Quality by Design

for

Process Understanding

Multivariate Studies on a Recombinant

E. coli Expression System

Master Thesis

Patrick Sagmeister, BSc

The thesis was performed from September 2010 to April 2011 at the Institute of Chemical Engineering, Vienna University of Technology, under the guidance of DI Patrick Wechselberger and the supervision of Prof. Dr. Christoph Herwig.

Although this may seem a paradox, all exact science is dominated by the idea of approximation.

<div style="text-align: right">Bertrand Russel</div>

Abstract

This contribution exemplifies a quality by design (QbD) approach for the characterization of a high density *E. coli* recombinant upstream process, investigating the multivariate interactions of three critical process parameters (CPPs) in respect to their impact on product quality using statistical experimental design and multivariate modelling. Intracellular production of alkaline phosphatase by means of the little studied positively regulated rhaBAD promoter was chosen as a model system. The multivariate study included the induction phase feeding strategy, expressed as the feed-exponent k (-0.007; 0; +0.007), induction temperature (20°C; 27.5°C; 35°C) as well as induction time (10 to 40 hours of induction) as presumptive critical process parameters (CPPs) under investigation. Based on a sequential approach for upstream design space development, the upstream critical quality attribute (CQA) was defined in respect to downstream demands (specific activity; $U/g_{biomass}$). Optimal operating conditions for the process were found at medium temperatures (27.5°C) and at a positive exponential feeding strategy. The result was the establishment of a knowledge space. The physiological impact of the critical process parameters on the culture was further evaluated by multi-linear regression of yields and specific rates with CPPs. This methodology revealed the protein release, expressed as the biomass specific protein release rate, to be correlating with the applied feeding strategy as well as product quality, making it critical for intracellular protein production. Lower induction temperatures increased the biomass yield, possibly due to a down regulation of recombinant protein expression. The biomass yield declined within induction time, indicating a rise in maintenance metabolism. This was found to go along with an increase in the carbon dioxide yield, independent of temperature and feeding strategy. The multivariate study gave evidence that specific activities are starting to decline after ten hours of induction, although SDS gel electrophoresis indicated that alkaline phosphatase levels were still increasing. Zinc, essential for folding and function of alkaline phosphatase (Torriani 1968), was monitored *via* ICP-OES, which gave evidence for a zinc limitation that explains the observed decrease in activities after 10 hours of induction.

1. Introduction

Innovation in the pharmaceutical industry is currently slowed down by rigid regulatory approaches and post approval regulatory aspects (Woelbeling 2008). The current procedure for determining manufacturing parameter ranges in process development for the registration file is the variation of one parameter at a time, fixing all other parameters at constant values. In surplus to this time consuming exercise, this way of proceeding does not take into account possible interdependencies between the parameters, which may result in quality issues (Woelbeling 2008). In addition, the process can only be operated in a narrow parameter range, limiting the possibilities for process optimization and adaptations to changing supplies. According to the ICH Q8 definition, "quality should be built-in or should be by design" rather than "tested into products" (ICH 2008a). This constitutes a major change in pharmaceutical development, replacing the very product centred traditional approaches with the more process focused and risk-based QbD approach. For the manufacturers, this opened new perspectives regarding the introduction of more flexible and cost effective processes.

Key demand from the side of the regulatory bodies is the science-based proof of process understanding (FDA 2004), achieved through sound pharmaceutical development. Following this major move towards more science for process und product understanding, several guidelines have been published, presenting a new leading strategy (ICH 2009): Based on the quality target product profile (QTPP), which summarizes the quality characteristics of the product, all properties of the drug product affecting its quality are to be identified (critical quality attributes, CQA). Process understanding is achieved, if the multivariate interactions between process parameters and the critical quality attributes are scientifically analyzed. If the variation of a parameter has an impact on a critical quality attribute, it is called a critical process parameter (CPP). If the manufacturer is able to provide evidence, that the CPPs are scientifically identified in course of process development and controlled alongside the manufacturing process, there is the possibility for real-time release, circumventing final product testing. The methodology of gaining process understanding includes process analytical technology (PAT) and quality by design (QbD), a systematic approach aiming at analyzing the multivariate interactions between CQAs and CPPs.

A major milestone for putting QbD into pharmaceutical practice was the introduction of design space (ICH 2008a). The establishment of a design space includes the clear differentiation between critical and non critical process parameters, followed by the scientific investigation of the multivariate interdependencies between the CPPs and the CQAs. The design space is composed of three areas: The knowledge space includes all information

gained during process development and the lifecycle of the product. The design space is an area inside the knowledge space where it has been scientifically proven, that the variation of CPPs in this range does not affect CQAs and therefore the QTPP. The normal operating space covers a small area inside the design space, where the routine manufacturing takes place. If CPPs fall- for any not intended or intended reason- outside the normal operating range into the area of the design space or move inside the design space, the production of product quality is still guaranteed. Besides the possibility of real-time release testing, the enhanced process and product knowledge gained through QbD can pave the way for more efficient and economical manufacturing processes, reducing product recalls compliance procedures and post approval modifications of change (Garcia et al. 2008).

This master thesis focuses on the emerging topic quality by design (QbD), performing a QbD in pilot scale for the understanding of an upstream process to produce a recombinant protein. *Escherichia Coli,* a well explored cell factory for the production of recombinant proteins, was chosen as host for the expression of recombinant alkaline phosphatase *via* the rhamnose inducible rhaBAD promoter. Since the specific activity ($U/g_{biomass}$) of the recombinant protein is possibly critical for a successful and economical subsequent downstreaming, it was defined as the presumptive critical quality attribute of the investigated upstream process.

CPPs such as the pH, temperature, dissolved oxygen tension and initial biomass concentration have been under investigation in recent publications (Rathore et al. 2008; Abu-Absi et al. 2010). Focusing on possible interactions with the energy metabolism in the induced culture, induction temperature as well as the induction phase feeding strategy, expressed as the induction phase exponential feeding exponent k, were under investigation in cause of this contribution. Besides the development of a knowledge space for the system of interest this thesis aims at providing a roadmap about how processed data such as specific rates and yields can be used to achieve process understanding.
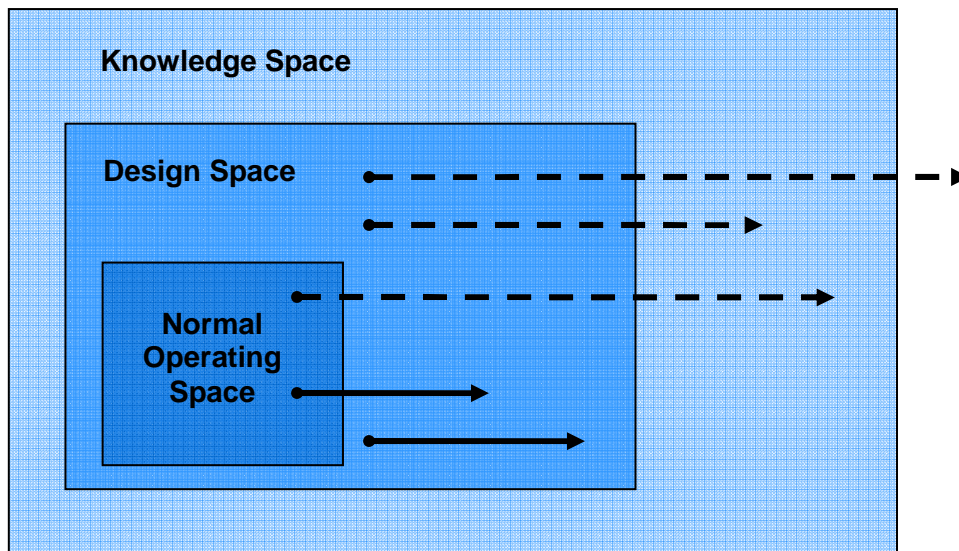
Figure 1: Illustration of the concept of design space. Movements from the normal operating space to the design space and movements inside the design space are not considered a change (drawn-through arrows). Any movement from the design or operating space into the knowledge space or beyond is considered a change and leads to a regulatory post approval change process (dotted arrows).

1.1 System of interest

1.1.1 *Escherichia coli* **for industrial applications**

For several reasons, the relevance of *Escherichia Coli* for life-sciences is exceptional. The rod shaped, gram negative intestinal bacterium served as a model organism ever since the emergence of modern microbiology, gene- and biotechnology, playing a key role for a variety of major scientific achievements such as bacterial conjugation (Lederberg and Tatum 1946), topography of gene structure (Benzer 1961) as well as transformation (Cohen and Chang 1973). Furthermore, the discovery of the endonuclease EcoRI, an enzyme cleaving DNA on defined sites producing sticky ends as well as the demonstration of cloning plasmids containing recombinant DNA into living organisms were both performed involving *E. coli*. These findings were essential for the development of genetic engineering, providing the basis for the biotech-industry. The role of *E. coli* as an omnipresent companion of biotechnology and its related sciences made it a well-characterized organism. The access to a multitude of strain-specific information, including fully sequenced genomes, has obvious advantages. Furthermore *E. coli* is outstanding regarding generation times, safety, ease of cultivation (media demands, possibility for high-density cultures) and the availability of a whole spectrum of genetic techniques for genome modification (Baneyx 1999). Nowadays *E. coli* serves as a host organism for the industrial production of a variety of recombinant proteins, including insulin, β-interferon and a vast number of technical enzymes.

Structure

E. coli cultures are typically comprised of rod-shaped cells measuring 3µm in length and 0.5 µm in diameter showing a volume of 0.6 - 0.7 μm^3. The availability of nutrients was shown to have an effect on cell size (Kubitschek 1990). As a gram negative bacterium, the cell envelope of *E. coli* is composed of two membranes and a peptidoglycan layer (Ruiz et al. 2006). The outer membrane (OM) is composed of phospholipids and lipopolysaccharides and faces the extra cellular space. The inner membrane (IM), solely composed of phospholipids, borders to the cytoplasm. The space spanning in between OM and IM is referred to as periplasmatic space. A layer of peptidoglycan lying between the membranes gives the cell mechanical stability. Figure 2 gives an illustration of the cell envelope of gram negative bacteria.
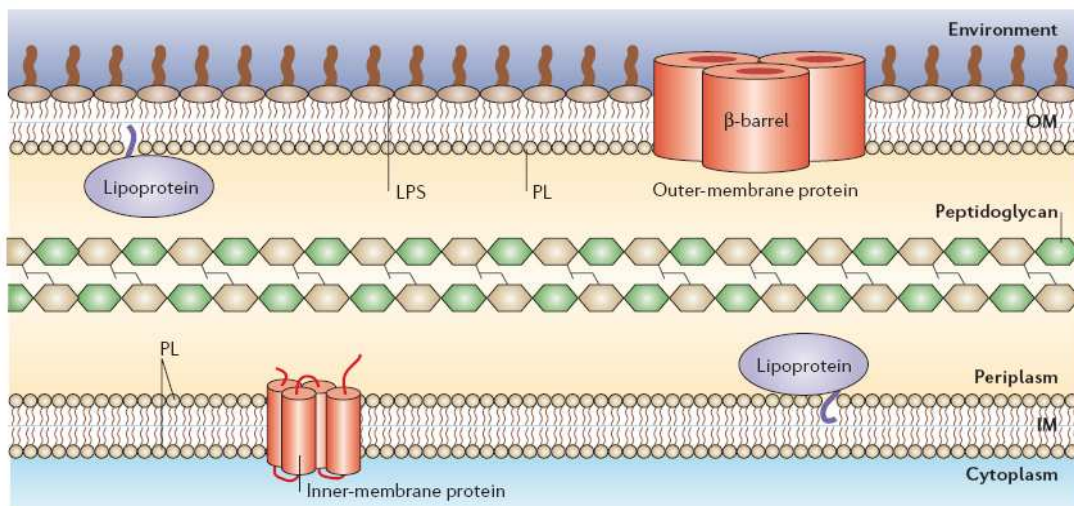
Figure 2: The cell envelope of *E. coli* is composed of two membranes as well as a peptidoglycan layer. While the peptidoglycan layer is fully permeable for molecules of various sizes, transport systems are necessary for the transfer of molecules though the inner- and outer membrane. Illustration is taken from Ruiz et al. (2006).

Fermentation strategies

Industrial processes for recombinant protein production with inducible promoters are carried out in a two step procedure: Before the start of recombinant protein production (induction), biomass is accumulated *via* an exponential fed-batch. Cultivation parameters such as temperature, pH, inducer concentration as well as the applied feeding strategy have a significant effect on productivity and product quality. Understanding of the biology of the process and the identification of biological key issues as well as their relation to process parameters is therefore essential for optimal process design.

State of the art: fed-batch processes for recombinant protein production

The formation of unwanted by-products due to carbon overflow reduces product- and biomass yields significantly and is therefore a major issue in fermentation processes. The control of the substrate available to the cell and consequently the specific growth rate by means of fed-batch processes has developed to become a standard technique for biotechnological fermentations since its first application for the production of baker's yeast in the 1950's. As regards the production of recombinant proteins, the availability of inducible promoters paved the way for the development of two step fed-batch fermentation techniques for maximal productivity: In the first step the culture is kept in a non-induced state and is exponentially grown under C-source limitation to achieve maximum high cell densities without the accumulation of overflow metabolites (Glick 1995). In industrial processes the maximum cell concentration achieved in this step is restricted by considerations regarding cooling of the reactor and oxygen mass transfer into the culture broth (Sanden et al. 2003). The second step is started by inducing the culture. Since the reactor already works at its limits the maximum feeding rate feasible in the post-induction phase is equal to the final feeding rate at the end of the non-induced fed-batch. As biomass tends to increase during induction phase feeding, the industrial two step strategy always goes along with a progressive decline of the specific growth rate (Sanden et al. 2003).

Carbon/energy depletion in post-induction feeding

Following the industrial two-step strategy with linear induction phase feeding, the specific availability of substrate per cell decreases within induction time. Carbon depletion results in the activation of the carbon stress system which is accompanied by the formation of stress proteins and the possible formation of subpopulations of respiring and non-respiring cells, referred to as "cell segregation" (Andersson et al. 1996). Furthermore, the concentration of RNA polymerase as well as the ribosome content are known to be correlated with the specific growth rate (Sanden et al. 2003) and therefore submitted to progressive decrease in cause of induction phase feeding.

Induction-phase limitations

Sanden et al. carried out an extensive studies on the post-induction limitations in recombinant *E. coli* for the production of *β*-galactosidase under high and low post exponential feeding rates (Sanden et al. 2003). Similar values of specific *β*-galactosidase messenger RNA indicated, that the initial production rate for both the high and low level feeding strategies is not limited

by transcription. However, by measurements of rRNA levels representing the ribosome content, translation was identified as initial bottleneck. Higher productivities were obtained applying the higher feeding rate. After 4 respectively 15 hours of induction, a decline of the production rate was observed at high respectively low feed rates. The change in physiological status during induction resulted in formation of guanosine tetraphosphate and acetic acid.

Plasmid loss

Another issue in fermentation processes aiming at recombinant protein production is plasmid instability. Next to the obvious loss of recombinant protein production capacity the non-recombinant cells show a reproduction advantage compared to plasmid bearing cells which can lead to overgrowth of the whole population (Kyslik et al. 1993; Summers 1991). Studies indicate that plasmid stability is enhanced in minimal medium compared to rich medium (Striedner 2001). Cells carrying a plasmid were reported to show reduced growth rates compared to non-transformed cells (Summers 1991). In absence of selective pressure, the plasmid can be lost, leading to overgrowth of the non-transformed cells and therefore reduced productivity.

Acetate production

Although growing under aerobic conditions, *E. coli* is known to produce significant amounts of acidic by-products under certain conditions. In analogy to the ethanol production of baker's yeast under aerobic conditions this is referred to as „bacterial Crabtree effect" (Rinas et al. 1989). In particular this can account for the significant formation of acetate. Andersen et al. (1980) suggested that the production of acetic acid under aerobic conditions is caused by an imbalance of glucose metabolism and cell respiration. If the glucose flux inside the cell exceeds a certain critical value, TCA activity is limiting and the organism may change to uncoupled metabolism (Luli and Strohl 1990). The protonated form of acetate, acetic acid, shows a slight lipophilic character, which enables transport through the membrane from the extracellular space into the cell (Luli and Strohl 1990). ATP synthase, a transmembrane protein located in the cell membrane, makes use of the H^+ gradient between intra- and extracellular space. The intracellular dissociation of the acetic acid in the cell interior decreases the intracellular pH and reduces the pH gradient used for ATP synthesis. The protonation of the acetate in the extracellular buffered environment has no significant effect on the overall pH of the medium. Therefore, if acetate is present in the medium, a net H^+ flux into the cell is assumed, counteracting the pH gradient contributing to ATP synthesis (Luli &

Strohl 1990). Moreover, acetate has a negative effect on recombinant protein production as well as on RNA, lipid and protein synthesis (Lee 1996), with a more detrimental impact on recombinant than non-recombinant cells (Koh et al. 1992). Several strategies with the goal of reducing acetate production can be found in literature. Due to the extent duration of fed-batch processes, acetate production is a key issue with detrimental effects on cell growth and protein productivity. Acetate formation can be suppressed by limiting and controlling the specific growth rate through the fed-batch feeding profile. Metabolic activity can also be degraded by lowering the cultivation temperature, resulting in a reduced nutrients uptake, cell growth, cellular oxygen demand, metabolic heat, toxic by-product production and acetate formation (Lee 1996). The reduced activity of proteases might contribute to higher protein productivity as well. The decreased oxygen demand at reduced temperature (26-30°C) makes high density cultures feasible without the application of pure oxygen (Lee 1996). Furthermore the reduced formation of inclusion bodies for cultures grown at low temperatures is reported (Kane 1988).

Metabolic load

The production of recombinant proteins interferes strongly with the physiology and biochemistry of the host cell. In recombinant production systems a significant fraction of the cell resources, such as ATP, GTP and amino acids, is directed towards the conservation and expression of recombinant DNA, a phenomenon referred to as "metabolic load" (Glick 1995). The severity of "metabolic load" or "metabolic burden" was shown to increase with plasmid size and plasmid copy number, resulting in a reduced relative specific growth rate and a reduced relative specific activity per vector (Birnbaum and Bailey 1991; Seo and Bailey 1985; Glick 1995). Furthermore the effects of metabolic load were demonstrated to be more adverse in minimal medium compared to rich medium (Seo and Bailey 1985) and to increase in oxygen limited conditions (Hopkins et al. 1987). In case the expressed recombinant protein is translocated from the cytosol, blocking of transport systems by the target protein can cause cell death (Skare et al. 1989). Another issue in context with metabolic load is the secretion of vast amounts of carbohydrates, referred to as "extracellular slime". This was reported to cause difficulties in cross-flow microfiltration in recombinant *E. coli* (Wood 1988). Multiple genetic approaches for the reduction of the metabolic load were reported, e.g. by reducing the number of plasmids (Seo and Bailey 1985), engineering of metabolic pathways (Flores et al. 2004), the application of non-ampicillin antibiotic resistance genes or the insertion of the recombinant DNA into the chromosome of the host organism (Hong et al. 1995; Glick 1995). Recently, the coupling of the expression of an essential gene to the presence of a plasmid was

presented as an alternative for the use of antibiotic resistance genes (Mairhofer et al. 2008). Following this strategy, a significant decrease in the metabolic load can be expected, since antibiotic resistance gene expression is not required. Fermentation technology and bioprocess development may offer complementary strategies for the reduction of metabolic load, such as the promoter down regulation *via* inducer concentration variation, reduction of temperature or the application of suitable feeding strategies. In order to achieve maximum productivity of the process and maximum quality of the desired product, the fermentation strategy should be tailored to the production host, the recombinant product and the applied promoter.

1.1.2 The model protein: alkaline phosphatase (AP)

Alkaline phosphatase (EC 3.1.3.1) is an ubiquitary hydrolase enzyme catalyzing the nonspecific phosphorylation and dephosphorylation of phosphomonoesters under alkaline conditions (see figure 3). Its biological function lies in the generation of phosphate under phosphate depletion (Horiuchi 1959). AP from *E. coli*, a homodimeric metalloenzyme, is coded chromosomally by the *phoA*-Gene and exported *via* the general secretory pathway (GSP) into the periplasm where it develops its active dimeric form.
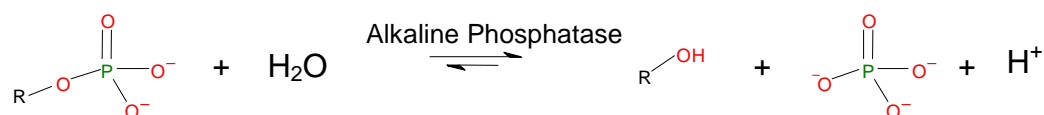

Figure 3: Phosphorylation and dephosphorylation of organic components under alkaline conditions by alkaline phosphatase. Hydrolization of a phosphomonoester yields one proton. Therefore the forward reaction requires alkaline conditions, shifting the equilibrium to the alcohol and the free phosphate.

Dimerization demands translocation into the periplasmatic space

Dimerization to the active enzyme requires the prior formation of disulfide bonds within the monomers (McCracken and Meighen 1980). In the cytoplasm of *E. coli*, to which the monomer is released from the ribosome, two thioredoxins (TrxA and TrxC) and three glutaredoxins are present which are theoretically capable of disulfide bond formation (Rietsch and Beckwith 1998). However, the presence of thioredoxin reductase (TrxB) and glutathione keeps the cytoplasmatic environment in a reduced state, repressing any oxidative activity. Disulfide bridges can only be formed in the periplasmatic space, where a more favourable redox environment along with the Dsb system leads to disulfide bridge formation (Bardwell et al. 1991). Consequently, transport through the inner membrane to the periplasmatic space is essential for final folding.

Translocation is achieved *via* the general secretory pathway

Translocation is achieved *via* the general secretory pathway (GSP), a transportation cascade involving a variety of proteins aiming at the secure transport of the pre-protein into the periplasmatic space. In the following, the key elements of this pathway are outlined (Sommer 2008). Proteins intended for translocation *via* the GSP are released from the ribosome carrying a secretion signal sequence, keeping the unfolded state of the amino acid sequence (MacIntyre et al. 1991). The transport-competent form is further stabilized by the binding of a trigger factor (Beck et al. 2000), which also enables the binding of the pre-protein to the chaperone SecB. The next protein involved in the GSP is the homodimeric shuttle protein SecA, located at the inner membrane and providing a binding site for the SecB/protein complex (Miller et al. 1998; Fekkes et al. 1997). Transport through the membrane takes place across the translocation pore SecYEG (Duong et al. 1997), which is coated by the trans-membrane proteins SecY and SecE, preventing hydrophobic interactions between the pre-protein and phospholipids (Joly and Wickner 1993). The active transport is driven by ATP hydrolysis and the proton motive force causing repeated insertion/deinsertion of SecA moving the target protein gradually through the membrane (van der Wolk et al. 1998; Economou 1999; Matsumoto et al. 1997). The secretion signal sequence is cleaved by a membrane-attached signal peptidase (Paetzel et al. 1998). Therefore the molar mass of the unprocessed monomer (49 kDa) differs from the molar mass of the processed monomer (46 kDa). Released into the periplasmatic space alkaline phosphatase starts folding and reaches its +native quaternary structure after disulfide bridge formation and dimerization, shown by figure 4. In contrast to the transport through the inner membrane, no protein transport system for the export of periplasmatic proteins into the extracellular space is reported (Sommer 2008).
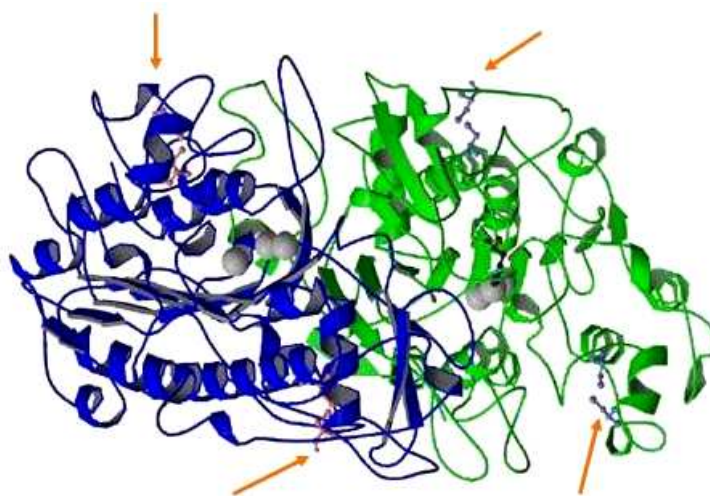

Figure 4: Quaternary structure of alkaline phosphatase from *E. coli*. The two identical subunits are marked in green and blue, disulfide bridges are highlighted with arrows (www.expasy.org).

Catalytic mechanism

The reported catalytic mechanism (Coleman 1992) of alkaline phosphatase involves two steps. In the first step the active centre of AP interacts with the phosphate group so that a serine residue (Ser102) finds itself in an apical position for a nucleophilic attack on the phosphorus, leading to a phosphate-enzyme intermediate (see figure 5). The preceding interaction involves the coordination of the phosphate *via* two zinc ions creating a phosphate bridge and *via* the formation of H-bonds between the guanidium group of Arg166 and the phosphate group. In the second step the previously formed phosphoseryl intermediate is hydrolyzed by a zinc coordinated water or hydroxide molecule.
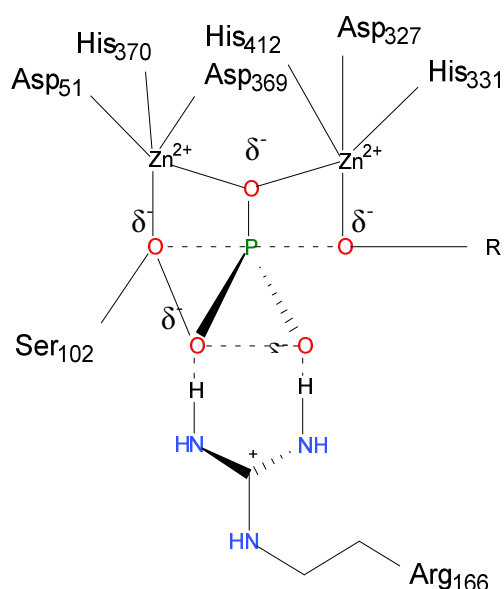

Figure 5: Phosphate-enzyme intermediate according to the transition-state model (Coleman 1992; Zalatan et al. 2007).

Relevance of metal ions for folding and function

Whereas the release of AP monomers from the ribosome was proven to be independent from the Zn^{2+} concentration, the subsequent folding in the periplasm is strongly inhibited under Zn^{2+} limitation (Torriani 1968). Although accumulated AP monomers can be re-dimerized *in vivo* as well as *in vitro* under elevated Zn^{2+} concentrations and moderate temperatures, their full activity can not be regained, which was probably due to instability (loss or denaturation) of free monomers (Torriani 1968). Noteworthy, *E. coli* cells are reported to possess the ability to accumulate intracellular Zn^{2+} against an existing concentration gradient (Torriani 1968). Magnesium is not essential for the catalytic activity of alkaline phosphatase, but decreased catalytic activities are found in the absence of Mg^{2+}, probably due to reduced interaction between the subunits (Orhanović and Pavela-Vrančič 2003). Although zinc and magnesium

can be substituted by other divalent metal ions, this goes along with a severe loss in catalytic activity (Wang et al. 2005).

Oligomeric quaternary structures of alkaline phosphatase

Three oligomeric forms of alkaline phosphatase are reported. Besides the earlier mentioned active dimers and inactive monomers, AP also comes in tetrameric form (Thomas et al. 1980), showing three to four times reduced activity compared to the dimer (Atyaksheva et al. 2007). These three oligomeric forms stand in equilibrium with each other, which is reported to be dependent on temperature, pH, buffer and enzyme concentration (Atyaksheva et al. 2007). Under elevated enzyme concentrations above 0.5 mg/L the equilibrium shifts toward tetramers; at acidic conditions and elevated temperatures monomers are prevailing (Zappa et al. 2004; Atyaksheva et al. 2007). Since the three oligomers show different specific activities, a change of equilibrium results in a change of overall observed specific activity.

$$AP_4 \Leftrightarrow 2AP_2 \Leftrightarrow 4AP_1$$

Figure 6: Three oligomeric forms of alkaline phosphatase stand in equilibrium with each other. (AP_4=Tetramer, AP_2=Dimer, AP_1=Monomer)

1.2 Quality by design for upstream process understanding

1.2.1 Positioning of the thesis within the framework of QbD

The first step for implementing QbD in pharmaceutical product development should be initiated after the discovery of a new and promising pharmaceutical drug (Rathore 2009). The quality characteristics ensuring a safe and efficacious use of the potential drug candidate should be thoroughly investigated and summarized in a quality target product profile (QTPP). Table 1 gives suggestions of quality attributes and characteristics that can be included in the QTPP.

Table 1: Possible quality included in the QTPP (Rathore 2009).

general attributes	pharmacokinetic characteristics	quality attributes
dosage form	aerodynamic performance	purity
administrative route	dissolution	sterility
therapeutic dose		
active substance release		

After the definition of the QTPP all microbial, chemical and biological characteristics or properties of the drug product should be assessed whether they have an impact on the QTPP.

If so, it is named a critical quality attribute (CQA) and has to be kept within a certain range, distribution or limit. In order to establish the product design space, the clear differentiation between critical and non critical quality attributes is necessary. This can be achieved using risk assessment tools (ICH 2006). Once the CQAs are identified and their acceptable ranges are determined, a multidimensional space with every CQA serving as one coordinate can be spanned, the so-called product design space. Possible interactions between the CQAs must be taken into account. The process design space is an essential part of the holistic QbD approach for pharmaceutical product development and should be included in the registration file (Rathore 2009). Data acquired through clinical and non clinical studies (*in vivo* assays, binding assays, cell-based assays) effect the size of the product design space (Mire-Sluis et al. 2004; Gupta et al. 2007; Barrett et al. 2007). Furthermore process robustness, applied process analytical technology (PAT) as well as the level of understanding of the multivariate interrelationships between CQAs and the QTPP determine the range of the design space (Rathore 2009). Hence, the more knowledge is gained in cause of product- and process development and the more data on the process is available in cause of production (PAT), the bigger the process design space can be laid out the more flexibility we gain in the manufacturing process. Table 2 summarizes the factors affecting the size of the product design space, subdivided in drug attributes and process characteristics.

Table 2: Attributes affecting the design space (Rathore 2009).

drug attributes	process characteristics
stability of the drug substance	process robustness
clinical data	process capability
non clinical data	understanding of the CQA/QTTP interrelationships
	PAT

After the identification of the product design space, process characterization studies (Rathore 2009; Harms et al. 2008) can be performed investigating the multivariate interdependencies between CQAs and the process parameters. The QbD approach demands the clear differentiation between non critical and critical process parameters (CPPs). Tools for the examination of process parameters are summarized in table 3.

Table 3: Tools for CPP and CQA assessment (Harms et al. 2008; Huang et al. 2009).

FMEA (Failure Mode Effect Analysis)
DOE (Design of Experiments)
Risk Filtering
Scale Down Modelling
Fishbone Diagram

According to Huang et al. (2009), one way to determine the criticality of quality attributes and process parameters is to pin down a manageable list of possible CQAs and CPPs using quality risk assessment (QRA) tools (FMEA, risk filtering, fishbone diagram), followed by screening DOEs. The process design space can then be established using process optimization DOEs in combination with multivariate techniques like principal component analysis or regression analysis using partial least squares (Huang et al. 2009).

The next step for the successful implementation of QbD constitutes in the definition of a control strategy. Per definition a control strategy is a "planed set of controls, derived from current product and process understanding that assures process performance and product quality" (Nosal and Schultz 2008; Garcia et al. 2008b). Following QbD principles, a dynamic control strategy can be applied (Rathore 2009). In contrast to fixed manufacturing controls in a traditional control strategy, this offers the possibility to adapt the manufacturing process to variable input factors as long as the variability takes place within the design space. According to the QbD approach, the product and process design space as well as the control strategy should be part of the regulatory filing. Before submitting the process filings to the regulatory authorities, it must be demonstrated that a quality drug product is guaranteed when working within the design space. As the design space is usually developed in bench or pilot scale, the developed models must be validated in full scale (Rathore 2009). Typical filing documents are given in table 4. The right column gives suggestions for additional information gathered within the framework of QbD, opening the possibility for future changes in the manufacturing process, without the necessity for tedious post approval regulatory proceedings.

Table 4: Typical filing documents (Rathore 2009).

general filing documents	additional information
description of product design space	comparability protocols
characterization of control strategy	expanded change protocols
proof of validation	

Settled within the complex framework of QbD, this thesis tries to establish an upstream process design space with a quality attribute (intracellular product activity/biomass) relevant to the subsequent downstream operations. This constitutes a sequential view on QbD, following the approach that the CQA of one unit operation is part of the CPPs of the following unit operation. This way of proceeding demands prior assessment of downstream requirements, capabilities and capacities, which impacts the acceptable ranges of the upstream CQA(s) and upstream design space.
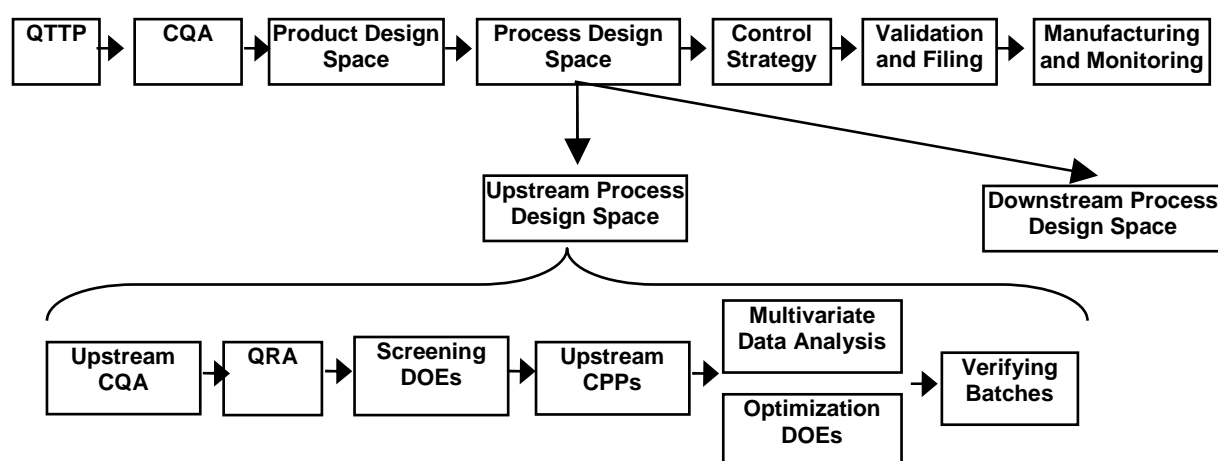

Figure 7: The QbD strategy for pharmaceutical product development and production.

1.2.2 PAT within the framework of QbD

According to the FDA definition, process analytical technology (PAT) is a "system for designing, analyzing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality" (FDA 2004). Hence, PAT provides the toolset for the identification, in-process measurement and control of CPPs or CQAs and can therefore be considered an *ab-initio* companion of quality by design. As the name implies, the PAT strategy aims at ensuring product quality by testing the process performance and identifying possible deviations that might alter product quality. The paradigm shifts away from final product testing towards a more process centered testing approach is illustrated in figure 8 and figure 9 (Hussain 2001).
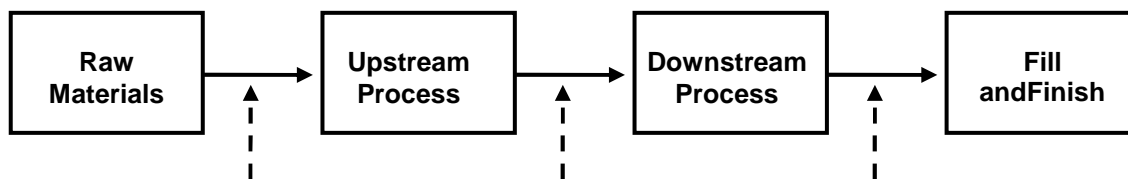
Figure 8: The traditional control- and testing strategy regards the processes as black boxes and only tests for the process outcome, as illustrated with dashed arrows.
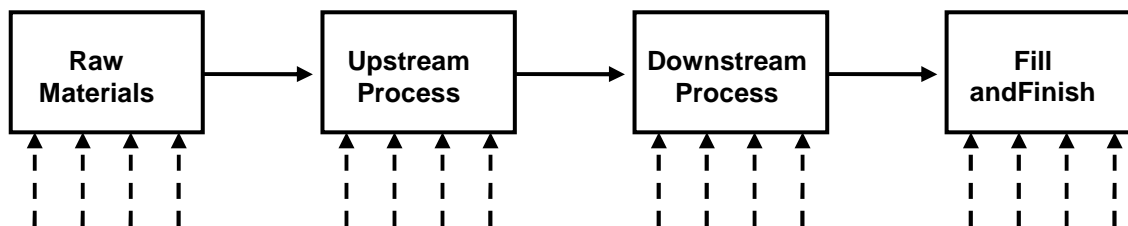

Figure 9: Within the PAT approach the CPPs of the (sub-) processes are continuously monitored. Since the multivariate relationships between CPPs and CQAs are investigated within the QbD process development, quality can be assured if the CPPs stay within the predefined design space. Alternatively CQAs can be directly monitored if appropriate analytical techniques are available. Dashed arrows hint to possible points of application for PAT tools.

1.2.3 Design of experiments as a tool for QbD

Since possible memory effects constrain the applicability of pulse- and transient experiments, process characterization for biotechnological processes demand a high number of fermentation runs; hence experimental work that is time- and labour intensive. Extracting the maximum amount of information from a minimum amount of experiments is therefore a key issue for biotechnological research and development. Design of experiments (DOE) provides a toolset for the statistical layout and the subsequent statistical evaluation of experiments in a predefined experimental range. DOE can be used for screening-, optimization- as well as robustness testing objectives (Eriksson 2000). The basic principle of DOE is the correlation of factors, hence the experimental conditions that are under investigation, with the target experimental outcomes, referred to as responses. Following this scheme the experimental process is regarded as a black box, as illustrated in figure 10.
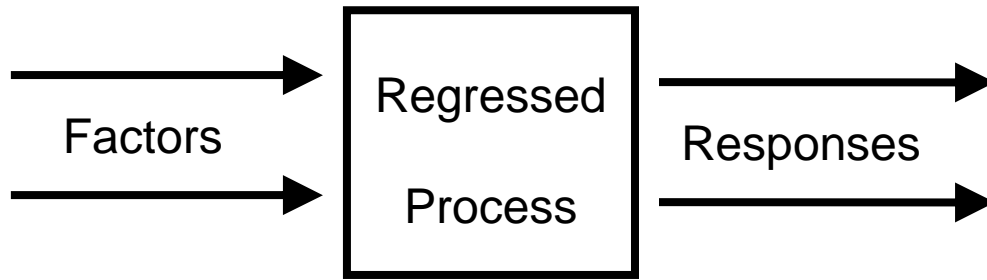
Figure 10: The black box nature of DOE. The mathematical correlation of factors and responses (regressed process) is called a model.

The mathematical correlation of factors and responses is referred to as a model. Linear models span a plane in the experimental area and take into account the linear effects of the investigated factors on the responses. Interaction models allow the estimation of factor/factor interactions; hence their interdependencies. The positioning of the experimental runs in the region under investigation is referred to as experimental design. The choice of an appropriate experimental design is determined by the requirements of the experimental objective. In screening tasks the experimenter deals with a high number of factors and needs to determine whether they have an impact on the responses or not. In this case the experimenter will readily make concessions regarding the applicable model for the benefit of a reduced number of experimental runs. On the other hand, for optimization or robustness testing the interactions of the different factors might play a key role and a more elaborated model might be needed.
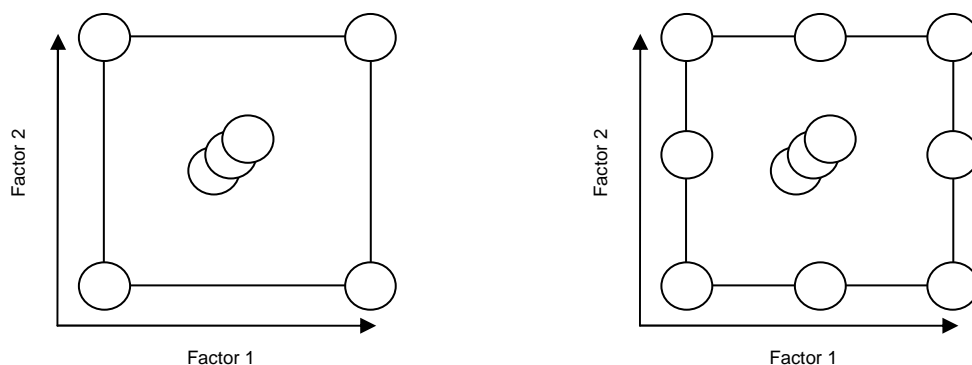

Figure 11: The left picture shows a screening design with four experimental points spaced equally around a center point which is performed in triplicate for the estimation of the experimental error. If the outcome should be a more elaborate model, such as in optimization or robustness testing objectives, more experiments are needed. A typical two factor design for optimization or knowledge space establishment objectives is shown on the right.

Within the framework of QbD, DOE screening designs can be used for the identification of CPPs. In the subsequent step, the investigation of the multivariate relationships between CPPs and CQAs which lead to the establishment of the knowledge-, design- and operating space, response surface modelling constitutes the only acceptable tool since one factor at a time approaches do not take into account possible factor/factor interactions. The regression between CPPs and CQAs is referred to as knowledge space, a mathematical display of the design region investigated. The establishment of the knowledge space demands the use of quadratic models, since the reality modelled can but not necessarily has to comprise quadratic relationships which are in need of display (figure 13).
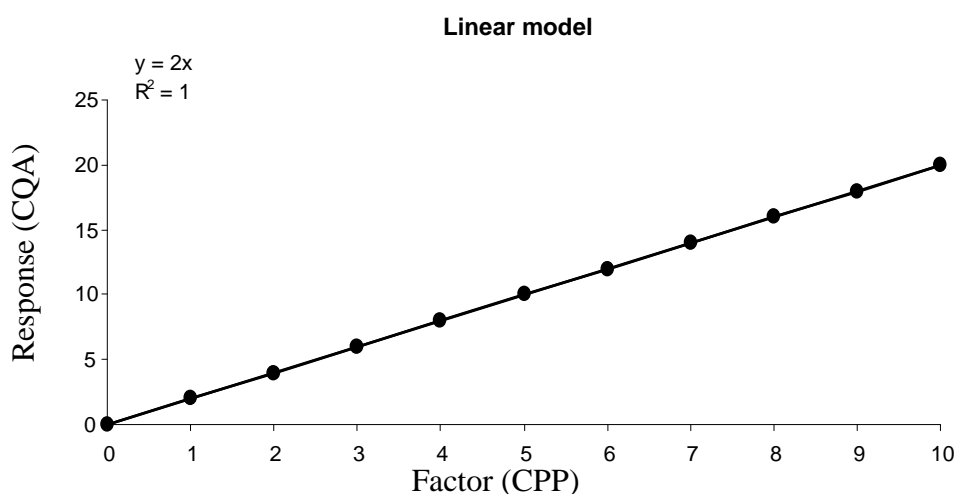
Linear model

$y = 2x$
$R^2 = 1$

Figure 12: The linear relationship between a factor and a response can be regressed using a linear term.

Quadratic model

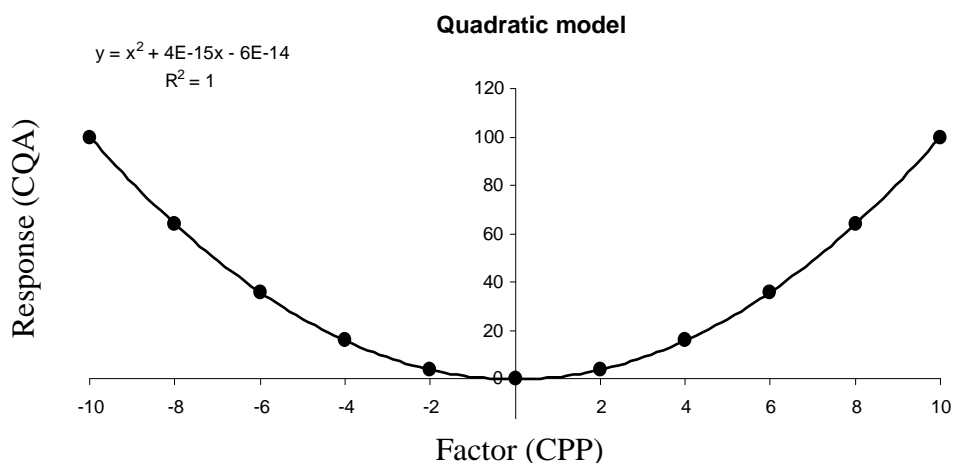$y = x^2 + 4E\text{-}15x - 6E\text{-}14$
$R^2 = 1$

Figure 13: The display of quadratic relationships demand the use of quadratic model terms.

The model terms for CPP and CQA regression can be chosen freely by the operator, but the inclusion of additional model terms does not necessarily result in a better model. The most important parameters for the assessment of the model fit are the *goodness of fit* parameter R^2

and the *goodness of prediction* parameter Q^2 (Eriksson 2000). R^2 gives the fraction of variation explained by the model and spans from 0 (no variance explained by the model) to 1 (all variance explained by the model). The inclusion of additional model terms helps to raise R^2, but this might impair the predictive power of the model. An important parameter for the assessment of the predictive power of the model is Q^2, which also spans from 0 to 1. Q^2 can be used for model tuning, since it decreases if non-significant terms are included in the model. Analysis of variance (ANOVA) provides the basis for the computation of the model validity.

1.2.4 Data exploitation methodology for QbD: calculation of yields and rates

Fermentation runs in fully instrumented bioreactors provide the researcher with huge amounts of offline- and online data. Processing the obtained data into specific rates and yields allows the extraction of scale-independent information on the fermentation process. For induction phase quantification cumulative rates served as the basis for the calculation of cumulative yields and balances. Cumulative rates can be useful to reduce the effect of measurement noise, since the calculation window is always at maximum for a given timepoint. Cumulative rates are sensitive to the starting value, because it is integer part of all the following calculations.

1.2.5 Process analysis: multilinear regression of specific rates and yields with CPPs

The establishment of the knowledge space demands the multilinear regression of pre-defined CPPs with pre-defined CQAs and provides information about their interaction. From this black box correlation no knowledge about the underlying biological mechanisms is accessible. Hence, the understanding of CPP/CQA interactions provided by the knowledge space is of a pure statistical nature and its biological interpretation remains speculation. Generally, only a small fraction of the data obtained during process development is processed for the establishment of the knowledge space. The majority of data routinely recorded during process development, such as off-gas analysis and balance signals, is not taken into account. Processing recorded data into the condensed form of specific rates and yields offers the possibility to extract scale independent information from the huge data set available. This information, obtained from each individual fermentation run, can be used to gain understanding of the CPP's impact on the CQA beyond the statistical regression provided by the knowledge space: The suggested methodology makes use of the multivariate study performed for the establishment of the knowledge space *via* subsequent multivariate linear regression of specific rates and yields with CPPs. The data processing cascade involves the

calculation of specific rates and yields out of online signals (off-gas analysis, balances) and offline data (extracellular protein concentration, biomass dry weight). The data necessary for further processing, displayed in table 5, is routinely recorded in bioprocess development. Hence, the extraction of information is solely a calculation exercise and does not demand the conduction of further experiments or the establishment of elaborate analytical procedures.

Table 5: Data necessary for data processing.

Quantified Analyte	Analytical Procedure	Type of Data
Base Addition	Balance	Non Invasive
Feed Addition	Balance	Non Invasive
Biomass Dry Cell Weight	Balance	Offline
Extracellular Protein	BCA	Offline
CO_2 Off-gas	NDIR	Non Invasive
O_2 Off-gas	Paramagmetic Sensor	Non Invasive
Air Flow in Reactor	Mass Flow Controller	Non Invasive
Oxygen Flow in Reactor	Mass Flow Controller	Non Invasive
Reactor Volume	Balances	Non Invasive
Molarity of Base	Titration	Prior to Fermentation
Feed Concentration	Density Correlation	Prior to Fermentation

The analytical methods are completely real-time with the exception of biomass dry weight and extracellular protein determination. A general scheme of the methodology is illustrated in figure 14. The statistical regression between CQAs and CPPs is referred to as knowledge space. Data exploitation of individual DOE runs result in the calculation of specific rates and yields. In analogy to the CQA/CPP regression performed for the establishment of the knowledge space, CPPs can also be regressed with specific rates and yields, where the CPPs serve as factors and specific rates and yields as responses. This methodology provides insight in the CPP's impact on the biological system. The obtained process knowledge can be used for a future control strategy or the demonstration of process understanding in the regulatory filing. The suggested methodology can also be applied to historical data, granting more insight in existing processes and opening new perspectives for advanced control strategies.
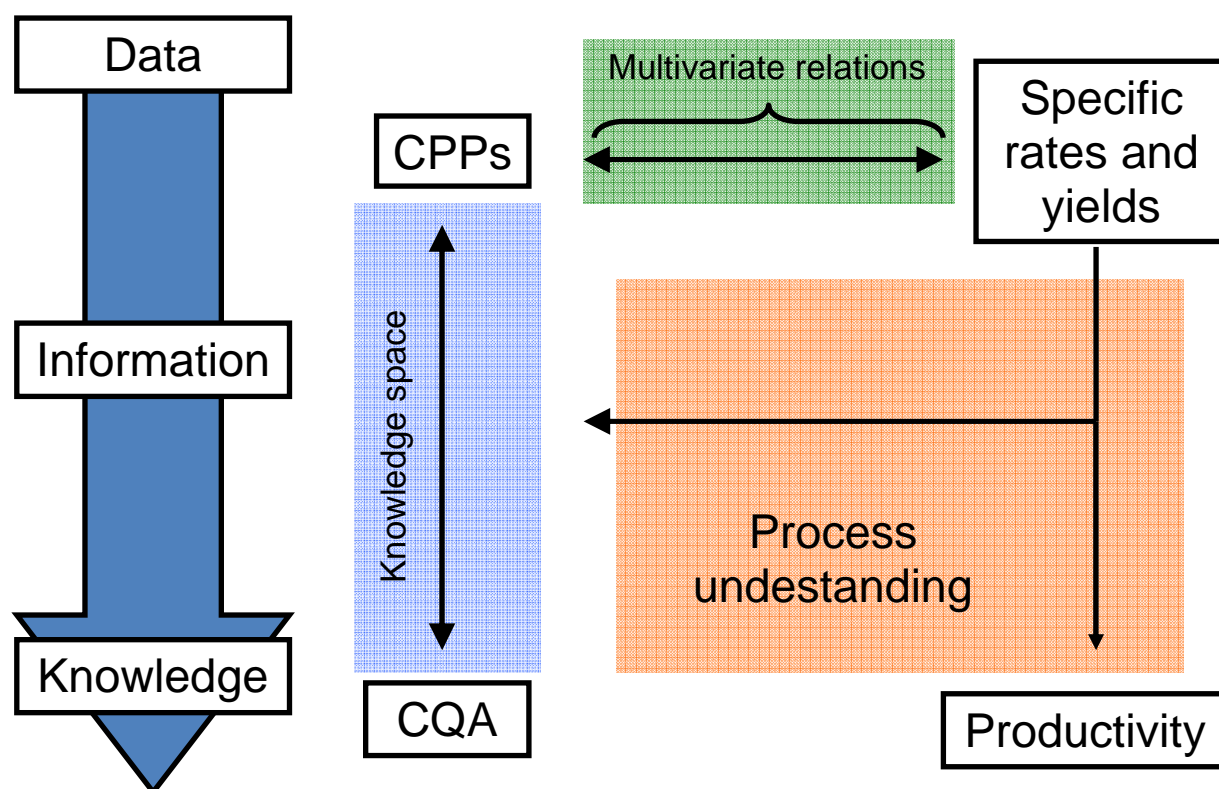
Figure 14: Roadmap to process understanding. Data is referred to as raw data obtained preferably online during the fermentation run. Information is created by processing of data into scale independent specific rates and yields. The correlation of the CPPs with specific rates and yields as well as the direct correlation of specific rates and yields with the CQA can be considered process knowledge, granting insight in the mechanisms of the biological system.

1.2.6 Bridging the gap between CPPs and CQAs by means of specific rates and yields

The presented QbD methodology for process understanding is to be regarded complementary to knowledge space development, offering a comprehensive roadmap for data exploitation and knowledge presentation that allows the understanding and interpretation of the knowledge space. From a mechanistic point of view, a critical process parameter does not directly affect the CQA, but triggers or changes multiple biological or chemical mechanisms that lead to a change in the CQA. What is observed as response of the knowledge space DOE is the combined impact of the individual affects impacting the CQA. Statistical exploitation of specific rates and yields offer a possibility to open up this hermetically sealed black box: By correlating CPPs with specific rates and yields, the CPP's impact on specific rates and yields is analyzed. If appropriate specific rates and yields are accessible, cumulative effects of CPPs observed in the knowledge space can be sorted out. For example, temperature might have a

positive effect on the expression rate of recombinant protein, so more protein is expressed at higher temperatures. On the other hand, the formation of inclusion bodies at elevated temperatures can decrease the active protein yield, hence the yield of active recombinant protein per total recombinant protein formed. If the specific activity is set CQA, the regression of the CPP temperature and the CQA specific activity reveals only the cumulative effect of the two mechanisms involved in the temperature dependence of the CQA, in this case a quadratic relationship. More insight in the biologic mechanism is granted if one investigates the recombinant protein expression rate and the active protein yield, which constitutes the mechanistic link between the critical process parameters varied in cause of the DOE and the observed critical quality attribute.

1.2.7 Use of ICP-OES and SDS-PAGE gel electrophoresis

Orthogonal analytical methods confirm the knowledge created within process development. The use of standard protein analytical techniques such as SDS-PAGE gel electrophoresis provides valuable qualitative information on the investigated process. ICP-OES as multi-element analytical technique can be used for the detection of trace-element limitations or unwanted metal ion accumulations in fermentation processes.

1.2.8 Knowledge space for the subsequent downstream operation: homogenizing

High pressure homogenizing constitutes a scaleable and reproducible method for cell disruption. Cell rupture is part of the subsequent downstreaming of the drug product and attention should be paid to the parameters of this unit operation that might impact product quality. Next to the key homogenizing parameters homogenizing pressure and number of passages (Kuboi et al. 1995), the effect of the biomass concentration on the activity obtained after homogenization was in focus of investigation.

1.3 Key elements of the study at a glance

- Statistical process understanding: multivariate study for the establishment of a knowledge space for a recombinant upstream process

- Roadmap to process knowledge: investigation of CPP/CQA interactions by means of multilinear regression of specific rates/yields with CPPs for the understanding of the CPPs impact on the biological system

- Use of ICP-OES and SDS-PAGE gel electrophoresis for the investigation of further process related issues

- Parameter investigation of cell disruption *via* high pressure homogenizing

2. Materials and methods

2.1 Strain

A proprietary *Escherichia coli* strain containing plasmids encoding for recombinant alkaline phosphatase was used. Recombinant protein production in the applied expression system is under the control of the rhamnose inducible rhaBAD promoter. Rhamnose addition was only necessary once for induction, since the rhamnose metabolism was genetically impeded in this strain (Wilms et al. 2001). Master, intermediate and working strains were stored at -80°C.

2.2 Media

Media according to Wilms et al. 2001 were used.

2.3 Bioreactor setup

The multiple fed-batch runs were performed on a Techfors-S pilot bioreactor (Infors, Bottmingen). In the following the main attributes of the Techfors-S bioreactor are outlined: The in-situ sterilizable stainless-steel vessel hosts a working volume of 10 liters and possesses multiple ports for online, inline and at-line analytics. The SOP for the steam generator used for in-situ sterilization (Infors, Bottmingen) is annexed in the appendix. Stirring is realized by a 50 to 1500 rpm „disk" type impeller, powered by an engine on top of the vessel, connected to the sterile interior by an end face mechanical seal. For sterile and effective gassing into the medium the inlet air is filtered by a membrane-type filter and dispensed by a ring sparger. To minimize water stripping the outlet air is cooled by a high efficiency exit gas cooler. Gas flow rates are measured by a thermal mass flow controller (Vögtlin Instruments, Switzerland). For temperature control a deep-profile double jacket supplied by an external chiller and an electrically powered heat exchanger is applied. For sampling and harvesting, the reactor is equipped with a steam-sterilizable bottom valve. Digital peristaltic pumps are used for pH adjustment. The Techfors-S bioreactor allows direct control of cultivation parameters such as pH, O_2 inlet flow, air inlet flow, pressure, temperature, agitation speed and DO_2 *via* an adjoining touch screen control panel.
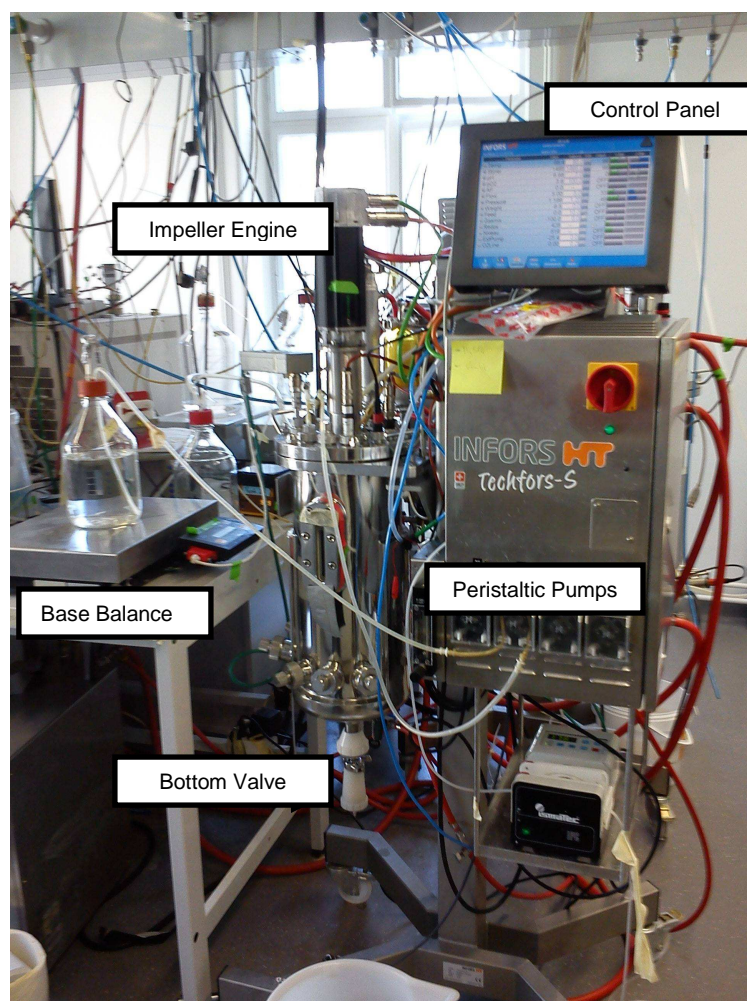
Figure 15: The Techfors-S Bioreactor. Impeller engine, membrane-type inlet air filter, control panel, digital peristaltic pumps and the steam-sterilizable bottom valve are highlighted

2.4 Bioreactor instrumentation

Quantification of base- and substrate feeding was achieved gravimetrically. Base- and feed solutions were placed on balances (Sartorius, Germany) and the weight loss was recorded. A peristaltic analogue pump (Lamda, Switzerland) assembled with silicon tubing was used for feed addition. Base addition was achieved *via* the Techfors-S integrated digital peristaltic pump assembled with Marpren and silicon tubings (3.2mm diameter; 0.8mm inner wall). Quick connect couplings (Roth, Germany) were used for sterile connections. The ports on the top plate of the reactor were used for a dissolved oxygen sensor (Hamilton, USA), pH probe (Yokogawa, Japan), pressure sensor (Keller, Switzerland), a septum and an overpressure valve (Infors, Switzerland). CO_2 and O_2 quantification in the off-gas stream was conducted *via* a gas analyzing unit (Müller Systems AG, Switzerland) based on NDIR (CO_2) and paramagnetic (O_2) principle. All the signals were recorded by the process management system Lucullus (Biospectra, Switzerland).

2.5 Fermentation parameters

The fermentation process followed the industrial two-step fermentation strategy. Fist, a pre-culture was grown in a 1000ml baffled Erlenmeyer flask containing 70ml of pre-culture medium for 8 hours at 35°C. After inoculation (20g/l glycerol) the C-source in the batch medium was consumed within 12 hours. At the time point when the CO_2 off-gas signal started to decrease, an exponential fed-batch for the accumulation biomass was started. The feed density was gravimetrically determined after sterile sampling of the feed solution in a laminar flow working bench and correlated with the concentration by means of linear regression. After 9 hours of fed-batch the induction was started by sterile addition of rhamnose. In dependence on the performed experiment, the post induction feed was carried out linear, exponential or negative exponential. For the latter the positive exponential feed was mirrored to the down side. The applied feeding profiles are displayed in figure 16.
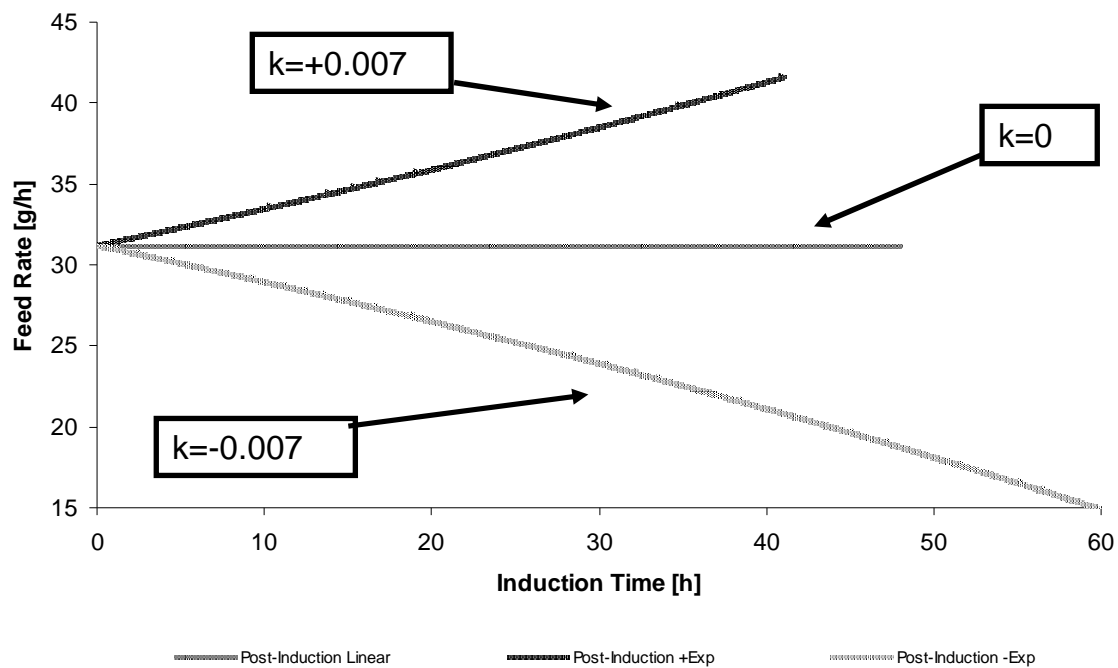

Figure 16: The applied feeding profiles.

$$F_{(t)} = F_0 * e^{k*t}$$

Formula 1: Calculation of the positive exponential induction phase feed rate. The induction phase feed exponent was included in the DOE.

$F_{(t)}$ Feed Rate as a Function of Induction Time [g/h]

F_0 Feed Rate before Induction   [g/h]

k   Induction Phase Feed Exponent   [h^{-1}]

$$F_{(t)} = 2F_0 * - F_0 * e^{k*t}$$

Formula 2: Calculation of the negative exponential feed rate: The positive exponential feed was mirrored to the down side.

Dissolved oxygen levels (DO_2) were kept over 20% (100% were set before inoculation at 35°C, 0.3 bar gauge, pH 7). The pH was kept constant by adding 12.5% NH_4OH, which also served as nitrogen source. For quantification of the base consumption the base was titrated prior to fermentation using potassium hydrogen phthalate as primary standard and bromotymol-blue as indicator. Fermentation parameters are listed in table 6.

Table 6: Fermentation parameters.

Parameter	Batch	Non-Induced Fed-Batch	Induced Fed-Batch
Temperature [°C]	35	35	35/27.5/20
Pressure [bar gauge]	0.5	0.7	0.5
Air Flow Rate [L/min]	4	4	4
Oxygen Flow Rate [L/min]	0	0.3	0.3
Stirring Speed [rpm]	1400	1400	1400
DO_2	>20%	>20%	>20%
pH	7±0.1	7±0.1	7±0.1

2.6 Analytical methods

Biomass dry weight concentration

After centrifugation (RZB 5171, 10min, 4°C) of 2ml of the cell suspension in pre-weighted glass tubes the supernatant was decanted and frozen at -20°C for further analysis. The pellets were washed twice using distilled water and dried at 95°C for 72 hours. The biomass dry weight concentration was determined in duplicate.

Cell rupture *via* homogenization

2ml of the cell suspension were centrifuged (RZB 5171, 10min), washed twice with distilled water and frozen at -20°C for further cell rupture. Samples were re-suspended in 20ml of 0.1M Tris buffer pH 8.4 and homogenized (Avestin EmulsiFlex©, Canada) at 700bar for 11 passages. The homogenizer was operated in continuous mode. The homogenizing internal volume was determined to be 20ml, the flow at homogenization 40ml/min. Plug flow was assumed and one passage determined to be equivalent to 30 seconds of homogenizing. The SOP for the homogenization procedure is annexed in the appendix.

Product activity

Product activity was determined by monitoring the absorption change (415nm) of the substrate para-Nitrophenylphosphate (pNPP) at 37°C. The Alkaline Phosphatase Yellow (pNPP) Liquid Substrate System for ELISA (Sigma, P7998) was implemented in an enzymatic analyzer robot (CuBiAn© XC, Innovatis). After dilution of the sample the reaction of the substrate was started by automatic pipetting of 10 µl of the sample to 120µl of the substrate system. The absorption change was monitored and a reaction rate calculated in the time window between 20 and 60 seconds after reaction start. For direct calculation of activities from the reaction rate with the CuBiAn© XC analytical robot, a calibration row in the activity range from 0–1.6 U/ml with alkaline phosphatase from *E. coli* (Sigma, P5931) as standard was established. Activities are given in respect to this standard.

Protein and metabolite concentrations

Acetate and glycerol concentrations in the supernatant were determined photometrically using commercial assay kits (ENZYTEC fluid acetate test kit, Id-N°5226; ENZYTEC fluid glycerol test kit, Id-N°5360) which were implemented in the CuBiAn© XC analytical robot. Since no significant amounts were detected the supernatant was cross-checked for metabolites using HPLC analysis (Supelcogel C-610, Sigma Aldrich) according to a method shown in table 7.

Extracellular as well as intracellular protein concentrations were measured using the Bicinchoninic Acid Kit for Protein Determination (Sigma, BCA1-1KT). The SOP for the total soluble protein determination is attached in the appendix.

Table 7: HPLC method.

Flow	0.5 ml/min isocratic
Eluent	0.1 H_3PO_4/NaN_3
Temperature	30°C
Injection Volume	10μl
Detector	Refractive Index Detector

SDS-PAGE gel electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used for the qualitative assessment of intra- and extracellular protein according to Laemmli 1970. Resolving and stacking gel solutions are given in table 8 and 9.

Table 8: Resolving Gel Solution.

Component	Amount [ml]
dd H2O	2.05
30% Acrylamide/Bis	1.65
1.5 M Tris-HCl buffer pH 8.8	1.25
10% SDS	0.05
10% APS	0.025
TEMED	0.0025

Table 9: Stacking Gel Solution.

	Amount [ml]
dd H2O	2.85
30% Acrylamide/Bis	0.85
0.5 M Tris-HCl buffer pH 6.8	1.25
10% SDS	0.05
10% APS	0.025
TEMED	0.005

ICP-OES analysis

The multi-element emission spectroscopy ICP-OES allows the quantification of trace elements of metals and non-metals in solution. The injected sample is nebulized and injected into argon plasma (6000-10000 K), where the elements are atomized and excited for the emission of element-characteristic radiation. The radiation is fragmented using a monochromator and the intensity of the radiation is recorded in dependence of the wavelength. The quantification of the elements is based on the radiation intensity measured. ICP-OES analysis was performed on the supernatant fraction obtained after centrifugation of the samples drawn from the reactor. Samples were diluted 1:50 to 1:200 with HNO_3 (1%). Blanks were prepared accordingly. Measurements were performed on an iCAP 6500 (Thermo Scientific, USA). Optimized parameters according to table 10 were used. Each sample was measured 6 times applying an integration time of 10 seconds.

Table 10: Instrumental parameters the ICP-OES analysis.

Parameter	Value	Parameter	Value
Rf power	1200 W	Nebulizer flow	0.65 l min^{-1}
Sample flow	0.75 ml min^{-1}	Auxiliary flow	1.0 l min^{-1}
Read delay	60 s	Coolant flow	12 l min^{-1}
Integration time	10 s	Number of replicates	5
Background correction	constant shift from analytical line		

Table 11: Wavelengths for the ICP-OES analysis.

Element	Wavelength [nm]		Element	Wavelength [nm]	
	λ 1*	λ 2		λ 1	λ 2
Al	308.215	167.079	Mn	257.610	259.373
B	208.959	249.773	Na	589.592	818.326
Ba	455.403	233.527	P	177.495	213.618
Ca	315.887	422.673	S	180.731	182.034
Cu	324.754	224.700	Si	212.412	251.611
Fe	234.349	259.940	Sr	407.771	421.552
K	766.490	769.896	Ti	334.941	308.802
Mg	285.213	202.582	Zn	202.548	213.856

Multi element calibration standards were used. Underground-corrected signal intensitities of two wavelengths were used for data exploitation. Reproducibility of the measurements was determined as 0.5 to 2% of the relative standard deviation. Limits of detection varied for the detected elements, but were generally determined to be between 0.5 to 5 µg/l. Under consideration of the applied dilution and the blanks concentrations were calculated for the investigated elements [mg/l].

2.7 Design of experiments

Definition of the CQA and CPPs

In a simplistic view, the upstream process aims at the production of a cell broth which is manageable for the subsequent downstream process. The specific activity, hence the fraction of obtained volumetric activity (U/ml) by the volumetric biomass dry weight concentration (g/l), was assessed to be possibly critical for the subsequent downstreaming. Since the specific activity is solely determined by the prior upstream process, it was assessed to be a presumptive upstream critical quality attribute. The specific activity is calculated by dividing the volumetric activity of the homogenized sample by the biomass dry weight, according to formula 3.

$$CQA = \frac{U}{x} * 1000$$

Formula 3: Calculation of the critical quality attribute (CQA).

CQA	Critical quality attribute	[U/g$_{biomass}$]
U	Volumetric activity	[U/ml]
x	Biomass concentration	[g/l]

In cause of this study, three presumptive critical process parameters for the induction phase were in focus of investigation: The induction phase feeding strategy as well as the induction phase temperature and the induction time. In recombinant protein production, the induction temperature possibly effects recombinant protein expression levels, inclusion body formation, the activity of proteases as well as the metabolic load posed on the system (Baneyx and Mujacic 2004). A broad experimental range from 20°C to 35°C was investigated. The applied feeding strategy impacts inclusion body formation (Baneyx and Mujacic 2004) as well as the extent of induction phase carbon depletion (Andersson 1996; Baneyx and Mujacic 2004). Applying a positive exponential feeding strategy counteracts the effects of progressive carbon depletion, whereas inclusion body formation is possibly enhanced. Hence, next to a positive

exponential feeding strategy, a linear as well as a negative exponential feeding strategy was applied (see figure 16). Initially, intracellular recombinant protein activities are expected to rise within induction time. On the other hand, effects detrimental for intracellular protein production such as proteolysis (Baneyx and Mujacic 2004), protein leakiness through the outer cell wall (Tong 2000), carbon depletion as well as plasmid loss are reported to be feed strategy and time dependent (Andersson 1996). Hence, the time point of process shut down possibly effects product quality and induction time was assessed to be a presumptive CPP.

$$F_{(t)} = F_0 * e^{k*t}$$

Formula 4: Calculation of the induction phase feed rate.

$F_{(t)}$	Feed rate as a function of induction time	[g/h]
F_0	Feed rate before induction	[g/h]
k	Induction phase feed exponent	[h^{-1}]

Experimental design

The establishment of knowledge space involves the investigation of CPP interactions and therefore demands the use of multivariate statistical experimental design. For the investigation of two factors aiming at response surface modelling, nine independent experiments are needed. The experimental design (figure 17) served as the experimental basis for the formulation of the knowledge space as well as for the subsequent data exploitation methodology. Each point illustrates one fermentation run with fixed parameters. The center point was performed in triplicate for the estimation of the experimental error, resulting in a total number of 11 fermentation runs. Induction time as third CPP under investigation opens a third axis to DOE, but demands no additional experiments since a time resolution can be obtained by timely measurements. Considering the biologic variation expected and on the basis of feasibility considerations this sampling interval was chosen to be five hours and the induction period monitored over a minimum of 40 hours.
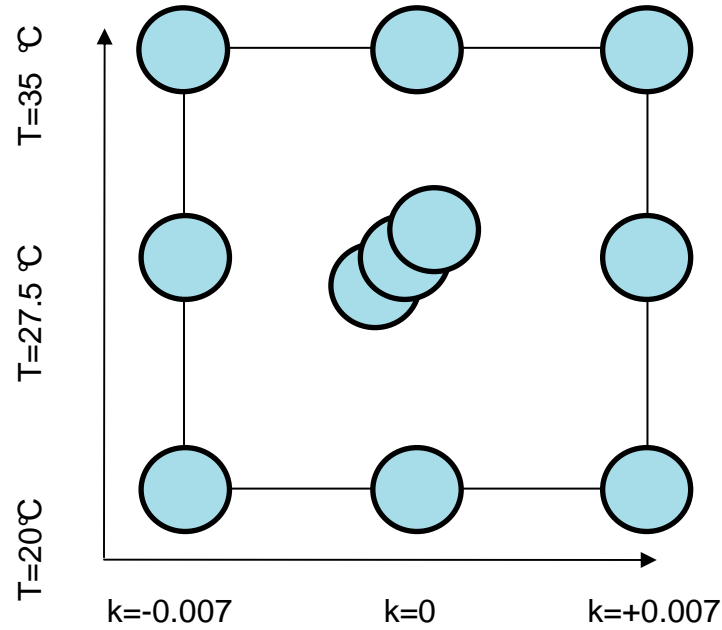
Figure 17: The orthogonal experimental design involved the variation of induction phase temperature (from T=20°C to T=35°C) as well as the variation of the induction phase feeding strategy, expressed as the feeding exponent k. Each circle represents one individual fermentation run that was monitored over 40 hours of induction time. The center point at T=27.5°C and k=0 was performed in triplicate, illustrated as three superimposed circles.

Process analyis DOEs

Next to the DOE for the development of the knowledge space (table 12), process analysis DOEs (table 13) were performed for the investigation of CPP/CQA interactions.

Table 12: Knowledge space development DOE.

Factors (CPPs)	Range	Response (CQA)
Induction phase temperature	20-35 °C	Biomass specific activity [$kU/g_{biomass}$]
Induction phase feed strategy (k)	-0.007 to +0.007	
Induction time	10 to 40 h	

Table 13: Process analysis DOEs.

Factors (CPPs)	Range	Responses
Induction phase temperature	20-35 °C	Carbon dioxide yield [Cmol/Cmol]
Induction phase feed strategy (k)	-0.007 to +0.007	Biomass yield [Cmol/Cmol]
Induction time	10 to 40 h	Mean specific protein release rate

3. Results

3.1 An exemplified data set

As an example for the richness of data obtained during one experimental run, a complete data set from one fermentation at 35°C induction temperature and a positive exponential induction phase feed is given in figures 18-21. The combined raw- and processed data used for knowledge space and process analysis DOEs is annexed in the appendix. After induction, alkaline phosphatase activities were found to increase until approximately 10 hours of induction. In compliance to previously published results (Wilms et al. 2001) the expression system was found to be very tightly regulated and almost no activity was detected in the non-induced culture. Intracellular total soluble protein levels were found to be 20% to 30% of the biomass dry weight, independent of induction time and the performed experiment. Extracellular total protein concentrations were found to rise up to 30 g/l. Within induction, biomass concentrations were still found to be increasing. Neither glycerol nor significant amounts of acetate accumulated in cause of the fermentation process (figure 18). The obtained data (figures 18 and 19) was processed into condensed information in the form of volumetric- and specific rates (figure 20) as well as yields (figure 21). Due to a decrease in the biomass yields, the specific substrate uptake rate q_s as well as the specific growth rate μ decreases within induction time, although a positive exponential feed is applied. This constant decrease in q_s is even stronger pronounced applying a linear and negative exponential feed (data given in the appendix). On the basis of the determined soluble protein concentration given in percent dry cell weight and the protein release rate calculated from the extracellular protein concentrations measured, lysis rates (r_{lys} and q_{lys}) can be calculated (figure 20). In cause of induction, the biomass yield dropped sharply compared to the previous non induced fed batch that showed a biomass yield of 0.59 C_{mol}/C_{mol}. A steep drop in the biomass yield was found at 10 hours of induction, independent of the performed experiment. The carbon dioxide yield $Y_{CO2/s}$ was found to increase in cause of induction (figure 21). The substrate specific ammonia yield $Y_{NH4/s}$ as well as the biomass yield $Y_{x/s}$ decreased within induction time, while the extracellular protein yield $Y_{(protein/s)}$ was found to increase (figure 21). DOR, C and N-balance served as consistency check (figure 21, right). The C, N and DOR recovery in the first 20 hours of induction was found to fit well. Independent of the performed experiment, a systematic decrease of the recovery values was detected.

Acquired data of fermentation 3

Fermentation parameters: induction temperature: 35°C, k=+0.007 (positive exponential feed).
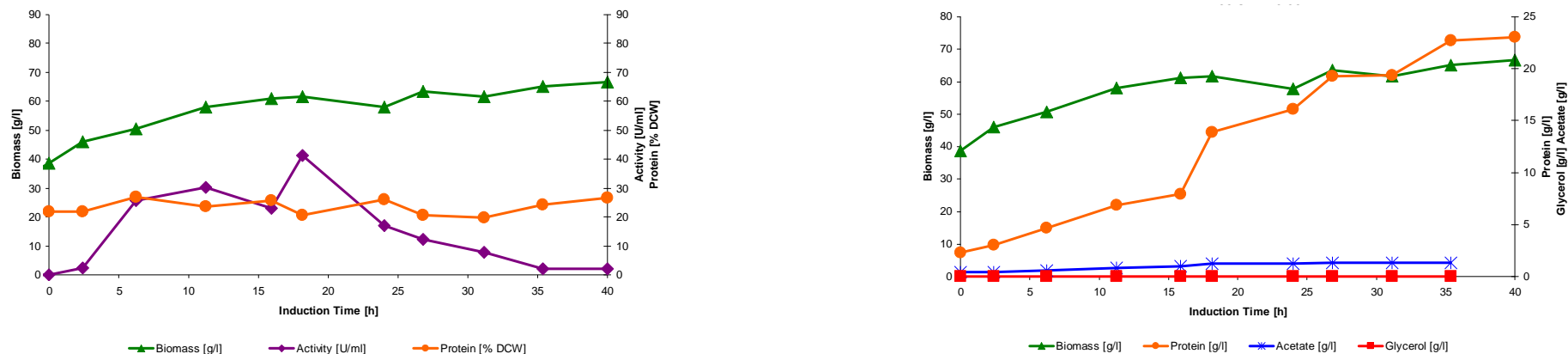

Figure 18: Fermentation at 35°C applying a positive exponential feed (k=+0.007). Left: intracellular protein concentrations (orange), intracellular activity (purple) and biomass concentrations (green). Right: extracellular protein- (orange), glycerol- (red), acetate- (blue) and biomass (green) concentrations. X-axis: induction time.
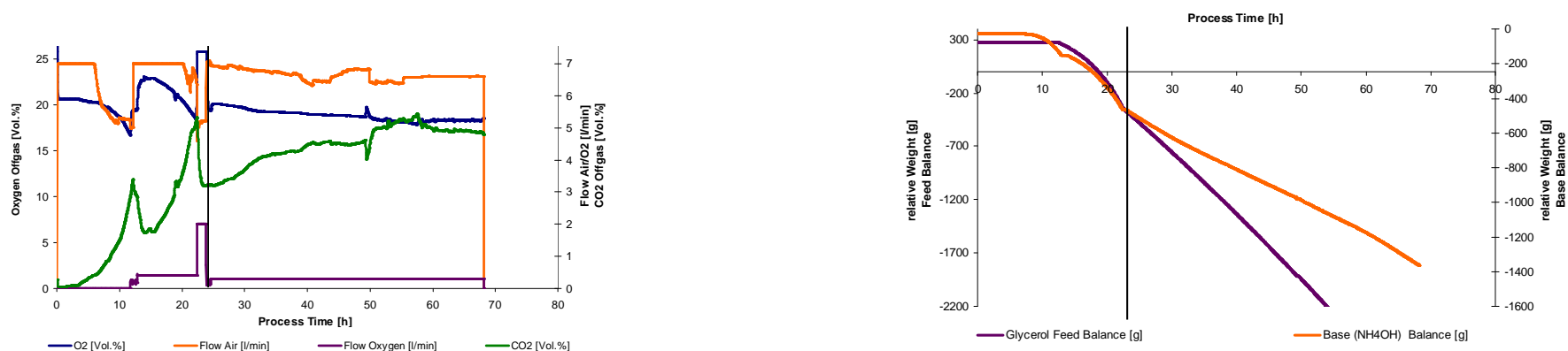

Figure 19: Fermentation at 35°C applying a positive exponential feed (k=+0.007). Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed data of fermentation 3

Fermentation parameters: induction temperature: 35°C; k=+0.007 (positive exponential feed).

Figure 20: Fermentation at 35°C applying a positive exponential feed (k=+0.007). Left: volumetric growth rate rx (blue), volumetric substrate uptake rate (rs), protein release rate $r_{Protein}$ (purple) and volumetric cell lysis rate (red). Right: specific growth rate μ (blue), specific substrate uptake rate qs (orange), specific protein release rate $q_{Protein}$ (purple) and specific cell lysis rate r_{Lys} (red). X-axis: induction time.

Figure 21: Fermentation at 35°C applying a positive exponential feed (k=+0.007). Left: biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen (green) balance.

3.2 Establishment of an upstream knowledge space

The contemporary perception of quality by design for upstream processes starts with the definition of CPPs and CQAs, followed by the variation of CPPs and the subsequent multivariate regression of CPPs with CQAs. The outcome of this approach is the formulation of a knowledge space, intrinsically the multivariate regression of CQAs and CPPs. For all fermentations included in the DOE the biomass dry weight concentration as well as the intracellular activity was measured in 5 hour intervals, as exemplified in figure 22 for one DOE run (35°C induction phase temperature; positive exponential induction phase feeding strategy). This data served as the basis for the calculation of the CQA, which was defined to be the specific activity given in ($kU/g_{biomass}$). Evaluation of the time dependence of the biomass yield (figure 21) revealed a steep drop after approximately 10 hours of induction, resulting in a decreased productivity of the process. Hence, the investigated time window included in the knowledge space DOE was chosen based on the productivity of the process (t=10 hours to t=40 hours of induction)
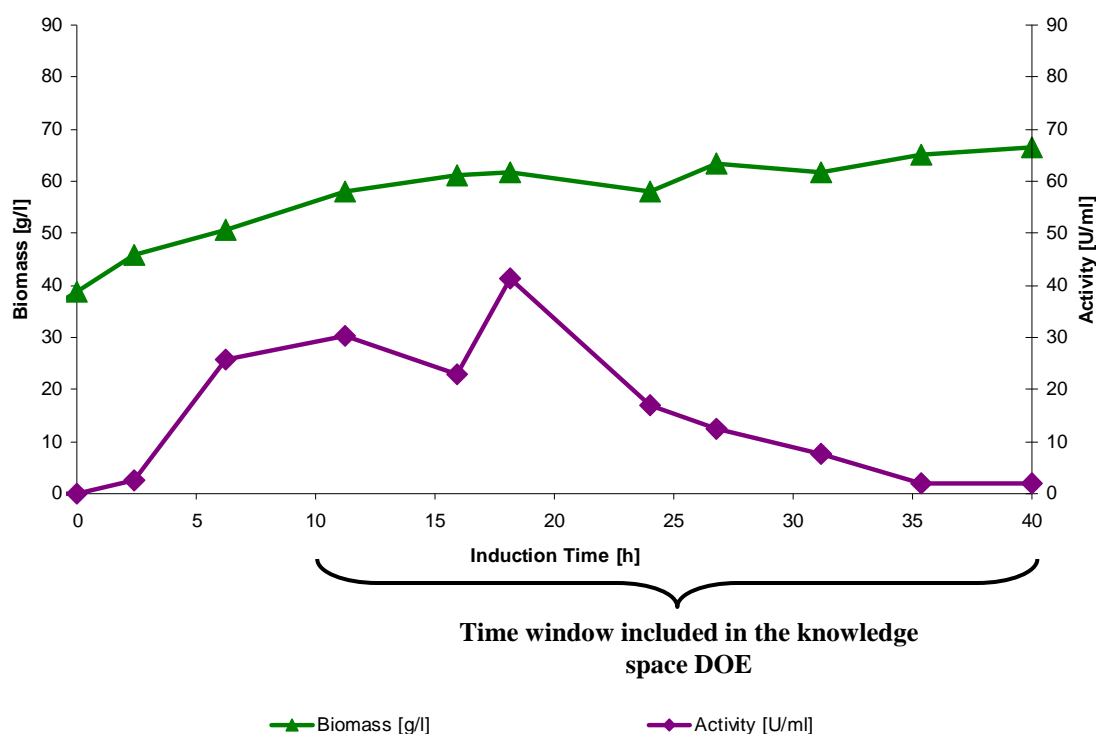
Figure 22: The fraction of biomass dry weight and volumetric activity (specific activity) was defined as a presumptive CQA of the investigated process. Biomass dry weight concentrations as well as the intracellular alkaline phosphatase activities were monitored in 5 hour intervals over induction time. The induction phase time window stretching from ten to forty hours of induction was included in the knowledge space DOE. Run at 35°C and k=+0.007.

The calculated specific activities (CQAs) between t=10 and t=40 hours of induction were multilinearly regressed with the CPPs induction time, induction temperature and induction phase feeding strategy expressed as the feed rate exponent k. Evaluation of the regression model revealed induction time, the post-induction feed exponent k as well as temperature as significant model terms (see figure 25). Considering the data from t=10 to t=40 hours of induction, induction time had a slight negative effect on the specific activity obtained. Induction temperature was considered with a linear and a quadratic term in the model, although only the quadratic term was found to be significant. The linear term can not be dismissed due to the hierarchic structure of the model. The induction phase feed strategy, expressed as the feed exponent k, was found to have a strong influence of the CQA and included as linear term. The most prominent interaction term was found to be the interaction of the quadratic temperature and the induction phase feeding exponent. Figure 23 displays the knowledge space for a fixed time point. Induction temperature is plotted on the right, the induction phase feeding exponent on the left and the specific activity, given in kU/g, is plotted on the towering axis. The linear time dependence of the CQA is illustrated in the multiple contour plots of figure 26. The basic model statistics are given in figure 24: The future prediction precision (Q^2) was computed as 0.627, the regression coefficient R^2 as model validity as 0.63. The multivariate study resulted in the detection of a temperature optimum at 27.5°C, applying a linear and positive exponential feed strategy. Applying a negative exponential feed, the model response maximum is found at low induction phase temperatures (20°C), as illustrated in the front corner of figure 23.
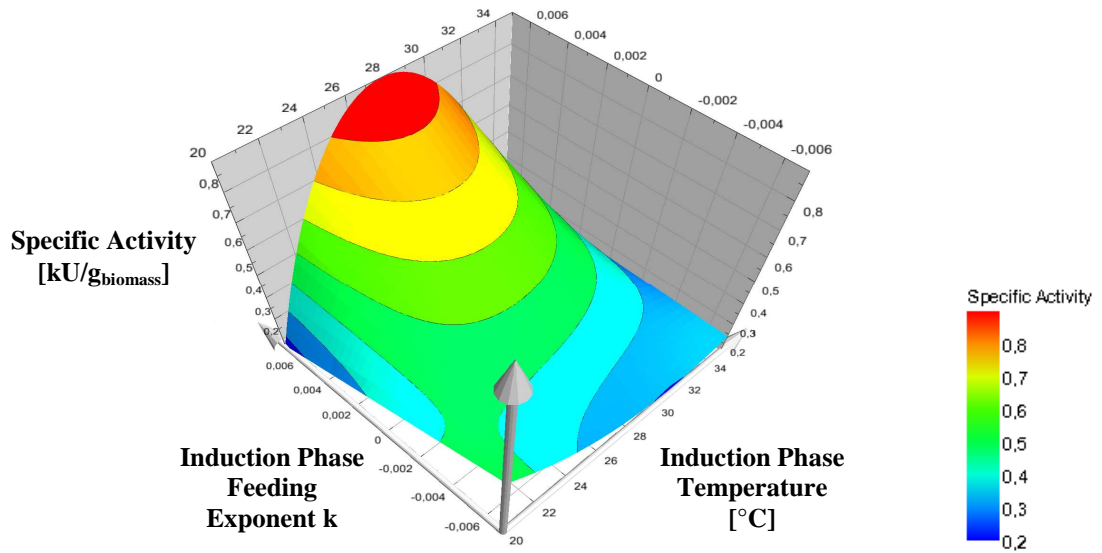
Figure 23: Contour plot of the obtained model at induction time=25h. Induction temperature [°C] is plotted on the right, the induction phase feeding exponent k [/] on the left and the specific activity [kU/$g_{biomass}$], is plotted on the towering axis.
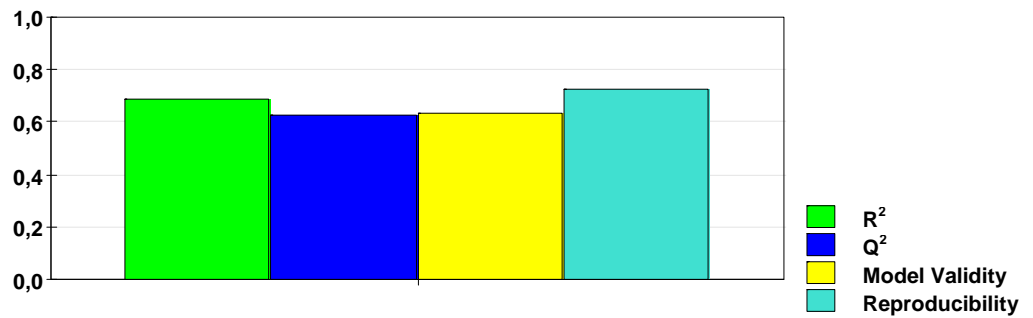
Figure 24: Basic model statistics of the regression model: Regression coefficient R^2, prediction precision Q^2, model validity and reproducibility of the regression model.
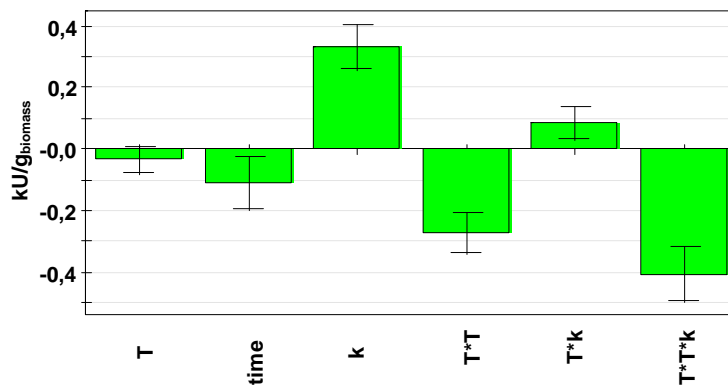

**Figure 25:** Scaled and centered coefficients of the factors included in the regression model. Induction temperature (T), induction time (time) as well as the induction phase feed exponent (k) were included as linear terms. Induction temperature (T) was included as quadratic term. The interaction terms of T and k were included as well. The interaction term of quadratic temperature (T*T) and k shows a high coefficient.
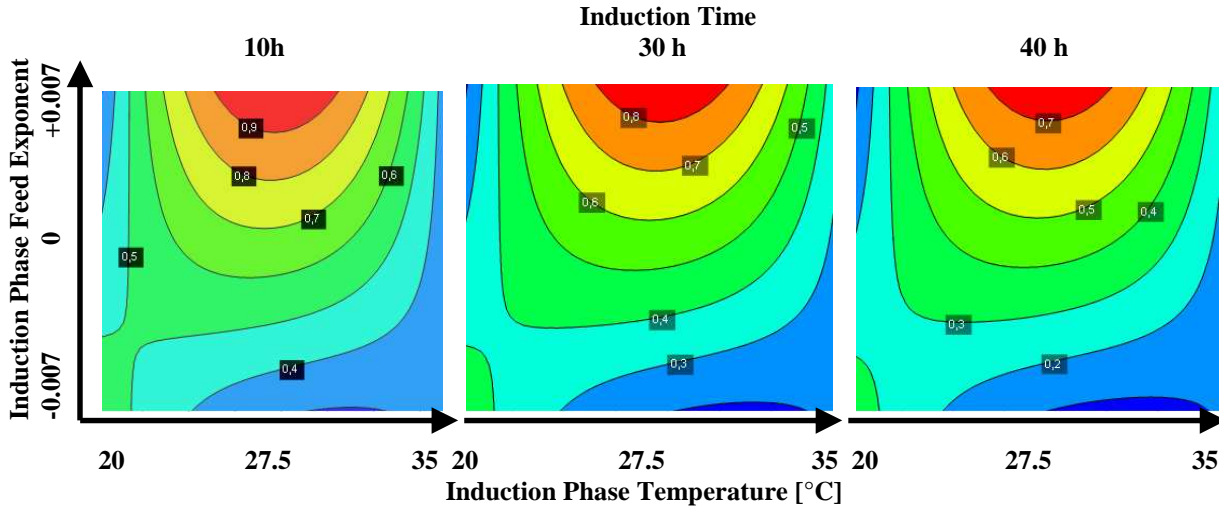
Figure 26: Contour plots for the knowledge space DOE at t=10 hours of induction (left), t=30 hours of induction (center) and t=40 hours of induction. For each individual plot the induction phase temperature is given on the x axis and the induction phase feed exponent on the y axis. The isolines are labelled with the model response, the specific activities given in kU/g_s (CQA).

3.3 Process analysis using multilinear regression of specific rates and yields with CPPs

For each fermentation run condensed information in the form of specific rates and yields was calculated, that can be considered a response of the biological system to the CPPs induction time, induction phase temperature and induction phase feed strategy. This chapter deals with the identification and quantification of the CPPs impact on specific rates and yields.

3.3.1 Analysis of CPP interactions with the mean specific protein release rate

$q_{protein}$

Extracellular protein concentrations rising from 0 g/l at time of induction to over 20 g/l were detected (data given in the appendix). The specific protein release rate for all 11 fermentations was calculated in cause of data exploitation. This served as the response for a 2 factor CCF design with the CPPs induction temperature (T) induction phase feed exponent (k) as factors. The response was entered as mean values for the time span from t=10 to t=40 hours of induction. Multivariate modelling identified the post-induction feeding coefficient k as sole significant influential factor (figure 27). The basic model statistics are given in figure 28: A prediction precision (Q^2) of 0.32, a regression coefficient (R^2) of 0.52, a model validity of 0.78 and a reproducibility of 0.67 was obtained. The obtained linear model is illustrated in figure 29.
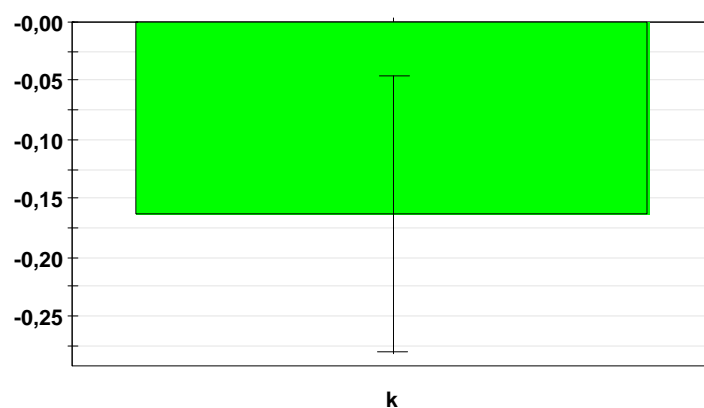
Figure 27: Scaled and centered coefficient for the mean specific lysis rate. The induction phase feed exponent k was identified as the sole influential factor for the mean specific lysis rate.
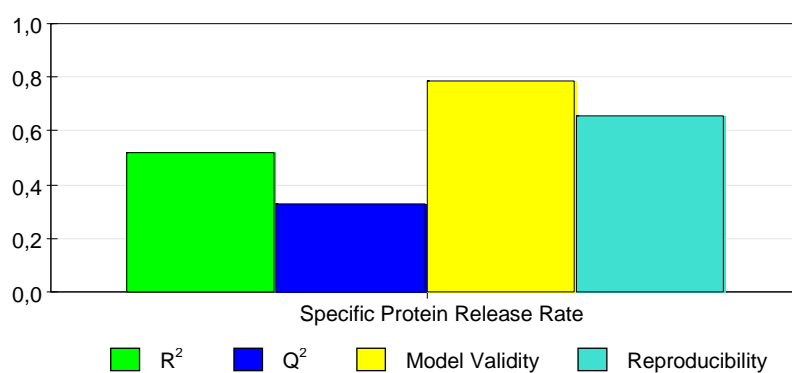
Figure 28: Basic model statistics of the regression model: Regression coefficient R^2, prediction precision Q^2, model validity and reproducibility of the regression model.
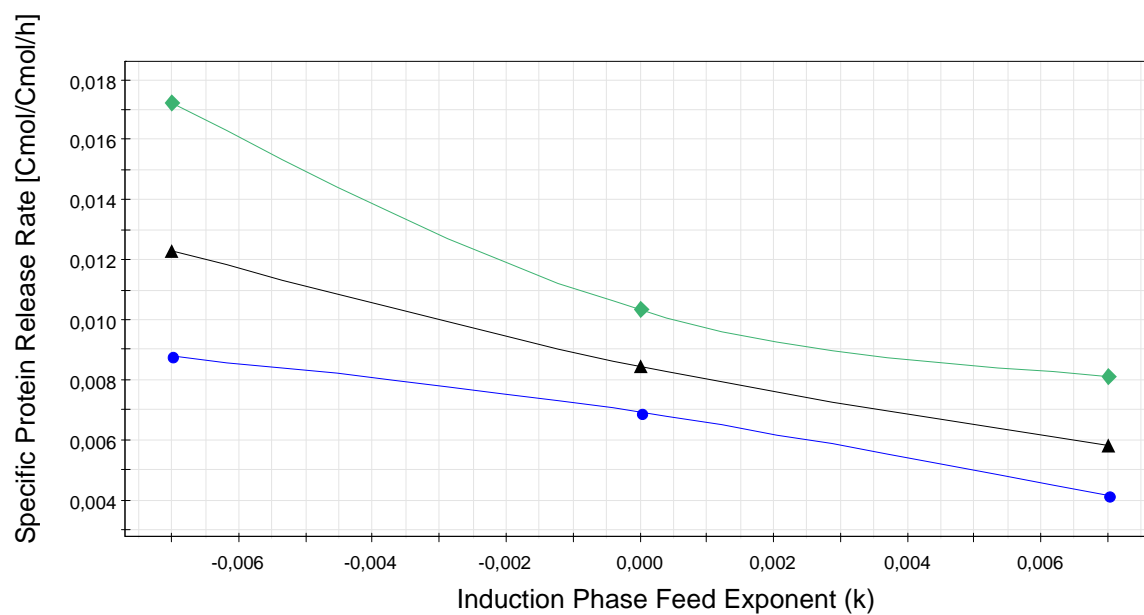

Figure 29: Prediction plot: The mean specific protein release rate (triangles) as a function of the induction phase feed exponent. The upper (diamonds) and lower (circles) confidence interval for the model with a confidence level of 0.95 are included in the plot.

3.3.2 Analysis of CPP interactions with the carbon dioxide yield $Y_{(CO2/s)}$

The carbon dioxide yield $Y_{(CO2/s)}$ served as a response of a three factor DOE with the CPPs induction time, induction phase temperature and induction phase feed profile as factors. Applying a CCF-design, 1 hour averages at the induction time points t=20 and t=40 were calculated from the processed data of all eleven fermentation runs used in the knowledge space DOE. The carbon dioxide yield at t=0 represents the carbon dioxide yield of the non-induced fed-batch. Model evaluation identified induction time as sole influential factor. No interaction with the CPPs induction phase temperature and induction phase feeding exponent was detected. The basic model statistics are given in figure 30: A future prediction precision (Q^2) of 0.64, a regression coefficient (R^2) of 0.72, a model validity of 0.904 and a reproducibility of 0.64 was obtained.
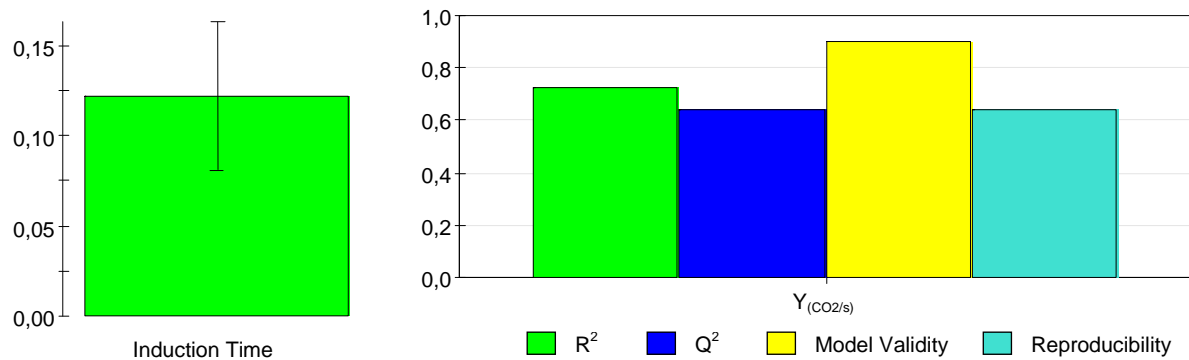
Figure 30: The induction time as model coefficient for the carbon dioxide yield $Y_{(CO2/s)}$ (left). Basic model statistics are given on the right: Regression coefficient R^2, prediction precision Q^2, model validity and reproducibility of the regression model (right).
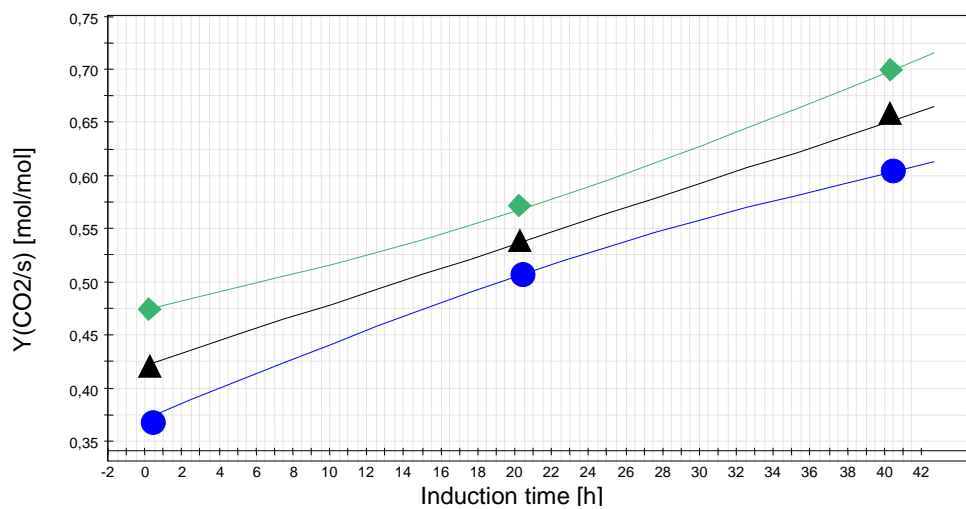
Figure 31: Prediction plot: The carbon dioxide yield $Y_{CO2/s}$ as a function of induction time (triangles). The upper (diamonds) and the lower (circles) confidence interval for the model with a confidence level of 0.95 are included in the plot.

3.3.3 Analysis of the CPP interactions with the biomass yield $Y_{(x/s)}$

In analogy to the carbon dioxide yield the biomass yield $Y_{x/s}$ was investigated as a function of induction time, temperature and post-induction feed. Applying a CCF-design, 1 hour averages at the induction time points t=20 and t=40 were calculated. The biomass yield at t=0 represents the biomass yield in the non-induced fed-batch. Model evaluation revealed induction time and induction temperature as well as their interdependencies as significant factors. For the regression model, induction time and induction temperature was included as linear terms. Furthermore, induction time as quadratic term as well as the interaction term of induction time and induction temperature was included in the model (figure 32). The induction phase feed exponent k was found to be non-significant. The obtained model showed a regression coefficient (R^2) of 0.92, a future prediction precision (Q^2) of 0.82, a model validity of 0.94 and reproducibility of 0.84. Figure 32 shows the scaled and centered coefficients for the model; figure 33 gives a summary of the basic model statistics. Reducing the induction phase temperatures had a positive effect on the obtained induction phase biomass yield $Y_{(x/s)}$. Furthermore, $Y_{(x/s)}$ was found to decrease in cause of induction time. The contour plot (figure 34) gives a graphical display of the obtained model.
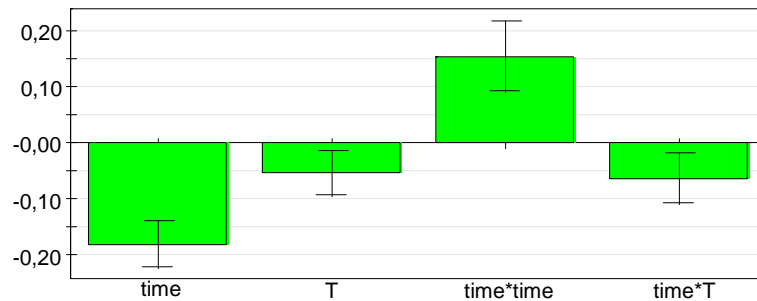
Figure 32: Scaled and centered coefficients for the biomass yield $Y_{(x/s)}$. Induction time (time), induction temperature (T) and the interaction term of induction time and induction temperature had a negative, the quadratic induction time term (time*time) a positive effect on the biomass yield $Y_{(x/s)}$.
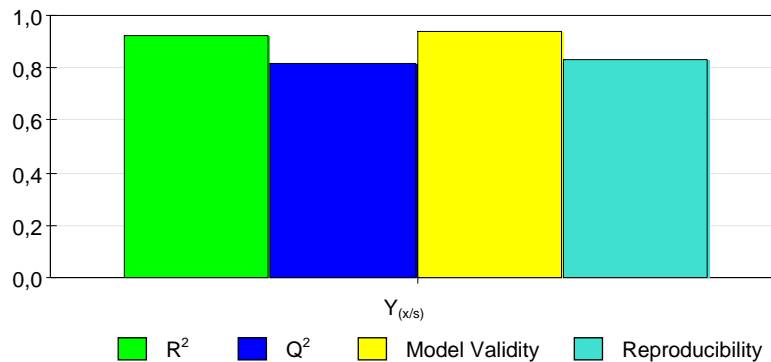

Figure 33: Basic model statistics for the modelling of the $Y_{(x/s)}$ yield: Regression coefficient R^2, prediction precision Q^2, model validity and reproducibility of the regression model.
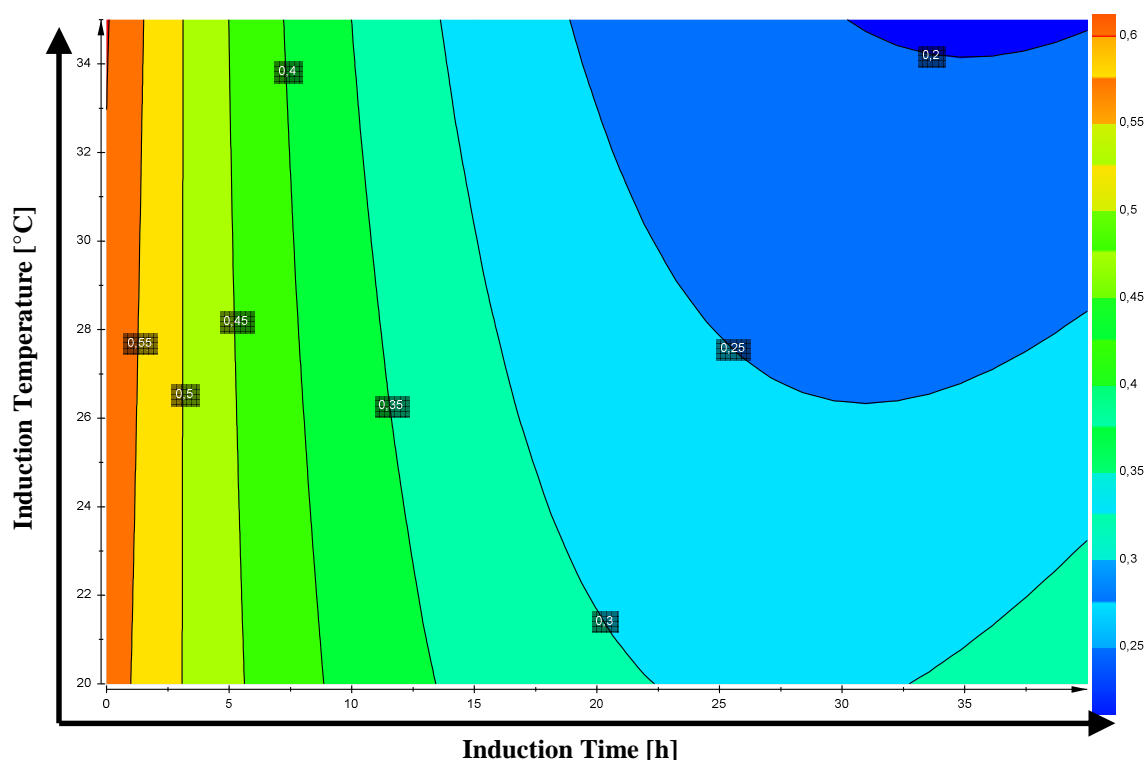
Figure 34: Contour plot for the $Y_{x/s}$ model obtained. Induction time is given on the x axis, induction temperature on the y axis. Isolines are labelled with the $Y_{x/s}$ given in Cmol/Cmol. The colour code on the right decodes the $Y_{x/s}$ within the plot, reaching from orange (high $Y_{x/s}$) to blue (low $Y_{x/s}$).

3.4 SDS-PAGE gel electrophoresis

SDS-PAGE gel electrophoresis was used to qualitatively monitor the intra- and extracellular protein levels. The target recombinant enzyme alkaline phosphatase, identified as progressively increasing band at ~49kDa, is highlighted in figure 35 and 36 (35°C induction temperature, k=+0.007). At the time point of induction, no alkaline phosphatase was detected in the homogenized intracellular fraction (figure 35, induction time 0) and no extracellular protein was detected at all in the extracellular medium (figure 36, induction time 0). Within induction, intracellular alkaline phosphatase levels were steadily increasing until an induction time of 27 hours. After 27 hours of induction no further rise in intracellular alkaline phosphatase levels was detected. A band at ~45kDa was also found to rise within induction time. Figure 36 shows a SDS-gel from the extracellular protein fraction. Extracellular protein levels were found to rise within induction time. Furthermore, the most prominent band was identified as alkaline phosphatase. Similar to the intracellular SDS-gel a band at ~45kDa was detected that increased within induction time.
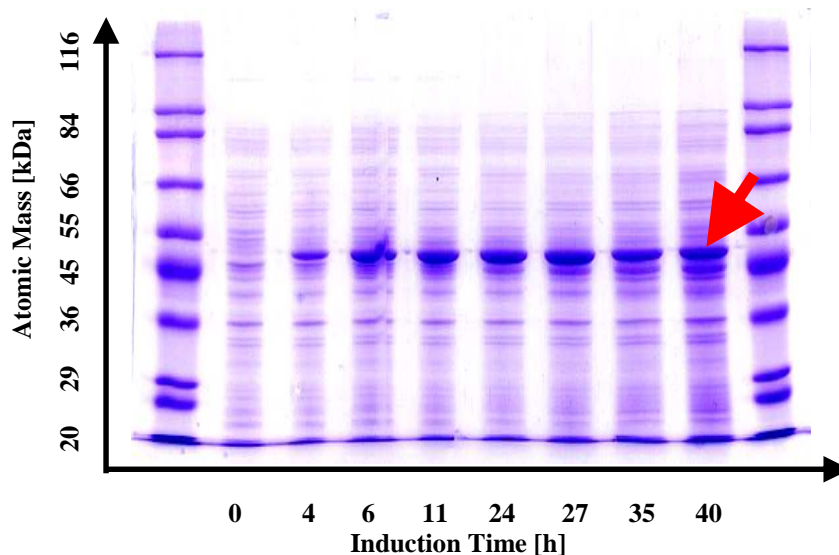
Figure 35: SDS gel electrophoresis of the disrupted cells. Alkaline phosphatase (~49kDa) as target enzyme is highlighted. On the basis of a protein standard (first and last column) the atomic mass of the protein bands is given on the y axis. Dilution: 1:10. Samples were taken from a fermentation run at 35°C induction phase temperature and a positive exponential feed exponent of +0.007.
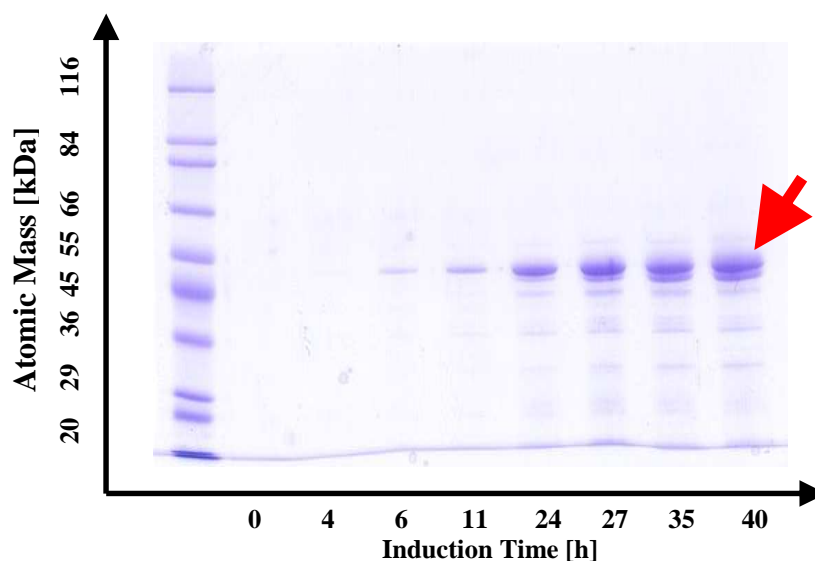

Figure 36: SDS gel electrophoresis of the extracellular protein fraction. Alkaline phosphatase as target enzyme is highlighted (~49kDa). Induction time is given on the x axis. On the basis of a protein standard (first column) the atomic mass of the protein bands is given on the y axis. Dilution: 1:20. Samples were taken from a fermentation run at 35°C induction phase temperature and a positive exponential feed exponent of +0.007.

3.5 ICP-OES of the extracellular medium

Extracellular ion concentrations were monitored using ICP-OES. Extracellular zinc and calcium levels are included in figure 37 (run at 35°C induction temperature, k=+0.007). Extracellular zinc concentrations were found to be decreasing until a time span between ten and 25 hours of induction. Within this time span the trend of zinc depletion in the extracellular medium is reversed and extracellular zinc levels were found to be increasing

again. The change pattern of the calcium concentration was found to be reversed: After a steady increase in calcium concentrations until the time window between 10 and 25 hours of induction calcium concentrations were found to be decreasing until the end of the fermentation.
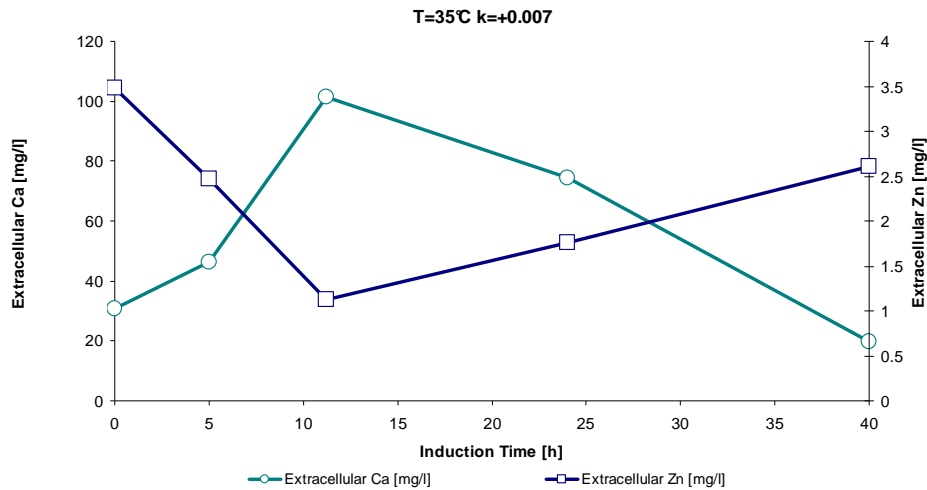

Figure 37: Extracellular zinc- and calcium concentrations are displayed for a fermentation run at 35°C induction phase temperature and a positive exponential induction phase feed strategy. Concentrations are given on the y axis; induction time is plotted on the x axis.

3.6 DOE for the evaluation homogenizing parameters

The disruption efficiency (figure 38) of the homogenizing step used for cell rupture was quantified by determination of colony forming units (CFU) before and after homogenization at different pressures (700 and 1100 bar) and different number of passages (11 and 20 passages). As expected, highest disrupting efficiencies were obtained at high pressures (1100 bar) and high number of passages (20), although almost 99% of all cells were disrupted even at low pressures (700 bar) and a few number of passages (10.5). Following the evaluation of the disrupting efficiency, a full factorial DOE was set up to investigate the effect of homogenization pressure, number of passages as well as biomass concentration on the protein activity obtained after homogenization. Evaluation of the model revealed that homogenization pressure, the number of passages as well as the biomass concentration is significant (figure 39). The created model showed a regression coefficient (R^2) of 0.73, a future prediction precision (Q^2) of 0.50, a model validity of 0.87 and reproducibility of 0.65. Figure 39 shows the scaled and centered coefficients for the model; figure 40 gives a summary of the basic model statistics. The obtained activity after homogenization displayed as relative activity in respect to the center point activity ($U_{measured}/U_{centerpoint}$) in the contour plots of figure 41.
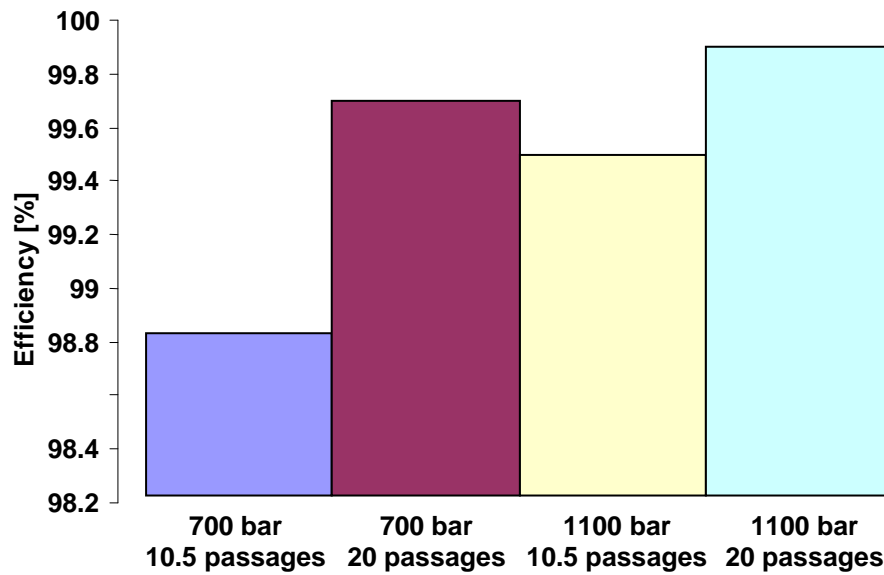
Figure 38: The disruption efficiency was quantified *via* monitoring the loss in CFUs before and after homogenization. The disruption efficiency is given on the y axis. Each column is labeled with the homogenizing parameters (homogenizing pressure and number of passages) of the respective homogenization run.
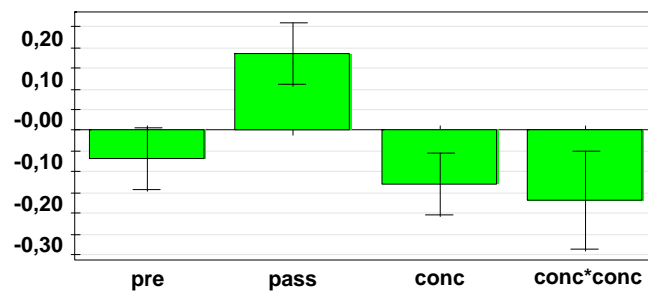

Figure 39: Scaled and centered coefficients for the full factorial homogenization DOE. Homogenization pressure (pre), sample biomass concentration (con) and the quadratic term of the sample biomass concentration (con*con) had a negative effect, the number of passages (pass) a positive effect on the relative activity in respect to the center point ($U_{measured}/U_{centerpoint}$).
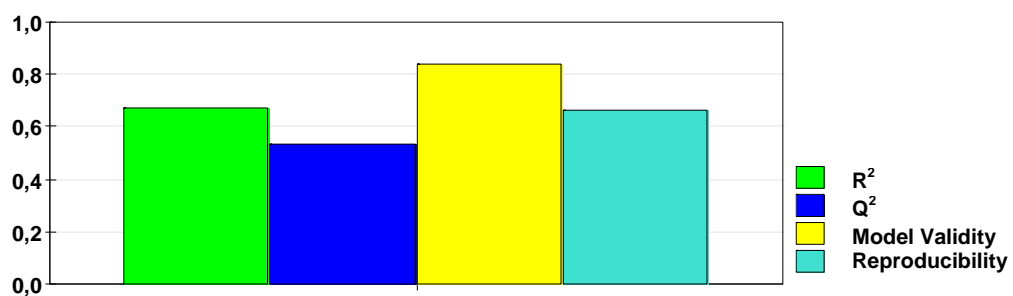
Figure 40: Basic model statistics: Regression coefficient R^2, prediction precision Q^2, model validity and reproducibility of the regression model.
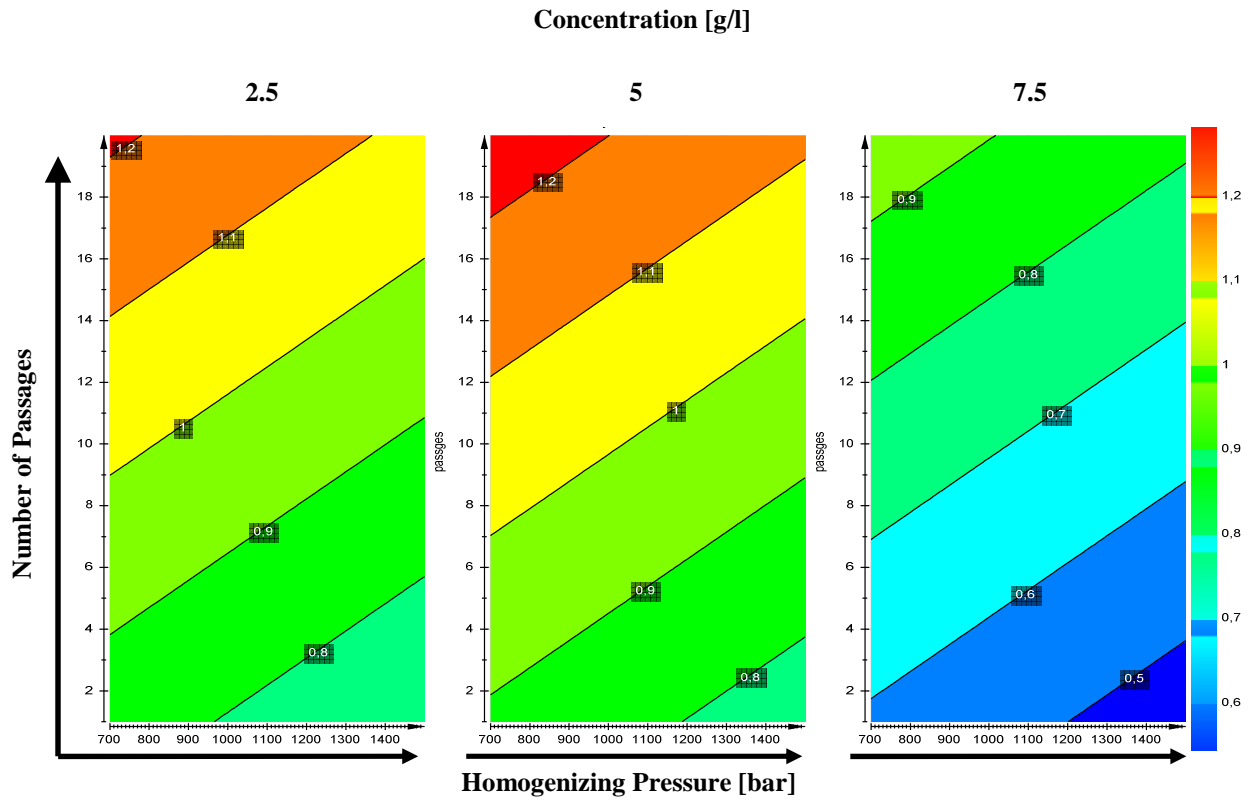
Figure 41: Contour plot of the three parameter homogenizing model, displayed for the investigated concentrations of 2.5, 5 and 7.5 g/l biomass (top). Homogenizing pressures are given on the x axis, the number of passages on the y axis. The colour code on the right decodes the obtained $Y_{x/s}$ within the plot, reaching from red (high relative activity in respect to the center point ($U_{measured}/U_{centerpoint}$) to blue (low relative activity in respect to the center point ($U_{measured}/U_{centerpoint}$). Highest relative activities were found at medium biomass concentrations, low pressures and high number of cycles. Isolines are labelled with the activity in respect to the center point ($U_{measured}/U_{centerpoint}$).

4. Discussion

4.1 Development of a knowledge space

Knowledge space development in this thesis followed the contemporary perception of quality by design for upstream processes: Followed by the definition of CPPs and CQAs, CPPs served as factors for DOEs. The CQA was monitored and set as the response of the DOE, followed by multilinear regression. The outcome of this approach was a mathematical model that displayed the CQA as a function of the CPPs. The mathematical model obtained can only be applied to interpolate within the experimental range. The process performed best at an induction of temperature of 27.5°C and a positive exponential feed exponent of +0.007. The finding that induction time had a negative effect on the specific activity indicates that the upstream process should be shut down after 10 hours of induction. Developing a design space out of the knowledge space demands the evaluation of the presumptive CQA's impact on the subsequent down stream operations, an exercise that has not yet been done at this stage of process development. Nevertheless, the high dependency of the presumptive CQA on induction phase temperature and induction phase feed strategy suggests the formulation of a rather narrow design space for these CPPs in case the presumptive CQA specific activity is effectively determined to be critical regarding its impact on product quality. The induction time dependency of the CQA is less pronounced; hence a future design space could be laid out generously concerning this axis. Following the contemporary perception of quality by design, process knowledge is solely demonstrated by the presentation of the knowledge space, intrinsically the multivariate regression of CQAs and CPPs. It must be outlined, that the understanding of CPP/CQA interactions provided by the knowledge space is of a pure statistical nature and its biological interpretation remains speculation. The amount of data necessary for the formulation of the design space (figure 22) is remarkably low and stands in high contrast to the richness of data routinely recorded during process development (figures 18 and 19).

4.2 Data exploitation methodology for knowledge space understanding

In cause of the presented data exploitation methodology, the recorded data was processed into specific rates and yields. In the next step, knowledge about the CPPs impact on the biological system was gathered by means of multilinear regression of specific rates and yields with the critical process parameters investigated in cause of the DOE. This methodology lays the basis for the following discussion aiming at a mechanistic understanding of the knowledge space.

4.2.1 Impact of the feed strategy on the biological system

The presence of high amounts of extracellular protein indicates cell lysis or periplasmatic protein release, since *E. coli* possesses no secretion pathway for alkaline phosphatase out of the periplasmatic space (Sommer 2008). SDS gel electrophoresis identified the total extracellular protein to be mainly composed of alkaline phosphatase, as illustrated in figure 42.
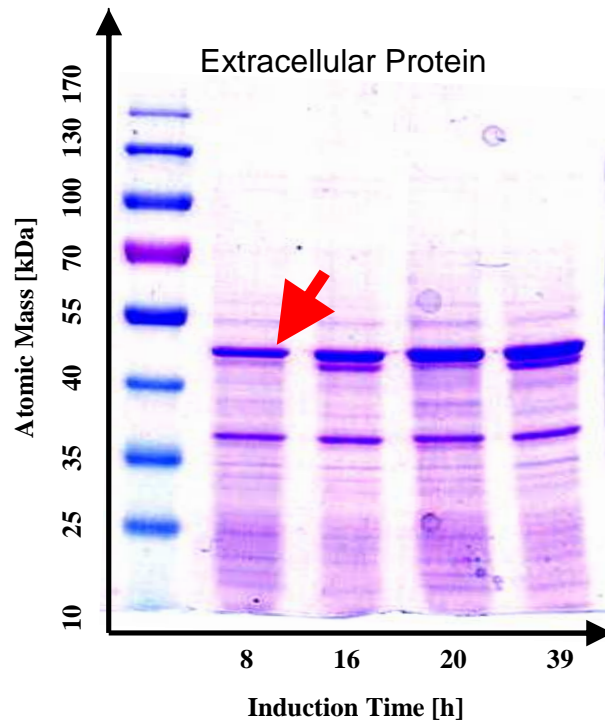

Figure 42: SDS gel electrophoresis of the supernatant broth. The band highlighted was identified as alkaline phosphatase (~49kDa). Dilution: 1:1; 1:2; 1:5; 1:10. Run at T=27.5°C, k=0.007.

From the knowledge space we obtained the information that the induction phase feeding exponent k had a positive effect on the specific intracellular activities obtained within the induction phase (figure 23). Process analysis identified the feeding exponent k as sole significant influential factor for specific protein release rate: A high induction phase feed exponent was found to have a negative effect on the specific protein release rate (figure 29). This finding underpins the positive effect of an increased feed rate on the specific intracellular activity: Higher intracellular protein levels are achieved, since less alkaline phosphatase is lost from the intracellular space due to protein release (formula 5). Hence, the feeding strategy affects the specific protein release rate, which stands in direct connection to CQA specific activity.

Cell lysis is reported for various induced systems (Wilms et al. 2001). Cell stress as the result of progressive carbon depletion as well as the metabolic load posed on the system in cause of

induction was assumed to be responsible for protein release. The reported negative effects of progressive carbon depletion in the induction phase include the activation of the carbon stress system, cell segregation (Andersson et al. 1996) as well as a reduction in RNA polymerase and ribosome content (Sanden et al. 2003). Spoken in terms of specific rates, carbon depletion means a declining specific substrate uptake rate (q_s). Applying a positive exponential feed rate counteracts the progressive decrease in q_s during induction time, while a negative exponential feed accelerates the decrease in qs in cause of induction. The relatively low model fit indicates that the entire variance of the specific release rate is not fully explained by the induction phase feed exponent k. Since the remaining CPPs as well as induction time were assessed to be non-significant for the specific protein release rate, other sources for the variation remain undetected in cause of this study. The obtained data suggests that the reproducibility of protein release is to query and that other factors influencing the specific protein release rate still remain undetected. It is likely, that the relatively high unexplained variance of the knowledge space is connected to the high unexplained variance of the specific protein release rate. This underlines the need for a PAT strategy for the monitoring of the extracellular protein, reducing the quality risk originating from protein release. Online quantification of extracellular protein would be feasible using infrared spectroscopy, inline sampling and enzymatic measurements as well as elemental-balance based soft sensors. The finding that the specific protein release rate is correlating with the feeding strategy does not provide any information on the mechanism of protein release. Possibilities include full cell lysis as well as selective protein release from the periplasmatic space, e.g. through local damages of the outer cell wall.

$$\frac{d(u)}{dt} * X + \frac{dX}{dt} * u = -\frac{q_{protein} * u * X * 100}{w_{protein}} + q_u * X$$

Formula 5: Activity balance of the intracellular activity; the change of the specific activity u [U/g] is dependent upon biomass formation (dX/dt), alkaline phosphatase formation and the loss of previously formed alkaline phosphatase due to protein release. The specific protein release rate $q_{protein}$, part of the loss term of the intracellular activity, is highlighted.

u	Specific activity [U/$g_{biomass}$]
X	Total biomass [g]
$q_{protein}$	Specific protein release rate [g/g/h]
q_u	Specific active alkaline phosphatase production rate [U/h/g]
$w_{protein}$	Weight fraction of protein of dry cell weight [% dry cell weight]

4.2.2 Can the carbon dioxide yield be exploited for a control strategy?

Since high amounts of protein were found to be released to the medium but no other metabolites could be detected in significant amounts, the question arose whether other cell components such as free amino acids and sugars are reused by the remaining viable cell population. Hypothetically, this should result in an increased CO_2/s yield referred to the substrate added, since the released cell components constitute a second substrate. In terms of PAT, the CO_2/s yield represents a robust, online accessible yield since CO_2 off gas analysis and feed quantification is routinely achieved in industrial fermentation processes. However, the effect of temperature and post-induction k as well as their interdependencies with time were found to be non-significant. Only the induction time as linear term was needed to fit the model. Except for induction time no correlation between CPPs and the CO_2/s yield was detected, hence no control strategy can be constructed upon this readily accessible yield. The finding that the CO_2/s yield is neither dependent upon temperature nor the post exponential feeding exponent k, from which the latter was found to influence cell lysis, allows the conclusion that no or just few of the components released from the cell are reused from the viable cell population, possibly due to the simple absence of free amino acids and metabolites in significant amounts. The rise in the CO_2/s yield indicates an increase in the maintenance metabolism in the induced culture, probably due to an increase in the metabolic load posed on the system as a consequence of recombinant protein production.

4.2.3 Time dependency of the energy metabolism of the induced system

The obtained knowledge suggests that the energy demand of the induced culture, hence their maintenance metabolism, is strongly dependent upon induction time. The already discussed time dependency of the CO2/s yield underpins this finding, which is also reflected by the decrease in biomass yield. It seems that the system is increasingly subjected to the effects of metabolic load; hence the direction of resources towards the promoter related activities (Glick 1995). This goes along with an increase in stress protein concentrations as well as a decrease in metabolic enzyme concentrations (Birnbaum and Bailey 1991). The biochemistry of the cell is possibly also submitted to the consequences of a stringent response triggered by the induction and subsequent carbon depletion, which is also reported to go along with increased concentrations of intracellular proteases (Voellmy 1980).

4.2.4 Induction phase temperature as adjustment screw for the expression system

The temperature dependence of the biomass yield indicates a reduced expression of recombinant protein at reduced temperatures: Protein synthesis capacity of native as well as recombinant proteins in E. coli is determined by the number of functioning ribosomes and the ribosome translation rate (Farewell and Neidhardt 1998). For induced systems, the bottleneck of protein synthesis is reported to be translation (Sanden et al. 2003). Within a temperature range of 25 to 37°C the ribosome concentration in E. coli is reported as constant and ribosomal activity reported to rise linearly with temperature and to correlate with growth (Farewell and Neidhardt 1998). However, the protein synthesis capacity is equal for native protein expression necessary for cell growth and recombinant protein production. In a carbon limited system the finding that more resources are directed towards biomass at reduced temperatures indicates that a higher fraction of the protein synthesis capacity is used for the synthesis of native proteins which are necessary for cell growth. The transcription rate of recombinant DNA seems to be reduced to a higher extent than the transcription rate of native proteins, possibly due to a down regulation of the rhaBAD promoter. Hence, the flux of cell resources to promoter related activities can be regulated by means of induction phase temperature.

4.2.5 Promoter down regulation positively impacts active protein formation

The evaluation of the specific protein release rate contributed to the understanding of the high CQA achieved at a high induction phase feed exponent. Furthermore, the discussion of temperature dependency of the biomass yield indicated a possible down regulation of the recombinant protein transcription rate at reduced temperatures. However, the question why the knowledge space revealed a temperature optimum for active protein at 27.5°C and why activities are tendentially decreasing after 10 hours of induction has not been discussed so far. Figure 43, offline data from a fermentation carried out at 35°C and a positive exponential feeding strategy, shows a strongly pronounced decrease in intracellular activity starting from approximately 20 hours of induction. Although intracellular activity is lost due to protein release, this does not account for all the decrease in intracellular activity: SDS gel electrophoresis (figure 44) of the intracellular homogenized samples provided evidence for an almost continuous increase in alkaline phosphatase concentrations. Some form of inactivation must therefore contribute to the decrease of alkaline phosphatase activities. In order to reach its active dimeric conformation, alkaline phosphatase needs to be translocated into the periplasmatic space for disulfide bridge formation (McCracken and Meighen 1980). This

system is sensitive to overload: Possible shortages in transport proteins involved in the Sec-pathway as well as shortages of folding modulators can lead to the cytoplasmatic accumulation of inactive protein in form of inclusion bodies (François Baneyx and Mujacic 2004). The presence of inclusion bodies in this system was confirmed by SDS gel electrophoresis of the insoluble fraction obtained after cell rupture (Gruber 2010), but this finding does not explain the loss of intracellular activity accompanied by an increase in soluble alkaline phosphatase concentrations. The obtained knowledge suggests that this observation can be traced to the consequences of high level recombinant alkaline phosphatase expression: Blocking of the sec-pathway (Skare et al.1989) is likely, which results in an accumulation of inactive alkaline phosphatase monomers in the cytosol. Furthermore non-realized disulfide bridge formation due to a bottleneck in mediators (Kadokura et al. 2003) possibly contributes to the presence of inactive alkaline phosphatase monomers as well. The investigation of the biomass yield revealed that the induction temperature is connected to the level of recombinant protein expression. For a maximum yield of active alkaline phosphatase, a moderate down regulation of the recombinant expression system seems favourable, explaining the temperature dependency of the knowledge space. Apparently the investigated system reaches its maximum active protein synthesis capacity at around 27.5°C, as illustrated in the knowledge space (figure 23).
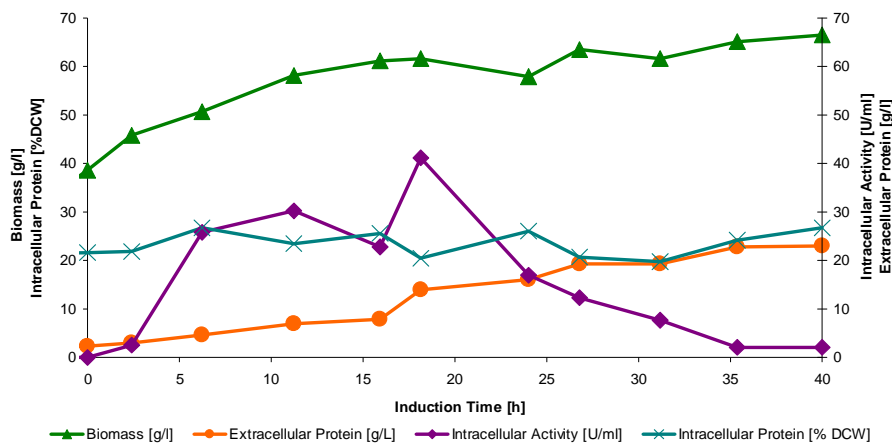
Figure 43: Extracellular protein concentration and intracellular activity of the discussed run (induction phase temperature 35°C; induction phase feeding exponent +0.007). Run at 35°C induction phase temperature and a positive exponential feed of +0.007.
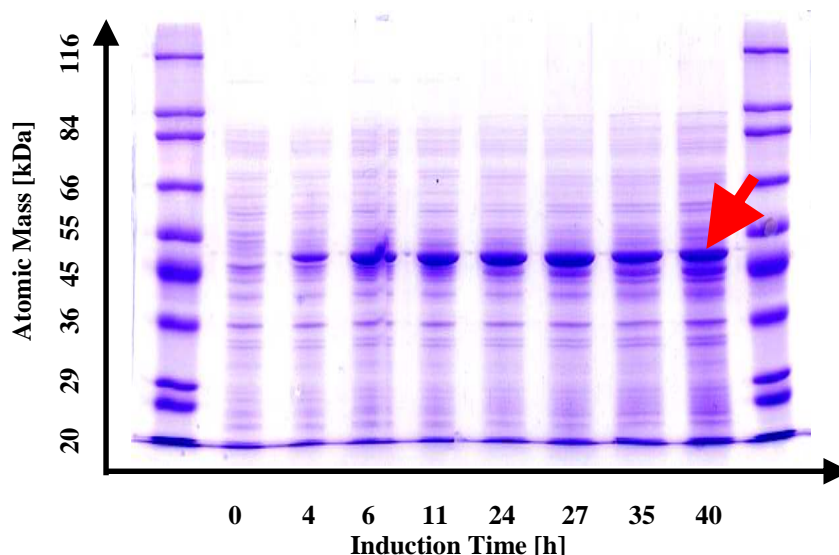
Figure 44: SDS gel electrophoresis of the disrupted cells. Alkaline phosphatase as target enzyme is highlighted (~49kDa). Discussion of the plot is given in the text. Dilution: 1:10. Run at 35°C induction phase temperature and a positive exponential feed of +0.007.

4.2.6 Detection of a possible zinc limitation using ICP-OES

The presence of zinc plays a key role in the successful folding and function of alkaline phosphatase in the periplasmatic space (Torriani 1968). The possible substitution of zinc by other divalent metal ions must be prevented because this goes along with a strong decrease in catalytic function (Wang et al. 2005). The extracellular zinc and magnesium concentrations were monitored using ICP-OES. By means of ICP-OES only total amount of extracellular zinc is accessible, hence the sum free zinc and zinc fixed in the active site of extracellular alkaline phosphatase. In the following the fluxes of zinc are discussed, which are composed of two terms: The intracellular accumulation of zinc and the export of zinc due to the release of alkaline phosphatase. The results are included in figure 45. Until ten hours of induction, extracellular zinc concentrations were found to decrease. This matches the assumption that zinc is readily taken up by the cell population for the formation of recombinant alkaline phosphatase. If this trend is linearly extrapolated, all zinc in the medium is used up at approximately 17 hours of induction. This time point correlates with the detection of the highest intracellular alkaline phosphatase activities. In the time span between ten and twenty-four hours of induction the trend of zinc uptake is reversed and a net efflux of zinc is observed. It is assumed, that from this time point onwards zinc as divalent metal ion in the active center of newly formed alkaline phosphatase is substituted by other metal ions. Similar results were reported by (Torriani 1968). The reversed behaviour of calcium fluxes allow the assumption that zinc is possibly substituted by this element. Parallel to the loss of intracellular zinc, a loss of intracellular activity is observed. SDS gel electrophoresis (figure 44) shows that

the protein synthesis machinery is still working and assumingly producing zinc depleted and therefore inactive alkaline phosphatase. The net efflux of zinc can be explained by the release of previously formed zinc containing alkaline phosphatase. Hence, the loss of intracellular activity after ten hours of induction was concluded to be the result of a net efflux of previously formed active zinc containing alkaline phosphatase. Zinc depleation is reported to inhibit dimerization (Torriani 1968); therefore the accumulation of inactive monomers is likely.
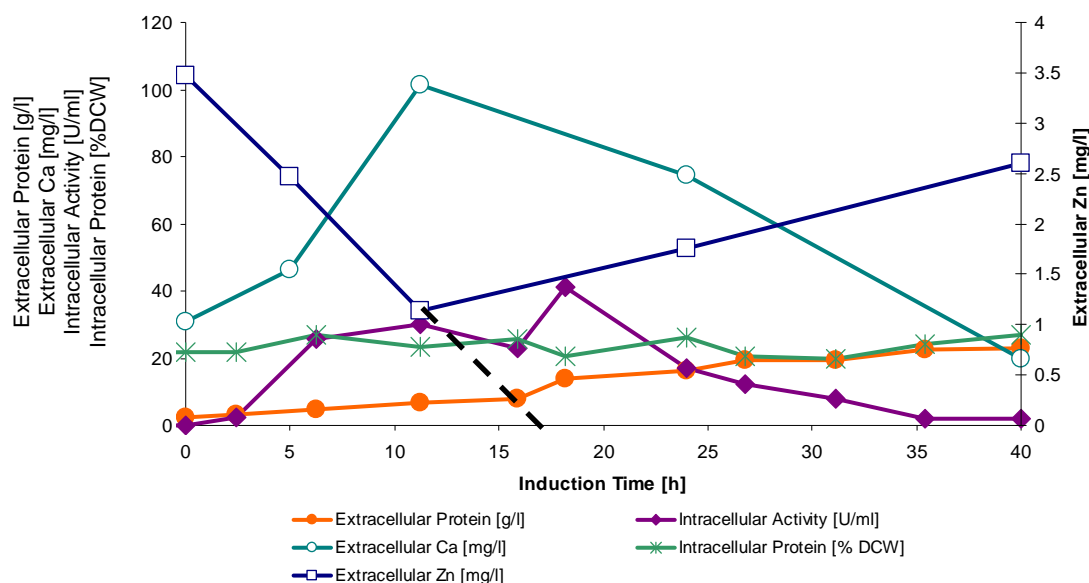
Figure 45: Extracellular zinc concentrations were detected using ICP-OES analysis. In the time span between 11 and 24 hours of induction the net influx of zinc is reversed. This finding correlates with the loss of intracellular activity (purple). An extrapolation of the decrease in zinc levels is indicated with a dashed line. Intracellular soluble protein levels stay at a constant level of approximately 20% of the total biomass dry weight. Run at 35°C induction phase temperature and a positive exponential feed of +0.007.

4.3 Multivariate study on the homogenizing downstream operation

Simply put, the upstream process aims at the production of a cell broth which is manageable for the subsequent downstream process. Since the target protein is located in the intracellular space, homogenization constitutes an integral part of the following downstreaming needed for product recovery. The obtained activity after homogenization is given as relative activity in respect to the center point activity ($U_{measured}/U_{centerpoint}$). The obtained model showed a Q^2 of >0.5 and can therefore be classified as a good model (Eriksson 2000). Model validity was computed as high. Evaluation of the model revealed that homogenization pressure, the number of passages as well as the biomass concentration is significant. Interestingly, higher pressures have a detrimental effect on the activity obtained, probably due to temperature degradation of the protein in cause of homogenization. Higher biomass concentration result in

decreased activities as well, an effect that can be traced to possible decreased homogenization efficiency at higher biomass concentrations. The most prominent model term was found to be the number of passages. Compared to reference values for *E. coli* homogenization (Kuboi et al. 1995) very high numbers of passages were needed for optimal product recovery. Possible explanations like improper functions of the device or low homogenizing efficiency due to polysaccharides in the broth (Wood 1988) are not part of this study. The multivariate model obtained constitutes a knowledge space for the homogenizing unit operation.

5. Conclusions

Exploitation methodology for quality by design purposes

Quality by design, a buzzword stroking pharmaceutical industry since its adaptation by the FDA in 2004 (FDA 2004), is about to lead pharmaceutical development in a direction where the display and proof of process understanding will be considered a key element for the regulatory submission of pharmaceutical processes. The question whether the display of the CPP/CQA regression in form of a design space is sufficient for the proof of process understanding is still not resolved. Multivariate studies are the only acceptable tool for the establishment of a knowledge space and demand a high number of experiments. Only a small fraction of the total data recorded is required for the knowledge space, leaving a plethora of valuable data unprocessed. Specific rates and yields represent condensed process information which is easily accessible by means of routinely recorded on- and offline data combined with data processing routines.

The presented work demonstrated that a high degree of biologic insight in the process can be extracted by means of statistical regression of specific rates and yields with the critical process parameters. The regression was carried out in the form of process analysis DOEs. The main interactions detected are summarizes in table 14. This additional information obtained during process development can be used for the interpretation of the knowledge space, creating a higher level of understanding which can be included into the regulatory filing. The applied methodology involving DOE execution, DOE data processing and DOE information processing is illustrated in figure 46. Given validated analytical methods for the estimation of extracellular protein and the biomass, the specific rates and yields used in this study are fully online accessible. This provides interesting opportunities as regards a future control strategy, since specific rates and yields offer insight in the biology of the system. The presented work demonstrated, that a high degree of biologic insight in the process can be extracted from the statistical regression of specific rates and yields with the critical process parameters

investigated in cause of knowledge space development. As regards the discussed system, the obtained data urge for a PAT system for the monitoring of the protein release rate, because the loss of the target protein is unwanted in terms of product quality and process productivity.
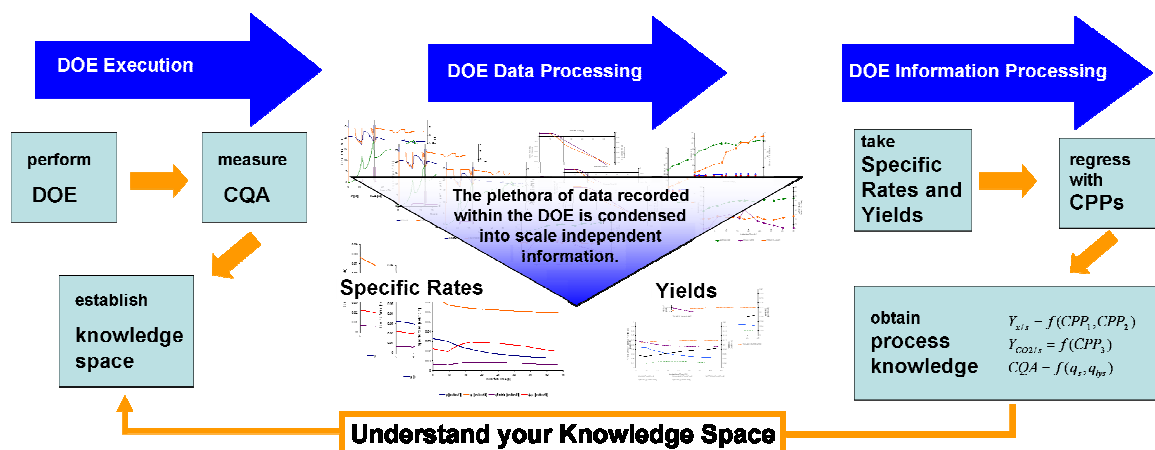

Figure 46: Schematic display of the applied data and information processing methodology aiming at the understanding of the knowledge space. The establishment of a knowledge space involves performing a DOE, the measurement of the CQA and the statistical regression of fixed CPPs and measured CQAs (DOE Execution). This yields only statistical information about CPP/CQA interactions. Processing available data into specific rates and yields (DOE Data Processing) yields scale independent information for each individual fermentation run. Subsequent regression of specific rates and yields with CPPs provides insight in the mechanistic impact of CPPs on the biological system (DOE Information Processing), allowing the interpretation of the knowledge space.

Table 14: Process analysis DOEs

	Carbon Dioxide Yield	Biomass Yield	Mean Specific Protein Release Rate
Induction Phase Temperature	no interaction	linear and quadratic	no interaction
Induction Phase Feed Strategy (k)	no interaction	no interaction	linear, positive
Induction Time	linear, positive	linear	x

Quality by design for upstream unit operations

The formulation of a critical quality attribute for this thesis was based on the assumption that the specific activity ($U/g_{biomass}$) is critical for the subsequent downstream operations. The intracellular soluble protein was found to account for 20% to 30% of the biomass dryweight and to be independent of CPPs and induction time. Hence, the specific activity linearly correlates with the activity per soluble protein. Whether or not this assumed CPP is critical or not for the downstream can only be found out within downstream process development. This underlines the expressed opinion that for successful process development following QbD

principles attention must be paid to the link between different unit operations, especially upstream and downstream.

Process specific conclusions

Zinc, essential for folding and function of alkaline phosphatase, was found to become limiting after ten hours of induction phase. The recombinant protein synthesis machinery was still found to be active, assumingly producing zinc depleted and therefore inactive alkaline phosphatase. Due to continuous protein release, active protein is washed out of the cells resulting in a decrease in the observed intracellular activity. For the production of alkaline phosphatase the feed medium should be zinc enriched.

Fermentation relevant characteristics of the rhaBAD promoter

Applying the rhaBAD promoter system for the production of alkaline phosphatase resulted in maximum specific actitities of $1000U/g_{biomass}$, a value which corresponds to 66 mg of active protein per gram biomass (specific activity [U/g AP] from sigma standard). The average intracellular soluble protein detected was determined as approximately 20 to 30% of the total cell mass. Hence, active alkaline phosphatase accounts for one third of the total soluble intracellular protein. This grade of quality without subsequent downstreaming is remarkable, since it almost matches the purity of the sigma standard, which is given as 40%. The host/vector combination was identified as prone to protein release which is unwanted in respect to product quality and productivity. The mechanisms of protein release in this system still remains in the dark. The extracellular protein detected largely contains alkaline phosphatase which is inactive and contributes to the low extracellular activities obtained. It must be outlined, that these findings are product and not host/vector specific and therefore a strategy for extracellular protein production might be feasible for another product. The detection of induction temperature as an adjustment screw for the regulation of recombinant protein expression represents a valuable tool for any process utilizing the rhaBAD promoter system. The finding that protein release linearly correspond to the applied feeding exponent was concluded to be due to progressive carbon depletion within induction time which can be counteracted applying a positive exponential feeding profile. This is probably host as host/vector specific and does not necessarily only account for the target protein produced.

Conclusions at a glance

- The establishment of a knowledge space provides only statistical understanding of CPP/CQA interactions and allows no insight in the underlying causes.

- Multilinear regression of specific rates and yields with the CPPs investigated in cause of the knowledge space DOE can provide insight in CPP/CQA interactions.

- The induction phase feed strategy affected the specific protein release rate, which was concluded to negatively impact the specific activity obtained in the induction phase.

- The carbon dioxide yield was not found to interact with the CPPs induction temperature and induction phase feed strategy; hence no control strategy can be created on the basis of this readily accessible yield.

- The decline in the biomass yield as well as the increase of the carbon dioxide yield in cause of induction time indicates a progressive increase in the culture's maintenance energy demand.

- The flux of cell resources to promoter related activities can be regulated by means of induction phase temperature. The investigated expression system performed best at 27.5°C induction temperature, indicating that the protein synthetic capacities of the host/vector combination are best exploited at this level of down regulation.

- A zinc limitation was detected using ICP-OES, explaining the decrease in alkaline phosphatase activities at late induction times.

- Homogenizing pressure, number of passages and biomass concentration were determined CPPs for the homogenizing operation and investigated in cause of a full factorial DOE.

6. Outlook

Extension of the investigated experimental area

The investigated region showed a decentered optimum as regards the feed strategy. Hence, it is possible that the process optimum lies beyond the investigated experimental area. The current multivariate study could be extended to a positive exponential feed of +0.014, demanding five more experimental runs.
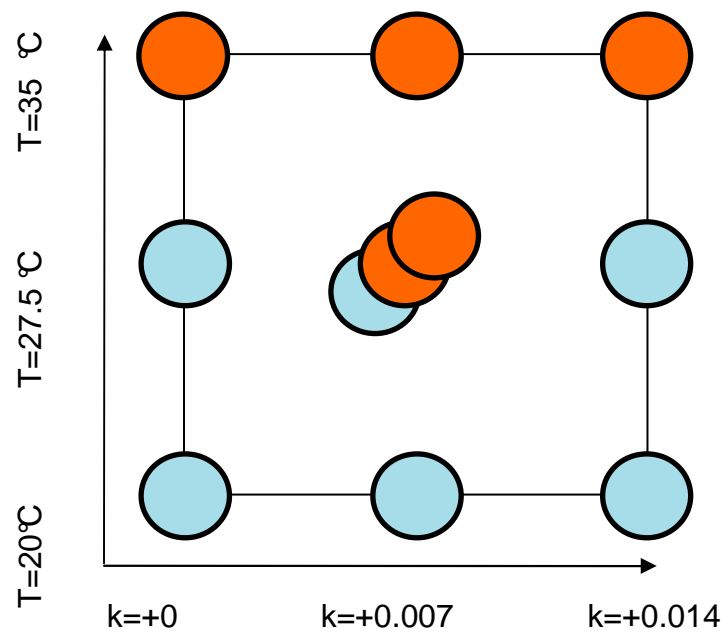
Figure 47: An extension of the experimental design to a positive exponential feed of 0.014 demands five additional fermentation runs (orange). The runs performed for the current DOE are illustrated as blue circles. Induction phase temperature is given on the y, the induction phase feed exponent on the x axis.

Media development using ICP-OES

The application of ICP-OES as multi-element technique for the monitoring of ion fluxes opens new perspectives as regards media development and optimization for recombinant protein production. Once the metal ion uptake rates are known, the fed-batch medium metal ion concentrations can be tailored to the exact demand of the culture.

Paradigm shift: Possible extracellular production of alkaline phosphatase?

The investigated process aims at the production of intracellular alkaline phosphatase. Protein release due to cell lysis or periplasmatic protein release was considered a loss term, detrimental for product quality as well as process productivity. Nevertheless, the high amount

of extracellular protein detected and the qualitatively high purity observed *via* SDS-gel electrophoresis (figure 35) suggest a possible paradigm shift towards a targeted extracellular protein production. For the target protein alkaline phosphatase this was assessed to be non-rewarding, since the phosphate buffered medium competitively inhibits extracellular alkaline phosphatase and a residence time in the periplasmatic space is necessary for folding and function (Torriani 1968). Extracellular activities measured did not exceed values above 0.05 $kU/g_{protein}$. Removal of salts *via* gel-permeation columns increased the active extracellular activity by the factor 8, assumingly due to the removal of the competitive inhibitors phosphate and EDTA. SDS gel electrophoresis indicates that at least 50% of the extracellular protein is alkaline phosphatase (figure 35). Consequently, only a small fraction of the alkaline phosphatase detected in the extracellular space can be recovered as active protein. Although not feasible for the target enzyme, the process shows potential for extracellular protein production in case the target products does not demand residence time in the periplasmatic space and is not irreversibly inhibited by the media components or degraded. Following this strategy, a thorough investigation of the mechanisms of protein release is necessary. Native gels can be used to distinguish between dimeric and monomeric alkaline phosphatase. Since dimerization takes place in the periplasmatic space (Torriani 1968) and therefore demands translocation *via* the Sec pathway, native gels can be used to discriminate between extracellular protein released from the periplasm and extracellular protein released from the cytosol.

Control strategy

In case the target process aims at intracellular protein production, the obtained results suggest that a positive exponential induction phase feed strategy is favourable due to a reduced protein release rate. As implication, a negative exponential feeding profile for maximum protein release seems promising in case the production target shifts from intra- to extracellular protein production. From an economic perspective extracellular production is favourable due to a simplified downstream, but issues concerning protein stability might arise. Induction phase temperature was identified as adjustment screw for the regulation of recombinant protein expression. For the investigated process a down regulation of the rha-BAD promoter improved quality of the product as well as process productivity. In industrial as well as pilot scale, the exponential feeding profile as well as the temperature control is routinely carried out using PID controllers. A positive exponential feed counteracts the progressive decline in the specific substrate uptake rate (q_s) during induction, which is favourable since the negative

effects of progressive carbon depletion are avoided. The direct control of the q_s is a valuable alternative to the positive exponential induction phase feed strategy. Following this strategy, a robust on-line measurement of the biomass concentration is necessary. The biomass concentration can be correlated to in-line turbidity measurement, in-line capacity measurement or fluorescence measurement (NADH). Furthermore soft-sensors on the basis of elemental balances can be used for the estimation of the biomass concentration.

7. Appendix

7.1 Standard Operating Procedures

7.1.1 Operation of Emulsiflex©-C3 Homogenizer

	Standard Operating Procedure	BioVT
TECHNISCHE UNIVERSITÄT WIEN / VIENNA UNIVERSITY OF TECHNOLOGY / TU VIENNA	**Operation of Emulsiflex©-C3 HOMOGENIZER**	
Research Division **Biochemical Engineering**	SOP Number: SOP027 Status: Effective	Date: 10.11.2010

Version	1.0
Replaced version	None
Author	Patrick Sagmeister
Date	20. 03. 2011
Signature	
Authorized by	
Date of authorization	21. 03. 2011
Signature	Ch. Herwig

Summary	Homogenization of E.Coli using the EmulsiFlex©-C3 homogenizer
Materials	

- 20ml 1M Tris Buffer (pH 8.4) per sample
- City Water for rinsing of homogenizer
- Syringe (20ml) and needle
- Ethanol 70%

Equipment
- EmulsiFlex©-C3 homogenizer
- Cooling Water Supply
- Compressed Air Supply (6 bar)

Procedure

Start-up

- Make sure that the red button is on the position off (depressed) before plugging in the instrument.
- Supply compressed air at a pressure of 6 bar.
- Plug in the instrument and switch on the main power supply on the back of the instrument.
- Open the cold water tap to provide cooling fluid for the heat exchanger.
- Check if there is any blockage in the homogenization valve or heat exchanger by flushing the system with compressed air. Put on the cap and connect the cap to the pressure hose.

Sample Preparation

- Frozen samples are suspended in 5ml of 1M Tris buffer pH 8.5 and kept on ice. After dilution to 20ml with Tris buffer pH 8.5 the sample is raised and released three times using a 20ml syringe with needle. The sample is inserted to the sample cylinder on top of the homogenizer.

Homogenization Procedure

- Put the transparent tubing into the sample cylinder, so that the sample can circulate.
- Twist the red mushroom button clockwise.
- Press the green start button.
- Continue circulating and increase the pre-pressure using the regulator to 2 bar, then fine-adjust the pressure so that the homogenization pressure on the manometer reaches a pressure of approximately 700 bar.
- After the pressure is correctly adjusted, the sample is circulated for 5min 15 seconds.
- Reset the pressure to 0 bar. Remove the sample and keep it on ice.
- Rinse the homogenizer with a full cylinder of tap water. Apply pressure to rinse the instrument.
- For turning off the instrument, press the red STOP button, shut off the air and gas supply and shut off the main power on the backside of the instrument.
- The device should be cleaned after each run. Rinse with water and use compressed air

afterwards.

- After use, ethanol (70%) is put into the sample line. The sampling cylinder is closed using the lid.

Troubleshooting

If the pump is not starting (no circulation of liquid after pressing the green button) stop the homogenizer, remove the sample and use once again a syringe to process the sample as described above. Clean the homogenizer as described above and start again. One possible reason for the pump not to start is the blocking of the valve situated under the sample cylinder by particles. This can be confirmed by watching the sample in the sample cylinder: If bubbles are rising from inside the homogenizer, this valve is blocked.

If the emergency shut-off is activated due to homogenizing pressure exceeding 2000 bar. If this is the case, shut off the machine for 30 seconds by turning the button for the main power supply at the backside of the instrument. Then start again.

If the homogenizing pressure can not be reached with 6bar pre-pressure, do NOT turn the pre-pressure adjustment wheel too far, you might damage the instrument. Instead check the seat of the homogenizing valve for wear. If you are not instructed how to do so, consult technical service.

Additional Information

A DOE was carried out using the EmulsiFlex©-C3 homogenizer. Sample concentration, homogenizing pressure as well as number of passages were determined as critical for yield of the released product.
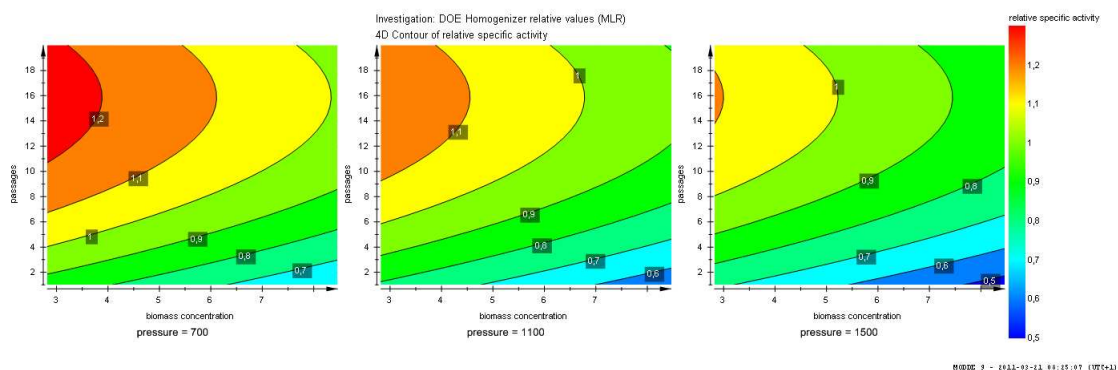
Since users experienced significant problems processing thawed samples and tried different techniques to overcome the problem of the pump not starting, here a quick list of strategies that did not work:

-	Sonication	
-	Waiting	
-	Applying Pressure (risk of sample spill)	

Literature

- www.avestin.com (You can find video instruction about how to operate the instrument)
- Avestin Operating Instructions for the Emulsiflex-C3
- See also: S:\BPT\BPT SHARED DATA\Equipment_Documentation\Homogenisator

7.1.2 Operation of steam generator "Infors HT"

TECHNISCHE **U**NIVERSITÄT **W**IEN / **TU VIENNA** / **V**IENNA **U**NIVERSITY OF **T**ECHNOLOGY	**Standard Operating Procedure**	**BioVT**
	Operation of steam generator "Infors HT"	Date: 13. 12. 2010

Version	1.0
Replaced version	none
Date	13. 12. 2010
Author	Patrick Sagmeister
Date	13. 12. 2010
Signature	
Authorized by	
Date of authorization	
Signature	

Summary	How to operate the steam generator "Infors HT"
Equipment	

- Steam Generator Infors HT

Procedure
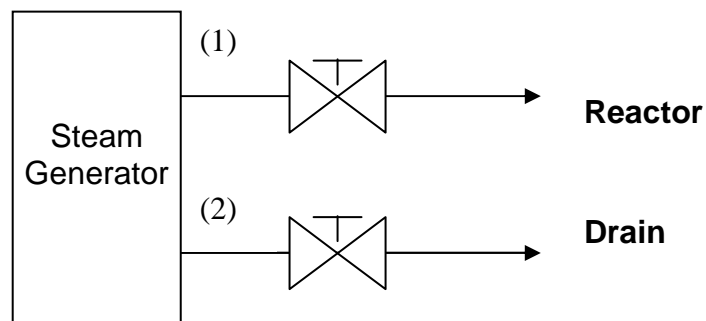
Starting Procedure

- valve 2 (to the drain) should be closed, valve 1 (to the reactor) and the valve at the reactor

(not shown) should be opened

- switch on the steam generator by using the green button, wait until the generator is heated up

- close valve 1

- wait until pressure is built up

Shutting down procedure

- switch off steam generator using the green button

- before carefully releasing pressure over valve 2 (to the sink), make sure that the hose leading

to the drain is tightly fixed and nobody stands close to the drain

    -carefully release pressure over valve 2

End

7.1.3 Protein determination *via* BCA

	Standard Operating Procedure	BioVT
TECHNISCHE UNIVERSITÄT WIEN **TU VIENNA** **VIENNA UNIVERSITY OF TECHNOLOGY**	**Use of the Bicinchoninic Acid Kit for Protein Determination (Sigma product number BCA1-1KT)**	
Research Division **Biochemical Engineering**		Date: 13.10.2010

Version	1.0
Replaced version	none
Author	Patrick Sagmeister
Date	13. 10. 2010

Summary	**Use of the Bicinchoninic Acid Kit for Protein Determination (Sigma product number: BCA1-1KT)**
Materials	
- BCA reagent A (Bichionic Acid Solution) - BCA reagent B (Copper II Solution) - Disposable plastic cuvettes	
Equipment	
- Spectrometer - Water bath	
Procedure	
i) Working reagent preparation: 1ml of reagent will be used per sample and standards. Reagent is prepared by mixing 50 parts of *Reagent A* with one part of *Reagent B*. For example, for 51 tests 50ml of *Reagent A* and 1ml of *Reagent B* are mixed. The solution should appear light green in colour. ii) Prepare a standard curve using 1ml BSA standard of 1mg BSA/ml. Use a 200µL piston	

pipette. As diluent, use the same buffer as used in your sample. Alternatively, BSA lyophilized standard can be used.

conc [µg/ml]	Standard [µl]	Diluent [µL]
200	40	160
400	80	120
600	120	80
800	160	40
1000	200	0

iii) Your samples should show concentrations in the calibration range. Otherwise dilute. For measurement of standards and samples, mix 50µl of protein sample with 1ml of the prepared BCA working reagent in an eppendorf tube and vortex. Choose one of the three possible incubation procedures:

a) 60°C using the water bath for 15 minutes.

b) at 37°C using the water bath for 30 minutes

c) at room temperature for a minimum of 2 hours or overnight

iv) After incubation, blank with working reagent and measure the absorbance of standards and samples at 562 nm.

v) Use your standard curve to determine concentrations of the unknown samples.

Literature
- Sigma BCA Protein Assay Kit Technical Bulletin
- (Sigma product number: BCA1-1KT)

End

7.2. Applied formulas

Calculation of reactor volume [l]

$$V_{R_{(t)}} = V_{R_{(t=0)}} + \frac{m_{Base_{(t=0)}} - m_{Base_{(t)}}}{\rho_{Base}} + \frac{m_{Feed_{(t=0)}} - m_{Feed_{(t)}}}{\rho_{Feed}} - f_{stripping} * (t - t_0) - f_{sampling} * (t - t_0)$$

Extracellular concentration correction [/]

$$f_{c.corr.} = x * \frac{WCW}{DCW} * \frac{1}{\rho_{mo}}$$

Water content in off gas [/]

$$x_{H2O} = \frac{x_{o2,IN} - x_{O2,wet}}{x_{o2,IN}}$$

Inert gas ratio [/]

$$r_{inert} = \frac{1 - x_{o2,IN} - x_{CO2,IN}}{1 - x_{o2,j} - x_{CO2,j} - x_{H2O}}$$

Volumetric growth rate [Cmmol/l/h]

$$r_x = \frac{x_t * V_{R_{(t)}} - x_{t=0} * V_{R_{(t=0)}}}{V_{R_{(t)}}} * \frac{1000}{M_X} * \frac{1}{(t - t_0)}$$

Volumetric substrate uptake rate [Cmmol/l/h]

$$r_s = \frac{m_{Feed_{(t)}} - m_{Feed_{(t=0)}}}{\rho_{Feed}} * x_{Feed} * \frac{1}{V_{R_{(t)}}} * \frac{1000}{M_X} * \frac{1}{(t - t_0)}$$

Volumetric protein release rate [Cmmol/l/h]

$$r_{protein} = \frac{c_{protein_{(t)}} * V_{R_{(t)}} - c_{protein_{(t=0)}} * V_{R_{(t=0)}}}{V_{R_{(t)}}} * \frac{1000}{M_{Protein}} * \frac{1}{(t - t_0)}$$

Oxygen uptake rate (OUR) [mmol/h]

$$r_{O2} = OUR = -\Gamma * 60 * \frac{1}{M_V} * (x_{O2,In} - r_{inert} * x_{O2,j}) * \frac{1000}{V_R}$$

Carbon dioxide evolution rate [Cmmol/l/h]

$$r_{CO2} = CER = \Gamma * 60 * \frac{1}{M_V} * (x_{CO2,j} * r_{inert} - x_{CO2,IN}) * \frac{1000}{V_R}$$

Ammonia uptake rate [mmol/l/h]

$$r_{NH_4^+} = \frac{m_{Base_{(t)}} - m_{Base_{(t=0)}}}{\rho_{Base}} * M_{Base} * \frac{1}{V_{R_{(t)}}} * \frac{1000}{(t - t_0)}$$

Biomass yield [Cmol/Cmol]

$$Y_{x/s} = -\frac{r_x}{r_s}$$

Extracellular protein yield [Cmol/Cmol]

$$Y_{Protein/s} = -\frac{r_{protein}}{r_s}$$

Carbon dioxide yield [Cmol/Cmol]

$$Y_{CO_2/s} = -\frac{CER}{r_s}$$

Ammonia yield [mol/mol]

$$Y_{NH_4^+/s} = -\frac{r_{NH_4^+}}{r_s}$$

C – balance [/]

$$Y_{x/s} + Y_{Protein/s} + Y_{CO_2/s} = 1$$

Degree of reduction balance [/]

$$\frac{r_x * \gamma_x + r_{protein} * \gamma_{protein}}{(r_{O2} * \gamma_{O2} + r_s * \gamma_s) * (-1)} = 1$$

Nitrogen balance [/]

$$\frac{r_x * x_n + r_{protein} * x_p}{- r_{NH_4}} = 1$$

Symbols

CER	carbon dioxide evolution rate	[mol/l/h]
fc. corr	concentration correction factor	[/]
Mi	molar weight	[g/mol]
MV	molar volume, gas	[l/mol]
MBase	molarity base	[mol/l[
mFeed	mass of feed flask	[g]
mBase	mass of base flask	[g]
OUR	oxygen uptake rate	[mol/l/h]
qi	specific conversion rate per gram of biomass	[Cmmol/g/h]
r	conversion rate	[Cmol/l/h]
rinert	inert gas ratio	[-]
RI	recovery of species I	[-]
RQ	respiratory quotient	[-]
t	expired time	[h]
VR	reactor volume	[l]
wi	weight from balance at time j	[g]
xi	molar fraction	[-]
xN,X	moles of N per mol biomass X	[-]
Ym/i	yield C-mol m per C-mole of i	[-]

Indices

Base	base addition for pH control
i	component ID
j	current time instant
n	component
N	nitrogen
Out	Exit envelope
Protein	extracellular protein
s	substrate

Greek Symbols:

Γ	gas flow	[l/min]
γi	degree of reduction coefficient of component I	[-]
μ	specific growth rate	[1/h]
ρ	density	[g/l]

7.3 QbD Glossary

	Definitions coined by the International Conference on Harmonisation *of Technical Requirements for Registration of Pharmaceuticals for Human Use* (ICH)	Reference
Control Strategy	"A planned set of controls, derived from current product and process understanding that ensures process performance and product quality. The controls can include parameters and attributes related to drug substance and drug product materials and components, facility and equipment operating conditions, in-process controls, finished product specifications, and the associated methods and frequency of monitoring and control."	ICH 2008b
Critical Process Parameter (CPP):	"A process parameter whose variability has an impact on a critical quality attribute and therefore should be monitored or controlled to ensure the process produces the desired quality."	ICH 2008a
Critical Quality Attribute (CQA):	"A physical, chemical, biological or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality."	ICH 2008a
Design Space:	"The multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality. Working within the design space is not considered as a change. Movement out of the design space is considered to be a change and would normally initiate a regulatory post approval change process. Design space is proposed by the applicant and is subject to regulatory assessment and approval."	ICH 2008a
Quality:	"The suitability of either a drug substance or a drug product for its intended use. This term includes such attributes as the identity, strength, and purity (ICH Q6A).*"*	ICH 1999.
Quality by Design (QbD):	"A systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management."	ICH 2008a
Quality Target Product Profile (QTPP):	"A prospective summary of the quality characteristics of a drug product that ideally will be achieved to ensure the desired quality, taking into account safety and efficacy of the drug product."	ICH 2008a
Real Time Release Testing:	"The ability to evaluate and ensure the quality of in-process and/or final product based on process data, which typically include a valid combination of measured material attributes and process controls."	ICH 2008a

7.4 Statistical notes

	Definition	Reference
DOE	Design of experiments. Statistical experimental design for the investigation of a predefined experimental objective.	Eriksson 2000
MLR	Multiple linear regression. Linear regression for multiple factors. The equation $Y = X * B + E$ is solved using singular value decomposition. Y…Response matrix, n*m matrix where m is the number of responses X…Design matrix, n*p matrix where p is the number of terms of the model B…Regression coefficients matrix E…Residual matrix	Eriksson 2000
PRESS	Predictive residual sum of squares. A measure for the future prediction capacity of the model and included in the calculation of Q^2. $$PRESS = \sum_{i=1}^{n} \frac{(y_n - \overline{y_n})^2}{(1 - h_i)^2}$$ h_i: The hat matrix' i^{th} diagonal element.	Eriksson 2000
SS	Sum of squares. The sum of the squared difference between the observed values and the predicted values. The residual sum of squares can be spitted into regression sum of squares (SS_{regr}) and residual sum of squares (SS_{resid}).	Eriksson 2000
SS$_{resid}$	The residual sum of squares (SS_{resid}) is composed of the pure error (SS_{pe}) and the lack of fit (SS_{lof})	Eriksson 2000
R^2	Goodness of fit parameter. Stretching from zero to one, R^2 indicates the fraction of variance explained by the model. $$R^2 = \frac{SS - SS_{resid}}{SS}$$	Eriksson 2000
Q^2	Goodness of prediction parameter. Stretching from minus infinity to one, Q^2 is a benchmark for the predictive power of the model. $$Q^2 = \frac{SS - PRESS}{SS}$$	Eriksson 2000

Model Validity	The "model validity" parameter estimates diverse model problems. Its calculation is based on the lack of fit test.	Eriksson 2000
Reproducibility	Measure for the replicate error. Two experiments are considered replicates if their factor values do not differ more than 5%.	Eriksson 2000

7.5. Acquired raw- and processed data

In the following raw- and processed data used for the knowledge space DOE as well as the process analysis DOEs is displayed. For a better orientation in the appendix each fermentation run was given a number, as illustrated in figure 48. This number neither reflects the timely experimental run order nor any other information.
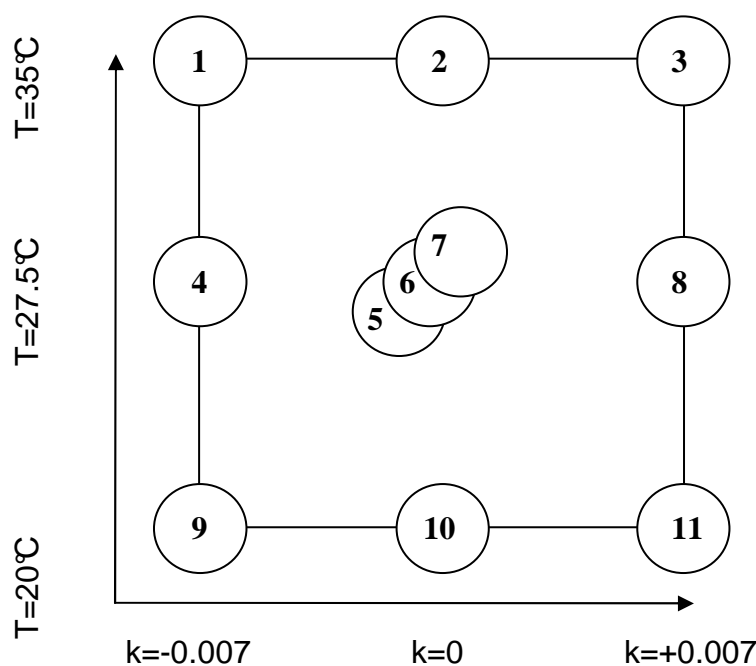
Figure 48: The experiments used for the establishment of the knowledge space as well as the process analysis DOEs. The induction phase feed exponent (k) is given on the x-axis, the induction phase temperature (T) on the y-axis. Each circle illustrates one fermentation run. For a better orientation within the appendix each fermentation run was given a number (numbers 1-11).

Acquired data of fermentation 1

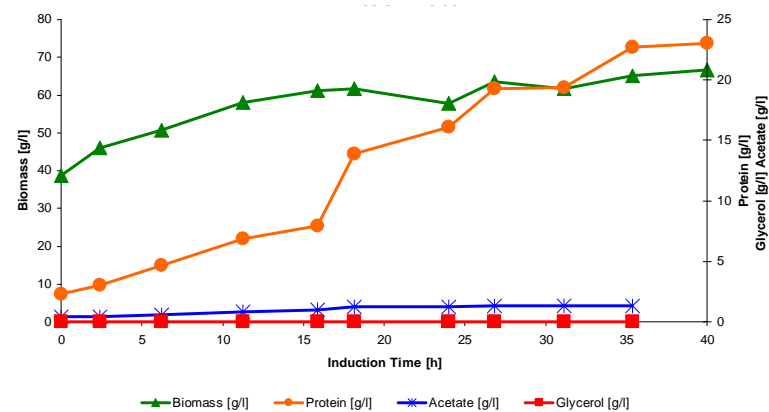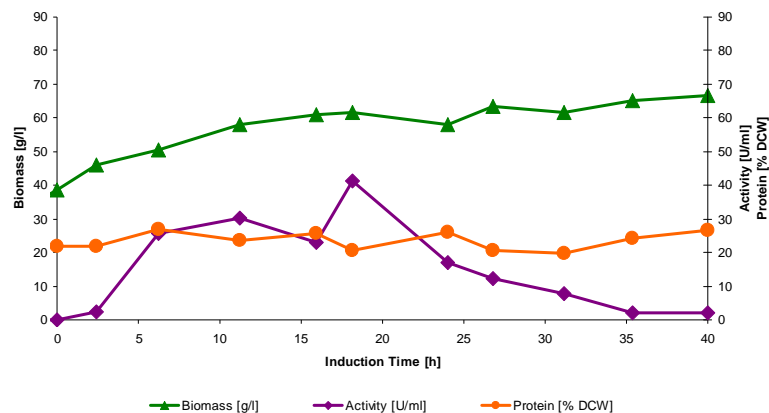Fermentation parameters: induction temperature: 35°C, k=-0.007 (negative exponential feed).
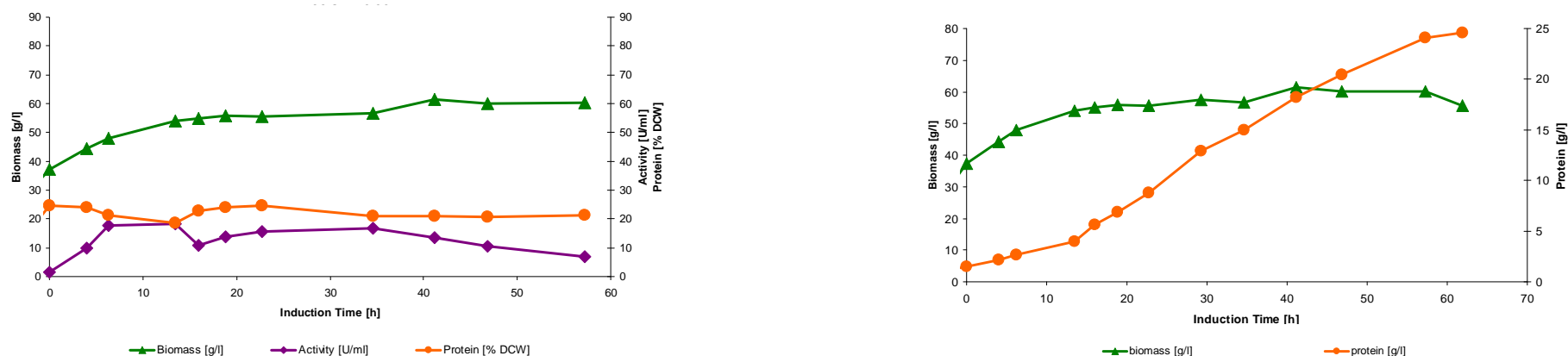
Figure 49: Fermentation at 35°C applying a negative exponential feed (k=-0.007). Left: **i**ntracellular protein concentrations (orange), intracellular activity (purple) and biomass concentrations (green). Right: extracellular protein (orange) concentrations and biomass concentrations (green). X-axis: induction time.
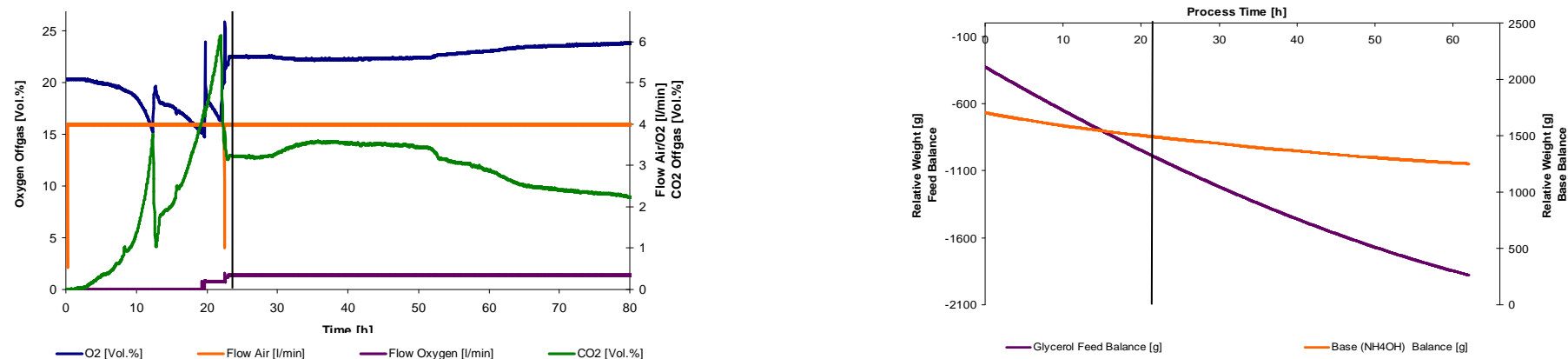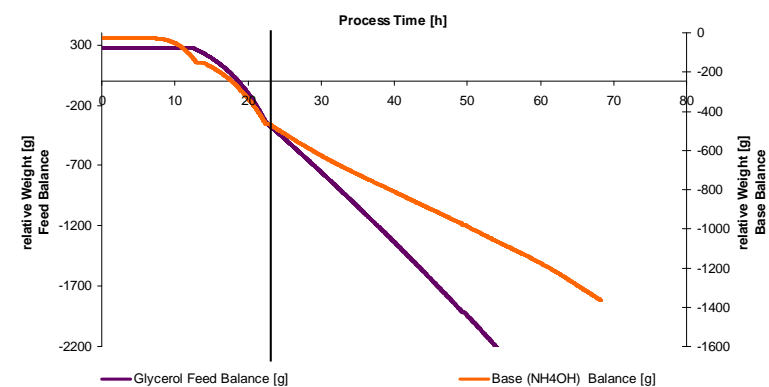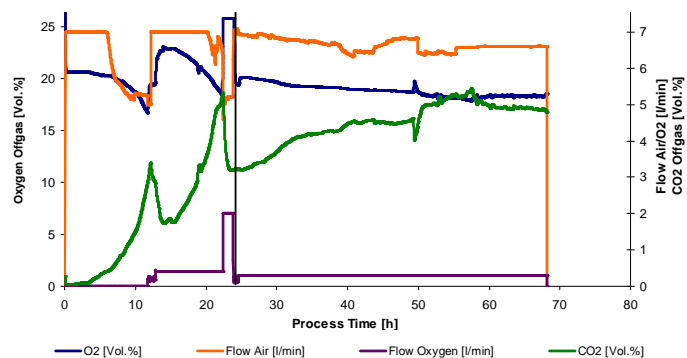

Figure 50: Fermentation at 35°C applying a negative exponential feed (k=-0.007). Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed data of fermentation 1

Fermentation parameters: induction temperature: 35°C; k=-0.007 (negative exponential feed).
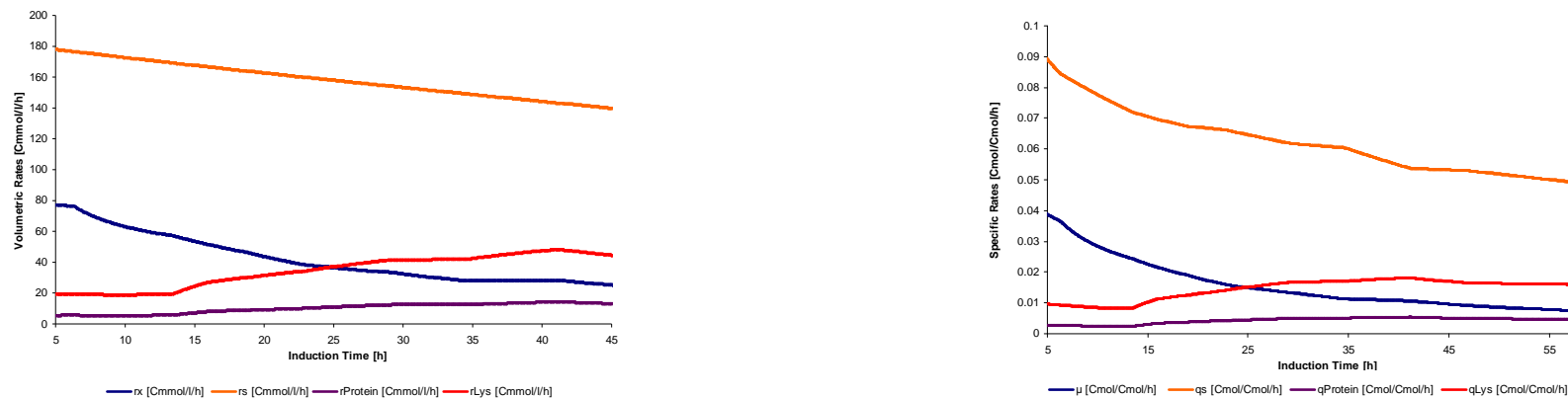
Figure 51 Fermentation at 35°C applying a negative exponential feed (k=-0.007). Left: **v**olumetric growth rate rx (blue), volumetric substrate uptake rate (rs), protein release rate r$_{Protein}$ (purple) and volumetric cell lysis rate (red). Right: **s**pecific growth rate μ (blue), specific substrate uptake rate qs (orange), specific protein release rate q$_{Protein}$ (purple) and specific cell lysis rate r$_{Lys}$ (red). X-axis: induction time.
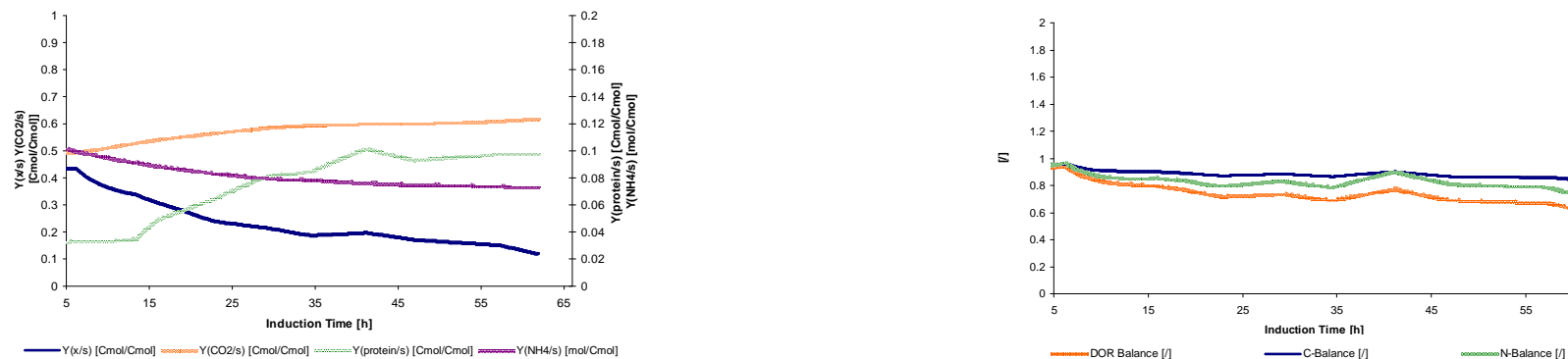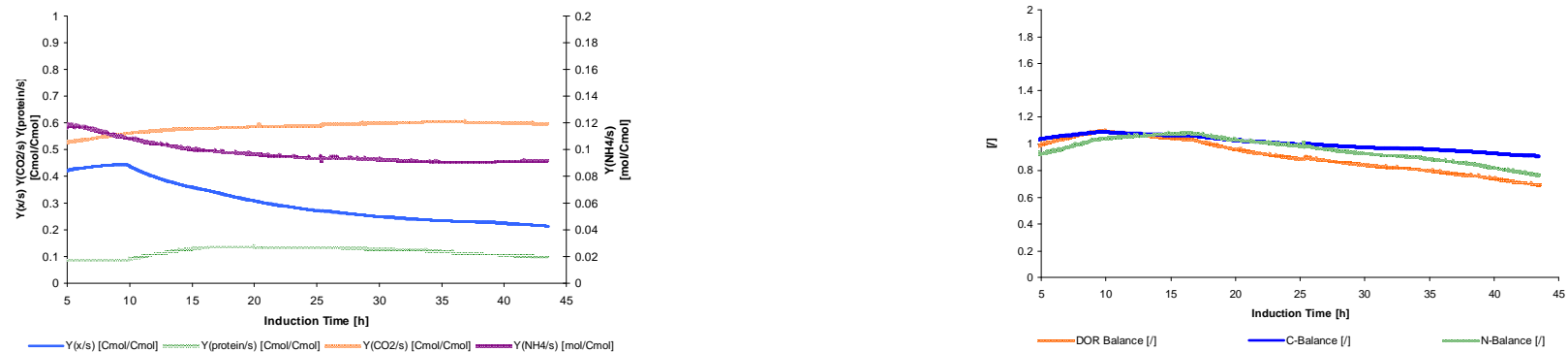

Figure 52: Fermentation at 35°C applying a negative exponential feed (k=-0.007). Left: biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen (green) balance.

Acquired data of fermentation 2

Fermentation parameters: induction temperature: 35°C, k=0 (linear feed).
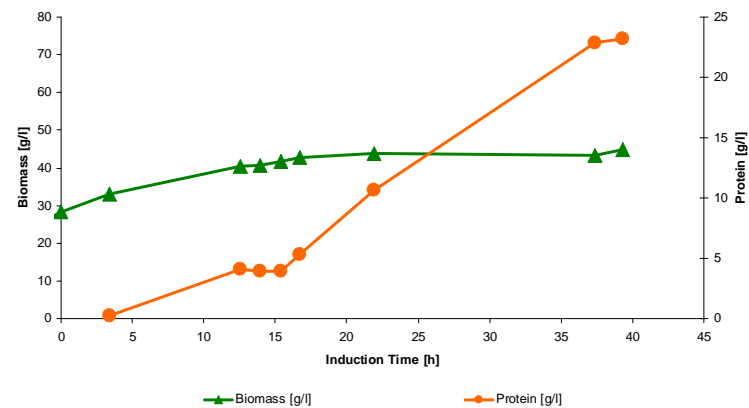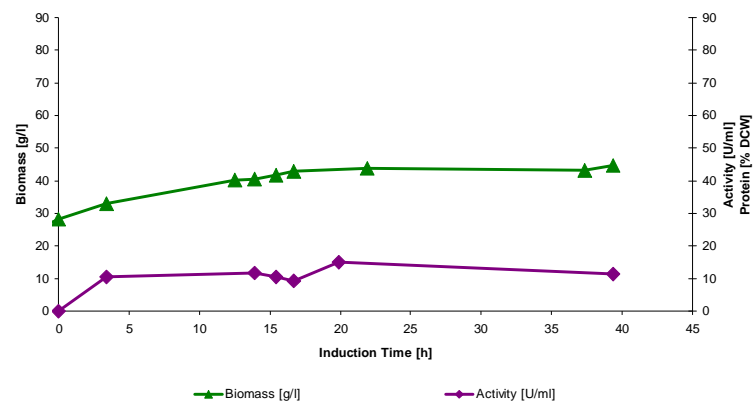
Figure 53: Fermentation at 35°C applying a linear feed (k=0). Left: intracellular activity (purple) and biomass concentrations (green). Right: extracellular protein (orange) concentrations and biomass concentrations (green). X-axis: induction time.
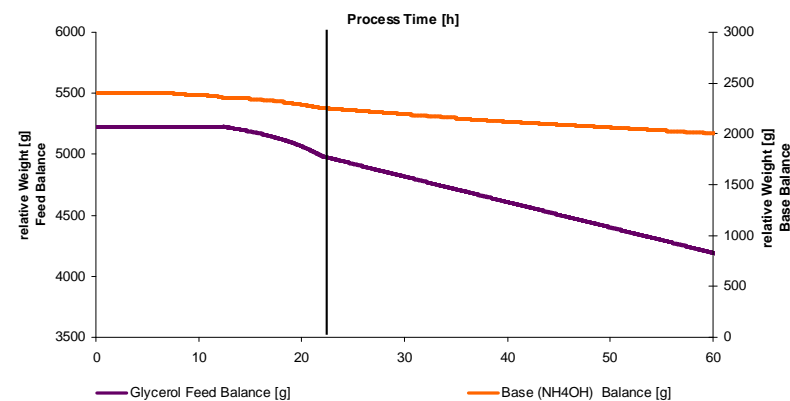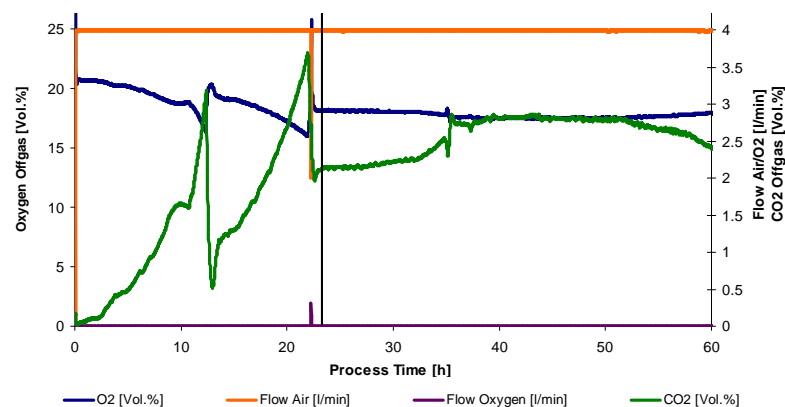
Figure 54: Fermentation at 35°C applying a linear feed (k=-0.007). Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed data of fermentation 2

Fermentation parameters: induction temperature: 35°C; k=0 (linear feed).

Figure 55: Fermentation at 35°C applying a linear feed (k=0). Left: **v**olumetric growth rate rx (blue), volumetric substrate uptake rate (rs), protein release rate $r_{Protein}$ (purple) and volumetric cell lysis rate (red). Right: **s**pecific growth rate µ (blue), specific substrate uptake rate qs (orange), specific protein release rate $q_{Protein}$ (purple) and specific cell lysis rate r_{Lys} (red). X-axis: induction time.
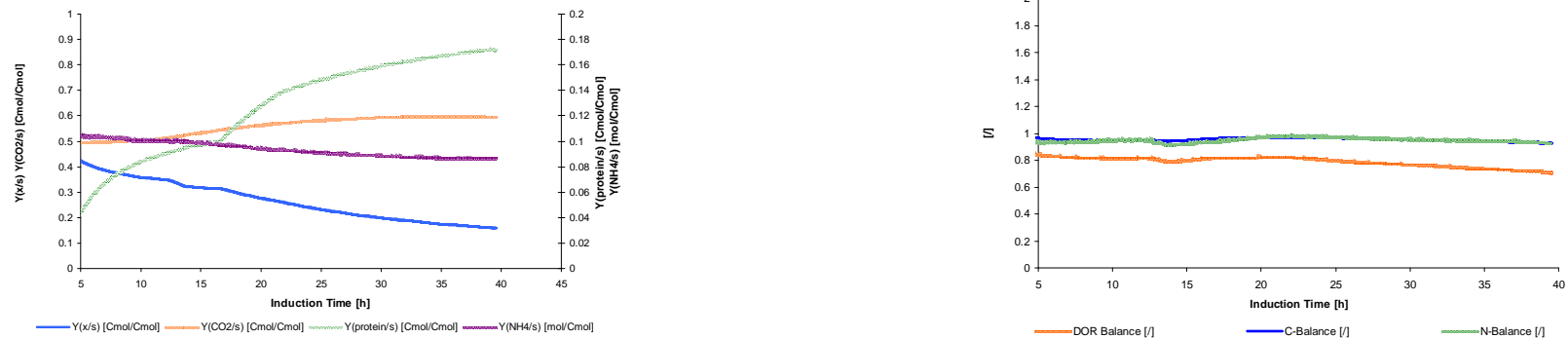
Figure 56: Fermentation at 35°C applying a linear feed (k=0). Left: biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen (green) balance.

Acquired data of fermentation 3

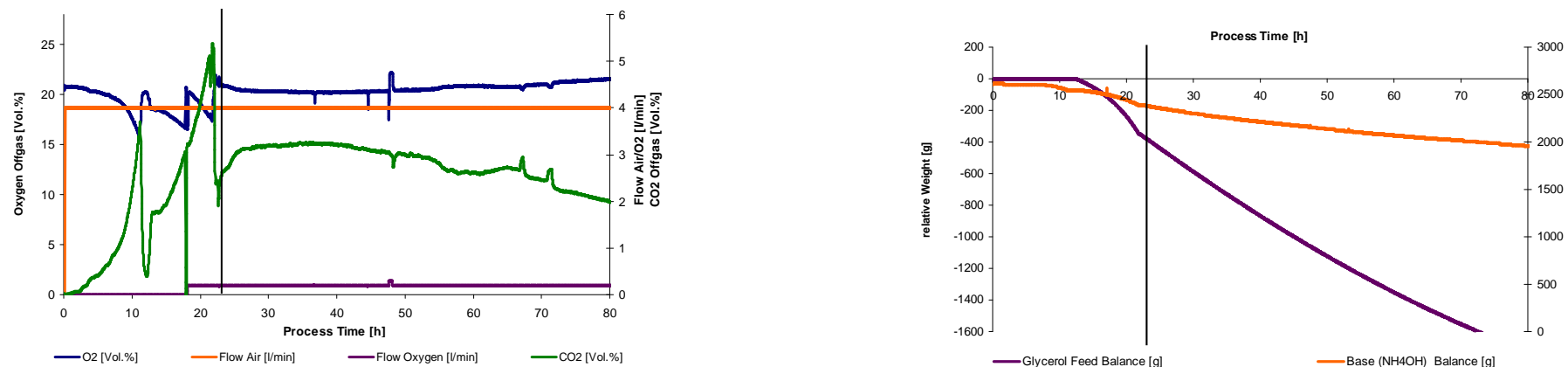Fermentation parameters: induction temperature: 35°C, k=+0.007 (positive exponential feed).

Figure 57: Fermentation at 35°C applying a positive exponential feed (k=+0.007). Left: intracellular protein concentrations (orange), intracellular activity (purple) and biomass concentrations (green). Right: extracellular protein- (orange), glycerol- (red), acetate- (blue) and biomass (green) concentrations. X-axis: induction time.

Figure 58: Fermentation at 35°C applying a positive exponential feed (k=+0.007). Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed data of fermentation 3

Fermentation parameters: induction temperature: 35°C; k=+0.007 (positive exponential feed).

Figure 59 Fermentation at 35°C applying a positive exponential feed (k=+0.007). Left: volumetric growth rate rx (blue), volumetric substrate uptake rate (rs), protein release rate $r_{Protein}$ (purple) and volumetric cell lysis rate (red). Right: specific growth rate μ (blue), specific substrate uptake rate qs (orange), specific protein release rate $q_{Protein}$ (purple) and specific cell lysis rate r_{Lys} (red). X-axis: induction time.
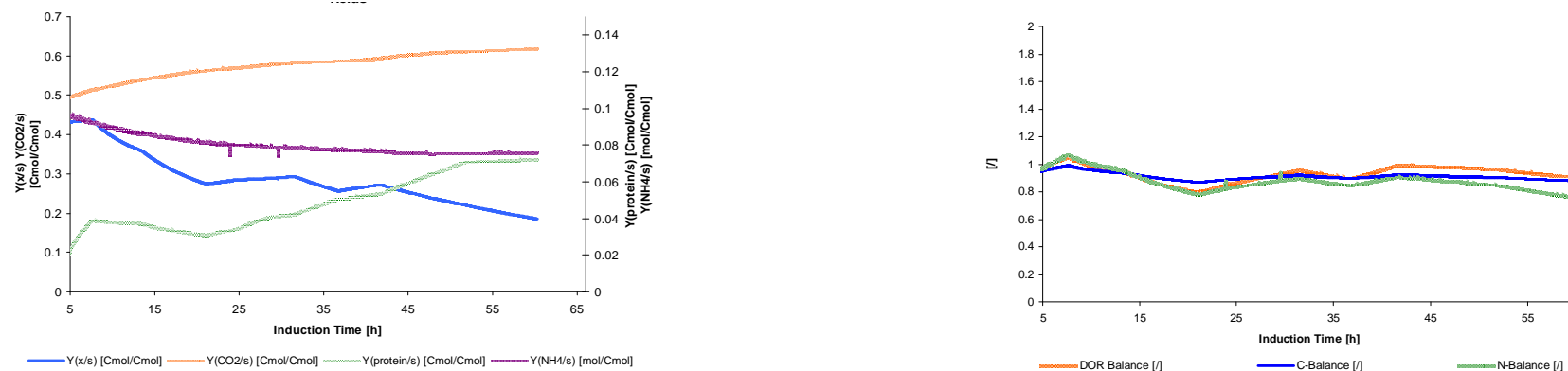

Figure 60: Fermentation at 35°C applying a positive exponential feed (k=+0.007). Left: biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen (green) balance.

Acquired data of fermentation 4

Fermentation parameters: induction temperature: 27.5°C, k=-0.007 (negative exponential feed).
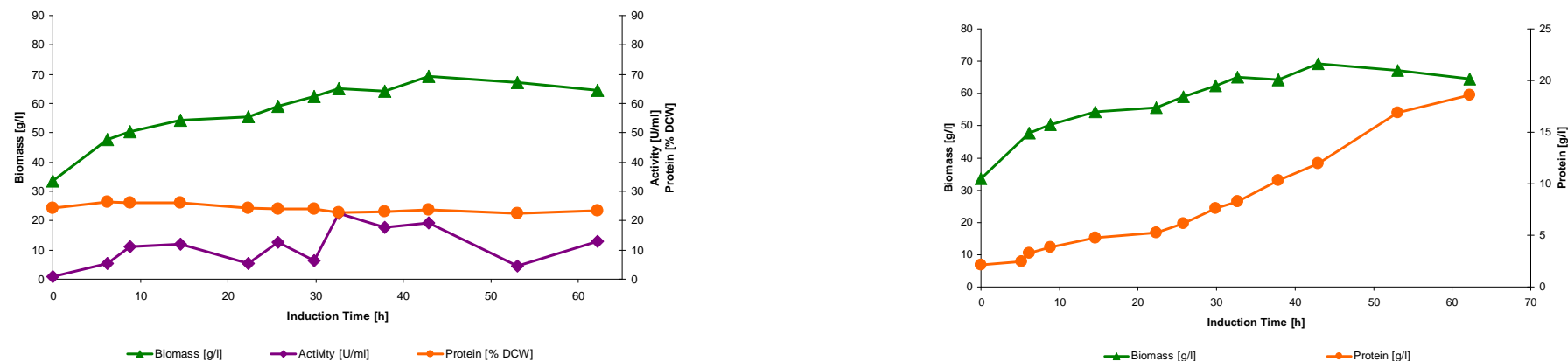
Figure 61: Fermentation at 27.5°C applying a negative exponential feed (k=-0.007). Left: intracellular protein concentrations (orange), intracellular activity (purple) and biomass concentrations (green). Right: extracellular protein- (orange) and biomass (green) concentrations. X-axis: induction time.

Figure 62: Fermentation at 27.5°C applying a negative exponential feed (k=-0.007). Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed data of fermentation 4

Fermentation parameters: induction temperature: 27.5°C, k=-0.007 (negative exponential feed).
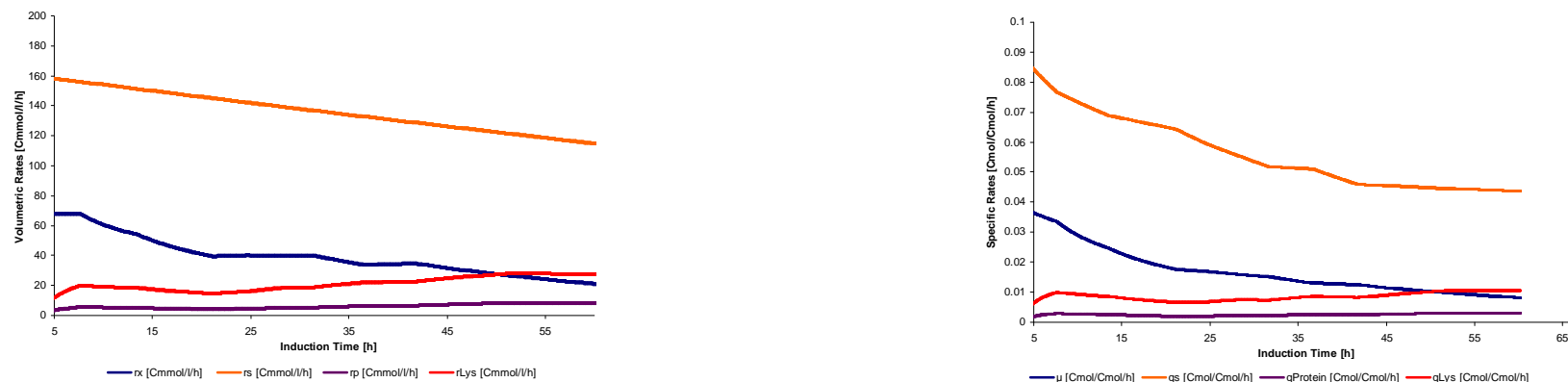
Figure 63: Fermentation at 27.5°C applying a negative exponential feed (k=-0.007). Left: **v**olumetric growth rate rx (blue), volumetric substrate uptake rate (rs), protein release rate r_p (purple) and volumetric cell lysis rate (red). Right: **s**pecific growth rate μ (blue), specific substrate uptake rate qs (orange), specific protein release rate $q_{Protein}$ (purple) and specific cell lysis rate r_{Lys} (red). X-axis: induction time.
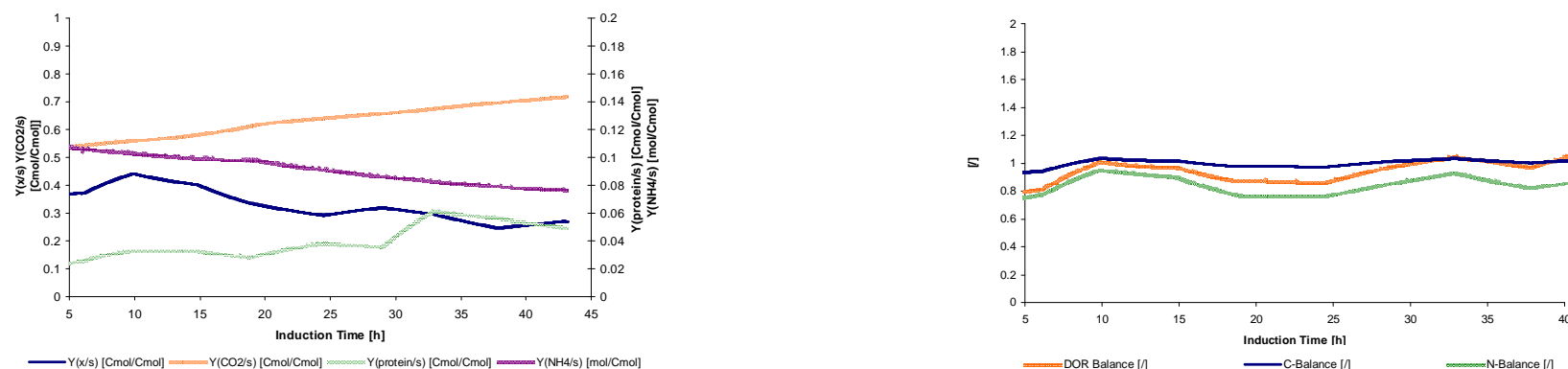

Figure 64: Fermentation at 27.5°C applying a negative exponential feed (k=-0.007). Left: Biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen balance (green).

Acquired data of fermentation 5

Fermentation parameters: Induction temperature: 27.5°C, k=0 (linear feed).
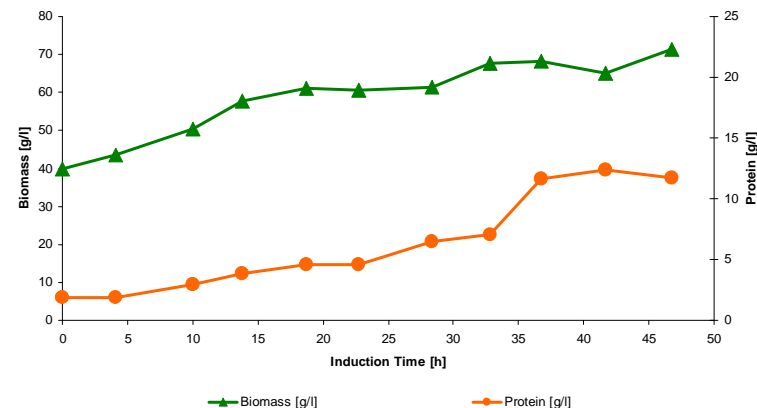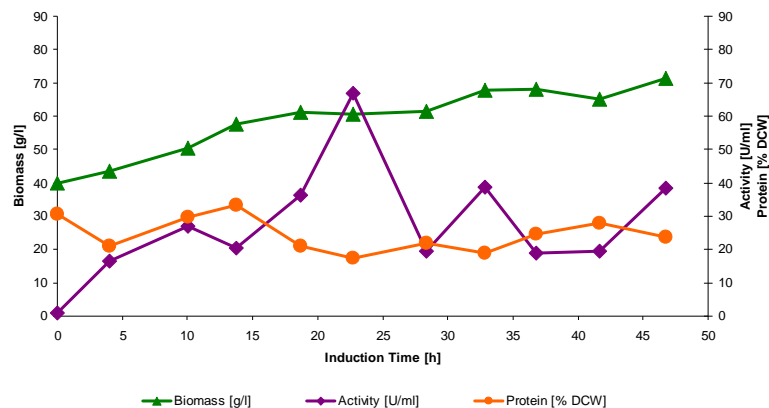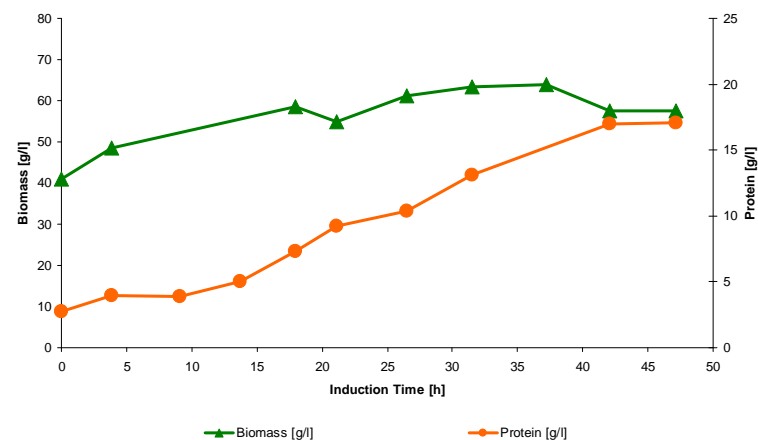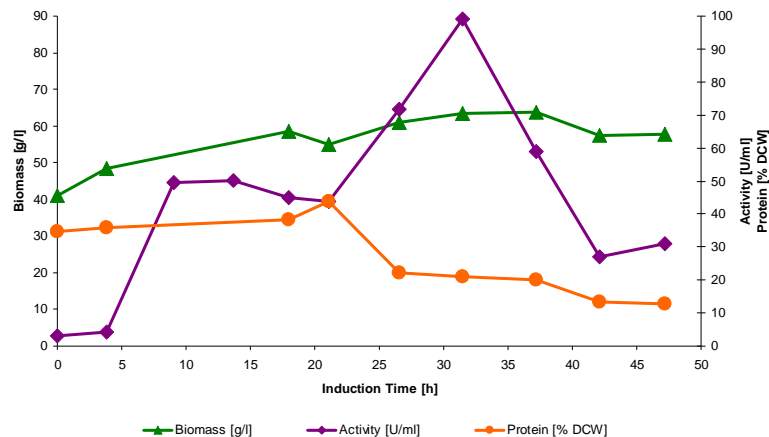

Figure 65: Fermentation at 27.5°C applying a linear feed (k=0). Left: intracellular protein concentrations (orange circles), intracellular activity (purple diamonds) and biomass concentrations (green triangles). Right: extracellular protein (orange circles) concentrations and biomass concentrations (green triangles). X-axis: induction time.
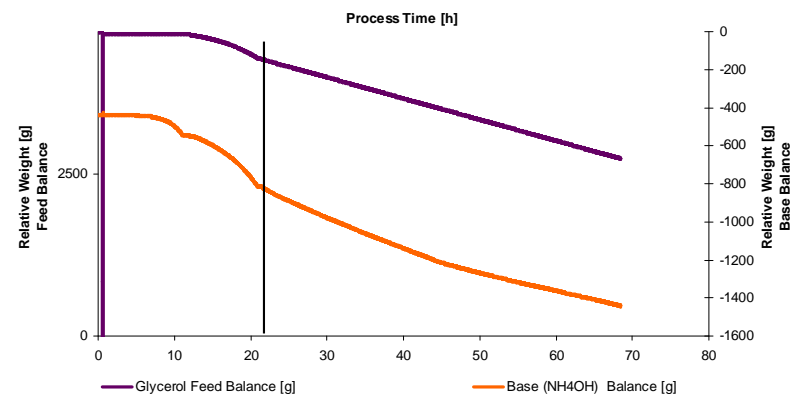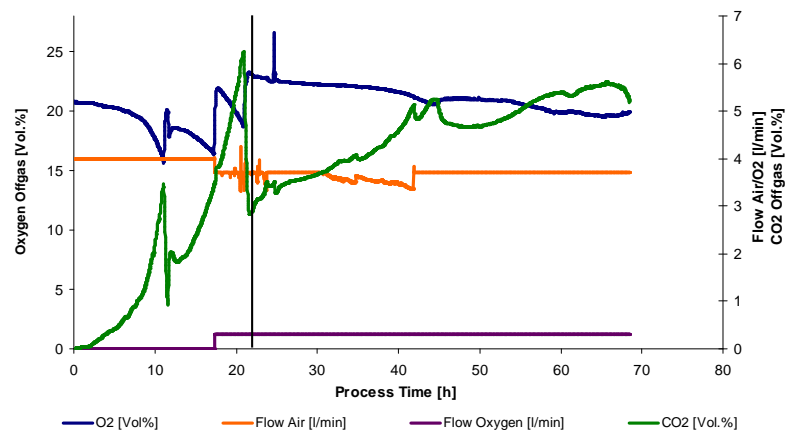
Figure 66: Fermentation at 27.5°C applying a linear feed (k=0). Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed data of fermentation 5

Fermentation parameters: induction temperature: 27.5°C; k=0 (linear feed).
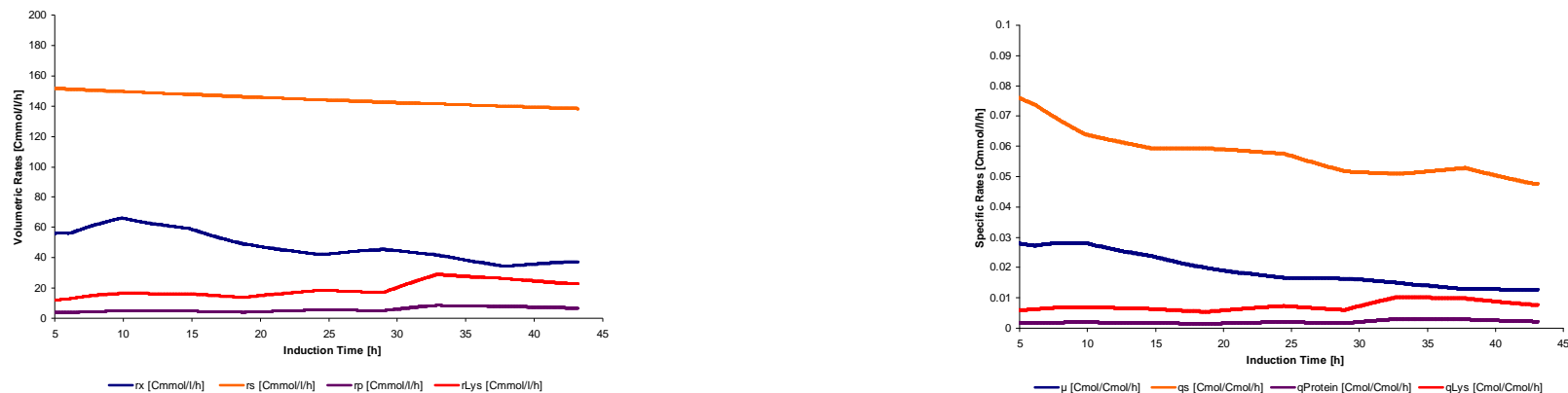
Figure 67: Fermentation at 35°C applying a linear feed (k=0). Left: **v**olumetric growth rate rx (blue), volumetric substrate uptake rate (rs), volumetric protein release rate r_p (purple) and volumetric cell lysis rate (red). Right: specific growth rate μ (blue), specific substrate uptake rate qs (orange), specific protein release rate $q_{Protein}$ (purple) and specific cell lysis rate r_{Lys} (red). X-axis: induction time.

Figure 68: Fermentation at 27.5°C applying a linear feed (k=0). Left: biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen balance (green).

Acquired data of fermentation 6

Fermentation parameters: induction temperature: 27.5°C, k=0 (linear feed).

Figure 69: Fermentation at 27.5°C applying a linear feed (k=0). Left: intracellular protein concentrations (orange circles), intracellular activity (purple diamonds) and biomass concentrations (green triangles). Right: extracellular protein (orange circles) concentrations and biomass concentrations (green triangles). X-axis: induction time.
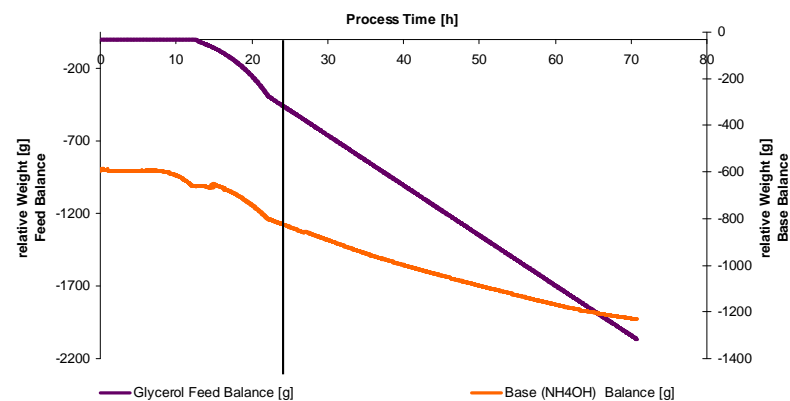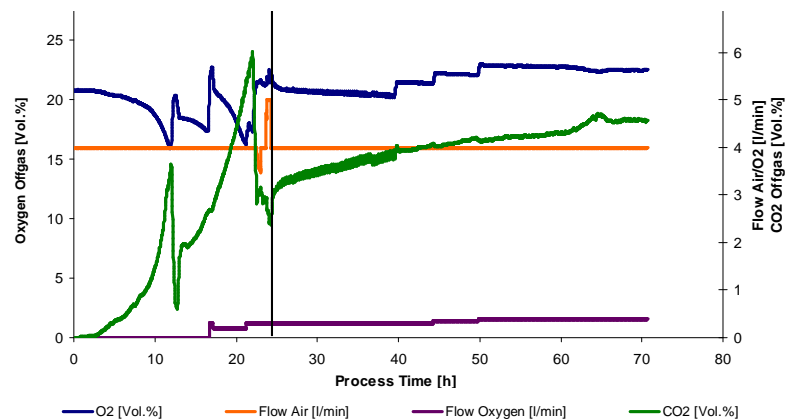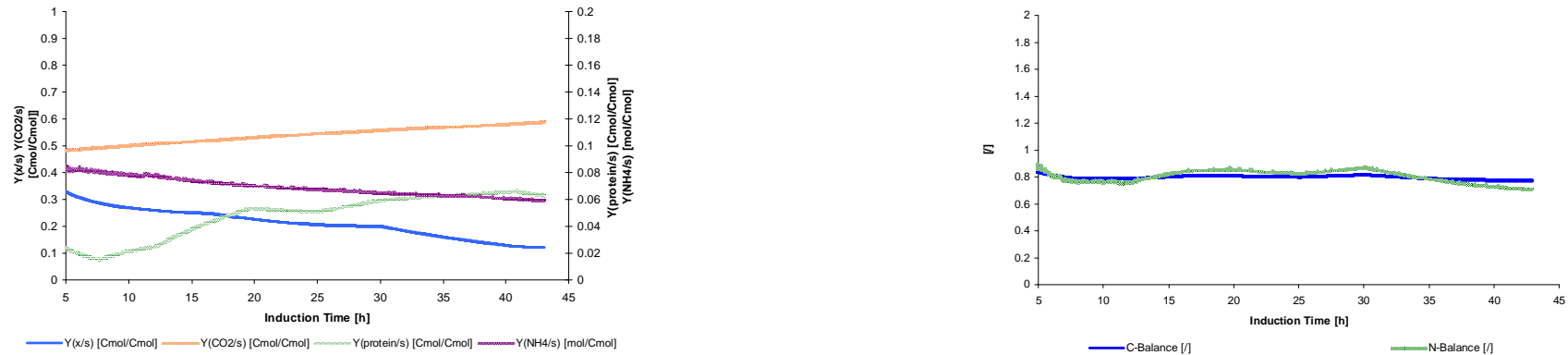

Figure 70: Fermentation at 27.5°C applying a linear feed (k=0). Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed data of fermentation 6

Fermentation parameters: induction temperature: 27.5°C; k=0 (linear feed).

Figure 71: Fermentation at 35°C applying a linear feed (k=0). Left: **v**olumetric growth rate rx (blue), volumetric substrate uptake rate (rs), protein release rate r_p (purple) and volumetric cell lysis rate (red). Right: specific growth rate μ (blue), specific substrate uptake rate qs (orange), specific protein release rate $q_{Protein}$ (purple) and specific cell lysis rate r_{Lys} (red). X-axis: induction time.

Figure 72: Fermentation at 27.5°C applying a linear feed (k=0). Left: biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen (green) balance.

Acquired data of fermentation 7

Fermentation parameters: induction temperature: 27.5°C, k=0 (linear feed).
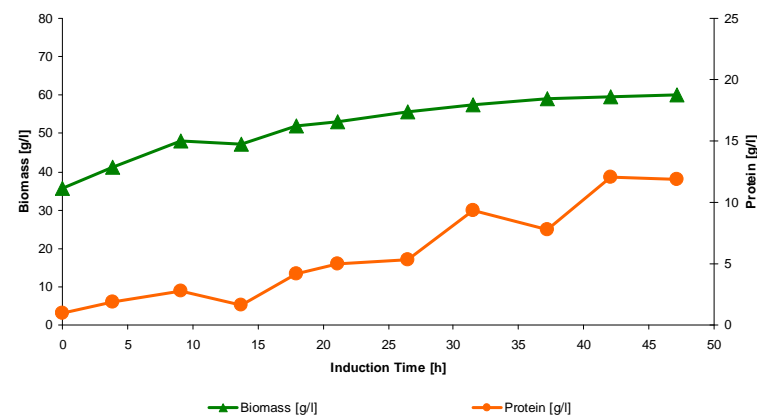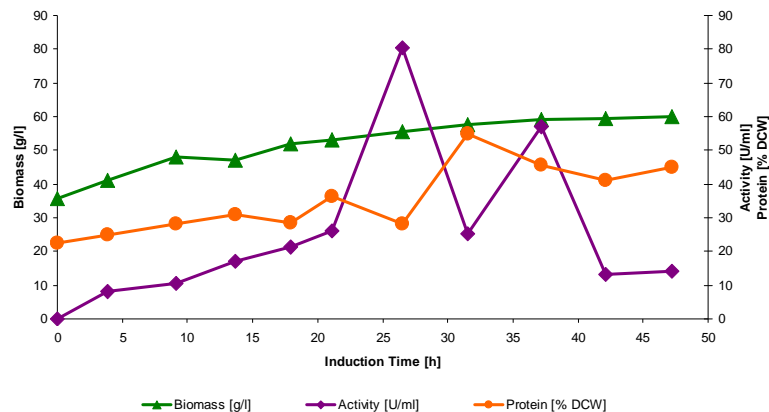

Figure 73: Fermentation at 27.5°C applying a linear feed (k=0). Left: intracellular protein concentrations (orange circles), intracellular activity (purple diamonds) and biomass concentrations (green triangles). Right: extracellular protein (orange circles) concentrations and biomass concentrations (green triangles). X-axis: induction time.
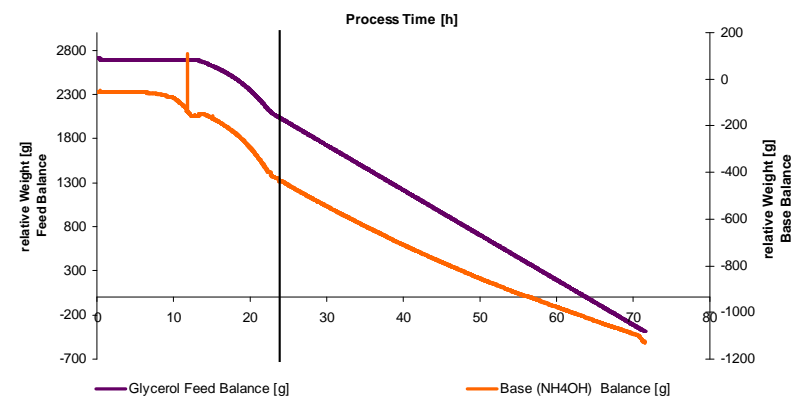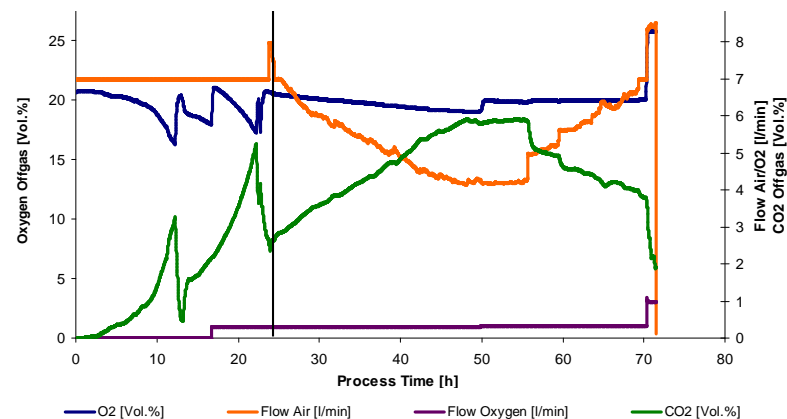

Figure 74: Fermentation at 27.5°C applying a linear feed (k=0). Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed data of fermentation 7

Fermentation parameters: induction temperature: 27.5°C; k=0 (linear feed).

Figure 75: Fermentation at 35°C applying a linear feed (k=0). Left: volumetric growth rate rx (blue), volumetric substrate uptake rate (rs), protein release rate $r_{Protein}$ (purple) and volumetric cell lysis rate (red). Right: specific growth rate μ (blue), specific substrate uptake rate qs (orange), specific protein release rate q_p (purple) and specific cell lysis rate r_{Lys} (red). X-axis: induction time.

Figure 76: Fermentation at 27.5°C applying a linear feed (k=0). Left: biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen (green) balance.

Acquired data of fermentation 8

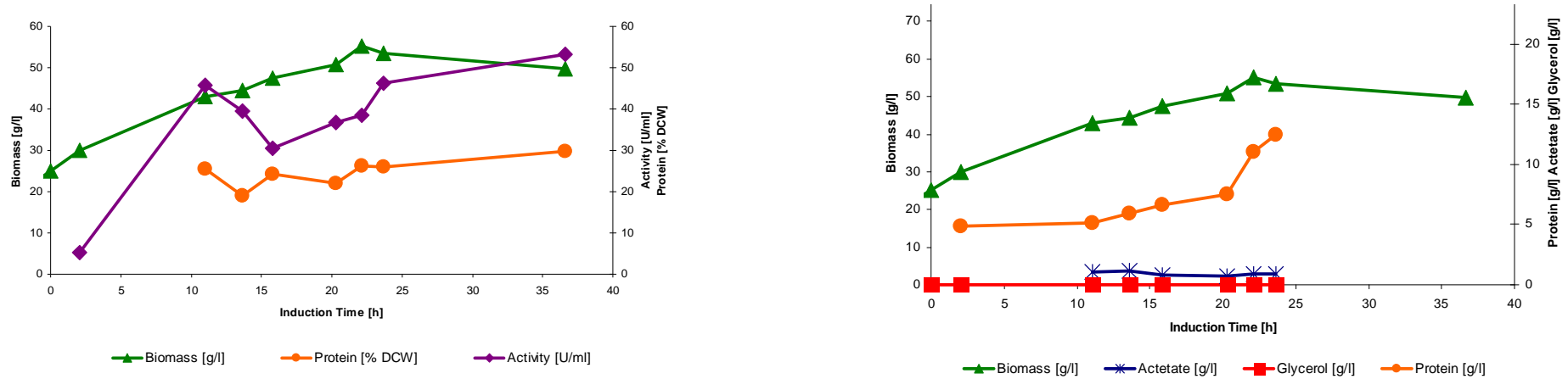Fermentation parameters: induction temperature: 27.5°C, k=+0.007 (positive exponential feed).

Figure 77: Fermentation at 27.5°C applying a positive exponential feed (k=+0.007). Left: intracellular protein concentrations (orange), intracellular activity (purple) and biomass concentrations (green). Right: extracellular protein (orange) concentrations and biomass concentrations (green). X-axis: induction time.
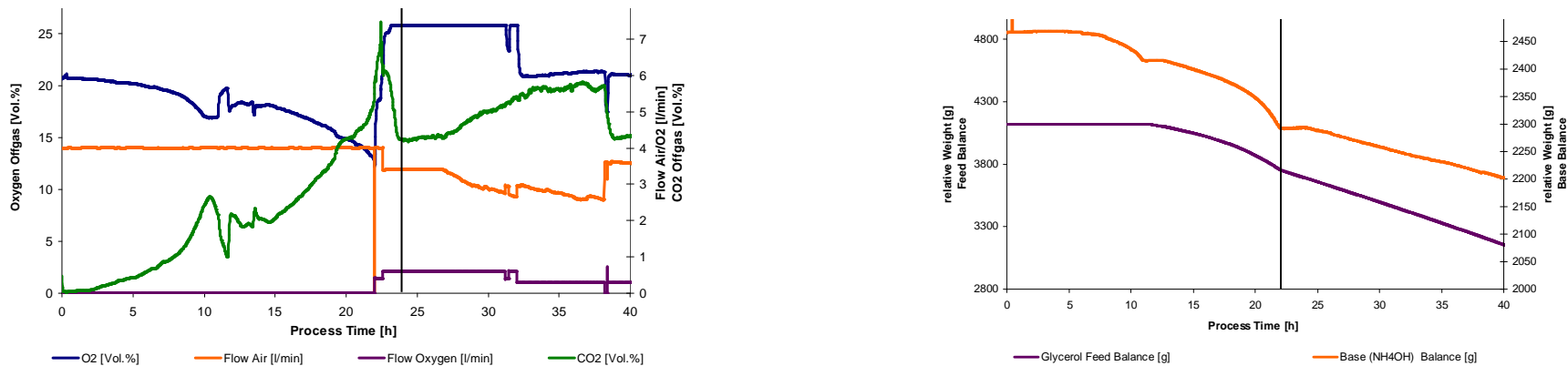

Figure 78: Fermentation at 27.5°C applying a positive exponential feed (k=+0.007) Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed data of fermentation 8

Fermentation Parameters: Induction Temperature: 27.5°C, k=+0.007 (positive exponential feed).

Figure 79: Fermentation at 27.5°C applying a positive exponential feed (k=+0.007). Left: volumetric growth rate rx (blue), volumetric substrate uptake rate (rs), protein release rate rp (purple) and volumetric cell lysis rate (red). Right: specific growth rate µ (blue), specific substrate uptake rate qs (orange), specific protein release rate $q_{Protein}$ (purple) and specific cell lysis rate r_{Lys} (red). X-axis: induction time.
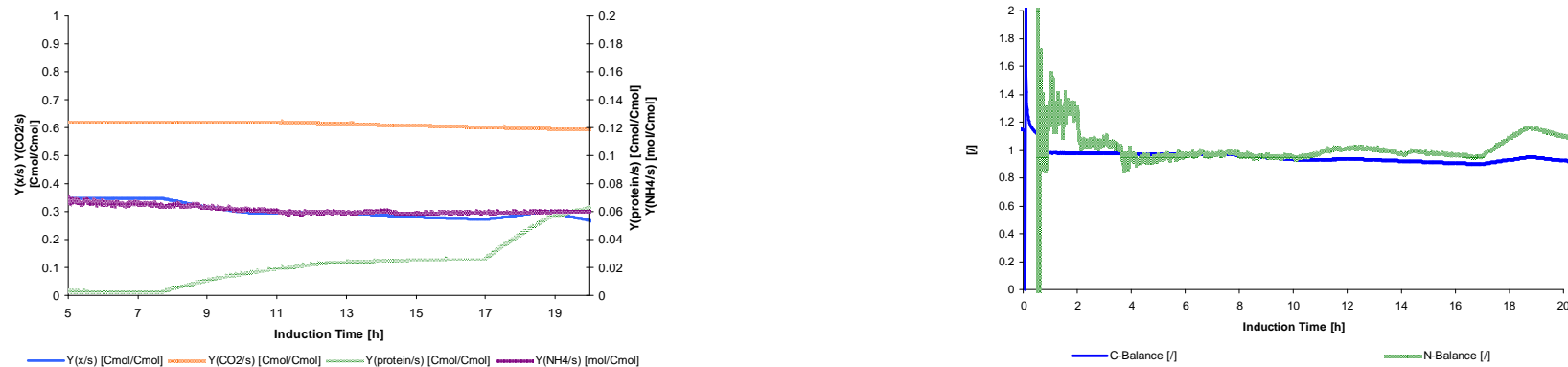

Figure 80: Fermentation at 27.5°C applying a positive exponential feed (k=+0.007). Left: biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen (green) balance.

Acquired data of fermentation 9

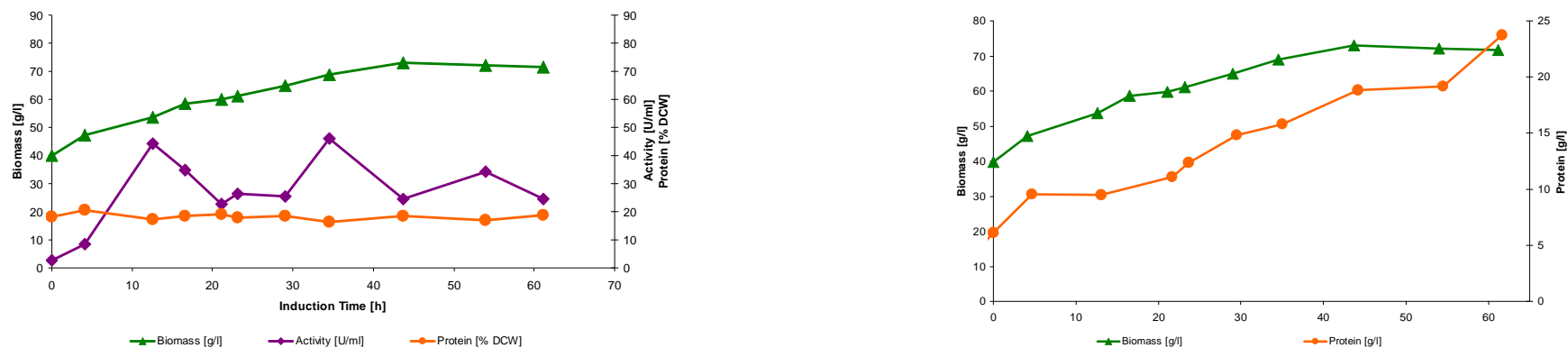Fermentation parameters: induction temperature: 20°C, k=-0.007 (negative exponential feed).

Figure 81: Fermentation at 20°C applying a negative exponential feed (k=-0.007). Left: intracellular protein concentrations (orange), intracellular activity (purple), acetate concentrations (blue) and biomass concentrations (green). Right: extracellular protein- (orange) and biomass (green) concentrations. X-axis: induction time.
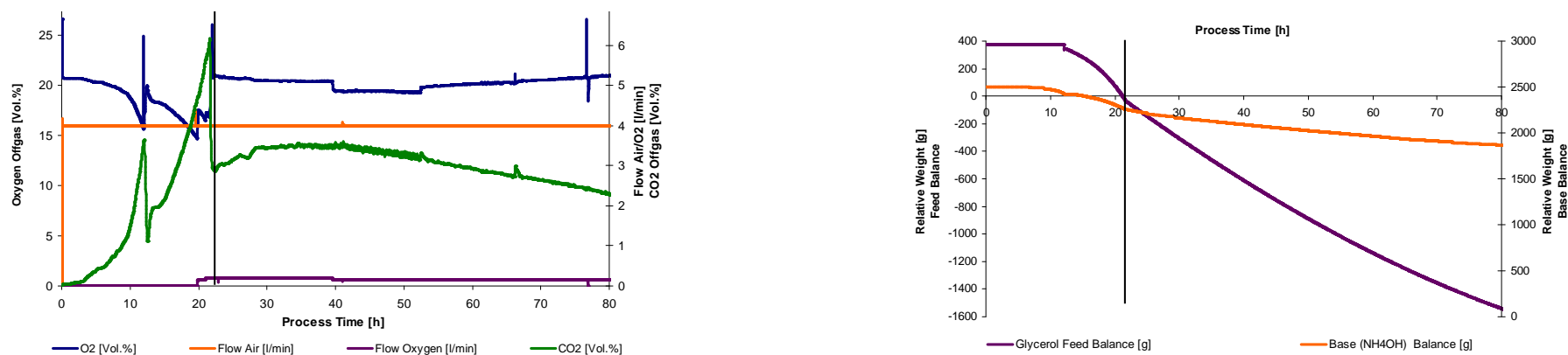

Figure 82: Fermentation at 20°C applying a negative exponential feed (k=-0.007). Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed data of fermentation 9

Fermentation parameters: induction temperature: 20°C, k=-0.007 (negative exponential feed).

Figure 83: Fermentation at 20°C applying a negative exponential feed (k=-0.007). Left: volumetric growth rate rx (blue), volumetric substrate uptake rate (rs), protein release rate r_P (purple) and volumetric cell lysis rate (red). Right: specific growth rate μ (blue), specific substrate uptake rate qs (orange), specific protein release rate q_P (purple) and specific cell lysis rate r_{Lys} (red). X-axis: induction time.
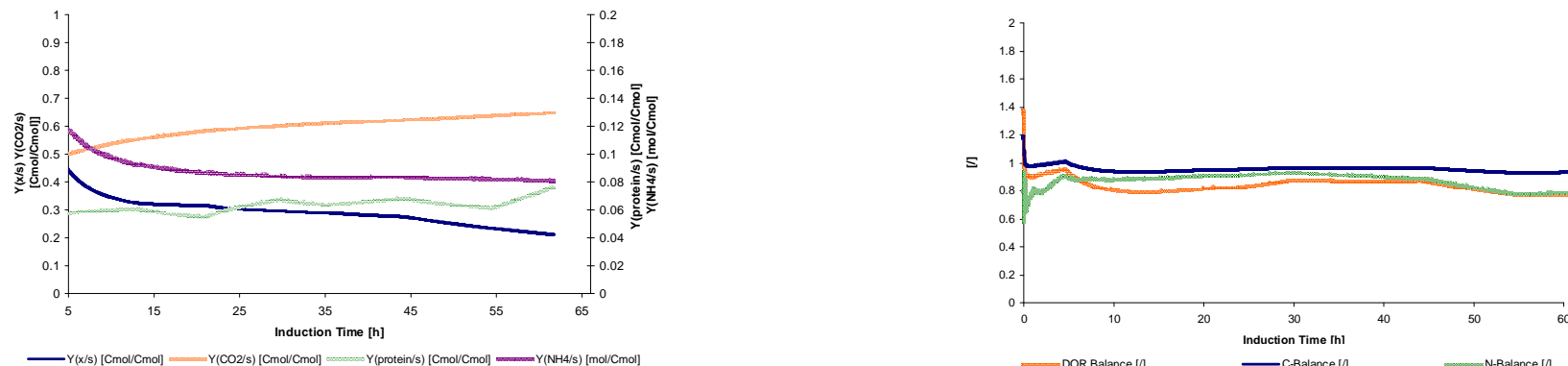

Figure 84: Fermentation at 20°C applying a negative exponential feed (k=-0.007). Left: biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen (green) balance.

Acquired data of fermentation 10

Fermentation parameters: induction temperature: 20°C, k=0 (linear feed).
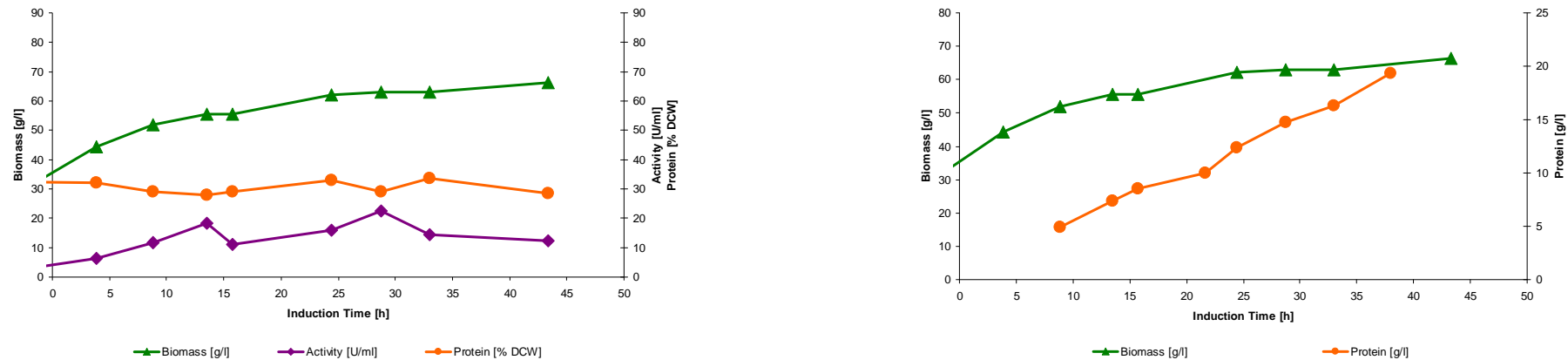

Figure 85: Fermentation at 20°C applying a linear feed (k=0). Left: intracellular protein concentrations (orange), intracellular activity (purple) and biomass concentrations (green). Right: extracellular protein- (orange) and biomass (green) concentrations. X-axis: induction time.
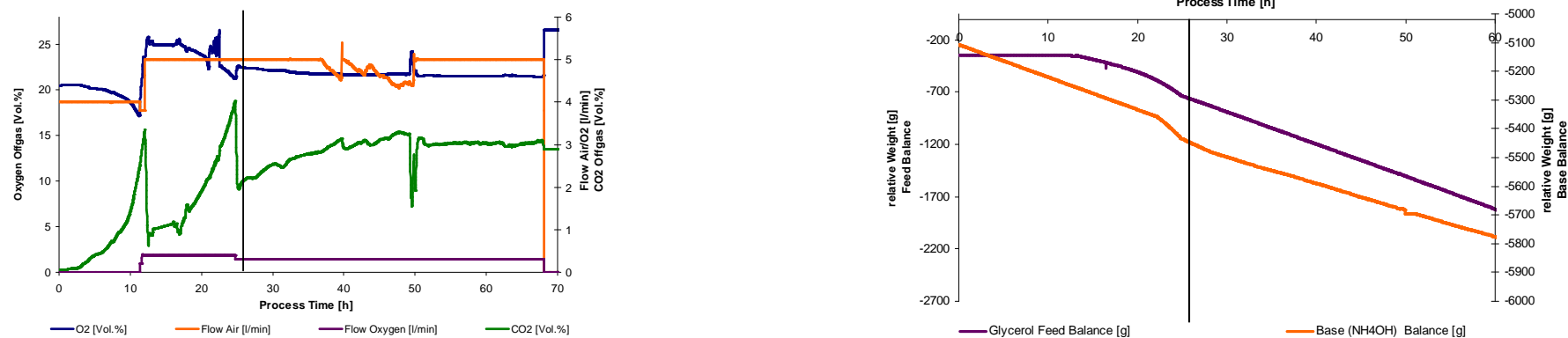

Figure 86: Fermentation at 20°C applying a linear feed (k=0). Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed Data of Fermentation 10

Fermentation parameters: induction temperature: 20°C, k=0 (linear feed).

Figure 87: Fermentation at 20°C applying a linear feed (k=0). Left: volumetric growth rate rx (blue), volumetric substrate uptake rate (rs), protein release rate r_P (purple) and volumetric cell lysis rate (red). Right: specific growth rate μ (blue), specific substrate uptake rate qs (orange), specific protein release rate $q_{Protein}$ (purple) and specific cell lysis rate r_{Lys} (red). X-axis: induction time.
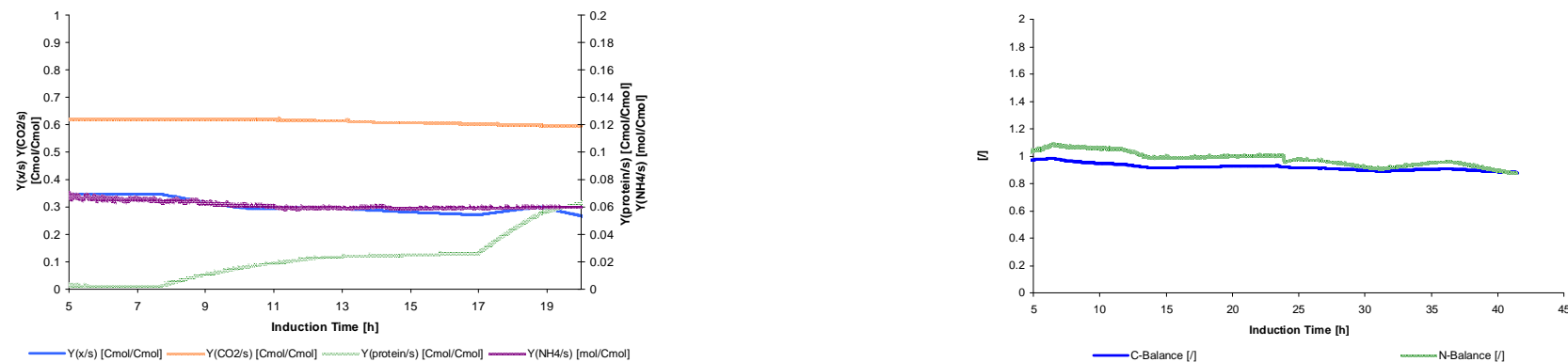

Figure 88: Fermentation at 20°C applying a linear feed (k=0). Left: biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen (green) balance.

Acquired data of fermentation 11

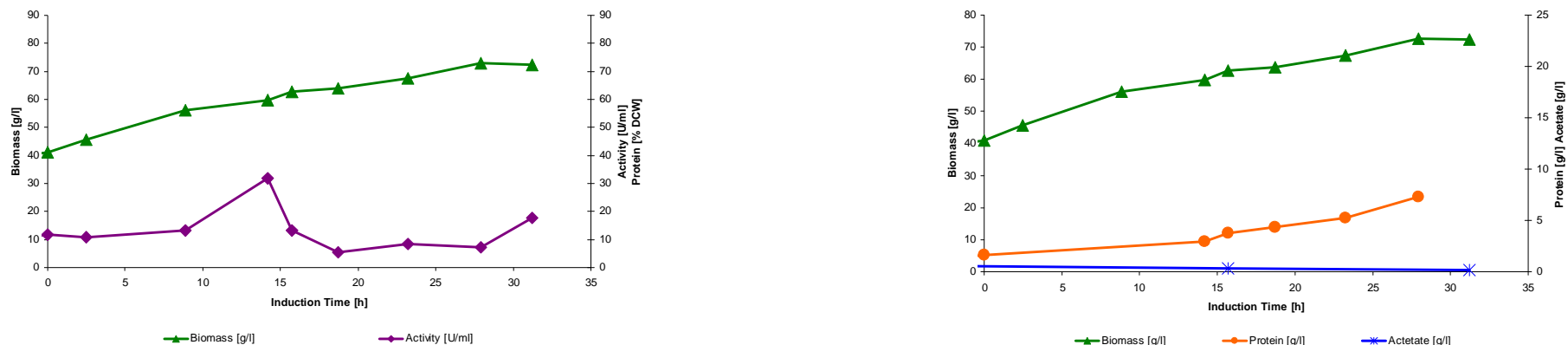Fermentation parameters: induction temperature: 20°C, k=+0.007 (positive exponential feed).

Figure 89: Fermentation at 20°C applying a positive exponential feed (k=+0.007). Left: intracellular protein concentrations (orange), intracellular activity (purple), acetate concentrations (blue) and biomass concentrations (green). Right: extracellular protein- (orange) and biomass (green) concentrations. X-axis: induction time.
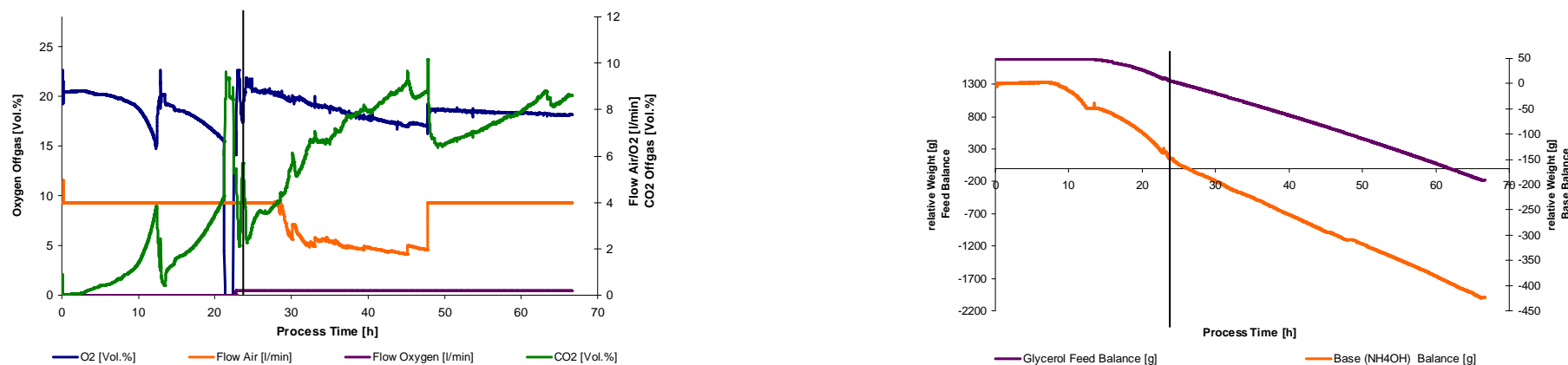

Figure 90: Fermentation at 20°C applying a positive exponential feed (k=+0.007). Left: off-gas measurements of CO_2 (green), oxygen (blue) as well as gas inlet flows of oxygen (purple) and air (orange). Right: signals recorded from the feed balance (orange) and base balance (purple). Time point of induction is indicated by a vertical bar. X-axis: process time.

Processed data of fermentation 11

Fermentation parameters: induction temperature: 20°C, k=+0.007 (linear feed).
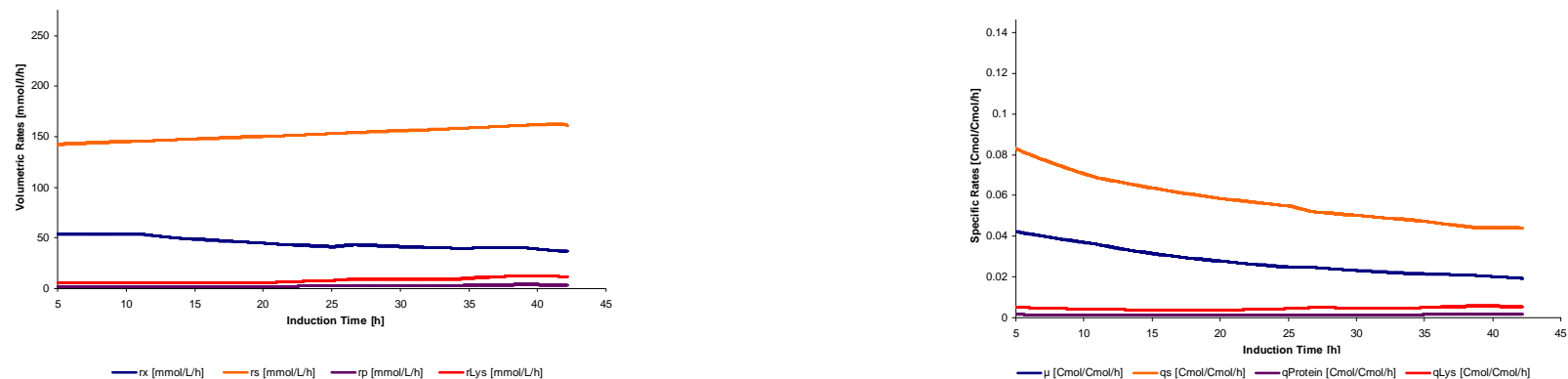
Figure 91: Fermentation at 20°C applying a positive exponential feed (k=+0.007). Left: Volumetric growth rate rx (blue), volumetric substrate uptake rate (rs), protein release rate $r_{Protein}$ (purple) and volumetric cell lysis rate (red). Right: Specific growth rate µ (blue), specific substrate uptake rate qs (orange), specific protein release rate q_P (purple) and specific cell lysis rate r_{Lys} (red). X-axis: induction time.

Figure 92: Fermentation at 20°C applying a linear feed (k=0). Left: Biomass- (blue), CO_2- (orange), extracellular protein- (green) as well as ammonia (purple) yield. Right: carbon- (blue), degree of reduction- (orange) as well as nitrogen (green) balance.

7.6 References

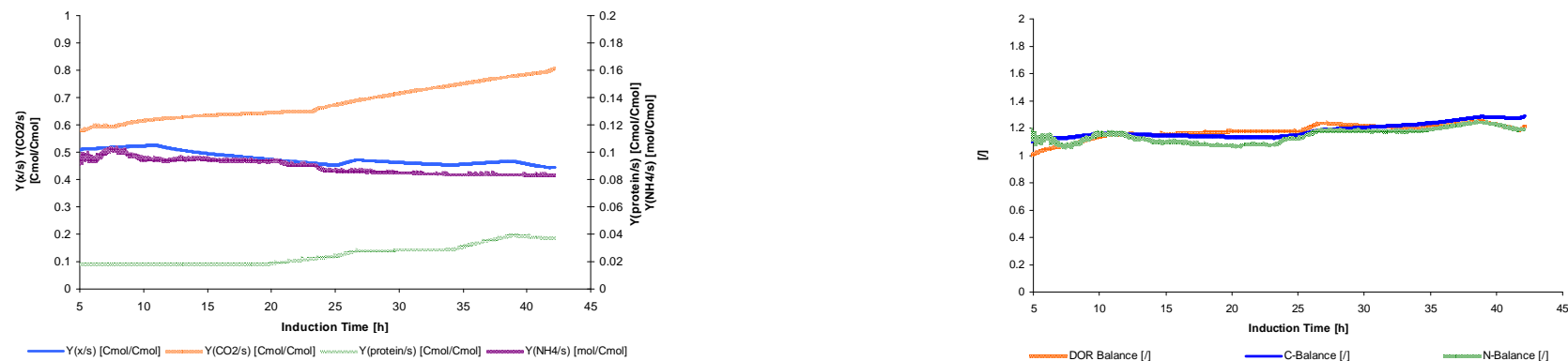Abu-Absi, S.F. et al., 2010. Defining process design space for monoclonal antibody cell culture. *Biotechnology and Bioengineering*, 106(6), pp. 894-905.

Andersson, L., 1996. Impact of plasmid presence and induction on cellular responses in fed batch cultures of Escherichia coli. *Journal of Biotechnology*, 46(3), pp. 255-263.

Andersson, L., Strandberg, L. and Enfors, S., 1996. Cell Segregation and Lysis Have Profound Effects on the Growth of Escherichia coli in High Cell Density Fed Batch Cultures. *Biotechnology Progress*, 12(2), pp. 190-195.

Atyaksheva, L., Chukhrai, E. C., Poltorak, O. M., 2007. The catalytic properties of alkaline phosphatases under various conditions. *Russian Journal of Physical Chemistry*, 82(11) pp. 1947-1951.

Baneyx, F., 1999. Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*, 10(5), pp. 411-421.

Baneyx, F. and Mujacic, M., 2004. Recombinant protein folding and misfolding in Escherichia coli. *Nature Biotechnology*, 22(11), pp. 1399-1408.

Bardwell, J.C., McGovern, K. and Beckwith, J., 1991. Identification of a protein required for disulfide bond formation in vivo. *Cell*, 67(3), pp. 581–589.

Barrett, Y.C. et al., 2007. Validation and implementation of drug-dependent antibody assays in clinical trials for safety monitoring of patients dosed with roxifiban, an orally bioavailable glycoprotein IIb/IIIa antagonist. *Journal of pharmaceutical and biomedical analysis*, 44(4), pp. 938–946.

Beck, K. et al., 2000. Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. *The EMBO Journal*, 19(1), pp. 134-143.

Benzer, S., 1961. On the topography of the genetic fine structure. *Proceedings of the National Academy of Sciences of the United States of America*, 47(3), pp. 403.

Birnbaum, S. and Bailey, J.E., 1991. Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant Escherichia coli. *Biotechnology and Bioengineering*, 37(8), pp. 736-745.

Cohen, S.N. and Chang, A.C., 1973. Recircularization and autonomous replication of a sheared R-factor DNA segment in Escherichia coli transformants. *Proceedings of the National Academy of Sciences of the United States of America*, 70(5), pp. 1293.

Coleman, J.E., 1992. Structure and mechanism of alkaline phosphatase. *Annual Review of Biophysics and Biomolecular Structure*, 21, pp. 441-483.

Duong, F. et al., 1997. Biogenesis of the gram-negative bacterial envelope. *Cell*, 91(5), pp. 567-573.

Economou, A., 1999. Following the leader: bacterial protein export through the Sec pathway. *Trends in microbiology*, 7(8), pp. 315-320.

Eriksson, L., 2000. *Design of experiments : principles and applications,* Stockholm: Umetrics AB, Umeå Learnways AB.

FDA, 2004. Pharmaceutical cGMPs for the 21st Century - A Risk-Based Approach. (http://www.fda.gov/Drugs/DevelopmentApprovalProcess/Manufacturing/QuestionsandAnswersonCurrentGoodManufacturingPracticescGMPforDrugs/ucm137175.htm)

Fekkes, P., van der Does, C. and Driessen, A.J., 1997. The molecular chaperone SecB is released from the carboxy-terminus of SecA during initiation of precursor protein translocation. *The EMBO Journal*, 16(20), pp. 6105-6113.

Flores, S. et al., 2004. Growth-rate recovery of Escherichia coli cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway. *Biotechnology and Bioengineering*, 87(4), pp. 485-494.

Garcia, T., Cook, G. and Nosal, R., 2008a. PQLI Key Topics - Criticality, Design Space, and Control Strategy. *Journal of Pharmaceutical Innovation*, 3(2), pp. 60-68.

Garcia, T., Cook, G. and Nosal, R., 2008b. PQLI Key Topics - Criticality, Design Space, and Control Strategy. *Journal of Pharmaceutical Innovation*, 3(2), pp. 60-68.

Glick, B.R., 1995. Metabolic load and heterologous gene expression. *Biotechnology Advances*, 13(2), pp. 247-261.

Gruber, Y., 2010. Process understanding in fermentation development. *MSc Thesis*, Vienna University of Technology, Austria.

Gupta, S. et al., 2007. Recommendations for the design, optimization, and qualification of cell-based assays used for the detection of neutralizing antibody responses elicited to biological therapeutics. *Journal of immunological methods*, 321(1-2), pp. 1–18.

Harms, J. et al., 2008. Defining Process Design Space for Biotech Products: Case Study of Pichia pastoris Fermentation. *Biotechnology Progress*, 24(3), pp. 655-662.

Hong, Y., Pasternak, J. and Glick, B., 1995. Overcoming the metabolic load associated with the presence of plasmid DNA in the plant growth promoting rhizobacterium Pseudomonas putida. *Can J Microbiol*, 41, pp. 624– 628.

Hopkins, D. J., Betenbaugh, M.J. and Dhurjati, P., 1987. Effects of dissolved oxygen shock on the stability of recombinantEscherichia coli containing plasmid pKN401. *Biotechnology and Bioengineering*, 29(1), pp. 85-91.

Huang, J. et al., 2009. Quality by design case study: An integrated multivariate approach to drug product and process development. *International Journal of Pharmaceutics*, 382(1-2), pp. 23-32.

Hussain, A. S., 2001. The ACPS's Process Analytical Technology Subcommittee. *(*www.fda.gov/ohrms/dockets/ac/01/slides/3804s1_02_hussain.ppt)

ICH, 1999. ICH Harmonised Tripartite Guideline: Q6A Specifications: Test Procedures and Acceptance Criteria for new Drug Substances and new Drug Products: Chemical Substances. (http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q6A/Step4/Q6A_G uideline.pdf)

ICH, 2005. ICH Harmonised Tripartite Guideline: Q9 Quality Risk Management. (http://www.ich.org/LOB/media/MEDIA1957.pdf)

ICH, 2008a. ICH Harmonised Tripartite Guideline: Q8(R1) Pharmaceutical Development. (http://www.ich.org/LOB/media/MEDIA4986.pdf)

ICH, 2008b. ICH Harmonised Tripartite Guideline: Q10 Pharmaceutical Quality Systems. (http://www.ich.org/LOB/media/MEDIA3917.pdf)

Joly, J.C. and Wickner, W., 1993. The SecA and SecY subunits of translocase are the nearest neighbors of a translocating preprotein, shielding it from phospholipids. *The EMBO Journal*, 12(1), pp. 255-263.

Kadokura, H., Katzen, F., Beckwith, Jon.,2003 Protein disulfide bond formation in prokaryotes. *Annual Review of Biochemistry*, 72, pp. 111-135.

Kane, J., 1988. Formation of recombinant protein inclusion bodies in Escherichia coli. *Trends in Biotechnology*, 6(5), pp. 95-101.

Koh, B.T. et al., 1992. Comparison of acetate inhibition on growth of host and recombinantE. coli K12 strains. *Biotechnology Letters*, 14(12), pp. 1115-1118.

Kubitschek, H.E., 1990. Cell volume increase in Escherichia coli after shifts to richer media. *Journal of Bacteriology*, 172(1), pp. 94-101.

Kuboi, R. et al., 1995. Optimal disruption methods for the selective recovery of [beta]-galactosidase from Escherichia coli. *Journal of Fermentation and Bioengineering*, 79(4), pp. 335–341.

Kyslik, P. et al., 1993. Plasmid burden in chemostat culture ofEscherichia coli: Its effect on the selection for overproducers of host enzymes. *Biotechnology and Bioengineering*, 41(3), pp. 325-329.

Laemmli, U., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227(5259), pp. 680-685.

Lederberg, J. and Tatum, E.L., 1946. Gene Recombination in Escherichia Coli. *Nature*, 158(4016), pp.558-558.

Lee, S., 1996. High cell-density culture of Escherichia coli. *Trends in Biotechnology*, 14(3), pp. 98-105.

Luli, G.W. and Strohl, W.R., 1990. Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations. *Applied and environmental microbiology*, 56(4), pp. 1004.

MacIntyre, S., Mutschler, B. and Henning, U., 1991. Requirement of the SecB chaperone for export of a non-secretory polypeptide in Escherichia coli. *Molecular and General Genetics: MGG*, 227(2), pp. 224-228.

Mairhofer, J. et al., 2008. A novel antibiotic free plasmid selection system: Advances in safe and efficient DNA therapy. *Biotechnology Journal*, 3(1), pp. 83-89.

Matsumoto, G., Yoshihisa, T. and Ito, K., 1997. SecY and SecA interact to allow SecA insertion and protein translocation across the Escherichia coli plasma membrane. *The EMBO Journal*, 16(21), pp. 6384-6393.

McCracken, S. and Meighen, E., 1980. Functional and structural properties of immobilized subunits of Escherichia coli alkaline phosphatase. *The Journal of Biological Chemistry*, 255(6), pp. 2396-2404.

Miller, A., Wang, L. and Kendall, D.A., 1998. Synthetic signal peptides specifically recognize SecA and stimulate ATPase activity in the absence of preprotein. *The Journal of Biological Chemistry*, 273(19), pp. 11409-11412.

Mire-Sluis, A.R. et al., 2004. Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. *Journal of immunological methods*, 289(1-2), pp. 1–16.

Nosal, R. and Schultz, T., 2008. PQLI Definition of Criticality. *Journal of Pharmaceutical Innovation*, 3(2), pp. 69-78.

Orhanović, S. and Pavela-Vrančič, M., 2003. Dimer asymmetry and the catalytic cycle of alkaline phosphatase from Escherichia coli. *European Journal of Biochemistry*, 270(21), pp. 4356–4364.

Paetzel, M., Dalbey, R.E. and Strynadka, N.C., 1998. Crystal structure of a bacterial signal peptidase in complex with a beta-lactam inhibitor. *Nature*, 396(6707), pp. 186-190.

Rathore, A.S. et al., 2008. Case study and application of process analytical technology (PAT) towards bioprocessing: Use of on-line high-performance liquid chromatography (HPLC) for making real-time pooling decisions for process chromatography. *Biotechnology and bioengineering*, 100(2), pp. 306–316.

Rathore, A.S., 2009. Roadmap for implementation of quality by design (QbD) for biotechnology products. *Trends in Biotechnology*, 27(9), pp. 546-553.

Rietsch, A. and Beckwith, J., 1998. The genetics of disulfide bond metabolism. *Annual Review of Genetics*, 32, pp. 163-184.

Rinas, U., Kracke-Helm, H. and Schügerl, K., 1989. Glucose as a substrate in recombinant strain fermentation technology. *Applied Microbiology and Biotechnology*, 31(2), pp. 163-167.

Ruiz, N., Kahne, D. and Silhavy, T.J., 2006. Advances in understanding bacterial outer-membrane biogenesis. *Nature Reviews Microbiology*, 4(1), pp. 57–66.

Sanden, A.M. et al., 2003. Limiting factors inEscherichia coli fed-batch production of recombinant proteins. *Biotechnology and Bioengineering*, 81(2), pp. 158-166.

Seo, J. and Bailey, J.E., 1985. Effects of recombinant plasmid content on growth properties and cloned gene product formation inEscherichia coli. *Biotechnology and Bioengineering*, 27(12), pp. 1668-1674.

Skare, J.T., Roof, S.K. and Postle, K., 1989. A mutation in the amino terminus of a hybrid TrpC-TonB protein relieves overproduction lethality and results in cytoplasmic accumulation. *Journal of bacteriology*, 171(8), pp. 4442.

Sommer, B., 2008. Neue Strategien zur extrazellulären Produktion rekombinanter Proteine mit Escherichia coli. *Ph.D Thesis,* University of Bielefeld, Germany.

Striedner, G., 2001. Metabolic approaches for the optimisation of recombinant fermentation processes. *Selected Articles from the Meeting of the EFB Section on Microbial Physiology*, Semmering, Austria.

Summers, D., 1991. The kinetics of plasmid loss. *Trends in Biotechnology*, 9(1), pp. 273-278.

Thomas, R., Kirsch, J. F., 1980. Kinetics and mechanism of inhibition of Escherichia coli alkaline phosphatase by permanganate ion. *Biochemistry,* 19(23), pp. 5328-5334.

Tong, L., 2000. Extracellular Expression, Purification, and Characterization of a Winter Flounder Antifreeze Polypeptide from Escherichia coli. *Protein Expression and Purification*, 18(2), pp. 175-181.

Torriani, A., 1968. Alkaline phosphatase subunits and their dimerization in vivo. *Journal of Bacteriology*, 96(4), pp. 1200-7.

Voellmy, Richard; Goldberg, Alfred L., 1980. Guanosine-5'-diphosphate-3'-diphosphate (ppGpp) and the regulation of protein breakdown in Escherichia coli. *Journal of Biological Chemistry,* 255(3), pp. 1008-14).

Wang, J., Stieglitz, K.A., Kantrowitz, E.R., 2005. Metal Specificity Is Correlated with Two Crucial Active Site Residues in Escherichia coli Alkaline Phosphatase. *Biochemistry*, 44(23), pp. 8378-8386.

Wilms, B. et al., 2001. High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the Escherichia coli rhaBAD promoter. *Biotechnology and Bioengineering*, 73(2), pp. 95-103.

Woelbeling, C., 2008. Creating Quality by Design/Process Analytical Technology (PAT/QbD) Management Awareness. *Pharmaceutical Engineering*, 28(3), pp. 1-9.

van der Wolk, J.P. et al., 1998. PrlA4 prevents the rejection of signal sequence defective preproteins by stabilizing the SecA-SecY interaction during the initiation of translocation. *The EMBO Journal*, 17(13), pp. 3631-3639.

Wood, D., 1988. Harvesting of E. coli cells using cross-flow membrane filtration. *MSc Thesis*, University of Waterloo, Ontario.

Zalatan, J.G. et al., 2007. Kinetic isotope effects for alkaline phosphatase reactions: Implications for the role of active-site metal ions in catalysis. *Journal of the American Chemical Society* 129(31), pp. 9789–9798.

Zappa, S., Boudrant, J., Kantrowitz, E.R., 2004. Pyrococcus abyssi alkaline phosphatase: the dimer is the active form. *Journal of inorganic biochemistry*, 98(4), pp. 575–581.