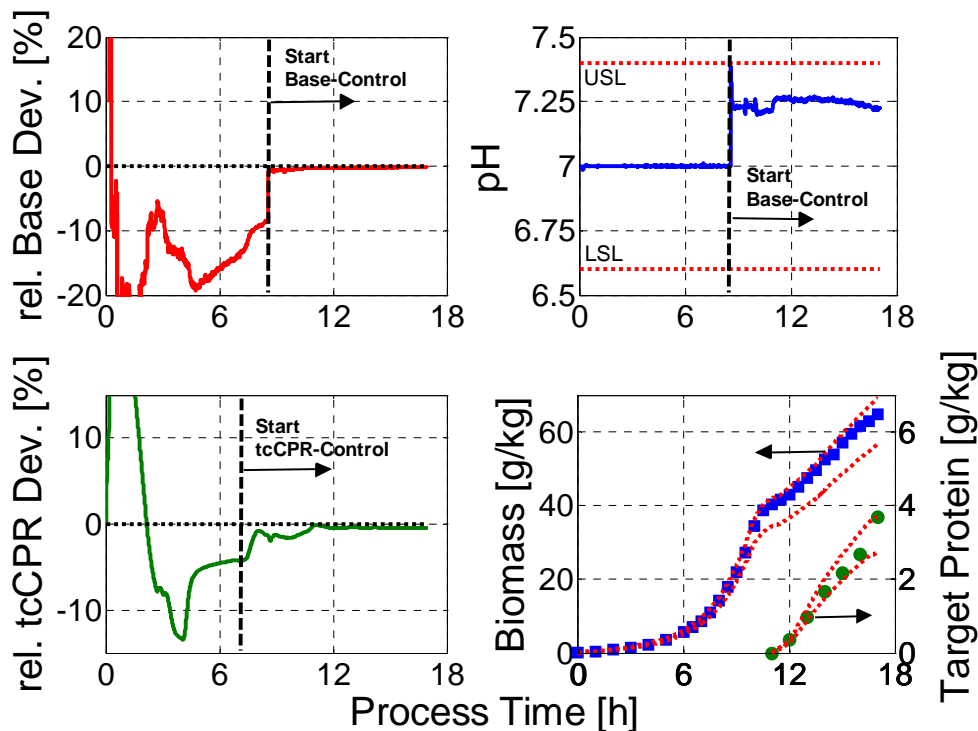


Figure 9. Trajectories of a fermentation where the feedback pH-controller was stopped at $t=8.5$ h and the fail-safe routine was started. The latter feeds the base for pH control along its reference profile. Consequently, the base signal deviation (*upper left*) is zero. The pH remains within its specification limits (*upper right*). The tcCPR controller was still working (*lower left*) and was able to keep biomass (*squares*) and product (*dots*) within their specification limits (*lower right*), which are marked by the *dashed* lines.

The procedure was tested in a fermentation where the feedback pH-controller was deactivated and replaced by an open-loop control mode at $t=8.5$ h. Here, the pH signal did not really fail, but was instead used to show the validity of this approach. The results are depicted in Figure 9.

As can be seen, the process could be kept within the specification limits up to the end of the production run. It is particularly interesting that the procedure worked during the entire product formation phase, i.e., after induction with IPTG.

Failures in the Off-gas Analyzer

All cultivations in the accompanying examples were kept precisely on track by means of a tcCPR controller (Jenzsch *et al.* 2007). This controller needs a tcCPR signal from the process, which is usually derived from the CO_2 -volume fraction measured by the off-gas analyzer. When this device fails, which, for instance, can be detected via the C- and the degree of reduction balance, two main possibilities for fail-safe routines exist.

- i) The faulty signal is replaced by an estimate based on other on-line signals, which are independent of the off-gas analyzer.
- ii) The feedback controller is shut down and the process is continued with an open-loop controller along the reference profile of the action variable, the substrate feed rate F_S .

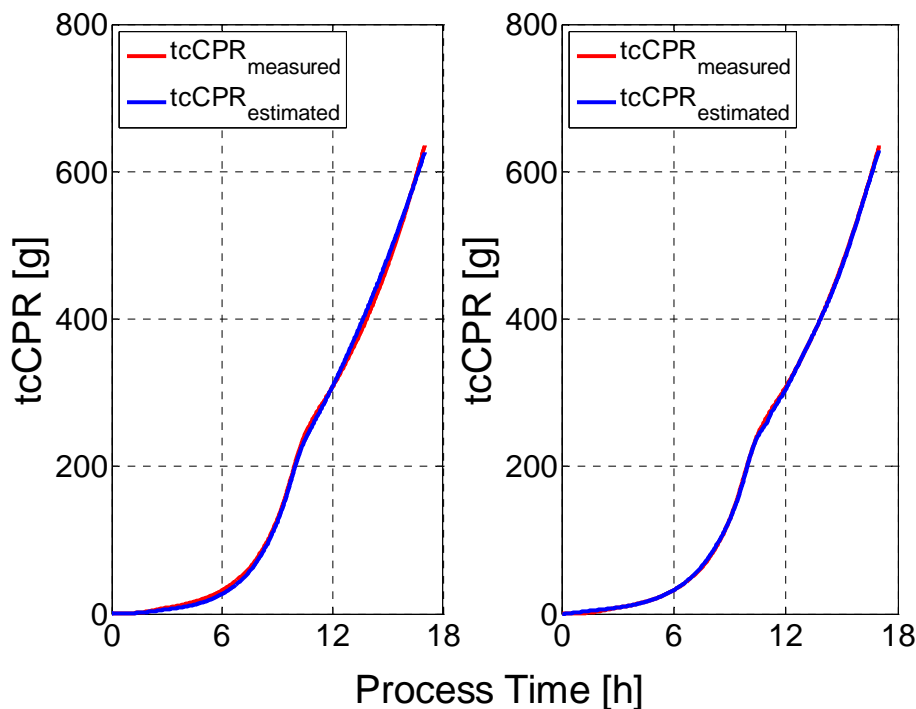


Figure 10. Measured tcCPR profiles and those reconstructed from a dynamic model (*left*) and an artificial neural network (*right*).

On-line estimates of a tcCPR signal can be derived in several ways. Good results were obtained with two approaches: The first is based on a black box model represented by an artificial neural network that directly estimates tcCPR from on-line signals of (a) the integral glucose consumption, (b) the total mass of base added to the process during pH control, and (c) the integral temperature difference of the coolant between output and input at the cooling jacket of the fermenter.

The second estimator is based on the mechanistic process model of Gnoth *et al.* (2010a), where the carbon dioxide production rate is additionally considered by means of a Luedeking-Piret-term. In this case, the on-line signals from the process are (a) the substrate feed rate, and (b) the total mass of base needed by the pH controller.

Figure 10 compares the measured tcCPR signal with the estimates of both approaches in a fermentation where no fault occurred. It becomes clear that both methods provide perfect estimates of the tcCPR signal under these conditions.

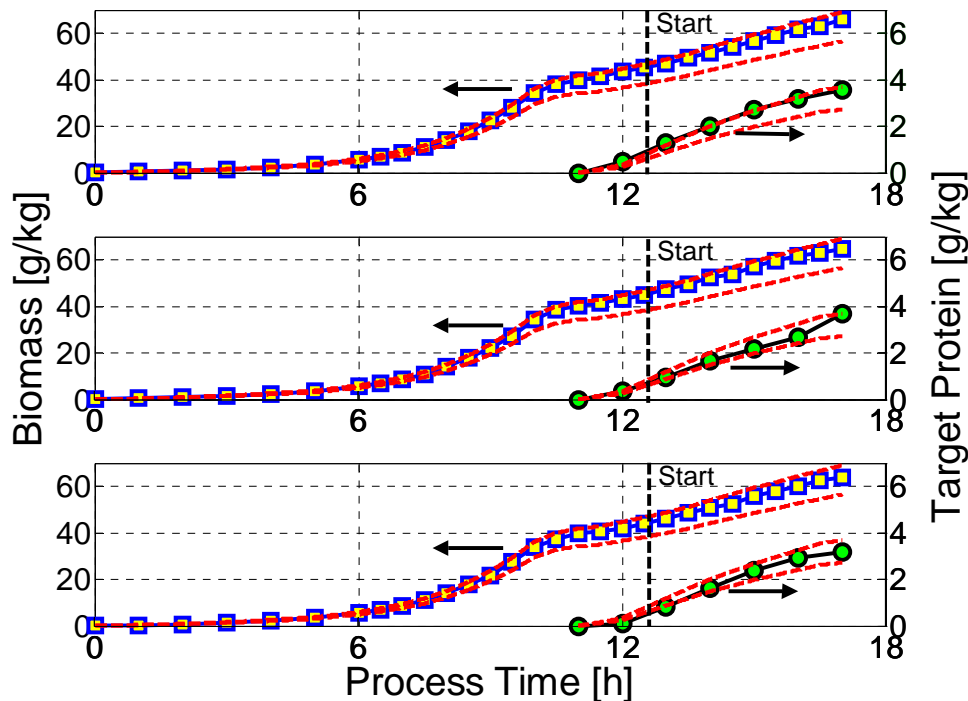


Figure 11. tcCPR controlled fermentations: results of different recovery techniques upon the detection of a failure in the off-gas analyzer at $t=12.5$ h. *Upper graph:* Switch to open-loop control of the reference feed profile, i.e., deactivation of the tcCPR controller. *Middle graph:* using the dynamic process model to estimate tcCPR for the tcCPR controller. *Lower graph:* Keeping the tcCPR controller running by reconstructing the tcCPR signal by means of an ANN using the integral temperature difference between coolant outflow and inflow, the total base consumption and the integral glucose consumption.

In a series of experiments, the three fail-safe routines (two estimators and one open-loop controller) were tested experimentally. The cultures were kept on track until $t=12.5$ h with a perfectly working tcCPR controller. Then, the normal tcCPR signal was either replaced by the estimates but keeping tcCPR control active, or the tcCPR feedback controller was shut down and switched to open-loop control mode.

The results are compared in Figure 11. In all three cases the process could be kept within its specification limits. Even the simplest approach, the open-loop control worked well. The best result was obtained with the tcCPR estimator based on the artificial neural network. If data of about 5 or more fermentations is available, such a network can be trained to give comparably good results.

CONCLUSIONS

In the course of the production of recombinant therapeutic proteins, the formation of the pharmaceutically active ingredients appears during the fermentation step. As this process is extremely complex, the only practical way to guarantee high product quality is to keep the cultivation process tightly under control, i.e., to assure a high batch-to-batch reproducibility. The best strategy of achieving this is to automatically compensate for randomly appearing disturbances by feedback control. Besides keeping pressure, temperature, pH and pO_2 at constant levels, the physiological state of the cells can be controlled via its key variable, the specific biomass growth rate. A very powerful way to perform this indirectly is controlling the total amount or mass of carbon dioxide produced. Jenzsch *et al.* (2007) showed that this tcCPR control is capable to keep the cultivation tightly on track.

Feedback controllers however, require accurate signals of the controlled variable from the cultures. When one of the measurement devices fails, the controllers are no longer able to work properly. Hence, the measurement signals must be supervised in order to make sure that they deliver valid information and in case of an error, fail-safe routines must be started preventing the process to violate its specification limits. Here, we showed some basic approaches to both measures, which were derived from a database and the experience of more than 500 fermentations of *E. coli*-based production processes for recombinant proteins.

With respect to fault detection, the experience is that classical univariate signal analysis, if installed at production fermenters within the standard process control systems, are able to do a very good job. As the number of on-line measurement signals at a single production fermenter is kept as small as possible in order to reduce the contamination risk and to avoid high expenses for device validation, the signals can be observed individually. Practically all distortions and disturbances were first detected by the supervision of the on-line measured variables.

If the signals leave the intervals of normal operating conditions, a measurement error can most often be detected by comparison with other simultaneously measured signals. This requires models relating the quantities to each other.

The easiest way to test on-line signals for consistency is making use of balances. We showed that three balances can routinely be employed for this purpose, the general mass balance around the culture as well as the carbon and the degree of reduction balance. As they consider different quantities, they already provide some information about erroneous signals in a failure situation.

Principal component analysis has often been proposed for supervising fermentation processes, since this linear approach has successfully been applied in extended chemical and petrochemical production plants. However, at single fermenters it cannot make use of its main advantage of data space reduction, because the number of signals to be taken into account is very small. Furthermore, the application of linear methods like PCA to intrinsically nonlinear fermentation processes is questionable. Hence, no practical improvement was observed in all tests of this approach.

Autoassociative artificial neural networks are the nonlinear variants of PCA. They allow to more sensitively detect inconsistencies, but cannot identify the responsible, erroneous signals as easy as PCA. However, since failures are easily detected by the univariate analysis, the aANNs can be employed to recover the distorted signals on-line.

In situations where measurement devices fail the question is how to keep the process within its specification limits. The two main approaches investigated were: (i) switching the process to an open-loop control mode. Here, the controlled variable is guided along a predefined trajectory, which is most often the reference profile of the action variable; and (ii) estimating the correct

value of the faulty signal from other independently measured on-line signals and replacing the disturbed one, thus keeping the controller running.

Both variants were tested at the accompanying example, the HumaX-production with *E. coli*, and both showed the general applicability in fermentation practice. pH could be kept within its specification limits by simply replacing the pH feedback controller by an open-loop control along the reference base feed rate profile. tcCPR control was continued running through replacement of the faulty signal by a neural network model estimate or by a value based on a dynamic process model calculation. In doing so, the fermentations could be successfully finished within the specification limits even in cases of severe process faults.

Acknowledgements

The financial support of "Excellence Initiative" of the German State of Sachsen-Anhalt is gratefully acknowledged.

NOMENCLATURE

Ac	Overflow metabolites	
E	Matrix of errors	
e	Vector of errors	
F_{AF}	Rate of addition of antifoam agent	[kg/h]
F_B	Rate of base addition during pH control	[kg/h]
F_{Evap}	Rate of water evaporation of the culture broth due to aeration	[kg/h]
F_{Sample}	Rate of sampling	[kg/h]
F_S	Substrate feed rate	[kg/h]
h	Statistical test function	
I	Number of batches	
J	Number of variables	
K	Number of time steps	
k	Current time step	
NH_3	Ammonia	
n_{CAc}	Amount of carbon in overflow metabolites	[mol]
n_{CCO_2}	Amount of carbon in the carbon dioxide	[mol]
n_{CF}	Amount of carbon that is added with the feeding	[mol]
n_{CP}	Amount of carbon in the target protein	[mol]
n_{CS}	Amount of carbon in the substrate in the culture medium	[mol]
n_{CS0}	Amount of carbon in the substrate being added at the start of fermentation	[mol]
$n_{CSample}$	Amount of carbon removed due to sampling	[mol]
n_{CX}	Amount of carbon in the biomass	[mol]
n_{CX0}	Amount of carbon in the initial biomass	[mol]
P	Concentration of target protein	[g/kg]
P^T	Transposed matrix of loadings	
r	Number of principal components	
S	Substrate concentration	[g/kg]
S_0	Initial substrate concentration at inoculation time	[g/kg]
SPE	Squared prediction error	
ΔT_{Jacket}	Temperature difference between coolant outflow and inflow	[K]
T	Temperature of the culture medium	[°C]
T	Matrix of scores	
t_{new}	Vector of scores of the fermentation to be monitored	

tCPR	Total carbon dioxide production rate	[g/h]
tCTR	Total carbon dioxide transfer rate	[g/h]
tOUR	Total oxygen uptake rate	[g/h]
tOTR	Total oxygen transfer rate	[g/h]
tcOUR	Total cumulative oxygen uptake rate	[g]
tcCPR	Total cumulative carbon dioxide production rate	[g]
W_0	Initial culture weight	[kg]
W_m	Measured culture weight	[kg]
X	Matrix of measurements (PCA)	
X	Biomass concentration	[g/kg]
X_0	Initial biomass concentration at inoculation time	[g/kg]
Y_{XS}	Yield coefficient of biomass formed per substrate	[mol/C-mol]
Y_{OS}	Yield coefficient of oxygen consumed per substrate	[mol/C-mol]
Y_{NH_3S}	Yield coefficient of ammonia consumed per substrate	[mol/C-mol]
Y_{PS}	Yield coefficient of target protein formed per substrate	[mol/C-mol]
Y_{CS}	Yield coefficient of carbon dioxide formed per substrate	[mol/C-mol]
Y_{H_2OS}	Yield coefficient of water formed per substrate	[mol/C-mol]
μ	Specific biomass growth rate	[1/h]

REFERENCES

- Albert, S., Kinley, R.D., 2001. Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision. *Trends Biotechnol.* 19: 53–62.
- Alvarez, A., Simutis, R., 2004. Application of Kalman filter algorithm in GMC control strategy for fed-batch cultivation process. *Informacinės Technologijos Ir Valdymas* 1: 7–12.
- Bastin, G., Dochain, D., 1990. On-line estimation and adaptive control of bioreactors. *Elsevier*, Amsterdam.
- Bicciato, S., Bagno, A., Solda, M., Manfredini, R., di Bello, C., 2002. Fermentation diagnosis by multivariate statistical analysis. *Appl. Biochem. Biotechnol.* 102-103: 49–62.
- Boudreau, M.A., McMillan, G., Wilson, G., 2006. Maximizing PAT benefits from bioprocess modeling and control. *Pharmaceutical Technology* Nov 1.
- Boyes, W., Hebert, D., O'Brien, L., McMillan, G.K., 2005. The light at the end of the tunnel is a train (virtual plant reality). *Control*.
- Breton, R.G., 2005. Chemometrics – data analysis for the laboratory and chemical plant. *John Wiley & Sons Ltd.*, Chichester.
- Chattaway, T., Demain, A.L., Stephanopoulos, G.N., 1992. Use of various measurements for biomass estimation. *Biotechnol. Progr.* 8: 81–84.
- de Assisa, A.J., Filho, R.M., 2000. Soft sensors development for on-line bioreactor state estimation. *Comp. Chem. Eng.* 24: 1099–1103.
- De Palma, A., 2004. Taking process monitoring to a next level. *Gen. Eng. News* 24: 46–47.
- Dong, D., McAvoy, T.J., 1996. Nonlinear principal component analysis – Based on principal curves and neural networks. *Comput. Chem. Eng.* 20: 65–78.
- European Medicines Agency, 2006. Reflection paper: Chemical, pharmaceutical and biological information to be included in dossiers when Process Analytical Technology (PAT) is employed. London.
- U.S. Food and Drug Administration, 2004. Guidance for Industry: PAT – a framework for innovative pharmaceutical manufacturing and quality assurance. Rockville.
- Fortuna, L., Graziani, S., Rizzo, A., Xibilia, M.G., 2007. Soft sensors for monitoring and control of industrial processes. *Springer*, London.
- Galvanauskas, V., Simutis, R., Lübbert, A., 2004. Hybrid process models for process optimization, monitoring and control. *Bioproc. Biosyst. Eng.* 26: 393–400.
- Glasse, J., Ignova, M., Ward, A.C., Montague, G.A., Morris, A.J., 1997. Bioprocess supervision: neural networks and knowledge based systems. *J. Biotechnol.* 52: 201–205.
- Gnoth, S., Simutis, R., Lübbert, A., 2010a. Selective expression of the soluble product fraction in *Escherichia coli* cultures employed in recombinant protein production processes. *Appl. Microbiol. Biotechnol.* 87: 2047–2058.
- Gnoth, S., Kuprijanov, A., Simutis, R., Lübbert, A., 2010b. Simple adaptive pH control using gain-scheduling methods. *Appl. Microbiol. Biotechnol.* 85: 955–964.
- Gregersen, L., Jørgensen, S.B., 1999. Supervision of fed-batch fermentations, *Chem. Eng. J.*, 75: 69–76.
- Gunther, J.C., Conner, J.S., Seborg, D.E., 2007. Fault detection and diagnosis in an industrial fed-batch cell culture process. *Biotechnol. Progr.* 23: 851–857.
-

- Han, L., Enfors, S.O., Häggström, L., 2003. *Escherichia coli* high-cell-density culture: carbon mass balances and release of outer membrane components. *Bioproc. Biosyst. Eng.* 25: 205–212.
- Hocalar, A., Türker, M., Öztürk, S., 2006. State estimation and error diagnosis in industrial fed-batch yeast fermentation. *AIChE J.* 52: 3967–3980.
- Jenzsch, M., Simutis, R., Lübbert, A., 2006. Optimization and control of industrial microbial cultivation processes. *Eng. Life Sci.* 6: 117–124.
- 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.
- Jobé, A.M., Herwig C., Surzyn, M., Walker, B., Marison, I., von Stockar, U., 2003. Generally applicable fed-batch culture concept based on the detection of metabolic state by on-line balancing. *Biotechnol. Bioeng.* 82: 627–639.
- Junker, B., Lester, M., Leporati, J., Schmitt, J., Kovatch, M., Borysewicz, S., Maciejak, W., Seeley, A., Heese, M., Connors, N., Brix, T., Creveling, E., Salmon, P., 2006. Sustainable reduction of bioreactor contamination in an industrial fermentation pilot plant. *J. Biosci. Bioeng.* 102: 251–268.
- Kramer, M.A., 1991. Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.* 37: 233–243.
- 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.
- Kuprijanov, A., Schaepe, S., Aehle, M., Gnoth, S., Simutis, R., Lübbert, A., 2010. Advanced control of fed batch cultures during recombinant protein production, *ESBES Symposium*, Bologna, publication in preparation.
- Liu, F., Zhao, Z., 2006. On-line batch process monitoring using multiway kernel independent component analysis. *Advances in Neural Networks-ISNN 2006*: 951–956.
- Levisauskas, D., Galvanauskas, V., Henrich, S., Wilhelm, K., Volk, N., Lübbert, A., 2003. Model-based optimization of viral capsid protein production in fed-batch culture of recombinant *Escherichia coli*. *Bioproc. Biosyst. Eng.* 25: 255–262.
- McMillan, G.K., Cameron, R., 2004. Models unleashed: Virtual Plant and model predictive control applications. *ISA*, Research Triangle Park (NC).
- McMillan, G.K., Boudreau, M.A., 2007. New directions in bioprocess monitoring and control. *ISA*, Research Triangle Park (NC).
- Nadri, M., Trezzani, I., Hammouri, H., Dhurjarti, P., Longin, R., Lieto, J., 2006. Modeling and observer design for recombinant *Escherichia coli* strain. *Bioproc. Biosyst. Eng.* 28: 217–225.
- Neeleman, R., van Boxtel, A.J.B., 2001. Estimation of specific growth rate from cell density measurements. *Bioproc. Biosyst. Eng.* 24: 179–185.
- Nielsen, J., Villadsen, J., 1994. *Bioreaction Engineering Principles*. Plenum Press, New York.
- Nomikos, P., MacGregor, J.F., 1994. Monitoring batch processes using multiway principal component analysis. *AIChE J.* 40: 1361–1375.

- Oliveira, R.M.F., 1998. Supervision, control and optimization of biotechnological processes based on hybrid models. *PhD thesis*, University of Halle-Wittenberg, Germany.
- Reilly, P.M., Carpani, R.E., 1963. Application of statistical theory of adjustment of material balances. 13th Congress Chemical Engineering Conference, Montréal, Quebec.
- Ripps, D.L., 1965. Adjustment of experimental data. *Chem. Eng. Prog. Symp. Ser.* 55, 61: 8–13.
- Roubos, J.A., Babuska, R., Krabben, P., Heijnen, J.J., 2000. Hybrid modeling of fed-batch bioprocesses; combination of physical equations with metabolic networks and black-box kinetics. *Journal A, Benelux Q J Automatic Control* 41: 17–23.
- Royce, P.N., 1992. Effect of changes in the pH and carbon dioxide evolution rate on the measured respiratory quotient of fermentations. *Biotechnol. Bioeng.* 40: 1129–1138.
- Royce, P.N., 1993. A discussion of recent developments in fermentation monitoring and control from a practical perspective. *Crit. Rev. Biotechnol.* 13: 117–149.
- Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 39: 1627–1639.
- Schölkopf, B., Smola, A., Müller, K.R., 1998. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* 10: 1299–1319.
- Shewart, W.A., 1931. Economic control of quality of manufactured product. *Van Nostrand*, Princeton (NJ).
- Shuler, M., Kargi, F., 2002. Bioprocess engineering. *Prentice Hall*, Upper Saddle River (NJ).
- Silva, R.G., Cruz, A.J.G., Hokka, C.O., Giordano, R.L.C., Giordano, R.C., 2000. A hybrid feedforward and neural network model for the cephalosporin C production process. *Braz. J. Chem. Eng.* 7: 4–7.
- Spérandio, M., Paul, E., 1997. Determination of carbon dioxide evolution rate using on-line gas analysis during dynamic biodegradation experiments. *Biotechnol. Bioeng.* 53: 243–252.
- Stoumbos, Z.G., Reynolds, M.R., Woodall, W.H., 2003. Control chart schemes for monitoring the mean and variance of processes subject to sustained shifts and drifts. *Handbook of Statistics 22, Statistics in Industry. Elsevier*, Amsterdam.
- Van der Heijden, R.T.J.M., Heijnen, J.J., Hellinga, C., Romein, B., Luyben, K.C.A.M., 1994a. Linear constraint relations in biochemical reaction systems: I. classification of the calculability and the balanceability of conversion rates. *Biotechnol. Bioeng.* 43: 3–10.
- Van der Heijden, R.T.J.M., Romein, B., Heijnen, J.J., Hellinga, C., Luyben, K.C.A.M., 1994b. Linear constraint relations in biochemical reaction systems: II. diagnosis and estimation of gross errors. *Biotechnol. Bioeng.* 43: 11–20.
- Van der Heijden, R.T.J.M., Romein, B., Heijnen, J.J., Hellinga, C., Luyben, K.C.A.M., 1994c. Linear constraint relations in biochemical reaction systems: III. Sequential application of data reconciliation for sensitive detection of systematic errors. *Biotechnol. Bioeng.* 44: 781–791.
- Wang, N.S., Stephanopoulos, G., 1983. Application of macroscopic balances to the identification of gross measurement errors. *Biotechnol. Bioeng.* 25: 2177–2208.
- Westerhuis, J.A., Gurden, S.P., Smilde, A.K., 2000. Generalized contribution plots in multivariate statistical process monitoring. *Chemometr. Intell. Lab. Syst.* 51: 95–114.
- Wold, S., Geladi, P., Esbensen, K., Ohman, J., 1987. Multi-way principal components and PLS analysis. *J. Chemometr.* 1: 41–56.

Zhang, Y., Qin, S.J., 2007. Fault detection of nonlinear processes using multiway kernel independent component analysis. *Ind. Eng. Chem. Res.* 46: 7780–7787.

Zheng, Z.Y., Yao, S.J., Lin, D.Q., 2005. Using a kinetic model that considers cell segregation to optimize hEGF expression in fed-batch cultures of recombinant *Escherichia coli*. *Bioproc. Biosyst. Eng.* 27: 143–152.

SUMMARY

Biopharmaceutical manufacturing processes have been for a long time operated under rigid regulatory conditions that prevented technological progress. However, the introduction of the process analytical technology (PAT) concept by the U.S. Food and Drug Administration in 2004 has led to fundamental changes with respect to the inspection and approval of manufacturing processes of biologics. Rather than keeping production processes fixed in any respect, it opens up the possibility to establish them on the basis of a comprehensive process understanding and leaves space for a continuous technological process improvement. A central role herein plays the use of modern process analytical tools, whose signals are not only utilized for process analysis and optimization but also to quickly counteract process inconsistencies during the process. Techniques and methods, that have been long term rejected due to regulatory concerns, but that have found a widespread use in other production industries, e.g., automobile and petrochemical industry, are now explicitly supported to be implemented in biopharmaceutical production processes. It is one of the major aims of the regulatory authorities in the near future to overcome the rigid mindset as well the technical standstill in biopharmaceutical production that have in the past hindered process innovations.

The presented work gives an insight into the practical implementation of the overall PAT concept in fermentation processes for production of recombinant therapeutic proteins. At several specific examples on different technological levels, it will be shown that all presented approaches result in improvements in terms of process robustness, performance as well as the degree of batch-to-batch reproducibility of important process variables.

First of all, it was shown that it is straightforward to intensively deal with quality of measurement data itself, meaning the improvement of the information content of recorded variables by a consequent reduction of measurement noise. In this context, parameters of pH and the dissolved oxygen in the liquid medium pO_2 were identified to critically affect fermentation performance, since they have a strong influence on biological as well as on physicochemical properties during fermentation. In accordance to the PAT approach, the dynamic effects were both qualitatively and quantitatively understood by means of mathematical models paving the way for a directed solution. The developed gain-scheduling approach, that adapts controller parameters to the nonlinear dynamics of the process, was first tested in numerical simulations and afterwards successfully applied in practical fermentations. The proposed technique proved to be beneficial directly for controller performance of the corresponding variables but also for the quality of measurements physico-chemically related to these variables. In this way, the accuracy prediction of advanced process models, requiring the measurement data in order to quantitatively describe complex relationships, was significantly improved, as well.

Besides that, special emphasis was put on the analysis of important state variables and their dynamical behavior during fermentation processes. Biomass and target protein herein were of paramount importance, as they crucially defined process quality and batch-to-batch reproducibility. Estimation of these variables with common methods was generally difficult due to their intrinsic nonlinear nature. Therefore, a hybrid modeling approach was set up. It combined artificial neural networks with general valid mass balances and enabled not only the reliable prediction of biomass and target protein but also the corresponding dynamics, as expressed by their specific formation rates. The technique worked completely data-driven and was therefore easily applicable to various expression systems, as it was shown at different types of product formation kinetics in *E. coli*.

In the next step, process models, either mechanistic or black box, were demonstrated to be quantitatively utilized for model-based optimization. At the example of a recombinant *E. coli* strain expressing the target in the soluble form, the hybrid modeling technique was firstly used to identify product formation kinetics as a function of central influencing factors. Subsequently, based on the trained model, optimal values of parameters having a major impact on target protein expression were calculated to redesign the process. It was then kept to its optimal trajectories using advanced feedback control. Therefore, process performance with respect to large amounts of target protein, as well a high degree of batch-to-batch reproducibility were increased simultaneously.

Finally, on the way to improve overall process robustness and reliability of process data on-line fault detection routines and supervisory systems were developed. Here, the main objective was to evaluate measurement data in terms of consistency and validity in comparison to reference batches run under normal operating conditions. However, these techniques were not only restricted to the identification of process errors, the methods established were designed to further activate automated fail-safe systems, which were able to intervene in classical control actions, if necessary. In this way, fermentation processes were finished within specification limits even in cases of fatal process deviations.

All approaches presented in this work basically share one common concept. It starts with the consequent analysis of measurement information during the fermentation process and its subsequent application for identification of fundamental cause-effect relationships. It continues with the quantitative description using process models and finally ends up with the use of the gathered know-how for process improvements in a directed way. Moreover, it was a major objective that the applied methods remain feasible, are easy to implement and are transferable to other hosts and production systems. Thus, the process analytical technology is not an abstract topic to be qualitatively discussed. Instead, the methods proposed herein showed to be valuable tools pushing fermentation processes forward toward a significantly improved performance and quality. The results demonstrated the potential to transfer the overall PAT idea to the real fermentation environment in the field of research and industrial manufacturing. It is now up to the manufacturers to adopt these state-of-the-art techniques and to overcome the technological backlog in biopharmaceutical production.

ZUSAMMENFASSUNG

Biopharmazeutische Herstellungsprozesse sind in der Vergangenheit über einen langen Zeitraum unter strengen regulatorischen Rahmenbedingungen durchgeführt worden, die eine technologische Weiterentwicklung verhinderte. Mit der Einführung der „*Process Analytical Technology*“ (PAT) Initiative im Jahre 2004 durch die US-amerikanische Aufsichts- und Zulassungsbehörde (FDA) hat sich jedoch eine grundlegende Wende hinsichtlich der Überwachung und Zulassung von Produktionsprozessen zur Herstellung biopharmazeutischer Wirkstoffe, sogenannter *Biologics*, vollzogen. Die PAT-Richtlinie bietet Herstellern die Möglichkeit, Produktionsprozesse basierend auf einem umfassenden qualitativen und quantitativen Prozessverständnis zu etablieren sowie auch nach der Zulassung kontinuierlich weiterzuentwickeln. Eine zentrale Rolle spielt dabei die Nutzung moderner Messtechnik, deren Signale nicht nur für die Analyse und Optimierung von Verfahren sondern ausdrücklich auch für die Bekämpfung von Prozessabweichungen verwendet werden sollen. Fortschrittliche Technologien und Methoden wurden früher aus regulatorischen Gründen nur zögerlich übernommen. Inzwischen wird jedoch der Einsatz bestimmter Verfahren, die bereits seit längerem in anderen Industriezweigen, wie z. B. der Automobilindustrie und Petrochemie, erfolgreich Einzug halten, in der biopharmazeutischen Produktion nun explizit gefordert. Es ist eines der erklärten Ziele der Zulassungsbehörden, mit der proklamierten PAT-Initiative in näherer Zukunft die starre, innovationshemmende Haltung sowie den technologischen Stillstand der Produktionsprozesse der vergangenen Jahre zu überwinden.

Die vorliegende Arbeit gibt einen Einblick in die praktische Umsetzung des PAT-Konzeptes in Bezug auf Fermentationsprozesse zur Herstellung rekombinanter, therapeutischer Proteine. Anhand konkreter Beispiele auf verschiedenen technologischen Ebenen wird gezeigt, dass alle vorgestellten Anwendungen zu einer Verbesserung hinsichtlich der Stabilität, Leistungsfähigkeit als auch der Reproduzierbarkeit wichtiger Kenngrößen des Fermentationsprozesses führen.

Das Hauptaugenmerk der Arbeit lag zunächst auf der intensiven Auseinandersetzung mit den während der Fermentation aufgezeichneten Messdaten. Hierbei konnte durch eine konsequente Reduzierung auftretender Messwertschwankungen der Informationsgehalt der Prozessvariablen sowie deren Messdatenqualität deutlich verbessert werden. In diesem Zusammenhang wurden der pH-Wert sowie der Gelöstsauerstoffpartialdruck pO_2 im Nährmedium als kritische Prozessgrößen identifiziert, da diese einen fundamentalen Einfluss sowohl auf biologische als auch auf physikochemische Effekte im Fermentationsprozess haben. Das auftretende dynamische Verhalten dieser Prozessgrößen wurde gemäß PAT zunächst qualitativ erfasst und anschließend mittels mathematischer Prozessmodelle quantitativ abgebildet. In numerischen Simulationen wurde dann ein Algorithmus entwickelt, der die Regelparameter automatisch an die nichtlineare Prozessdynamik anpasst. Anschließend wurde die Methode erfolgreich in Fermentationsexperimenten angewendet. Dadurch konnte zum einen die Regelgenauigkeit und zum anderen die Messdatenqualität zentraler Prozessvariablen erhöht werden. Der gesteigerte Informationsgehalt der Messdaten führte dazu, dass die im Zuge dieser Arbeit entwickelten Prozessmodelle in ihrer Vorhersagequalität wesentlich verbessert werden konnten.

Ein weiterer Schwerpunkt stellte die Analyse wichtiger Zustandsgrößen sowie deren zeitliche Entwicklung im Fermentationsprozess dar. Von zentraler Bedeutung waren hierbei die Konzentration der Biomasse sowie des exprimierten Zielproteins, da diese in besonderem Maße die Qualität und Reproduzierbarkeit eines Fermentationsprozesses bestimmten. Grundsätzlich gestaltete sich mit den herkömmlichen Methoden die Zustandsschätzung dieser beiden Größen wegen des grundlegenden nichtlinearen Verhaltens schwierig. Der in diesem Zusammenhang

entwickelte Hybrid-Modellansatz vereint künstliche neuronale Netze mit allgemein gültigen Massenbilanzen. Dadurch war es möglich, zuverlässige Schätzungen für Biomasse und Zielproteinwerte vorzunehmen und darüber hinaus auch die zu Grunde liegenden Kinetiken zu identifizieren. Das vorgestellte Verfahren arbeitete dabei vollständig datenbasiert und war somit sehr leicht auf unterschiedliche Arten von Expressionssystemen und Bildungskinetiken übertragbar.

Die formulierten Prozessmodelle, entweder mechanistisch oder datenbasiert, wurden in einem nächsten Schritt konsequent für eine modellgestützte Prozessoptimierung eingesetzt. Am Beispiel eines heterologen, löslich exprimierten Proteins wurde ein hybrides Modell verwendet, um zunächst die Produktbildungskinetik in Abhängigkeit zentraler Einflussgrößen quantitativ zu erfassen. Basierend auf dem trainierten Prozessmodell wurden optimale Parameter für eine verbesserte Zielproteinexpression berechnet und die Prozessführung dahingehend angepasst. Mit Hilfe höherwertiger Regelalgorithmen wurde sichergestellt, dass der Prozess entlang seiner optimalen Trajektorie geführt wurde. Durch diese Strategie wurden die absolute Zielproteinmenge und gleichzeitig die Reproduzierbarkeit des Fermentationsprozesses wesentlich erhöht.

Um die Robustheit und die Zuverlässigkeit der Fermentationsprozesse insgesamt zu verbessern, wurden in einem weiteren Schritt Routinen zur Fehlererkennung und Prozessüberwachung entwickelt. Ihre Hauptaufgabe bestand darin, Messdaten hinsichtlich ihrer Konsistenz sowie ihrer Gültigkeit in Bezug auf Referenzprozesse zu überprüfen. Neben der primären Aufgabe der Fehlererkennung waren die entwickelten Methoden in der Lage, wenn nötig, zusätzlich automatische Sicherungssysteme zu aktivieren. Dadurch war es möglich, Fermentationen auch bei Auftreten schwerwiegender Prozessabweichungen innerhalb ihrer definierten Spezifikationsgrenzen zu beenden.

Allen im Rahmen dieser Arbeit vorgestellten Methoden liegt ein gemeinsames Grundkonzept zu Grunde. Dieses beinhaltet im Wesentlichen die konsequente Nutzung von im Fermentationsprozess anfallender Messinformation, um die fundamentalen Zusammenhänge der Variablen untereinander zu identifizieren. Es setzt sich fort in der Beschreibung der zu Grunde liegenden Beziehungen in Form quantitativ auswertbarer Prozessmodelle und der gezielten Anwendung dieser Modelle für eine gerichtete Prozessoptimierung. Darüber hinaus ist es ein Hauptanliegen, dass die präsentierten Techniken praktikabel sowie einfach auf verschiedene Produktionssysteme übertragbar sind. Das PAT-Konzept ist insofern nicht nur als ein abstraktes, qualitativ zu diskutierendes Thema anzusehen. Vielmehr wird im Rahmen dieser Arbeit gezeigt, dass die entwickelten Methoden Fermentationsprozesse technologisch und qualitativ signifikant verbessern. Dadurch ergibt sich ein großes Potenzial für einen verbreiteten Einsatz nicht nur in der Forschung sondern auch im routinemäßigen, industriellen Produktionsbetrieb. Es liegt nun an den Herstellern, die neuen Werkzeuge einzusetzen und damit den technologischen Rückstand in der Produktion von *Biologics* aufzuholen.

PUBLICATIONS

Published Papers

- [1] Jenzsch, M., **Gnoth, S.**, Beck, M., Kleinschmidt, M., Simutis, R., Lübbert, A., **2006.**
Open-loop control of the biomass concentration within the growth phase of recombinant protein production processes. *J. Biotechnol.* 127: 84–94.
- [2] Jenzsch, M., **Gnoth, S.**, Kleinschmidt, M., Simutis, R., Lübbert, A., **2006.**
Improving the batch-to-batch reproducibility in microbial cultures during recombinant protein production by guiding the process along a predefined total biomass profile. *Bioproc. Biosyst. Eng.* 29: 315–321.
- [3] Jenzsch, M., **Gnoth, S.**, Simutis, R., Lübbert, A., **2006.**
Einsatz robuster Regelungsstrategien zur Steigerung der Reproduzierbarkeit von Fermentationsprozessen. *Chem. Ing. Tech.* 78: 1374.
- [4] **Gnoth, S.**, Jenzsch, M., Simutis, R., Lübbert, A., **2006.**
Product Formation kinetics in a recombinant production process. *10th International IFAC Symposium on Computer Applications in Biotechnology*: 197–202, Cancun, Mexico.
- [5] 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.
- [6] **Gnoth, S.**, Jenzsch, M., Simutis, R., Lübbert, A., **2007.**
Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control. *J. Biotechnol.* 132: 180–186.
- [7] **Gnoth, S.**, Jenzsch, M., Simutis, R., Lübbert, A., **2007.**
Control of cultivation processes for recombinant protein production: a review. *Bioproc. Biosyst. Eng.* 31: 21–39.
- [8] **Gnoth, S.**, Jenzsch, M., Simutis, R., Lübbert, A., **2007.**
Product formation kinetics in genetically modified *E. coli* bacteria: inclusion body formation. *Bioproc. Biosyst. Eng.* 31: 41–46.
- [9] 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.
- [10] 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.
- [11] **Gnoth, S.**, Kuprijanov, A., Simutis, R., Lübbert, A., **2010.**
Simple adaptive pH control in bioreactors using gain-scheduling methods. *Appl. Microbiol. Biotechnol.* 85: 955–964.
-

- [12] **Gnoth, S.**, Simutis, R., Lübbert, A., **2010**.
Selective expression of the soluble product fraction in *Escherichia coli* cultures employed in recombinant protein production processes. *Appl. Microbiol. Biotechnol.* 87: 2047–2058.
- [13] **Gnoth, S.**, Simutis, R., Lübbert, A., **2010**.
Fermentation process supervision and strategies for fail-safe operation – a practical approach. *Eng. Life Sci.*, accepted for publication.

Oral Presentation/Poster

- [14] ESBES 6, Salzburg, Austria (oral presentation), **2006**.
6th European Symposium on Biochemical Engineering.
Batch-to-batch-reproducibility of fermentation processes by robust process operational design and control
- [15] Brač Engineering Course, Supetar, Croatia (oral presentation, poster), **2006**.
Improvements of the reproducibility of fermentation processes by robust process operational design and control
- [16] Gesamtarbeitsbesprechung Exzellenznetzwerk Biowissenschaften, Halle, Germany (oral presentation, poster), **2008**.
Design of fermentation processes for recombinant protein processes on highest quality and productivity levels.
- [17] ESBES 7, Faro, Portugal (poster), **2008**.
7th European Symposium on Biochemical Engineering.
Investigating product formation kinetics of recombinant proteins in cultivation processes of *E. coli*.

CURRICULUM VITAE

General

Name	Stefan Gnoth
Date of Birth	22.03.1981
Place of Birth	Neubrandenburg
Gender:	Male
Nationality	German

School Education

09/1987 – 08/1990	Polytechnische Oberschule, Neubrandenburg
09/1990 – 08/1993	Realschule, Neubrandenburg
09/1993 – 07/1999	Lilienthal-Gymnasium, Neubrandenburg

Military service

11/1999 – 08/2000	Basic military service in Bad Segeberg
-------------------	--

University Education

10/2000 – 09/2005	Martin-Luther-University of Halle-Wittenberg Study: Bioengineering
09/2005	Graduation in Bioengineering Degree: “Diplom-Ingenieur” (A-Level, Mark 1.3) Title of Diploma Thesis: „Untersuchung von Strategien der Zustands- schätzung und Regelung eines industriellen Fermentationsprozesses mit rekombinanten <i>Escherichia coli</i> “
10/2005 – 06/2009	PhD student within the Excellence Initiative of the Federal State of Sachsen-Anhalt at the Centre of Bioprocess Engineering of the Martin-Luther-University of Halle-Wittenberg

DANKSAGUNG

Diese Dissertation sowie alle praktischen Arbeiten wurden im Zeitraum von Oktober 2005 bis Juni 2009 unter Leitung von Herrn Prof. Dr. Andreas Lübbert am Institut für Biochemie/Biotechnologie, Zentrum für Bioverfahrenstechnik der Martin-Luther-Universität Halle-Wittenberg angefertigt.

Im Voraus bedanke ich mich bei Herrn Prof. Andreas Lübbert für die Bereitstellung des Themas, seine wissenschaftliche Betreuung sowie die stets konstruktive Kritik. Ganz herzlich möchte ich mich bei Herrn Prof. Dr. Rimvydas Simutis bedanken. Sein wissenschaftlicher Beitrag sowie seine zielführende Diskussionsbereitschaft bei komplexen Problemstellungen haben entscheidend zum Gelingen dieser Arbeit beigetragen.

Der Landesexzellenzinitiative Biowissenschaften Sachsen-Anhalt danke ich für die finanzielle Unterstützung. In diesem Zusammenhang danke ich auch Herrn Prof. Rainer Rudolph für sein stetes Interesse an meiner Arbeit und die Förderung des wissenschaftlichen Nachwuchses. Darüber hinaus bleiben seine wissenschaftliche Leistungen und seine Verdienste am Forschungsstandort Halle unvergessen.

Daneben bedanke ich mich bei allen Mitarbeitern des Zentrums für Bioverfahrenstechnik. Besonderer Dank gilt hierbei Frau Martina Anwand, die mich intensiv bei der Proteinanalytik, der Vorbereitung von Fermentationen und bei der Organisation des Laborbetriebes unterstützte. Unserem Werkstattmeister Herrn Frank Ullmann danke ich für die praktische Umsetzung von technischen Innovationen in und am Bioreaktor, die stets mit einem Höchstmaß an Präzision und Funktionalität durchgeführt wurden. Weiterhin will ich nicht Frau Cornelia Homolya vergessen, die sich um die Verwaltung unserer Arbeitsgruppe kümmerte. Vielen Dank dafür.

Ich danke insbesondere Christian Sieblist für die große Hilfe bei Problemen rund um die Computerhardware und Software. Er hatte auch stets ein offenes Ohr für die kleinen und großen Probleme, die im Zuge der Arbeit auftraten. Artur Kuprijanov möchte ich für die enge Zusammenarbeit und die programmiertechnische Umsetzung der entwickelten Ideen in das Prozessautomatisierungssystem SIMATIC PCS7 danken. Ohne ihn und seinen unermüdlichen Einsatz wären viele technologische Neuerungen nicht umsetzbar gewesen. Mathias Aehle und Sebastian Schaepe danke ich für das angenehme Arbeitsklima, ihre Unterstützung und ihre Bereitschaft zur wissenschaftlichen Diskussion.

Ich möchte mich bei den zahlreichen Projektstudenten, die ich im Laufe meiner Zeit als Doktorand betreuen durfte, bedanken. Mein besonderer Dank aber gilt allen von mir betreuten Diplomanden, die mit ihrem Einsatz und wissenschaftlichen Arbeiten einen wesentlichen Beitrag zu der vorliegenden Arbeit leisteten. Deshalb herzlichen Dank an Susi Freund, Kathrin Knof, Sebastian Schaepe, Dorina Schäfer, Ireen Schubert und Franziska Wilk.

Auf keinen Fall möchte ich vergessen, mich bei meinen Eltern Margot und Martin Gnoth sowie meinem Bruder Marcel Gnoth zu bedanken. Durch ihre zuverlässige Unterstützung haben sie einen nicht unwesentlichen Anteil an der Vollendung dieser Arbeit gehabt.

Ich bin froh, während meiner Zeit als Doktorand einen ganz besonderen Menschen kennengelernt zu haben. Meine Partnerin Kathrin Gans hat mich mit viel Liebe in allen Lebenslagen begleitet. Sie hatte immer ein offenes Ohr und ließ mich so alle kleinen und großen Probleme meistern – ich danke dir von ganzem Herzen.

Schließlich bin ich stolz und glücklich, dass mittlerweile unsere gemeinsame Tochter das Licht der Welt erblickt hat – willkommen im Leben kleine Prinzessin!

ERKLÄRUNG

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Hiermit erkläre ich weiterhin, dass ich mich mit der vorliegenden Arbeit erstmals um die Erlangung des Doktorgrades bewerbe. Die Arbeit wurde noch keiner anderen Prüfungskommission vorgelegt.

Halle (Saale), den 07. September 2010

Stefan Gnoth