University of Natural Resources and Life Sciences, Vienna

Vienna University of Technology

Institute of Chemical Engineering

Research Division Biochemical Engineering

CD Laboratory for mechanistic and physiological methods for improved bioprocesses

Master Thesis

Dynamic physiological experimentation

for the assessment of physiological maxima and influencing inclusion body properties & the specific product titer

Submitted by: **Markus Brillmann Bsc.**

Vienna 2015

Supervisors: **Univ. Prof. Dipl.-Ing. Dr.techn. Christoph Herwig**

Mag.rer.nat. Wieland Norbert Reichelt

Time period: September 2014 – February 2015

Degree program: Master program Biotechnology (066 418)

Institution: University of Natural Resources and Life Sciences, Vienna

Acknowledgements

I hereby express my honest gratitude to my supervisors Christoph Herwig and Wieland Reichelt for giving me the opportunity to work in such a motivating environment, as well as for their widespread guidance and the challenges they gave me.

Special thanks go to my seatmate and lab colleague Peter Thurrold who always supported me with a helping hand and words of advice. It is furthermore important to acknowledge the backup of many team members of the working group Herwig, in particular Julian Kager, Andreas Kaineder, Sophia Ulonska and Nicole Mahler.

On a final note I thank my partner in life Melanie Vrchoticky, my brother René Brillmann and my parents Wolfgang Brillmann & Angelika Horwitz who supported me, each in his/her unique way.

Abstract

The established approach of separating up- and downstream development has evoked a bottleneck of downstream efficiency in pharmaceutical bioprocesses. Hence this bottleneck a real necessity for integrated bioprocess development regarding overall efficiency increase of protein production processes has yet to come in pharmaceutical industry.

In this context this work tried to bridge the gap between USP and DSP development by trying to affect inclusion body properties in an *E. coli* process, namely solubility, purity as well as the specific product titer through physiological USP parameters. For approaching that problem controlled dynamic experimentation is used in terms of controlled oscillations in the specific substrate uptake rate q_s. Control of q_s was carried out using a first principle softsensor based on elemental balances and off-gas analysis. The oscillations were described by their q_s mean, q_s amplitude and frequency, which were used as factors in a 2 level full factorial design of experiment (DoE). As responses of this DoE the specific product titer, together with the solubility kinetics, the solubilisation yield and the IB purity were chosen.

With the controlled q_s oscillations a significant correlation between the decline in maximal physiological capabilities (q_s crit) and the q_s mean was shown. The knowledge about this correlation could potentially be used in novel control strategies limiting the danger of substrate accumulation. Regarding IB properties a correlation of the solubilisation kinetics and the amplitude of q_s oscillations was shown within this thesis. Furthermore the q_s mean of the oscillations negatively affected the specific product titer.

Table of Content

1. Introduction

1.1. Motivation

In the major part of bioprocesses downstream-processing (DSP) embodies a bottleneck for productivity which affects costs and time. This is a consequence of the faster technological advances in upstream-processing (USP) compared to DSP (Gottschalk 2008). Hereby directly influencing product properties in USP might potentially yield a significant efficiency increase in DSP. For this debottlenecking of DSP through USP operating parameters or process state variables the identification of controllable quality attributes is critical (Lionberger et al. 2008).

Opposing to such an integrated approach for bioprocess development outlined above, efficiency increase in up- and downstream processing for industrial processes is nowadays performed separately as it's far simpler (Gottschalk 2008). The demand for a holistic approach in bioprocess development is likely to be intensified by the market pressure applied by the manufacturing of biosimilars (Calo-fern & Mart 2012).

A sound science based approach to bioprocess development is also motivated in the ICH Q8(R2) (ICH 2009) guideline which was elaborated as part of the Quality by Design initiative. The aim of bioprocess development based the QbD approach is to design bioprocesses ensuring a high quality product (FDA 2004). As quality cannot be tested into the product it should be an inherent part of the process design (ICH 2009).

1.2. Framework of this study

1.2.1. E. coli as production host for recombinant proteins as Inclusion Bodies

In 2009 46% of the recombinant protein pharmaceutical products on the market were produced in microbial host organisms. These 46% are almost exclusively shared between the Yeast Saccharomyces cerevisiae (15%) and Eschericha coli (31%)(Walsh 2010).

Although microbial cells as hosts for recombinant protein production lack for the molecular apparatus necessary for human-like post-translational modifications these expression systems come with some inherent advantages. Some of these advantages are: high growth rates, higher overall product yields and low media costs compared to mammalian cell culture production processes.

A widely used strain is *E. coli* BL21 (DE3) for high cell density (>100 $g*L^{-1}$ DCW) production processes (Choi et al. 2006). Furthermore the fact that *Eschericha coli* is one of the most intensively studied microorganisms eases even extensive genetic modifications of *E. coli* (Swartz 2001). The high growth rates used in *E. coli* processes are often accompanied with by-product formation – mostly acetate, which leads to suboptimal product yields (Heyland et al. 2011). Unlike K12, the tendency of acetate production of *E. coli* BL21 (DE3) is very low (Waegeman et al. 2012).

Recombinant proteins produced in *E. coli* can either be located directly in the cytosol (Sørensen & Mortensen 2005) or directed to the periplasm by addition of a signal peptide sequence (Schlegel et al. 2013). Apart from their localisation recombinant proteins are differentiated by means of their solubility. As prokaryotic organism, *E. coli* lacks the ability to correctly fold complex secondary/tertiary molecule structures of eukaryotic derived recombinant proteins. This often results in the formation of intracellular protein aggregates – so called inclusion bodies (IB) (Williams 1982).

Inclusion body formation is especially favourable in case of the expression of toxic proteins. By adding a fusion tag (called N^{pro}) to the N-Terminus these toxic proteins aggregate in the cytosol. Upon protein aggregation toxicity is reduced. Upon a pH shift the N^{pro} tag becomes native and acts as an auto-protease, cleaving itself off the target protein and leaving behind a native N-terminus (Dürauer et al. 2010). This N^{pro} technology renders the advantages of *E. coli* processes accessible for a larger variety of recombinant protein products.

IBs mostly consist of highly pure but inactive target protein. Inclusion body formation requires less purification but the necessary solubilisation and refolding significantly impacts the overall protein yield. Hereby, especially high concentrations of chaotrope reagents in the solubilisation step negatively affect the subsequent refolding yield. Since refolding is based on dilution of the chaotropic solution huge refolding tanks and high refolding times make the refolding step cost intensive and require complex engineering solutions (Pan et al. 2014).

A broad spectrum of literature covering the linkage of USP process parameters with IB properties exists (Ami et al. 2006; Margreiter, Messner, et al. 2008; Margreiter, Schwanninger, et al. 2008; Upadhyay et al. 2012). Various authours elaborated the effect of temperature and expression levels on the secondary structure elements of IBs (Margreiter, Schwanninger, et al. 2008, Ami et al. 2006), which is of little relevance for increasing DSP efficiency. Another study of Margreiter et al (Margreiter, Messner, et al. 2008) measured IB size and density with sedimentation field flow fractionation and electron microscopy. Upadhyay et al (Upadhyay et al. 2012) investigated the susceptibility of IBs to denaturation agents and proteolysis in relation to their aggregation behaviour. Luo et al (Luo et al. 2006) showed that IB size varied with the cultivation temperature and the induction time, which seems reasonable as IB size ought to correlate with product titer. Summarizing most publications on IBs properties in general focus on method establishment and the study of IB as such than on directly trying to increase DSP efficiency through affecting IB attributes through USP parameters.

1.2.1. Physiological process control – a key to influence IB properties?

As a biological phenomenon IB formation should underlie the stoichiometric restrictions imposed by the cellular framework. From a theoretical point of view IB properties should be controllable by physiological process control strategies. In respect of physiological process control Wechselberger et al (Wechselberger et al. 2012) demonstrated the initial specific substrate uptake rate at time point of induction q_s init being the process parameter of highest importance for feeding profile optimizations.

$$q_S = \frac{g_{Substrate}}{g_{BM} * h}$$

Equation 1

The specific substrate uptake rate q_s (Equation 1) is defined as the amount of substrate which is fed per biomass and hour. Approaches for obligatory real time biomass estimation are discriminated into hardtype sensors or model based sensors. Hard type sensors measure physical properties e.g. absorbance, fluorescence, permittivity and make use of correlations of these physical properties with the biomass concentration (Kiviharju et al. 2008). Model based sensors are subdivided into first principle softsensors, which use elemental balances to estimate biomass from process data like off-gas measurements (Wechselberger & Sagmeister 2013), and data driven softsensors, which use historical training data sets for calibration of their underlying models (Lu 2006).

The control scheme of q_s control strategy using a so called first principle softsensor is depicted in figure 1 (Sagmeister et al. 2013). This softsensor uses the substrate feed rate and the carbon dioxide evolution rate (CER) to estimate the biomass formation rate r_x under the assumption of a closing carbon balance. The closing carbon balance implies an absence of substrate accumulation. The elemental biomass composition, the substrate concentration in the feed as well as an initial biomass concentration are obligatory input parameters. According to biomass growth the feed supply is adapted in order to maintain the q_s of interest.

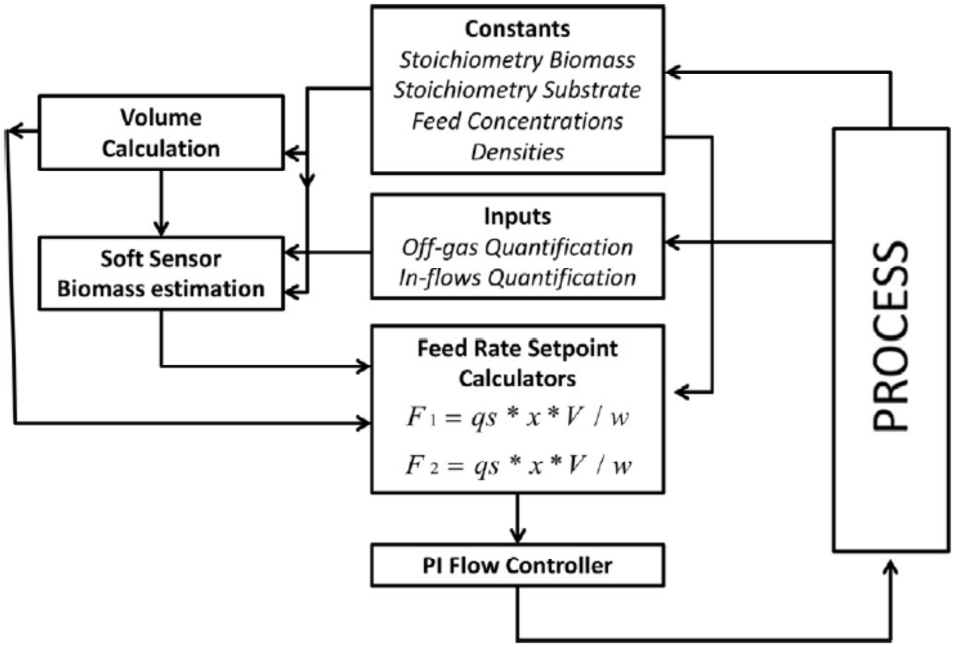


Figure 1: Flow diagram depicting the q_s control based on first principle softsensor; Constants (Biomass elemental composition, Substrate elemental composition, Feed concentration, Densities), and online process signals (off-gas measurements and substrate inflow) are used as inputs for total biomass estimation; From this feed-rate set-points to maintain a certain q_s are calculated (Sagmeister et al. 2013).

1.2.2. Dynamic experimentation – imposing physiological changes to investigate physiological differences

Dynamic experiments are typically performed to estimate strain specific descriptors like the maximum specific growth rate, or maxima of cellular capacities like substrate- or oxygen uptake rates which are essential descriptors needed in bioprocess development (Dietzsch et al. 2011)(Lin et al. 2001). Another common motivation for dynamic experiments found in literature is to investigate inhomogeneities encountered upon reactor scale-up (Sunya et al. 2013). The main distinguishing parameter is the frequency of the applied dynamics.

Three general types of common dynamic experimental setups can be distinguished. Ramp experiments, pulse experiments and shift experiments. In ramp experiments like accelero- and decelerostats usually the specific growth rate µ follows a pre-defined ramp (Paalme et al. 1995). In pulse experiments shots of C-source are imposed upon an otherwise substrate limited culture that is run in fed-batch mode or as a chemostat (Sunya et al. 2013). For shift experiments sudden changes in e.g. temperature are performed to study the response of a culture (Soini et al. 2005).

Recent studies of Sunya et al investigate the stress responses of *E. coli* to glucose pulses (Sunya et al. 2012; Sunya et al. 2013). These studies showed that the respiratory culture response differed in dependency of number pulses. They further illustrated a correlation of stress duration and pulse intensity, in terms of general stress related reporter gene activation. However the first response on a transcriptional level was triggered 2 minutes

after the perturbation, independent on pulse intensity. The same time ranges for the first response on a genetic level are also reported by Neubauer et al (Neubauer et al. 1995). In terms of full adaptation to new conditions Jozefczuk et al (Jozefczuk et al. 2010) investigated different stressors on a metabolomic/transcriptomic level. This study concluded that full metabolic adaptation require a time window of 10-40 minutes depending on the stress-type.

Caspeta et al (Caspeta et al. 2013) used a novel approach to induce process dynamics. They used a thermos-inducible promoter system for recombinant protein production and performed temperature oscillations (30 min - 2 h per temperature level) in combination with different post induction feeding profiles. Performing these temperature oscillations they were able to show improved productivities (Caspeta et al. 2013). Interestingly comparing their oscillation frequencies with the adaption times of 10-40 minutes found by Jozefczuk et al Caspeta chose time spans that allowed for cellular adaptation to the distinct temperature levels.

Recapitulating, current literature on process dynamics is lacking an investigation of the timely decline of physiological maxima, although the methods for investigation of such are well established. Furthermore the use of controlled dynamics as an alternative to i.e. constant feeding strategies have thus far not been studied in respect of their effect on IB quality attributes.

1.3. Goals

The goal of this master thesis is to quantify the decline of physiological maxima by using oscillatory post induction q_s profiles. This work is further striving to investigate the effects of these q_s oscillations on inclusion body properties and the specific product titer.

1.4. Hypotheses

- physiological maxima are accessible and quantifiable by oscillating the culture discreetly between defined levels of qs during post induction
- oscillatory post induction q_s profiles influence the specific product titer
- oscillatory post induction q_s profiles influence IB Solubility & IB Purity

1.5. Roadmap / Approach

The experiments performed to test the hypotheses of this work are conducted as a DoE (Design of Experiment). In this section the rationales for the experimental design are presented.

The following descriptors are used to describe distinct q_s oscillation trajectories:

- q_s mean:

  the arithmetic mean of the q_s trajectory; To effectively oscillate the q_s above and below the physiological maxima q_s mean values ranging from 0.234 – 0.4 g/g/h were used. The high q_s mean boundary was set in accordance with historical pulse experiments using the same strain (Keil 2014).

- q_s amplitude:

  The amplitude of the q_s oscillations, defined as positive/negative deviation from the q_s mean compared to the different q_s levels are based on the error of the softsensor based q_s control (10%). Based on the q_s mean of the center points of 0.317 this error of 10% was triplicated to give the low q_s amplitude level of 0.1 g/g/h and taken sixfold to give the high q_s amplitude level of 0.2 g/g/h.

- Frequency:

The oscillation frequency ranging from 0.25 h-1 – 1 h-1allow for cellular adaptations to the distinct qs levels is based on the findings of Caspeta et al (Caspeta et al. 2013), ranging from 0.25 h-1 – 1 h-1, allowing for cellular adaptations to the distinct qs levels.

Figure 2 shows the DoE resulting from the above described factors.

To assess the IB quality the following descriptors are chosen as responses of the DoE, apart from the specific product titer:

- Solubility kinetic constant and solubilisation yield:

As solubilisation is a critical process step in the DSP of inclusion bodies, the kinetics of IB solubilisation are assessed with a kinetic assay monitoring the solubilisation (Walther et al. 2014). Furthermore the yield of this solubilisation will be a descriptor for the efficiency of the solubilisation.

- Purity:

Inclusion body purity as the ratio of target protein to host cell proteins makes up for a meaningful parameter for increasing DSP efficiency based on the fact that less host cell protein (HCP) in the IBs potentially reduces the effort needed for product purification.

An overview of the resulting experiments for the DoE can be found in the materials and methods section.

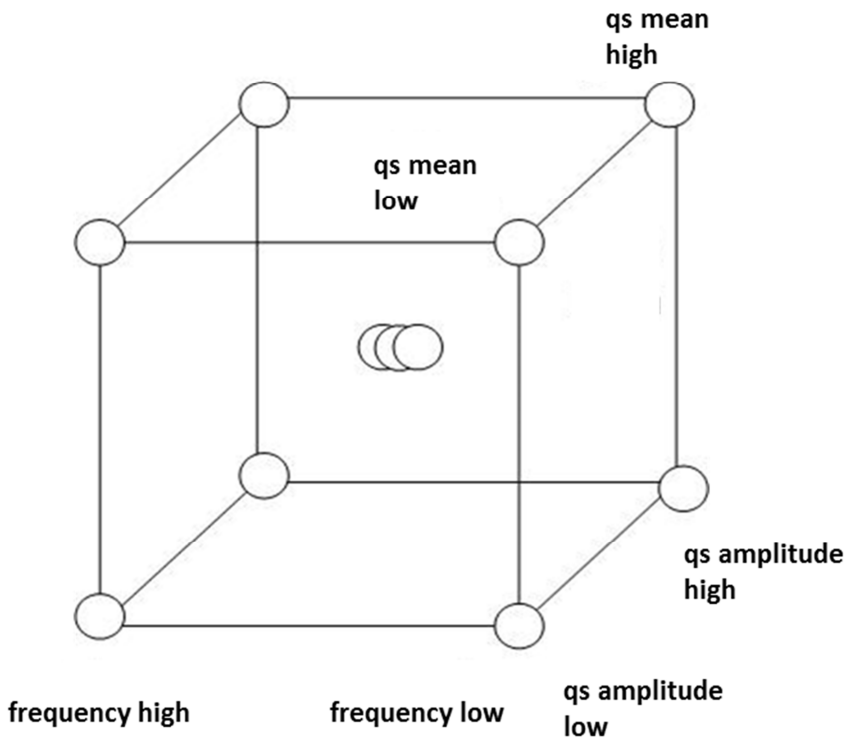


Figure 2 Design of experiment. The resulting DoE to investigate the effects of oscillatory q_s, trajectories on IB quality attributes and the specific product titer. Factor scaling: q_s mean: 0.234 – 0.4 g/g/h, q_s amplitude: 0.1 – 0.2 g/g/h, frequency: 0.25 – 1 h^{-1}.

1.6. Novelties of this work

- The novelty of this work lies in the use of a physiological process control to facilitate controlled oscillations of the post induction specific substrate uptake rate with the goal of assessing physiological maximas.
- Investigation of the impact of USP PP associated with the q_s oscillations on IB quality and product titer which represent DSP response variables, constitutes the second novelty of this work.

2. Materials & Methods

2.1. Design of Experiment

The set-points of the experiments resulting from the rationales for DoE design outlined in the introduction (1.2.2) are shown in table 1. pH and temperature set-points were selected in accordance to point specifications of the former DoE project (pH 6.9, 29°C). The responses of the DoE are the specific product titer, the kinetic constant of IB solubilisation, IB solubilisation yield and the IB purity.

Table 1 Set-points of the DoE factors (q_s mean, q_s amplitude, frequency of the oscillations) as well as the set-points for pH and Temperature, for the fermentation experiments.

Experiment No.	qs mean [g/g/h]	qs Amplitude [+/- qs]	freq. [1/h]	pH	Temp [°C]
1	0.234	0.1	1	6.9	29
2	0.234	0.1	0.25	6.9	29
3	0.234	0.2	1	6.9	29
4	0.234	0.2	0.25	6.9	29
5	0.4	0.1	1	6.9	29
6	0.4	0.1	0.25	6.9	29
7	0.4	0.2	1	6.9	29
8	0.4	0.2	0.25	6.9	29
9	0.317	0.15	0.625	6.9	29
10	0.317	0.15	0.625	6.9	29
11	0.317	0.15	0.625	6.9	29

2.2. Fermentations

2.2.1. Strain

A modified *E. coli* BL21 DE3 (provided by an industrial partner) was used for the experiments. The strain features an IPTG inducible promoter) for production of a Npro-fusion protein. The recombinant protein is expressed intracellular in the form of inclusion bodies.

2.2.2. Reactor setup

A special master slave reactor system, consisting of a 10 L working volume master reactor (Sartorius BIOSTAT® Cplus, Sartorius, Germany), equipped with a 3-way port for base and feed addition, pH probe (Mettler Toledo USA), Pt_{100} temperature an optical dO_2 sensor (Mettler Toledo USA), a double jacket for temperature regulation and an off-gas cooler. Mixing was done with a 3 rushton turbine stirrer and 4 additional baffels inside of the reactor. Process control was done using Lucullus PIMS (Seucurecell Switzerland). The 4 (2L working volume) slave reactors (DASGIP® Parallel Bioreactor System, Eppendorf, Germany) each featured Pt_{100} temperature sensors, pH probes (Mettler Toledo USA), optical dO_2 sensors (Mettler Toledo USA), 3 rushton-turbine stirrers and 4 baffles for mixing, heating pads and cooling fingers for temperature control. Further they featured a 3-way port for base/feed addition and sampling. Reactor contents were monitored using 4 scales (Sartorius, Germany). Feed/base addition as well as off-gas analysis and in-gas mixing were done using the respective modules of the DASGIP® system (off-gas: GA4, temperature & stirrer: TC4SC4, gas-mixer: MX4/4, pH & pO_2: PH4PO4, pumps: MP8). Process control was performed with DASGIP Control software (Eppendorf, Germany).

2.2.3. Preculture

The preculture was obtained from shake flask cultures (100 ml, 3 1L Erlenmeyer flasks in total) inoculated with frozen cryo-stocks (-80 °C) using a chemically defined media (recipe not shown due to confidentiality constrains). After autoclavation of the prefilled flasks ampicillin (selective pressure), sterilised glucose glucose solution and trace element solution were added to the media before inoculation. Incubation was done in a shaker at 230 rpm and 30 °C. After reaching an OD600 of 1.5-3 (approx. 17h) the preculture was used for batch inoculation.

2.2.4. Batch

10L batch media (recipe not shown) was sterilized in situ. After autoclavation the pH of 6.7 was set with NH_4OH (~12,5% (w/w)) solution and sterile glucose solution was added to the media. 250 ml of the described preculture was added through a septum using sterile syringes to 10 L batch media. Temperature and pH was held constant at 30 °C and 6.7 respectively. Aeration was done using pressurised air at a flow rate of 1.4 vvm. Stirrer speed was 400 rpm. Batch end was monitored by off-gas CO_2 content analysis (duration approx. 10-12h). Final biomass concentration was approx. 2.4 g/L. Batch end was indicated by a peak in CO_2 offgas signal followed by a drop to 0% CO_2.

2.2.5. Fed-batch and induction phase

After the batch phase 1 L of fermentation broth was transferred to each of the pre-sterilized slave reactors. Glucose feed solution was added using an exponential feeding profile with a specific growth rate of 0.2 h^{-1}. pH and temperature were kept at 7 and 30 °C respectively, aeration was done using a flow rate of 1.4 vvm. Stirrer speed was kept at 1400 rpm. dO_2 was controlled to stay above 30% by mixing pure oxygen to the in-gas using a step controller.

At 30 g/L BM concentration (calculated based on a feed forward profile with a constant $Y_{x/s}$,yield of 0.4 g/g) an adaption phase of 30 minutes was started. During adaption phase pH and temperature were set according to the set-points (Table 1 Set-points of the DoE factors (q_s mean, q_s amplitude, frequency of the oscillations) as well as the set-points for pH and Temperature, for the fermentation experiments.. The feed rate was set to meet the intended q_s mean for induction phase. After adaption phase sterile IPTG solution was added (1mM final concentration) and the q_s control script was started using a first principle softsensor (as described elsewhere(Wechselberger & Sagmeister 2013)). The code for data reconciliation of this softsensor is shown in the appendix. The oscillations were controlled using the following Visual Basic code:

osc_time … *time of constant qs in hours*
.InoculationTime… *timer started at induction*

if (.InoculationTime_H/osc_time) mod 2D > 1
qs = qs low

else if (.InoculationTime_H/osc_time) mod 2D < 1
qs = qs high
end if

Within induction phases 7-8 manual samples were taken for supernatant and product analytics.

2.3. Analytics

2.3.1. Biomass dry cell weight

Throughout the induction phase BM samples were taken every 30 min using an automated sampling device consisting of 2 pump modules (2 tubes per module), and an autosampler with a cooling block holding up to 50 vials. This autosampler consisted of 4 tubes each connected to one sampling port of one of the slave reactors, 2 tube pumps (2 tubes per pump) and a robotic arm to manoeuvre the tubes to the corresponding sampling vials which have been placed in the cooling block. The sampling procedure featured a tube flushing step of 2 minutes followed by the actual sampling into the sampling vials (30 seconds). Upon activation the robotic arm automatically counted the number of previous samples and therefore moved to the next sampling vial for the subsequent sampling cycle. The automated sampling device was controlled using Lucullus PIMS. Samples were stored at 4 °C until the end of the fermentations. Vial volume was measured gravimetrically based on a density of 1 kg/m^3, afterwards the suspension was centrifuged (5000 rpm) and the pellet was washed with deionized water, centrifuged again (5000 rpm) and finally dried at 110 °C for at least 72 h before weighing on an analytical scale.

2.3.2. Manual fermentation samples

6 ml aliquots of fermentation broth were centrifuged, 2 aliquots of 1 ml of supernatant stored in Eppendorf tubes (rest discarded) and both (Pellets and supernatant tubes) stored on -20 °C until further analysis.

2.3.3. Supernatant analysis

Glucose and acetate contents in fermentation supernatant were measured using enzymatic test kits for Cedex Bio HT Analyzer (Roche, Switzerland) for the manually taken fermentation samples.

2.3.4. Product analytics

Homogenisation and Inclusion body washing

The pellets of 6 ml fermentation samples were thawed, and re-suspended in 40 ml of ice-cold lysis buffer (100mM Tris, 10mM Na_2EDTA, pH 7,4) and homogenized using a high-pressure homogenizer (Avestin EmulsiFlex; Canada) at 1400 $\pm$100 bar in 6 passages. The pellets of 15 ml homogenized samples (15 min, 13000 rpm) were resuspended in 15 ml washing buffer A (50 mM Tris, 0,5 M NaCl, 0,02% Tween 80 (w/v) pH 8), centrifuged again

(15 min, 13000 rpm) and finally resuspended in 1.5 ml washing buffer B (50 mM Tris, 5mM EDTA, pH 8). Samples were then stored at -20 °C until further usage.

Solubilisation

300 µl of washed IB samples (see 2.3.4) were mixed with 1.2 ml solubilisation buffer (endconcentration 6 M GHCl, 50 mM Tris, 100 mM DTT pH 8). The mixture was vortexed every 10 minutes for a total of 20 minutes and kept at room temperature. Afterwards the mixture was filtered through a 0.2 µm syringe filter into HPLC glass vials and stored on -20 °C.

RP-HPLC product quantification

The product was quantified from the solubilized IB samples with HPLC (Thermo Scientific Dionex Ultimate 3000, USA) using a C8 column (NUCLEOSIL® 300-5 C8, Machery Nagel, Germany) and an acetonitrile gradient elution profile (Buffer A: MilliQ-H_2O, 0.1% TFA; Buffer B: Acetonitrile HPLC grade, 0.1% TFA), with a flow rate of 3 ml/min. Detection was at 290 nm. As calibration standard the target product was used (provided by Sandoz GmbH. Austria) (standard concentration range: 50-500 µg/ml). Quality control for HPLC runs was done using injections of one standard solution (225 µg/ml) after every fermentation sample injection (acceptance range +/- 10% peak area for these standard injections). Injection volume for samples was 10 µL, for standard solutions 2 different injection volumes were used (10 µL & 20 µL).Evaluation was done using Chromeleon 7 (Thermo Scientific Dionex, USA).

IB solubilisation kinetic assay

IB solubilisation kinetics as well as IB solubilisation yields were measured using an adapted version of a published (Walther et al. 2014) Tecan plate reader (Tecan Group AG, Switzerland) assay (Solubilisation buffer: 4 M Urea, 1 mM NaCl, 50 mM Tris, 5 mM EDTA, 0,1 M DTT pH 8; Reference buffer: 6 M Urea, 1 mM NaCl, 50 mM Tris, 5 mM EDTA, 0,1 M DTT pH8). 50 µl of washed IB samples (highest titer samples for the individual fermentations) were therefor mixed with 200 µl solubilisation/reference buffer in a 96 well plate. OD_{600} decline of the samples was monitored over 110 min. Evaluation of the assays was done using Matlab R2015a (MathWorks, USA) byfitting a logarithmic curve to the resulting OD_{600} curves (for kinetics). In order to calculate the solubilisation yield the OD_{600} difference over time was divided by the OD_{600} difference of the reference buffer.

IB purity measurements

Purity was measured electrophoretically by analysing the samples after solubilisation (samples after IB solubilisation kinetic assays 2.3.4) with Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., USA) on a p230 chip. Purity was calculated using Matlab R2015a (MathWorks, USA) by calculating the ratio of product peak area to whole electropherogram area. An example of the process is shown in this section:

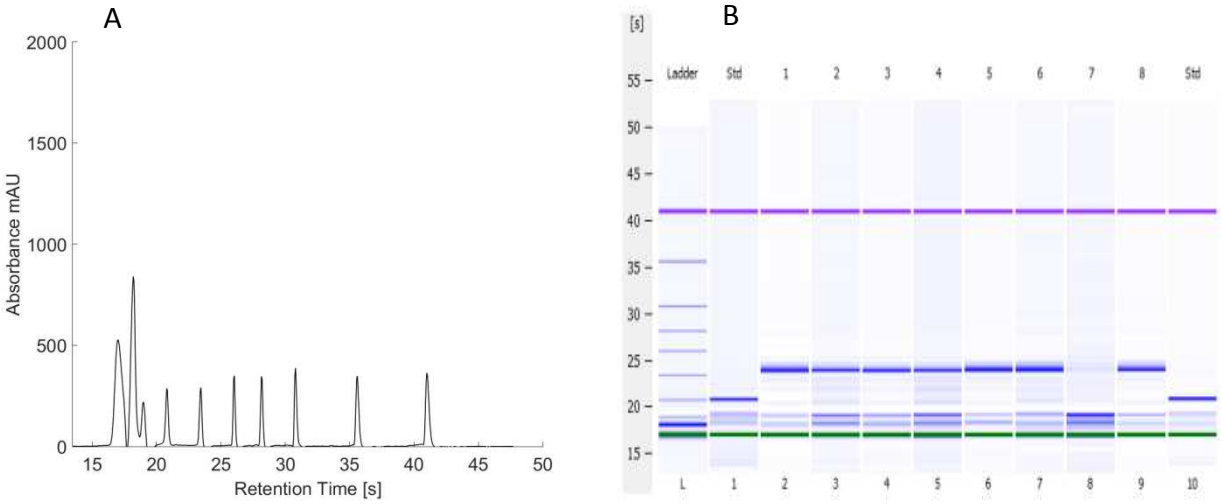


Figure 3 Protein Ladder Electropherogram and Gel of the Bioanalyzer. (A) shows the Electropherogram of the Protein Ladder (x-axis: Absorbance [mAU], y-axis: retention time). The corresponding molar masses for the protein ladder are (in ascending order): 4.5,6.0,7.0,15,28,46,63,95,150,240 kDa; (B) shows the digitally generated gel from the electropherogram data for one Chip (10 samples total).

From the electropherogram data and the known molar masses of the protein ladder (Figure 3) a quadratic regression is calculated for subsequent size identification of the sample proteins. The sizes for the target protein are ~16.7 kDa (for the standard) and 31.7 kDa (for the product in the solubilized IB samples). The size difference between standard and product can be attributed to the missing fusion protein (~18.4 kDa) in case of the standard.

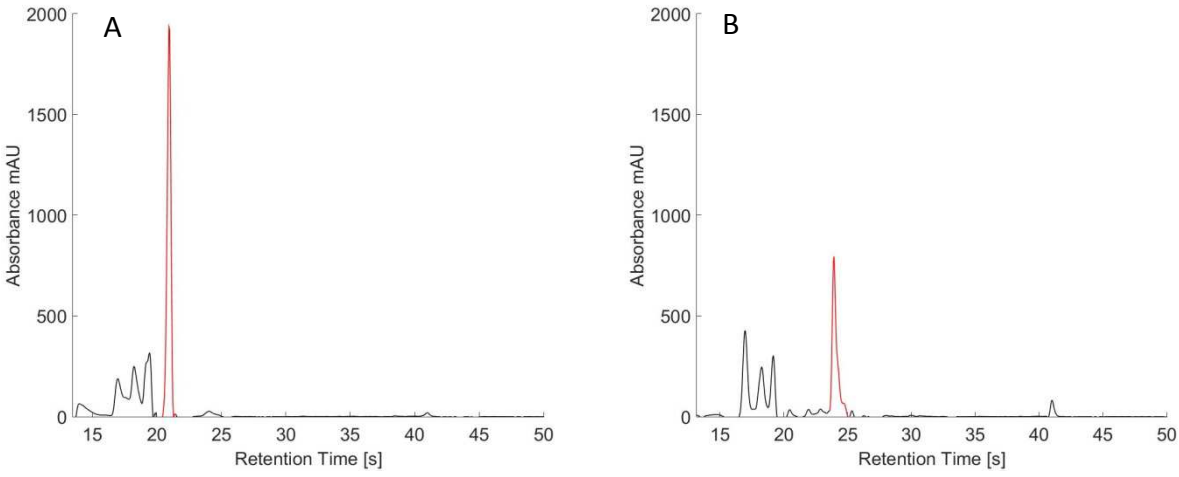


Figure 4 Electropherogram-example for the samples. (A) shows an electropherogram for a standard; (B) shows an electropherogram for the solubilized IBs from a fermentation sample. The red peaks are the standard protein and the target protein respectively. The high and low limits for integration of the product peaks were 13/18 kDa (standard) and 29/37 kDa (sample).

The purity is calculated integrating the target peaks (red peaks in Figure 4) and divide the peak area by the total area beneath the electropherogram. For the total area calculation only peaks within regression range for the size/retention time regression are integrated. Additionally the first 3 peaks in the sample electropherograms are excluded from integration because they originate from the solubilisation buffer (data not shown).

2.4. Data evaluation

2.4.1. Fermentation data analysis

Matlab R2015a (MathWorks, USA) was used for calculation of rates and yields from online data (feed rate, gassing rate, off-gas analysis) and offline data (BM dry cell weight, Glucose and acetate measurements in SN). Reaction rates were all normalized by the reaction volume. For calculation of r_X a quadratic fit for the total biomass was used to minimize the effect of error propagation which could lead to artifacts caused by sampling interference (shown in Figure 5). The novel approach used for calculation of q_s based on off-gas measurements and offline biomass is further discussed in the results section (3.1).

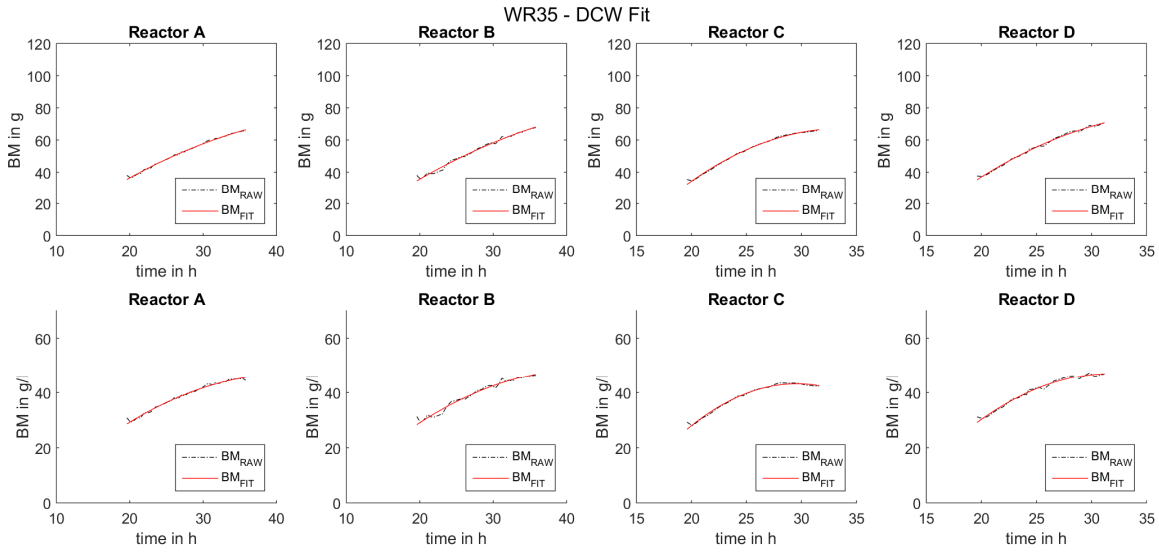


Figure 5 Biomass fitting for rate smoothing. The plots show the quadratic fits used for total biomass (top row) and biomass concentration (bottom row) used to give a smoothened r_X upon rate calculation. The red line shows the fit and the black dashed line shows the actual measured offline biomass.

2.4.2. DoE evaluation and statistics

DoE evaluation in terms of multiple linear regressions (MLR) was performed using MODDE 10 (Umetrics, Sweden). Basic statistics (i.e. t-tests) were carried out using Matlab R2015a (MathWorks, USA).

3. Results & Discussion

3.1. Fermentation data evaluation

3.1.1. Profiles and calculation of qs

Calculation approach for qs

For q_s and metabolized substrate calculation a novel approach was used. The textbook approach for calculation of q_s (Equation 14) takes into account the feed flow rate $\dot{F}_{feed}$(L/h), the substrate concentration in the feed solution c_s(g/L), and the biomass X (g).

$$q_s = \frac{r_S}{X} = \frac{\dot{F}_{S,V} \times c_S}{X}$$

Equation 14

As long as physiological maximas are not exceeded and no accumulation occurs r_S correlates to the feed flow rate.

$$\dot{F}_S = r_S$$

Equation 15

But physiological maximas are variable and a function of time (Lin et al. 2001). Exceeding q_{smax} consequently constitutes an increasing risk during fermentation.

This risk prohibits a direct correlation between substrate flow $\dot{F}_S$ and the substrate conversion rate r_s (Equation 16).

$$\dot{F}_S \neq r_S$$

$$\dot{F}_S = r_S + r_{acc}$$

Equation 16

Growth of *E. coli* on glucose as carbon source follows the reaction Equation 17. It can be seen that the carbon atoms of the substrate can be used for building biomass and for energy production (CO_2). The ratios between these routes are defined by the yield coefficients.

$$CH_2O + Y_{O_2/S} \cdot O_2 \underset{\leftarrow}{\rightarrow} Y_{CO_2/S} \cdot CO_2 + Y_{X/S} \cdot CH_{1,82}O_{0,5} + Y_{H_2O/S} \cdot H_2O$$

Equation 17

Concerning the calculation of q_s the substrate conversion rate r_s is critical. All possibilities for the substrate usage and consequently all means of calculation are outlined in Equations 5-7, with Equation 18 being the carbon balance equation. Wanting to quantify the oxidative metabolism Eq. 4 can be simplified since accumulated substrate is not oxidized. Hereby follows:

$$r_S = r_{CO_2} + r_X$$

Equation 18

$$r_{CO_2} = Y_{\frac{CO_2}{S}} \times r_S$$

Equation 19

$$r_X = Y_{\frac{X}{S}} \times r_S$$

Equation 20

Combing Equations 8, 5 and 6 via the C balance leads to:

$$1 = Y_{\frac{CO_2}{S}} + Y_{\frac{X}{S}}$$

Equation 21

So even if the yields are dynamically changing r_s can be calculated from simple summing up r_{CO_2} and r_x. Consequently q_s is calculated from Equation 22.

$$q_s = \frac{r_{CO_2} + r_X}{X}$$

Equation 22

The q_s calculated with Equation 22 is independent of accumulation resulting from an exceedance of q_s crit with the q_s setpoint as it only depends on off-gas CO_2 measurement and the biomass.

As the CO_2 equilibrium between the liquid and the gas phase changes with pH the r_{CO_2} is prone to error under highly dynamic experimental conditions that lead to pH changes. To prevent such errors r_s can also be calculated from r_{O2} by taking into account the RQ (Equation 23).

$$RQ \mathrel{\widehat{=}} \frac{r_{CO_2}}{r_{O_2}} = \frac{Y_{\frac{CO_2}{S}}}{Y_{\frac{O_2}{S}}}$$

Equation 23

The $Y_{O2/S}$ is calculated from the DoR balance if (also on a C-molar basis) if the biomass yield is known (Equation 24).

$$Y_{\frac{O_2}{S}} = \frac{\gamma_S - Y_{\frac{X}{S}} \times \gamma_X}{4}$$

Equation 24

With the known $Y_{O2/S}$ the r_{CO_2} can be substituted in Equation 22. Hereby follows for the q_s:

$$q_s\ bal = \frac{r_{O_2} \times RQ + r_X}{X}$$

Equation 25

3.1.2. Fermentation data of pH/Temp/dO2 affecting q_s bal approach

Figure 12 shows the dynamic response of the process data of temperature, dO_2 and pH to the oscillations in feed rate (respectively q_s). These real time data were used for quality control and subsequent selection of experiments for further analysis. The online data (offgas signals) as well as offline data (BM and Supernatant Measurements) was used for calculation of specific rates and yields is shown and discussed in the publication part of this thesis.

The temperature profile (Figure 12B) shows temperature control limitations of the DASGIP system that were discovered during the dynamic experiments performed within this work. These cooling problems basically appeared at high cell densities (see BM DCW data in the appendix) together with high q_s plateaus. 9 hours after induction an additional cooling device was set in place. The pH fluctuations (Figure 12D) are a result of the pH control strategy (only with base).The oscillations in dO_2 are resulting from the control strategy for dO_2 control as O_2 was gradually increased in case of dO_2 limitation (< 30%).

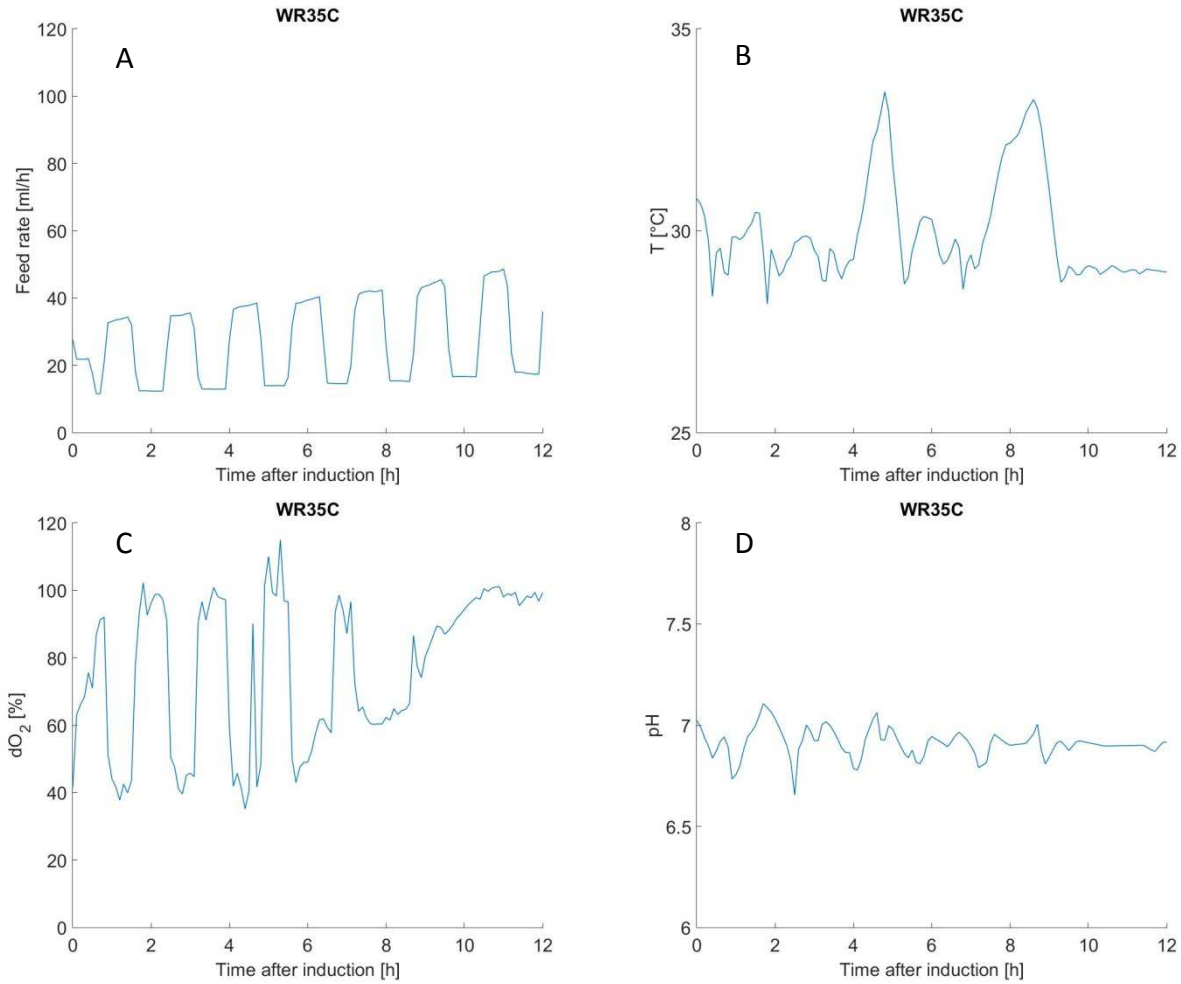


Figure 12: Online process data exemplary shown for one fermentation. Feed Rate [ml/h] (A), Temperature [°C] (B), dO_2 [%] (C) and pH (D) plotted over the time after induction. The plots show the dynamic behaviour of these process signals in response to the q_s oscillations (oscillations can be deduced from the Feed Rate data). The complete data set for all experiments passing the quality control (see Marterial and Methods) is deposited in the appendix.

q_s trajectories – q_s from feed rate vs. q_s bal

The difference between the intended q_s profiles and the actual achieved q_s profiles are illustrated in Figure 13. The cause for the accumulation of glucose and acetate, namely the decline in the critical q_s over process time and its dependency on q_s mean is discussed in the publication part of this thesis. This accumulation (Figure 13B) caused an overestimation of biomass by the softsensor, which consequently overfed the culture even further. This effect is illustrated in the q_s from feed rate in Figure 13A starting at 7 hours after induction.

Owned to the decline the actual achieved q_s amplitudes within the DoE strongly deviate from the set-points. The implications of this deviation on the design space of the DoE are further discussed in section 3.3.1.

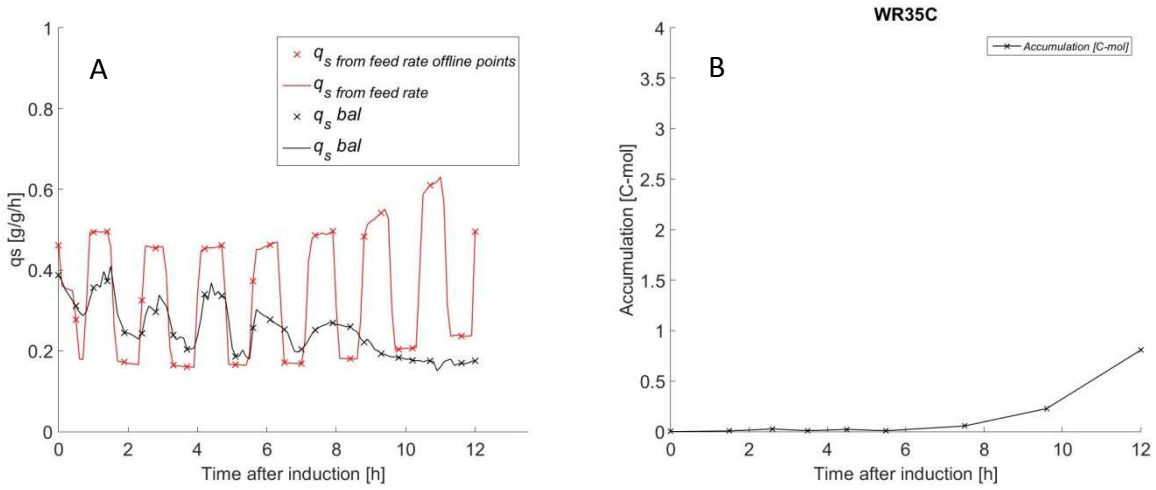


Figure 13: Exemplary q_s trajectory. (A) shows the comparison of the q_s calculated only from the Feed Rate and the offline Biomass (red line) and the q_s calculated from the offgas-Signal and offline Biomass (yellow line), marks (x) depict the sampling points for offline Biomass. (B) shows the Accumulation in C-mol as the sum of Glucose and Acetate measurements in the supernatant; the data (in A&B) is plotted over time after induction. The whole data set is shown in the appendix.

3.2. Publication part – quantification of the decline in q_s crit

In the following section the current draft of the publication covering one aspect of the results of this thesis, the decline in the critical specific substrate uptake rate q_s crit, is deposited. The direct contributions of the work accompanied with the writing of this thesis to the following publication lie in the execution and evaluation of the oscillation experiments, calculation of the q_s crit decline over time and it's correlation with q_s mean, re-evaluation of the pulse experiments to quantify q_s crit, preparation of all figures and figure captions and the preparation and description of the equations needed for the q_s bal approach.

Physiological bottlenecks – the only constant is change

Wieland N. Reichelt[1], Markus Brillmann[1], Peter Thurrold[2], Peter Keil[2], Jens Fricke[1], Christoph Herwig[1,2*]

[1] Christian Doppler Laboratory for Mechanistic and Physiological Methods for Improved Bioprocesses, Institute of Chemical Engineering, Vienna University of Technology, Getreidemarkt 9/166, A-1060 Vienna, Austria

[2] Research Division Biochemical Engineering, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorfer Strasse 1A/166-4, 1060 Vienna, Austria;

Introduction

Motivation

Escherichia coli is one of the most exploited organisms for industrial production of recombinant proteins (Terpe 2006, Walsh 2010). For a recently developed drug time to market plays a crucial role for success especially in a highly competitive environment as in pharmaceutical industry. As a consequence, besides productivity, transferability of the process knowledge from strain to strain as well as from scale of scale is of great interest. Hereby the usage of product independent platform technologies e.g. vector systems and strains have proven highly beneficial in comparison to process development from scratch. Unfortunately, despite congruent technical key-parameters different reactor scales often display different productivities, implying physiologic differences induced by the difference in scale. Concerning bioprocess development this circumstance substantiates why instead of addressing pure technical parameters a more physiological approach has emerged (Levisauskas, Simutis et al. 1996, Levisauskas 2001, Henes and Sonnleitner 2007, Gnoth, Jenzsch et al. 2008). Instead of focusing on technical parameters this physiological bioprocess development approach puts the physiology of the actual producer into focus - the cells.

Problem statement

Physiology can be described in time dependent manner using physiological variables e.g. rates and yields or process phase specific using physiological descriptors (cite). To elucidate the role of physiologic variables in a bioprocess, physiologic variables have to be controlled at discrete levels within the physiological design space. Of all physiological variables specific rates are most frequently target of control approaches, since specific assure biomass independent comparability and transferability. An increasing number of scientific contributions has been using physiological feeding profiles for the control of specific rates (Levisauskas, Simutis et al. 1996, Levisauskas 2001, Henes and Sonnleitner 2007, Gnoth, Jenzsch et al. 2008). But regardless of the specific rate of interest e.g. the specific substrate uptake rate (qs) or specific growth rate (μ), the physiological process development approach elicits two main challenges: (1) physiological control and (2) physiologic design space

(1) Physiological process control

For accurate control of specific rates during induction phase, real time biomass estimation is obligatory. Especially in industrial relevant high cell density fermentations monitoring of biomass growth is challenging. In general literature favors data driven models or hybrid models for real time biomass estimation (de Assis and Filho 2000, Jenzsch, Simutis et al. 2006). Nevertheless, in the setting of bioprocess development, historic process data is scarce, which impairs the use of data based algorithms. Consequently, hard type sensor or first principle mass balance based approaches are more feasible by means of transferability and simplicity of the control approach (Reichelt et al 2015).

(2) Physiologic design space: Maximum physiologic capacities

Experiments are commonly designed within the common space of the technical and physiological feasible space. Hereby, the technical feasible space is scale specific but indifferent towards product changeover, while the physiological feasible space is product specific but indifferent towards scales. The impact of technical process parameters (e.g. pH, temperature) on the physiological space is known on a qualitative level but the quantitative impact on the physiological space remains to be investigated. Consequently the quantitative strain characterization as a response to the technical process parameters is obligatory for each product setting up a physiology based design of experiment.

Maximum productivity is more often correlated to high specific growth rates (de Hollander 1993). High specific growth rates in turn require high substrate supply (Varma, Boesch et al. 1993) increasing the risk of exceeding the physiologic capacities. But exceeding physiologic capacities leads to accumulation of substrate or metabolites, which inhibits growth (Luli and Strohl 1990) and protein production (Jensen and Carlsen 1990). Physiologic capacities feature two definitions: qs_{max} defines the total cellular capacity to metabolize substrate. qs_{crit} is defined as the cellular capacity of metabolism without accumulation (Åkesson, Hagander, & Axelsson, 1999) therefore, a combination of anabolism and oxidative catabolism. Subsequently the physiological feasible space is defined by qs_{crit}, as within this border substrate and metabolite accumulation is avoided.

To quantify physiologic capacities various approaches have been outlined; all aiming to generate a spontaneous perturbation of C-source availability in an otherwise C-source limited process (Hunter and Kornberg 1979, Åkesson, Hagander et al. 1999, Lin, Mathiszik et al. 2001, Henes and Sonnleitner 2007). In a setting of a fixed yield exponential feed forward strategy Åkesson et al used periodical “up-pulsing” of the feed rate to trigger a transient surplus of C-source. On the basis of the response of the DO2 signal the exponential feed rate was adapted (Akesson, Karlsson et al. 1999). This approach did not yield a saturation of the glucose uptake system (Lin, Mathiszik et al. 2001) and only allows qualitative conclusions towards the qs_{crit}. This problem was tackled by Lin et al by using the DO2 response on C-source shots imposed on an otherwise volumetric constant feeding rate, leading to a metabolic saturation. But the inherent time dependent decline of μ in the context of a volumetric constant feeding rate raises questions concerning the conclusion of a correlation of μ and qs_{crit} (Lin, Mathiszik et al. 2001). Using a fixed yield exponential intermittent feed forward strategy “down-pulsing” and an DO2 response has been used for process control (Henes and Sonnleitner 2007). Hereby, the question remains unanswered whether the observed decline in μ_{crit} is actually correlated to time or more to an overestimation of produced biomass as a result of the assumption of a constant biomass yield. More recently

an approach based on a DO2 response and a dynamic feeding profile “up-pulsing” has led to the conclusion that physiological capacities decline over induction time. The outlined approach has consequently been used to timely resolve the trajectory of the maximum physiologic capacities of a respective strain. The illustrated decline in qs_{crit} was attributed to the metabolic load, imposed gradually by heterologous protein expression over induction phase (Schaepe, Kuprijanov et al. 2014). Ultimately, the question of the fundamental cause of the observed physiologic limitations is of less concern than the timely resolution of this trajectory and the interrelation with other process parameters.

Goals

Summarizing, published work on the quantification of physiologic capacities is relies widely on the highly sensitive DO2 signal as response while biomass estimation was either neglected, inaccurate or based on data driven models. Within this contribution we want to assess the workflow of investigating and quantifying the physiological feasible space. Moreover, we want to demonstrate the cross correlation of physiological variables at hand of qs_{mean} and the decline in qs_{crit}.

Materials and methods

Bioreactor system

Fed-batch experiments conducted out in a DASGIP multi-bioreactor system with a working volume of 2 l each (Eppendorf; Hamburg, Germany). The reactors are equipped with baffles and three disk impeller stirrers. The DASGIP control software v4.5 revision 230 was used for control: pH (Hamilton, Reno, USA), pO2 (Mettler Toledo; Greifensee, Switzerland; module DASGIP PH4PO4), temperature and stirrer speed (module DASGIP TC4SC4), aeration (module DASGIP MX4/4) and pH (module DASGIP MP8). CO2, O2 concentrations in the off-gas were quantified by a gas analyzer (module DASGIP GA4) using the non-dispersive infrared and zircon dioxide detection principle, respectively.

Cultivations

A recombinant BL21 DE3 E.coli strain was cultivated, producing an intracellular protein (~30 kDa) in form of inclusion bodies, after a one-time induction with IPTG (1 mM). The synthetic media was based on the recipe of Korz, Rinas et al. (Korz, Rinas et al. 1995), where the limiting C-source was glucose.

The starting feed rate (F_0) was set using a constant biomass yield ($Y_{X/S}$), the formula F_0reactor volume (V_0), the starting biomass (x_0) calculated from a feed forward profile and the concentration of the feed solution (c_{feed}).

$$F_0 = \frac{x_0 * V_0 * \mu}{c_{feed} * Y_{xs}}$$

Equation 1: Start feed rate for µ maintanance

Subsequently an exponential feeding profile given by $F_{(t)} = F_0 * e^{\mu \cdot t}$ was used for feed rate adaption ($F_{(t)}$).

$$F_{(t)} = F_0 * e^{\mu \cdot t}$$

Equation 2: Exponential feed rate adaption according to μ set point

The $qs_{(t)}$ for a given feed rate is calculated via the feed concentration and the biomass at a given time point (Equation 3). The biomass $X_{(t)}$ needed for q_s calculation was estimated using a first principle softsensor based on carbon and DoR-balance as described elsewhere (CITE). With this biomass estimation q_s was controlled to follow a predefined trajectory using a feed forward control strategy.

$$qs_{(t)} = \frac{F_{(t)} * c_{feed}}{X_{(t)}}$$

Equation 3: Specific substrate uptake rate at t

Pulse experiments

Process parameters

Precultures were incubated at 30°C and 170 rpm to an OD_{600} of approx. 1.5 in 150 mL batch media and 2.5% batch volume aliquots were used for inoculation. Both strains were cultivated at controlled pH (7), DO2 (>30 %) and temperature (A = 30°C and B = 35°C). After depletion of the C-source in an initial batch phase, the pre-induction fed-batch was started. The pre-induction feeding strategy was based on an exponential feed forward profile (Equation 2) to maintain a predefined growth rate. On attainment of the predefined biomass the cultures were induced after 30 min adaption time and the feed forward q_s control (Equation 3) was started. Stirrer speed was set to 1400 rpm and aeration to 1.4 v/v/m for the whole process. The pH was maintained by adding 12.5% NH_4OH, which also served as nitrogen source. The dissolved oxygen (DO2) was kept over 30% by supplementing oxygen to the air.

Data processing and Data Analysis

Metabolic rates and yield coefficients were calculated with matlab r2013 b (Mathworks; Natick, Massachusetts, USA). Software was used for the calculation of specific rates and yield coefficients, as we described elsewhere (Sagmeister, Wechselberger et al. 2012).

For quantitative analysis of estimation accuracy the coefficient of variation of the root mean squared error (cvRMSE) was used (Willmott 1981). Analysis of variance (Anova) was used for the assignment of significances in the differences of the average cvRMSE from experimental sets. Several biomass estimation methodologies were combined by using a weighted average including the determined error of the single methodology (Aehle, Simutis et al. 2010).

Analytics:

Biomass dry weight (CDW)

Biomass concentrations were gravimetrically quantified after drying at 105°C for min. 72 h. Therefore 2 mL of culture broth were centrifuged (4500 x g, 10 min, 4°C) in a pre-weighted glass tube and the pellet was washed once with 5 mL RO water. The determination was done

in duplicates. After drying in the drying oven the biomass dry weight was measured on a scale.

Substrate conc. and small metabolites

The C-source concentration in the feed media was calculated using the gravimetrically determined density. NH_4OH concentration was determined by titration with 1 M HCl. Acetate concentrations were quantified from the supernatant by enzymatic photometric principle in a robotic system (Cedex BioHT, Roche, Switzerland). The analysis was used as a quality control to exclude possible acetate production due to oxygen limitation or overflow metabolism.

biomass specific substrate uptake rate calculated with C-Balance and DoR-Balance: qs bal

For q_s and metabolized substrate calculation a novel approach was used. The textbook approach for calculation of q_s (Equation 4) takes into account the feed flow rate $\dot{F}_{feed}$(L/h), the substrate concentration in the feed solution c_s(g/L), and the biomass X (g).

$$q_s = \frac{r_S}{X} = \frac{\dot{F}_{S,V} \times c_S}{X}$$

Equation 4: Specific substrate uptake rate from feed flow rate

As long as physiological maxima are not exceeded and no accumulation (r_{acc}) occurs the substrate conversion rate r_S correlates to the feed flow rate.

But physiological maxima are variable and a function of time and qsmean. The prediction of $qs_{crit}(t)$ is challenging and in real time prone to errors. Exceeding qscrit consequently constitutes an increasing risk of accumulation (r_{acc}) during fermentation. This risk prohibits a direct correlation between the molar substrate flow $\dot{F}_S$ and the substrate conversion rate r_s (Equation 5).

$$\dot{F}_S = r_S + r_{acc}$$

Equation 5: Partition of the molar feed rate into metabolized substrate and accumulation

Growth of E. coli on glucose as carbon source follows the reaction specified by Equation 6.

$$CH_2O + Y_{O_2/S} \cdot O_2 \underset{\leftarrow}{\rightarrow} Y_{CO_2/S} \cdot CO_2 + Y_{X/S} \cdot CH_{1,82}O_{0,5} + Y_{H_2O/S} \cdot H_2O$$

Equation 6: E. coli growth equation for carbon

Nitrogen and other elements are negligible as further considerations only elaborate on the carbon balance. From the reaction equation it can be seen that the carbon atoms of the substrate can be used for building biomass and for energy production (CO_2). The ratios between these routes are defined by the yield coefficients.

Looking at the substrate conversion rate r_s (Equation 7) needed for q_s calculation all possibilities for the substrate usage are given by $r_S = r_{CO_2} + r_X$. oxidation to carbon dioxide (r_{CO2}) and biomass growth (r_X).Wanting to quantify the oxidative metabolism accumulation must not be considered in this equation. Hereby follows that the CO_2 ($Y_{CO2/S}$) and biomass yields ($Y_{X/S}$) sum up to 1(Equation 8).

$$r_S = r_{CO_2} + r_X$$

Equation 7: Partition of the substrate conversion rate into carbon dioxide production and biomass growth

$$1 = Y_{\frac{CO_2}{S}} + Y_{\frac{X}{S}}$$

Equation 8: Sum of the yield coefficients according to the carbon balance

So even if the Yields are dynamically changing r_s can be calculated from simple summing up r_{CO2} and r_x. Consequently $q_{sbal}q_s$ is calculated from dividing (Equation 9) by biomass (X).

??

$$q_s = \frac{r_{CO_2} + r_X}{X}$$

Equation 9: Molar specific substrate uptake rate from balancing approach

The CO_2 equilibrium between the liquid and the gas phase changes with pH the r_{CO2} is prone to error under highly dynamic experimental conditions that lead to pH changes. To prevent such errors, r_s can also be calculated from r_{O2} by taking into account the RQ. The RQ (Equation 10) as the ratio between CO_2 production and O_2 consumption and can be expressed by the respective rates (r_{CO2}, r_{O2}) as well as well as by the corresponding molar yields ($Y_{CO2/S}$, $Y_{O2/S}$).

$$RQ \widehat{=} \frac{r_{CO_2}}{r_{O_2}} = \frac{Y_{\frac{CO_2}{S}}}{Y_{\frac{O_2}{S}}}$$

Equation 10: Respiratory quotient from rates and from yields

The oxygen yield on substrate ($Y_{O2/S}$) is calculated from the DoR balance if the biomass yield ($Y_{X/S}$) is known (Equation 11).

$$Y_{\frac{O_2}{S}} = \frac{\gamma_S - Y_{\frac{X}{S}} \times \gamma_X}{4}$$

Equation 11: Correlation of the biomass yield with the oxygen yield on basis of the DoR balance

Hereby follows a calculation approach (Equation 12) for the specific substrate uptake rate (q_s) which uses on an oxygen uptake rate derived from off-gas measurements (r_{O2}) and offline biomass (X) as inputs.

$$q_s = \frac{r_{O_2} \times RQ + r_X}{X}$$

Equation 12: Molar specific substrate uptake rate from oxygen uptake, respiratory quotient and biomass

List of Symbols

Abbreviations:

HCD

Variables

CH_2O	c-molar substrate composition
$CH_{1,82}O_{0,5}$	c-molar biomass composition without nitrogen
c_{feed} …	substrate concentration in feed [g/L]
$dSn_{(t)}$ …	fed substrate normalized by the CDW at the end exp. fed-batch [g/g]
$F_{(t)}$ …	feed flow rate [L/h] after time (t)
$F_{S,V}$ …	flow rate of feed solution [L/h]
F_S …	substrate feeding rate [c-mol/h]
q_S…	biomass specific substrate uptake rate [c-mol/c-mol/h] or [g/g/h]
$q_{S(t)}$…	biomass specific substrate uptake rate [g/g/h] at time point (t)
qs_{mean} …	average qs within a predefined window of dSn or time [g/g/h]
$q_{S\ bal}$ …	biomass specific substrate uptake rate calculated with C-Balance and DoR-Balance [c-mol/c-mol/h]
q_{Scrit} …	the critical specific substrate uptake rate as defined by Åkesson, Hagander, & Axelsson, 1999 [g/g/h]
qs_{glc}	specific substrate uptake rate calculated from the glucose concentration gradients in the pulse experiments [g/g/h]
qs_{glc} fit	linear fit to determine the slope of the decline in q_s glc of the pulse experiments
RQ …	respiratory quotient [mol/mol]
r_{acc} …	rate of accumulating substrate and acetate [c-mol/h]
r_{CO2}	CER, carbon dioxide evolution rate [mol/ h]
r_s	substrate conversion rate [c-mol/h] or [g/h]
r_{O2} …	OUR, oxygen uptake rate [mol/h]
r_x	biomass conversion rate [c-mol/h]
γ_S …	Degree of Reduction (DoR) of substrate [mol/mol]
γ_X …	Degree of Reduction (DoR) of biomass [mol/mol]
μ …	specific biomass growth rate [1/h]
V_0 …	volume at t = 0 [L]
X…	CDW [g] at (0) batch end or after time (t)
x…	CDW concentration [g/L] at (0) batch end or after time (t)
$Y_{X/S}$…	biomass yield on substrate [g/g] or [c-mol/c-mol]
$Y_{CO2/S}$…	carbon dioxide yield on substrate [c-mol/c-mol]
$Y_{O2/S}$…	oxygen yield on subtrate [c-mol/c-mol]
$Y_{H2O/S}$…	water yield on subtrate [mol/c-mol]

indices:

t…	process time [h]
SP …	setpoint; the intended value of a given process parameter
PV …	process value; the actual value of a given process parameter based on measured quantities

w SN …	addition of w SN indicates values calculated with taking glucose and acetate measurements in supernatant into account
w/o SN …	addition of w/o SN indicates values calculated without taking glucose and acetate measurements in supernatant into account
bal …	denotes values calculated with the balancing approach (Equation 5 - Equation 12)

Results

Physiological maxima for physiologic DoE design

Targeting a physiological experimental design (qs/T), physiological capacities (qs_{crit}) have to be known, since exceeding the physiologic feasible space causes unwanted substrate accumulation. Investigating the physiological feasible space for the extreme points of the design FIGURE 1 displays the residual glucose concentration in the supernatant of repetitive pulse experiments. Compared at the start of induction the results indicate an qs_{crit} of 0.33 [g/g/h] for 20°C (FIGURE 1 A) while for 35°C qs_{crit} is reached at 0.55 [g/g/h] (FIGURE 1 B). The results indicate that qs_{crit} is heavily impacted by the process parameter temperature and time after induction. In accordance to literature the maximum physiological capacity of the oxidative metabolism is a function of time and dependent on the conventional process parameter: temperature.

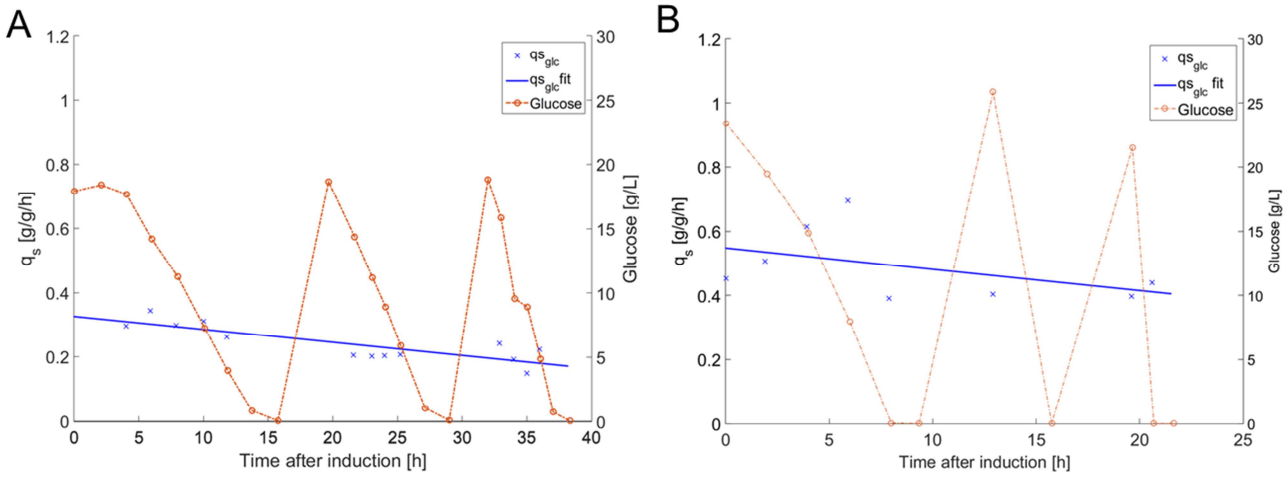


FIGURE 1: The maximum physiologic capacity for the oxidative metabolism (qscrit) declines as function of time qs=f(t); three repetitive substrate pulses (20 g/L) were administered to fully saturate the oxidative metabolism; in between pulses a time delay of 30 min was scheduled to facilitate full clearance of accumulated substrate; offline samples are indicated as (x) and were fitted by a linear function; Acetate accumulation was at all times below 0.3 g/L (data not shown); (**A**) 20°C (qs_{glc}fit = -0.004*t+0.33) indicating a maximum qscrit at start of induction of 0.33 [g/g/h]; (**B**) 35°C (qs_{glc}fit = -0.0066*t+0.55) indicating a maximum qscrit at start of induction of 0.55 [g/g/h]; Hereupon a trajectory for 29°C was deducted mathematically (qs_{glc}fit = -0.0056*t+0.46)

Sample interval independent data evaluation based on oxidative metabolism

Based on the positive correlation of qs and productivity (data not shown) the qs maximization leads to maximum productivity. In accordance to the function of qs as of time (FIGURE 1) the experiment was designed not to exceed qscrit within 10 h of induction phase (FIGURE 2). Accumulation occurred in the first phase of induction phase already 6 hours after induction (FIGURE 2 A), despite careful experimental design and tightly controlled qs. While

over 5 g/L glucose was accumulated, acetate was at all times below 0.3 g/L. This finding challenges the preciously deducted qs_{crit} trajectory of FIGURE 1 and raises the question of transferability and scope of such a trajectory.

The ill closing mass balance based on substrate rs, rx and r_{CO2} (FIGURE 2 A) indicates an accumulation of C-source elsewhere. This substrate accumulation hinders data processing, since the substrate accumulation needs to be accounted for. As a consequence the balancing approach grows increasingly dependent on the sampling interval. Calculating qs_{bal} based on our model instead of the conventional qs, takes accumulation into account and leads to a better closing mass balance (FIGURE 2 B). Using the available high resolution off gas data to calculate the rate of oxidative metabolism (r_{ox}) makes data processing less dependent on offline sampling and consequently reduces the noise inflicted by analytical errors. Subsequently, on the basis of r_{ox} , qs_{bal} as well as mass balances were calculated FIGURE 2 B..

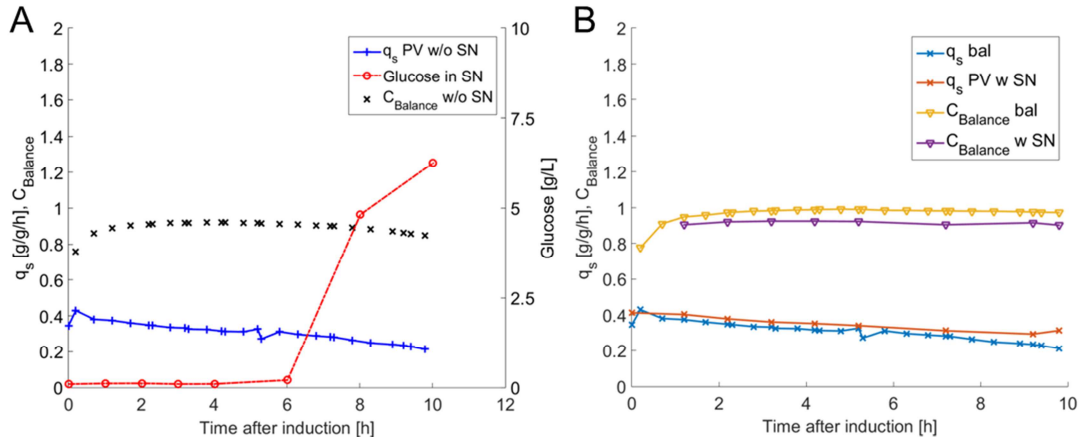


FIGURE 2: Exemplary process supposedly beneath the qscrit displays accumulation; **A**: Negative qs ramp experiment (qs_{sp} = -0.009*t+0.4) On the left x-axis qs (x) & carbon balance (+) calculated without taking supernatant measurements (SN) into account, on the right x-axis the glucose concentration (circles) is shown. Glucose is accumulating ongoing from 6 hours after induction– consequently the C-balance without SN measurements (C-Balance w/o SN) does not close to 1; Acetate accumulation throughout the process was below 0.3 g/L (data not shown); **B**: qs (solid lines) & C-balance (dashed lines) for the same process; Both (qs & C-balance) are calculated with the balancing approach (+ and circles) as well as with taking into account SN measurements (x and triangles).

Better closing C-balances substantiate qsbal approach

To quantify the benefit of the qsbal approach we analyzed the average level as well as the noise of C-balance and compared them by statistical means. The mean level of C-balance (FIGURE 3 A) as well as the average standard deviation (FIGURE 3 B) were calculated and averaged over induction phase. By calculating r_s via the off-gas data the mean C-balance value significantly increased indicating a better defined system. Additionally the standard deviation was significantly decreased (p(f)=XXX) which indicates a lower level of noise and consequently substantiates the benefit of the approach.

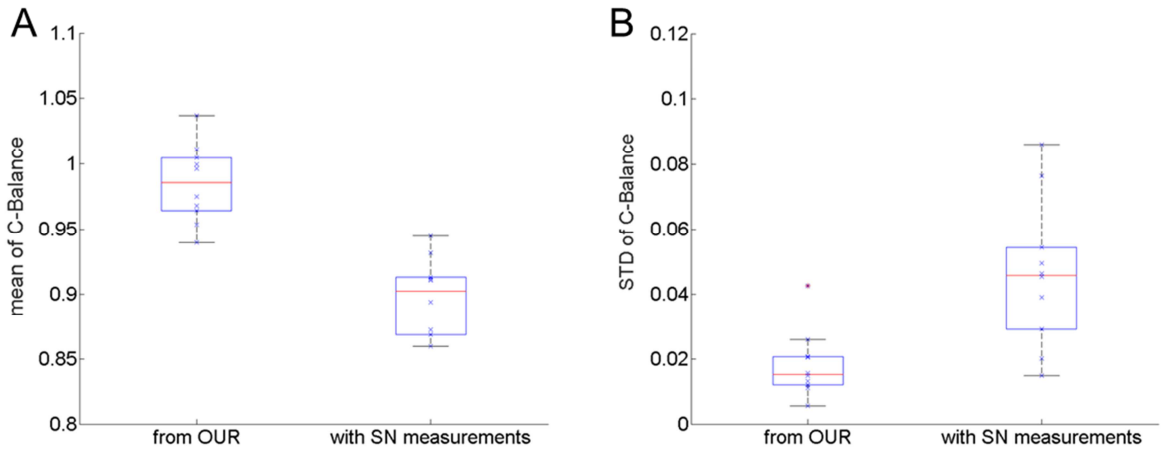


FIGURE 3 Massbalance calculation based on the model approach yields a better defined system;
A: The boxplots show the mean values of the C-balance calculated at every offline BM (n=10). (from OUR) corresponds to the C-balance based on the model approach; (with SN measurements) corresponds to the C-balance including SN measurements. The means for latter is significantly ($p= 7.9*10^{-5}$) lower than for the model based approach. This can be explained by not quantified components in the SN (refer to Supplemental 1). **B**: Comparison of the standard deviation of the two balancing approaches; the STD is significantly (p=0.0023) lower for the balancing approach (from OUR) compared to the C-balance including SN measurements. This can be attributed to analytical error of offline measurements which impact the conventional calculation to a greater extend.

Offgas data quality is critical for qsbal

The consecutive experiment was designed to challenge the robustness of the outlined approach. The experiments illustrated in FIGURE 4 underline the sensitivity of the model based approach to highly dynamic process conditions. The comparison of the steady feeding profile (FIGURE 4 A) to the oscillatory feeding profile (FIGURE 4 B) to shows a substantial difference in noise of the c-balance. This phenomenon can be attributed to controller actions to maintain DO2 within bounds FIGURE 4 A/C). The regulation of the oxygen partial pressure leads to spikes in OUR. This technical cause hinders the correct estimation of rox and consequently leads to increased noise on the C-balance.

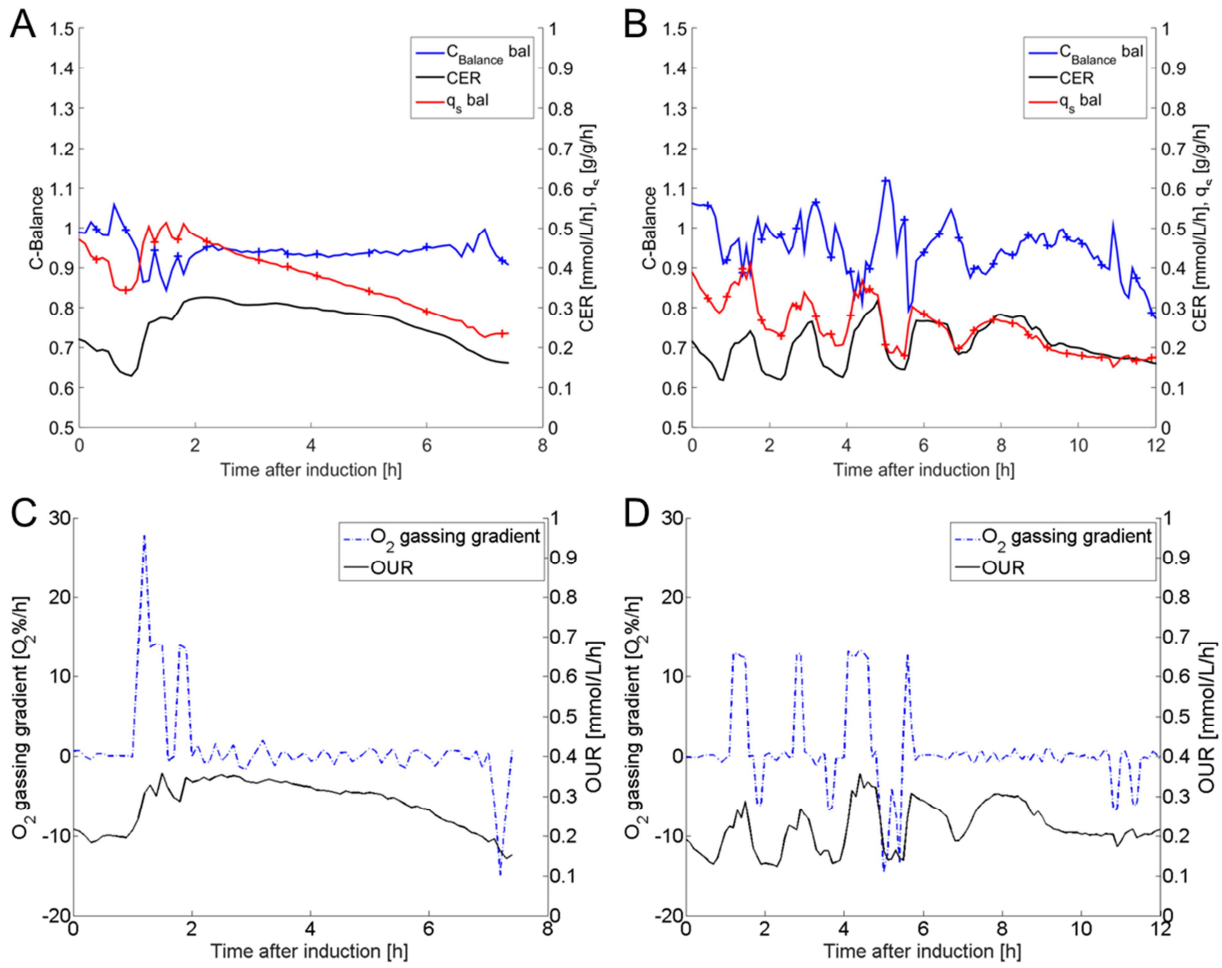


FIGURE 4 Process dynamics calling for frequent DO2 controller action cause noise in C-balance

A/B: C-balance calculated with balancing approach (left y-axis), qs and CER (right y-axis) for two different processes showing the limitations of the balancing approach, offline sampling points are indicated as (x); C/D: The OUR (right y-axis) as well as the change in partial pressure of O2 by the step controller ((left y-axis) is displayed. A/C: Dynamic q_s profile of low dynamics displays little noise on the C-balance; B/D A highly dynamic process (oscillatory q_s profile) is shown spikes in OUR directly affect the C-balance. The respective spikes coincident step controller actions for the partial pressure of n=2 in the gassing.

qsbal can predict accumulation of unknown metabolites

FIGURE 5 illustrates the benefit of estimating the oxidative metabolism in terms of sampling interval and metabolite identification. Offline metabolite quantification thereby is limited in temporal resolution by the sampling interval of supernatant. In general it is regarded as more feasible to sample biomass in high frequency than supernatants, given the biologically dynamic nature of metabolites. In comparison to the conventional approach, estimating the oxidative metabolism relieves from the dependency of supernatant sampling interval granting a higher temporal resolution (FIGURE 5 A/B. Additionally the use of the model consistently predicted a higher level of accumulation than substantiated by measuring the expected metabolites as acetate and accumulating substrate (FIGURE 5 A). Only further going analysis of the supernatant revealed the release of additional C sources as organic

acids (Supplemental 2). This difference in accumulation is the presumed reason for the difference of the average level of the C-balances. Owned to the higher level of estimated accumulation r_{ox} is smaller which in turn affects the yields of biomass and of CO2 (FIGURE 5 B). While in the offline calculation of each yield (Yxs/YCO2) the error of two offline measurements is propagated, in case of the estimation only one offline measurement is used. The consequent reduction in noise in combination with the higher level of accumulation leads to a better closing C-balance.

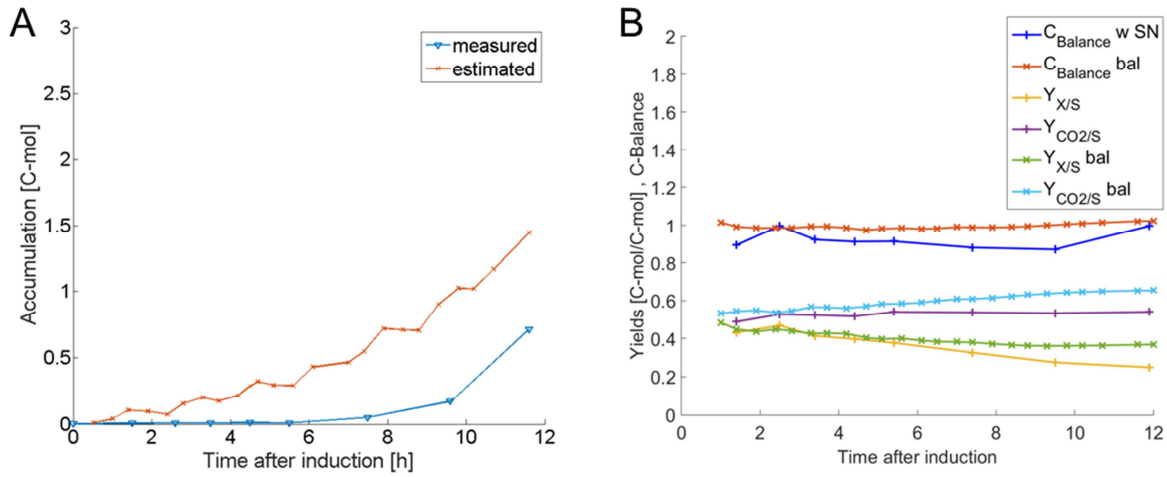


FIGURE 5: Model leads to overall improvement of data quality and indicates accumulation of a not quantified metabolite
A: Comparison of the theoretical accumulation calculated using the balancing approach (x depicts the offline BM sampling points) and actual measured accumulation (acetate and glucose) in the SN (triangles are the SN sampling points) for the two processes displayed in Fig. 4. The calculated accumulation is constantly higher than the actual measured one, this can be attributed to the non-quantified compounds in the SN (see Supplemental 2). **B**: cumulative molar yields (Yx/s; Yco2/s) as well as C-balance calculated cumulatively with both methods (balancing approach and SN measurements).

qsbal reveals dynamic decline of physiologic maxima

At hand of FIGURE 6 the dependency of qscrit on time and moreover on metabolic activity (qsmean) shall be brought to the reader's attention. Based on the outlined strategy of quantifying the oxidative metabolism, the qs sp can be put in relation to a qs pv (FIGURE 6 A/C). In any case the pv of qs is not congruent with the sp the metabolic state of the culture is not in control. If the sp is smaller than the pv accumulated C Source is being oxidized. If the sp is greater than the pv the maximum physiological capacity to metabolize substrate (qscrit) is exceeded. Under the circumstance qscrit is trespassed, substrate/metabolites are being accumulated. By fitting a linear function a time dependent decline of qscrit becomes apparent (FIGURE 6 A/C; Supplemental 3). Nevertheless, this time dependent decline does not appear to be transferable, indicating a further significant but underlying variable. FIGURE 6 C illustrates the significant (p=0.00065) correlation of the qscrit slope in dependency qsmean.

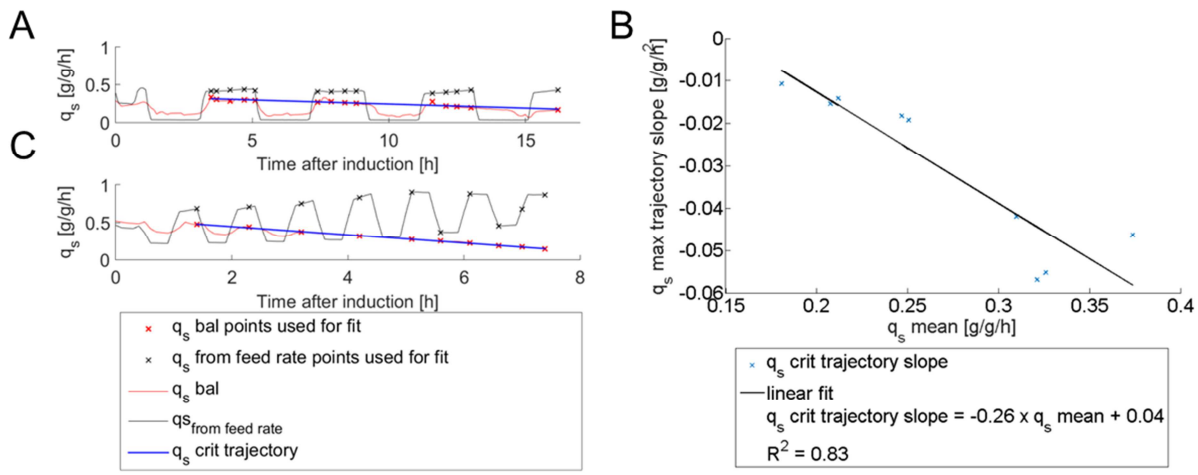


FIGURE 6: The critical specific substrate uptake rate is a function of time and metabolic load (qsmean)

A & **C**: Two different q_s profiles of oscillation processes, to determine the q_s max trajectories, are shown. Marks (x) depict the data points used for calculation of the q_s max trajectories. The used data points are all corresponding to offline samples for which the q_s calculated from the balancing approach is lower (including a safety margin) than the q_s calculated from the feeding rate. A linear curve was fitted into these data points describing the decline of q_s max – quality criteria for this fit was $R^2 > 0.8$. **B**: Dependency of the decline of the maximum q_s on the q_s mean for 9 different oscillation experiments (with different q_s means). The slope of the q_s max trajectories (y-axis) has units of an acceleration ($g/g/h^2$) which is negatively correlated with the q_s mean (g/g/h) of the culture (x-axis), using the same analogy the q_s mean stands for the speed of cellular metabolism.

Discussion

To quantify the physiological capacity of the cellular metabolism repetitive up-pulsing experiments have been utilized for the experimental part of the workflow. The pulse experiments (FIGURE 1) substantiated the impact of time of induction as well as the impact of the process parameter temperature on qscrit. Nevertheless, in an experiment to maximize qs the deducted trajectory of qscrit lead to substrate accumulation 4 h earlier than predicted. Subsequently qscrit is a function of time but the decline over time displays a dependency on the historic metabolic activity during induction phase. This impact of the physiologic descriptor of qsmean infers a descriptive function towards the memory effect. Presumably qscrit was overestimated in the up-pulsing experiments since in accordance to literature only glucose and acetate had been quantified (*Akesson, Karlsson et al. 1999*, *Schaepe, Kuprijanov et al. 2014*) while the accumulation of other organic acids as oxalic acid and ketoglutarate weren`t quantified. Potentially the continuously high metabolic activity during the ramp (FIGURE 2) experiment led to a faster decline in qscrit in comparison to the repetitive pulse experiments which comprise phases of recovery in-between pulses.

In differentiation to literature (*Akesson, Karlsson et al. 1999*) a higher final pulse concentration of 20 g/L glucose was utilized. The pulse concentration was 50% in comparison to Phue et al (*Phue and Shiloach 2005*) to assure saturation while minimizing pule duration. DO2 levels below 30% have been reported to alter transcription and impact the metabolism (*Phue and Shiloach 2005*). In this respect a drop of DO2 below 15%, as occurred during up-pulsing of Lin et al (*Lin, Mathiszik et al. 2001*), can significantly impact qscrit, since the metabolism is restrained by oxygen availability. This fragile metabolic state in-between glucose saturation (*Akesson, Karlsson et al. 1999*) and oxygen limitation (*Lin, Mathiszik et al. 2001*) questions the experimental setup of up-pulsing while relying on DO2 as response. This problem can be avoided by down pulsing (*Henes and Sonnleitner 2007*) or by decoupling data analysis from DO2.

Regarding the workflow of quantifying the physiologic capacities the subsequent step of data evaluation is crucial and highly dependent on data quality. Especially in the setting of HCD bioprocesses, transitions from growth to limitation can occur within 15-30 s (*Schaepe, Kuprijanov et al. 2014*) which emphasizes the necessity of high frequency sampling. In contrast to biomass quantification the volatility of metabolites and substrate in the supernatant constitutes a significant risk for metabolite/substrate quantification. Irrespective of the frequency, in-between offline samples data can merely be interpolated. This circumstance impairs proper physiological characterization in-between sampling points. Errors imposed by sample handling and analytical errors cannot be fully avoided, consequently a certain level of uncertainty remains. By basing data evaluation on biomass and off-gas data as model inputs only verified by analytical data, noise is reduced and overall data quality is increased (FIGURE 3). Since the model provides an estimate of the total accumulation of C-source the necessity of identification and quantification of each metabolite is reduced.

Model performance was challenged at hand of a dynamic oscillation of the feeding rate (FIGURE 4). The clear correlation of model disturbance with DO2 controller actions, illustrates the limits of the model for qsbal quantification, as adaptations of pO2 lead to perturbations gas equilibrium. These controller actions should either be avoided or accounted for by the means of correction algorithms. Since such process dynamics are not

completely avoidable in HCD fermentations with high metabolic activity, any approach based on the DO2 signal appears too sensitive for robust process development and process control.

By controlling the qs within the oscillation experiments we facilitated the correlation of qsmean and qscrit slope. A positive correlation of qscrit and the specific growth rate has been suggested before (*Lin, Mathiszik et al. 2001*). But in the latter contribution C-source was provided by a constant feeding rate. Consequently the growth in biomass leads to an inevitable time correlated decline in µ. Not controlling the physiology in terms of µ hereby makes the differentiated identification of the cause of qscrit decline between time and µ impossible. A sound science based identification of the cause requires the control of one of the two variables of interest. This generally applicable principle emphasizes the necessity for physiological process control in the context of physiological process development.

Conclusions

The goal of this paper was to outline a workflow enabling the quantification physiological capacities which are a pre-requisite for the design of experiment within physiological process development.

In accordance to the state of the art pulse up-pulsing experiments were conducted which were decoupled from the DO2 signal as response. Within this workflow the following points were made:

- ➔ According to the conducted up-pulsing experiments qscrit is dependent on induction time and on the process parameter temperature. Changing the objective of the workflow to define a time dependent trajectory of qscrit.
- ➔ Model based data evaluation with off-gas and biomass as only time resolved input parameters increases data quality and decreases dependency on offline sample handling.
- ➔ Owned to the model based approach we found for the first time that the decline of qscrit is closely correlated to metabolic activity as qsmean, suggesting a metabolic memory effect induction phase.
- ➔ Based on the impact of qsmean on the decline of qscrit we propose to utilize qsmean as descriptor for the metabolic memory effect.

The latter findings change the initial objective of the workflow to quantify the physiologic capacity of respective strain. The time dependency of qscrit calls for a time resolved qscrit trajectory instead of a single numeric value. But since this trajectory is additionally dependent on process parameters and memory effects a large number of pre-curser experiments is required in order to avoid accumulation. As a consequence we propose to utilize the outlined model for a real time feedback control on physiological variables. Hereby the model would require real biomass estimation and offgas analysis as input and deliver a process value of the current qs as output. Thereby a simple step controller would facilitate the feedback control of qs while sensing qscrit in real time. This would eliminate the need for pre-curser experiments to determine the qscrit trajectory for each setting of process parameters as whole.

Acknowledgements

We are grateful for the financial support of the Sandoz AG and the Christian Doppler Society Austria.

Conflicts of Interest

The authors declare no conflict of interest.

References

Aehle, M., R. Simutis and A. Lubbert (2010). "Comparison of viable cell concentration estimation methods for a mammalian cell cultivation process." Cytotechnology **62**(5): 413-422.

Åkesson, M., P. Hagander and J. Axelsson (1999). "A probing feeding strategy for Escherichia coli cultures." Biotechnology Techniques **13**(8): 523-528.

Akesson, M., E. N. Karlsson, P. Hagander, J. P. Axelsson and A. Tocaj (1999). "On-line detection of acetate formation in Escherichia coli cultures using dissolved oxygen responses to feed transients." Biotechnol Bioeng **64**(5): 590-598.

de Assis, A. J. and R. M. Filho (2000). "Soft sensors development for on-line bioreactor state estimation." Computers & Chemical Engineering **24**(2–7): 1099-1103.

de Hollander, J. A. (1993). "Kinetics of microbial product formation and its consequences for the optimization of fermentation processes." Antonie Van Leeuwenhoek **63**(3-4): 375-381.

FDA (2004). "Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance."

FDA (2006). "Guidance for Industry Q9 Quality Risk Management."

Gnoth, S., M. Jenzsch, R. Simutis and A. Lubbert (2008). "Control of cultivation processes for recombinant protein production: a review." Bioprocess Biosyst Eng **31**(1): 21-39.

Henes, B. and B. Sonnleitner (2007). "Controlled fed-batch by tracking the maximal culture capacity." Journal of Biotechnology **132**(2): 118-126.

Hunter, I. S. and H. L. Kornberg (1979). "Glucose transport of Escherichia coli growing in glucose-limited continuous culture." Biochem J **178**(1): 97-101.

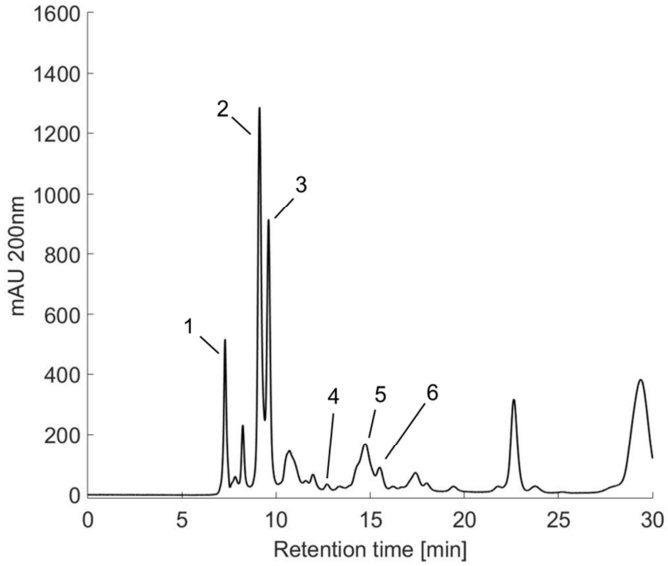
Jensen, E. B. and S. Carlsen (1990). "Production of recombinant human growth hormone in Escherichia coli: expression of different precursors and physiological effects of glucose, acetate, and salts." Biotechnol Bioeng **36**(1): 1-11.

Jenzsch, M., R. Simutis and A. Luebbert (2006). "Generic model control of the specific growth rate in recombinant Escherichia coli cultivations." J Biotechnol **122**(4): 483-493.
Korz, D. J., U. Rinas, K. Hellmuth, E. A. Sanders and W. D. Deckwer (1995). "Simple fed-batch technique for high cell density cultivation of Escherichia coli." Journal of Biotechnology **39**(1): 59-65.
Levisauskas, D. (2001). "Inferential control of the specific growth rate in fed-batch cultivation processes." Biotechnology Letters **23**(15): 1189-1195.
Levisauskas, D., R. Simutis, D. Borvitz and A. Lübbert (1996). "Automatic control of the specific growth rate in fed-batch cultivation processes based on an exhaust gas analysis." Bioprocess Engineering **15**(3): 145-150.
Lin, H. Y., B. Mathiszik, B. Xu, S. O. Enfors and P. Neubauer (2001). "Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of Escherichia coli." Biotechnol Bioeng **73**(5): 347-357.
Luli, G. W. and W. R. Strohl (1990). "Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations." Appl Environ Microbiol **56**(4): 1004-1011.
Phue, J.-N. and J. Shiloach (2005). "Impact of dissolved oxygen concentration on acetate accumulation and physiology of E. coli BL21, evaluating transcription levels of key genes at different dissolved oxygen conditions." Metabolic Engineering **7**(5–6): 353-363.
Sagmeister, P., P. Wechselberger and C. Herwig (2012). "Information Processing: Rate-Based Investigation of Cell Physiological Changes along Design Space Development." PDA J Pharm Sci Technol **66**(6): 526-541.
Schaepe, S., A. Kuprijanov, R. Simutis and A. Lübbert (2014). "Avoiding overfeeding in high cell density fed-batch cultures of E. coli during the production of heterologous proteins." Journal of Biotechnology **192, Part A**: 146-153.
Terpe, K. (2006). "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical

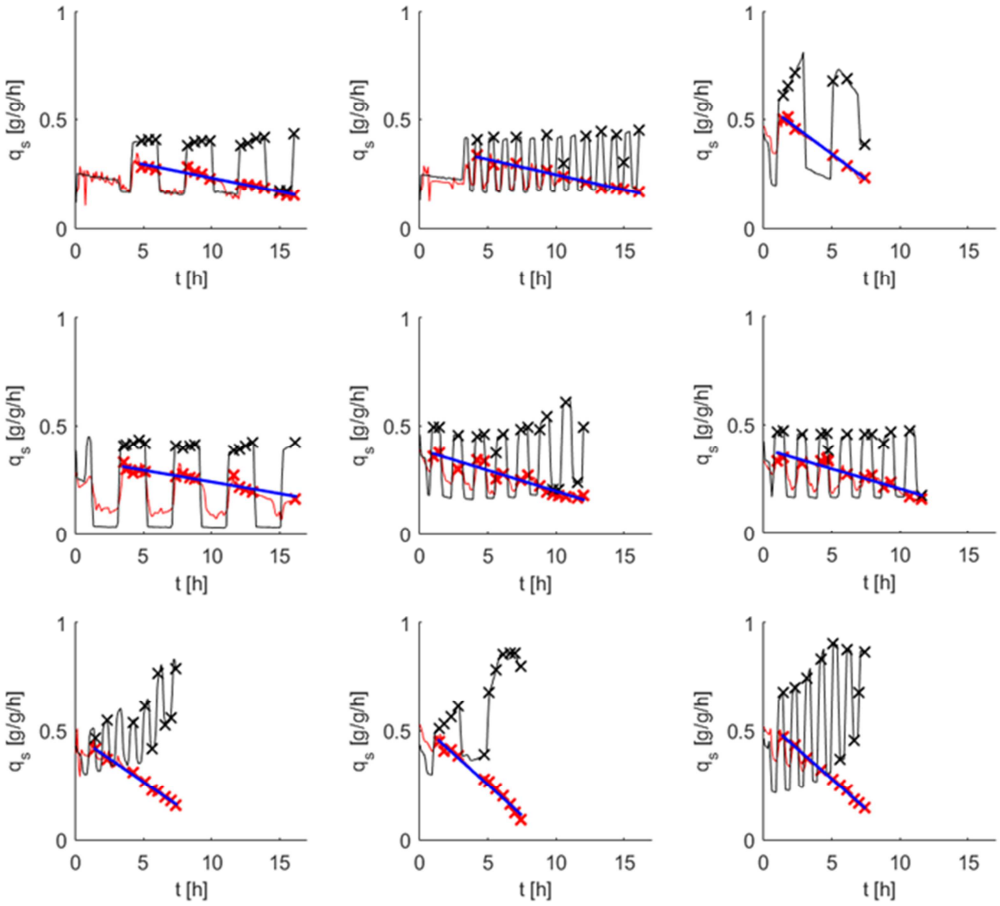
fundamentals to commercial systems." Applied Microbiology and Biotechnology **72**(2): 211-222.
Varma, A., B. W. Boesch and B. O. Palsson (1993). "Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates." Appl Environ Microbiol **59**(8): 2465-2473.
Walsh, G. (2010). "Biopharmaceutical benchmarks 2010." Nat Biotech **28**(9): 917-924.
Willmott, C. J. (1981). "On the validation of models." Physical geography **2**(2): 184-194.

Supplemental

Supplemental 1: Controller actions for DO2 control are the cause of noise in OUR. The discrepancy between OUR and CER leads to a misestimation of rox and consequently to a ill-closing C-balance



Supplemental 2: Supernatant chromatogram using a HPLC method for determination of organic acids. Various organic acids/intermediates of metabolism are present in the SN: 1: Oxalic acid, 2: Oxalacetic acid, 3: Ketoglutaric acid, 4: Glyceric acid, 5: Methyl-Succinic acid.



Supplemental 3: decline of qscrit over time (all included experiments)

Summarizing discussion - decline of qs crit

The dependency of the decline of qs crit on the q_s mean of the culture is a novel finding of this work, although a principal decline of maximum physiological capabilities is already published by other researchers (Schaepe et al. 2014; Lin et al. 2001). In comparison to these previous publications on declining physiological capabilities this work offers a more comprehensive investigation of the decline of q_s crit.

The fitness of the culture decreases with process time, especially for an IB production process – as IB formation imposes a high level of stress to the cells (Fahnert et al. 2004). Also for non IB processes induction of product formation withdraws energy and carbon from other physiological pathways (metabolic burden)(Heyland et al. 2011). The dependency of the q_s crit decline on the q_s mean of the culture could furthermore be explained by the increase in overall physiological activity at higher q_s mean values.

The decline of q_s crit is associated with a decline in r_x and μ respectively as the calculation approach for r_s presented is based on the summation of the oxidized substrate and the substrate used for building biomass (Equation 25). This implies a declining biomass yield over process time and therefore dynamic changes in the underlying stoichiometry.

3.3. Evaluation of the DoE to investigate the effects of the q_s oscillations on IB quality attributes and the specific product titer

3.3.1. Calculation of DoE factors & factor variability

For DoE evaluation the actual process values instead of the set-points were used. Figure 14 illustrates the calculation of the q_s amplitude and q_s mean. The q_s amplitude levels (red dashed lines) are calculated as the mean of q_s values 15% higher/lower than the q_s mean (pink dashed line). The q_s amplitude is consequently calculated as the average of the qs process values above/below the latter limits. For experiments that show no real oscillation profile of q_s the calculated amplitude can thus be regarded as over time steadily declining q_s. The frequency was calculated from the time intervals of the feed rate changes between low and high q_s levels.

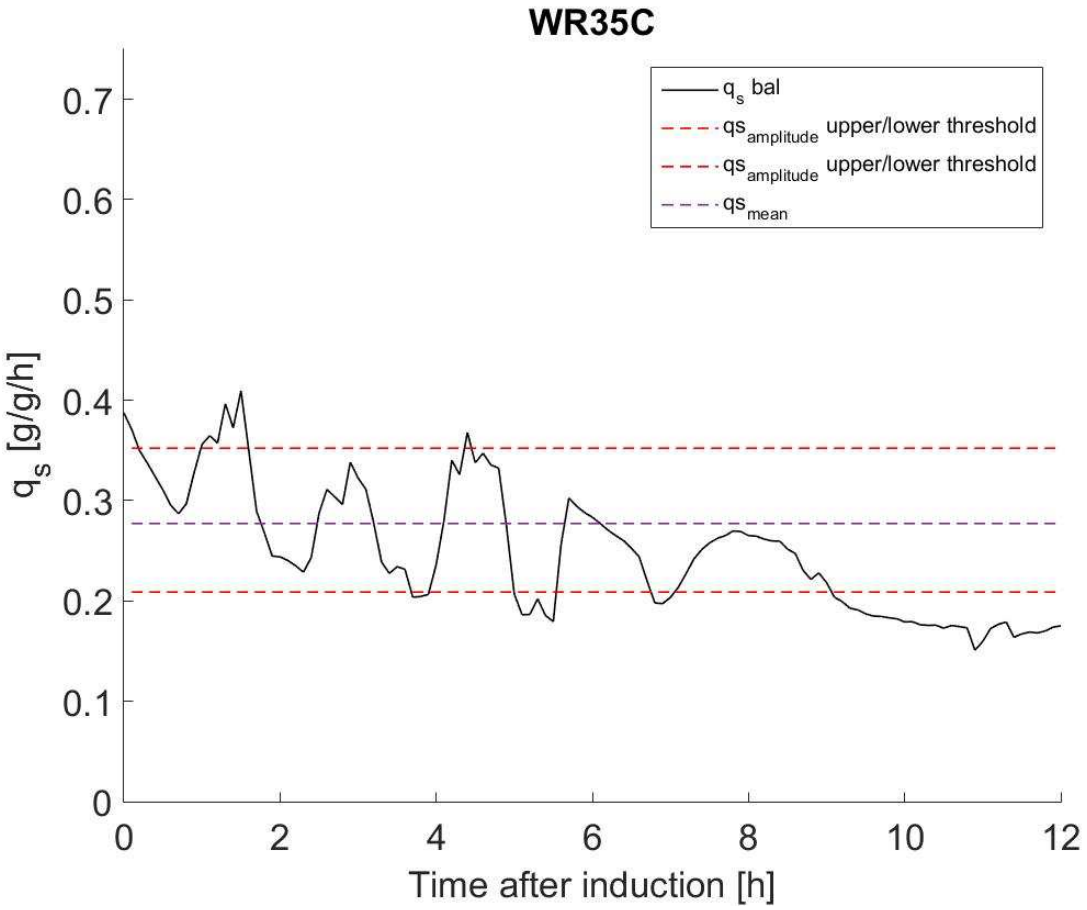


Figure 14: Calculation of the DoE factors q_s mean and q_s amplitude. The Figure shows the q_s calculated from the offgas Signal and offline BM (black line) over time after induction. The pink and red dashed lines show the q_s mean (pink) and the low and high q_s amplitude levels (red) used for q_s amplitude calculation.

Figure 15 shows set-points and process values of the DoE factors. By not reaching the intended q_s amplitude levels this dimension of the DoE is clipped, spanning only ~50% of the intended factor scaling in this dimension (from 0.1-0.2 clipped to 0.05-0.11). Furthermore

the q_s amplitude only reaches the intended lower boundary of 0.1 g/g/h as the actual high boundary. This is a consequence of the q_s crit decline over time and it's relation to q_s mean as shown in the publication part.

The achieved q_s mean values cover the whole setpoint range, but are generally lower (mean at high level: 0.35 g/g/h, mean at low level 0.21 g/g/h) and also have a relative error of ~12% (std/mean) at the low and high level of the DoE. As the frequency is calculated from the feed rate change the process values are equal to the set-points for this DoE factor.

The lacking orthogonality of the design caused by the declining q_s crit is reflected in the relatively high condition number of 13.14.

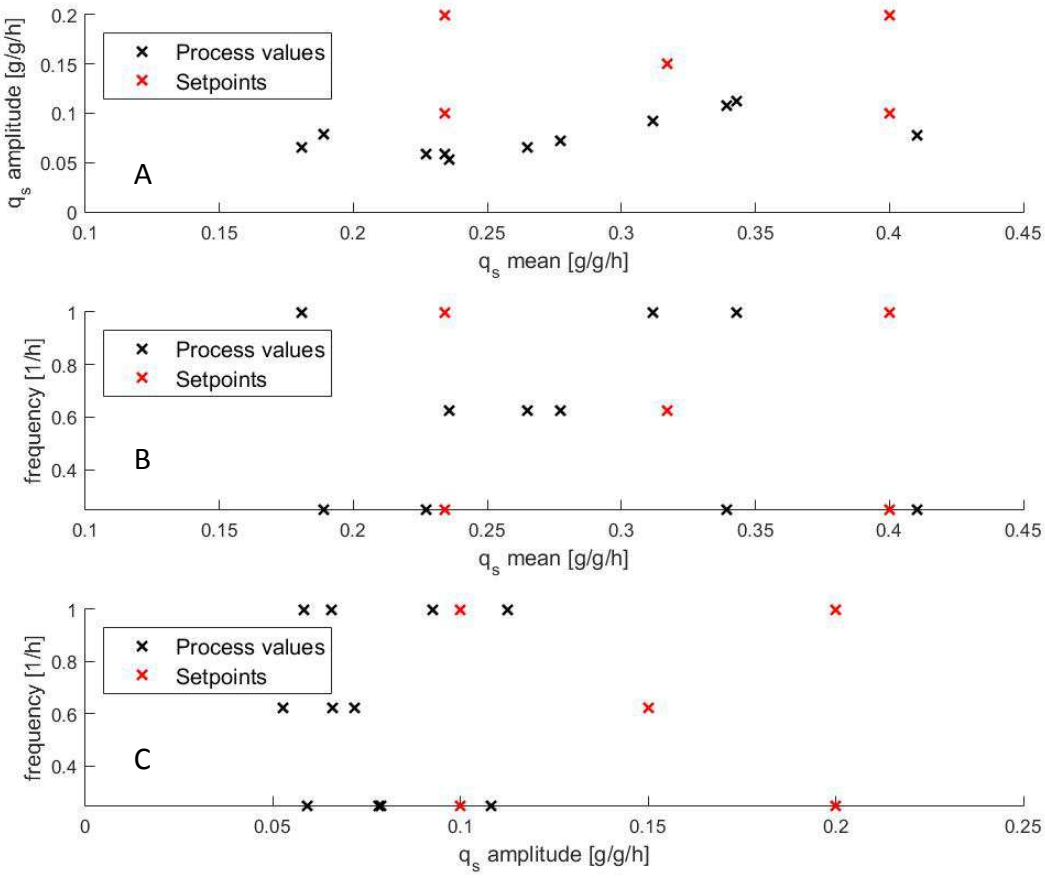


Figure 15: Comparison of the set-points and process values of the DoE factors. The figure shows the combination of the set-points (red x) and process values (black x) for 2 out the DoE factors q_s mean, q_s amplitude and frequency at a time. (A) q_s amplitude over q_s mean. (B) Frequency over q_s mean. (C) Frequency over q_s amplitude. The condition number of the DoE is 13.14 which indicate the lack in orthogonality of the design. There are no center point replicates (selection criteria: within a 10% range of the middle of the factor scaling for all factors)

3.3.2. DoE evaluation

Table 2 shows the Set-points and process values of the DoE factors as well as the responses (Specific product titer [g/g/h], solubility kinetic constant [OD/min], IB Purity [%] and Solubility Yield [%]).

Table 2: Process values and set-points of the DoE factors & responses.

q_s mean [g/g/h] PV	q_s mean [g/g/h] SP	q_s amp [g/g/h] PV	q_s amp [g/g/h] SP	Freq. [1/h] SP/PV	Specific titer [g/g]	Solubility kinetic constant [OD/min]	Purity [%]	Solubility Yield [%]
0,41	0,40	0,08	0,20	0,25	0,03	-0,0007	90,6	67,8
0,23	0,23	0,06	0,10	0,25	0,08	-0,0008	86,2	74,3
0,23	0,23	0,06	0,10	1	0,07	-0,0003	84,8	112,7
0,24	0,32	0,05	0,15	0,625	0,08	-0,0003	74,0	74,6
0,18	0,23	0,07	0,20	1	0,06	-0,0026	88,3	72,7
0,19	0,23	0,08	0,20	0,25	0,07	-0,0036	74,1	92,1
0,28	0,32	0,07	0,15	0,625	0,06	-0,0046	39,2	57,7
0,27	0,32	0,07	0,15	0,625	0,06	-0,0030	75,3	78,3
0,31	0,40	0,09	0,10	1	0,06	-0,0031	77,6	62,9
0,34	0,40	0,11	0,10	0,25	0,05	-0,0028	76,3	44,0
0,34	0,40	0,11	0,20	1	0,05	-0,0054	86,3	50,8

Investigation of the effects of the q_s oscillations on IB quality attributes and specific product titer

The coefficients for all responses of the MLR models are shown in Figure 16. The effect of a given factor on a response is significant if the estimated scaled and centered coefficient is bigger than its confidence interval (at α =0.05).

Two significant (confidence level = 0.95) correlations were found in the DoE evaluation (Figure 16A & B), which are a negative correlation between q_s mean and the product titer as well as a negative correlation of q_s amplitude and the solubility kinetic constant.

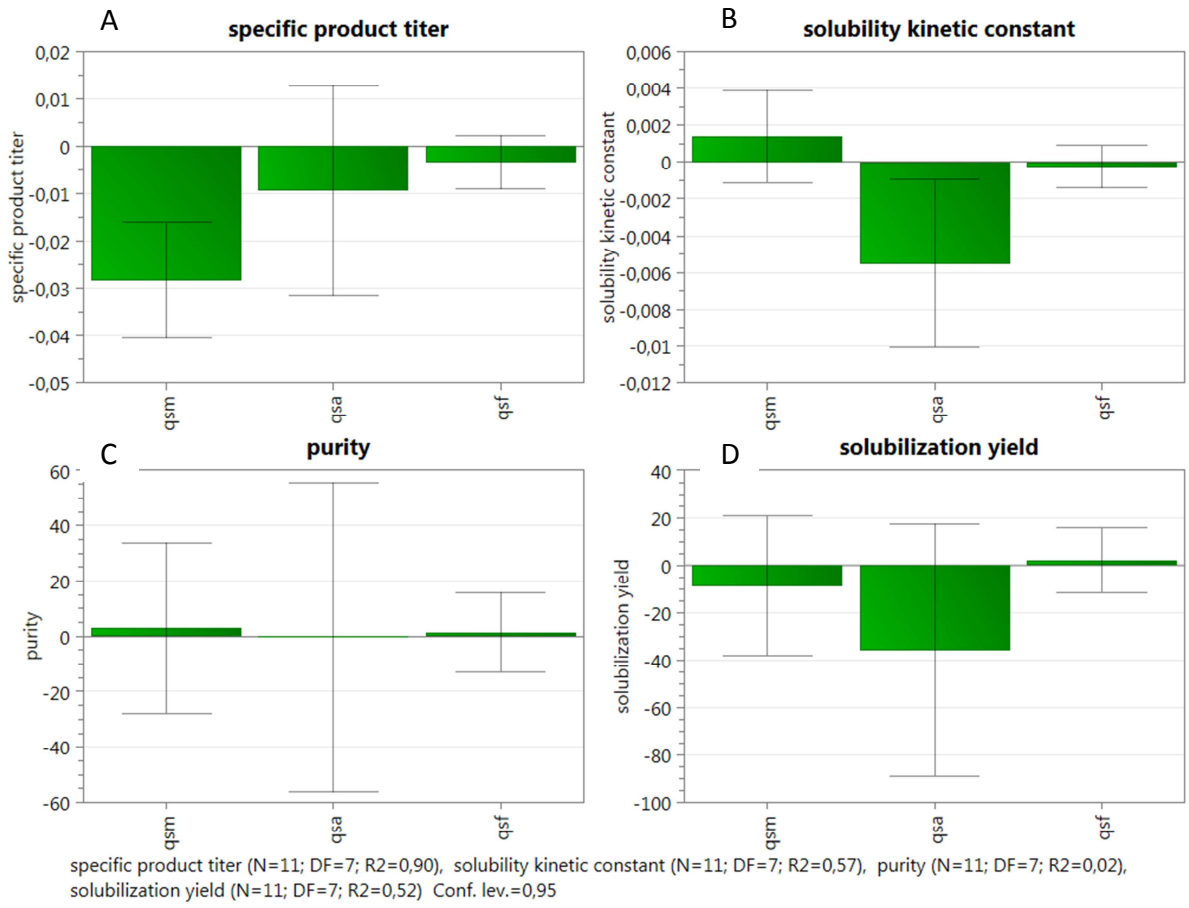


Figure 16: Coefficient plot for the DoE evaluation. The plot shows the scaled and centred coefficients (and the 0.95 confidence interval as error bars) for the MLR of the DoE factors (q_s mean…qsm, q_s amplitude…qsa and frequency…qsf) with the responses (A) specific product titer, (B) solubility kinetic constant, (C) purity and (D) solubilisation yield. R^2 values, number of experiments and degrees of freedom of the models are written below the graphs.

The significant correlations are shown in more detail in Figure 17. Figure A shows the negative correlation between the qs mean and the specific product titer. In Figure 10 B the correlation between the qs amplitude and the solubility kinetic constant is shown.

The found correlation between the q_s amplitude and the solubility kinetic constant, implying that higher amplitudes lead to slower solubilizing IBs, has a relatively low R^2 (0.57) as well as Q^2 (0.21).

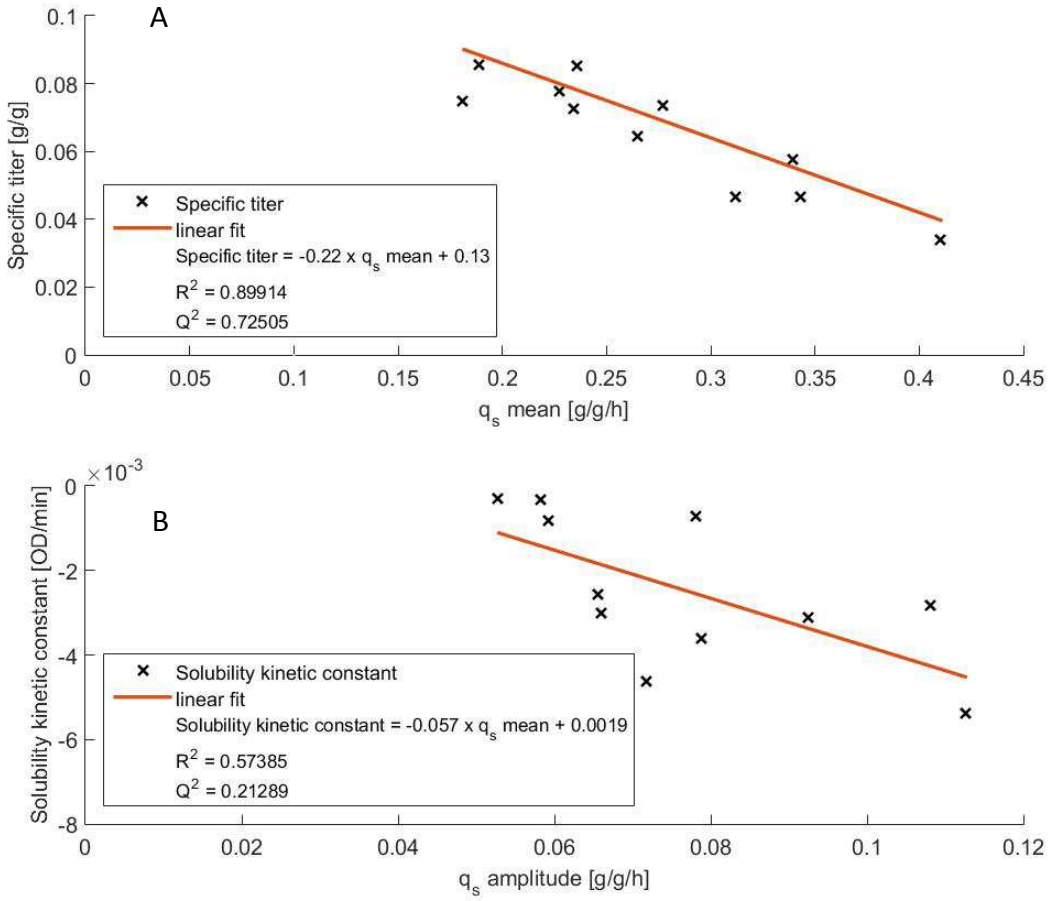


Figure 17: Significant correlations found in the DoE evaluation. (A) Specific product titer vs q_s mean; (B) Solubility kinetic constant vs q_s amplitude.

Summarizing discussion - effects of q_s oscillations on product quantity and IB quality

- negative correlation of the specific product titer with the q_s mean

In literature examples for positive (Hollander 1993; Cheng et al. 2003) and for negative correlations (Norsyahida et al. 2009) of q_s, and respectively the specific growth rate μ extist. In this thesis a negative correlation between q_s mean and the specific product titer was found for q_s oscillation experiments, whereas for constant and q_s ramp experiments a positive correlation has been found beforehand.

The reason for these opposing findings could either lie in the pH fluctuations or in the higher accumulation for oscillation experiments. This higher accumulation is based on the biomass overestimation as a result of substrate accumulation in the supernatant shown in Figure 13 together with timely recurring high q_s phases.

- negative correlation of the solubility kinetic constant and the q_s amplitude

Known correlations between USP parameters and IB attributes are correlations of the induction regime with secondary structure motives (Margreiter, Schwanninger, et al. 2008) and induction time and temperature with the IB size (Luo et al. 2006). A direct investigation of the effect of process dynamics on IB properties has not been done yet.

The linkage of the q_s amplitude, which represents the dynamic changes in the q_s, with the IB solubility is thus a novel finding in contrast to current literature linking USP parameters to IB properties.

4. Conclusion & Outlook

Decline of qs crit

The first goal of this master thesis was to quantify the decline of physiological maxima by using oscillatory post induction q_s profiles.

By conducting controlled oscillatory q_s profiles at different q_s mean and q_s amplitude levels this work found and quantified a dependency of the physiological process descriptor q_s mean on the decline in q_s crit over the time after induction. The quantification of the decline of physiological maxima at different q_s mean levels by utilising physiological process control embodies the novelty of this work.

The found dependency of the q_s crit decline on q_s mean must be considered for future DoE designs and strain characterizations, as the gained knowledge about the strain specific decline of q_s crit would limit the danger of running into substrate accumulation and by-product formation.

The presented approach of comparing q_s bal with q_s calculated from feed rate could furthermore be used for real-time recognition of potential exceedances of q_s crit and a respective control approach. The sensitivity of the presented approach to the dO_2 control should be decreased for such a potential control strategy.

Effect of oscillations on product quantity and IB quality

The second goal of this thesis was the investigation of the effects of the q_s oscillations on inclusion body properties and the specific product titer.

In this regard a negative correlation between the q_s mean and the specific product titer was found as well as a negative correlation of the q_s amplitude and the solubility kinetic constant.

An impact of dynamics in an USP parameter on an IB quality attribute, the solubility kinetic constant has thus far not been investigated in literature. Directly influencing the IB solubility kinetics could potentially be of high value for increasing DSP efficiency.

Further experiments for verification and reproducibility regarding the correlation of the solubility kinetic constant and the q_s amplitude would be beneficial as there might be an underlying effect of the pH fluctuations on the solubility kinetics which could potentially be the reason for the low R^2 and Q^2.

Regarding the correlation of q_s mean and the specific product titer further experiments would lead to a more reliable conclusion as an explanation for the opposite findings in q_s ramp and q_s constant experiments is still lacking. These additional experiments should focus on an investigation of the effect of pH fluctuations and the substrate accumulation on the specific product titer.

5. Appendix

List of Symbols

CH_2O	c-molar substrate composition
$CH_{1,82}O_{0,5}$	c-molar biomass composition without nitrogen
c_{feed} …	substrate concentration in feed [g/L]
$dSn_{(t)}$ …	fed substrate normalized by the CDW at the end exp. fed-batch [g/g]
$F_{(t)}$ …	feed flow rate [L/h] after time (t)
$F_{S,V}$ …	flow rate of feed solution [L/h]
F_S …	substrate feeding rate [c-mol/h]
q_S…	biomass specific substrate uptake rate [c-mol/c-mol/h] or [g/g/h]
$q_{S(t)}$…	biomass specific substrate uptake rate [g/g/h] at time point (t)
qs_{mean} …	average qs within a predefined window of dSn or time [g/g/h]
$q_{S\ bal}$ …	biomass specific substrate uptake rate calculated with C-Balance and DoR-Balance [c-mol/c-mol/h]
$q_{S\ crit}$ …	the critical specific substrate uptake rate as defined by Åkesson, Hagander, & Axelsson, 1999 [g/g/h]
qs_{glc}	specific substrate uptake rate calculated from the glucose concentration gradients in the pulse experiments [g/g/h]
RQ …	respiratory quotient [mol/mol]
r_{acc} …	rate of accumulating substrate and acetate [c-mol/h]
r_{CO2}	CER, carbon dioxide evolution rate [mol/ h]
r_s	substrate conversion rate [c-mol/h] or [g/h]
r_{O2} …	OUR, oxygen uptake rate [mol/h]
r_x	biomass conversion rate [c-mol/h]
γ_S …	Degree of Reduction (DoR) of substrate [mol/mol]
γ_X …	Degree of Reduction (DoR) of biomass [mol/mol]
μ …	specific biomass growth rate [1/h]
t…	process time [h]
V_0 …	volume at t = 0 [L]
X…	CDW [g] at (0) batch end or after time (t)
x…	CDW concentration [g/L] at (0) batch end or after time (t)
$Y_{X/S}$…	biomass yield on substrate [g/g] or [c-mol/c-mol]
$Y_{CO2/S}$…	carbon dioxide yield on substrate [c-mol/c-mol]
$Y_{O2/S}$…	oxygen yield on subtrate [c-mol/c-mol]
$Y_{H2O/S}$…	water yield on subtrate [mol/c-mol]
SP …	setpoint; the intended value of a given process parameter
PV …	process value; the actual value of a given process parameter based on measured quantities
w SN …	addition of w SN indicates values calculated with taking glucose and acetate measurements in supernatant into account
w/o SN …	addition of w/o SN indicates values calculated without taking glucose and acetate measurements in supernatant into account
bal …	denotes values calculated with the balancing approach

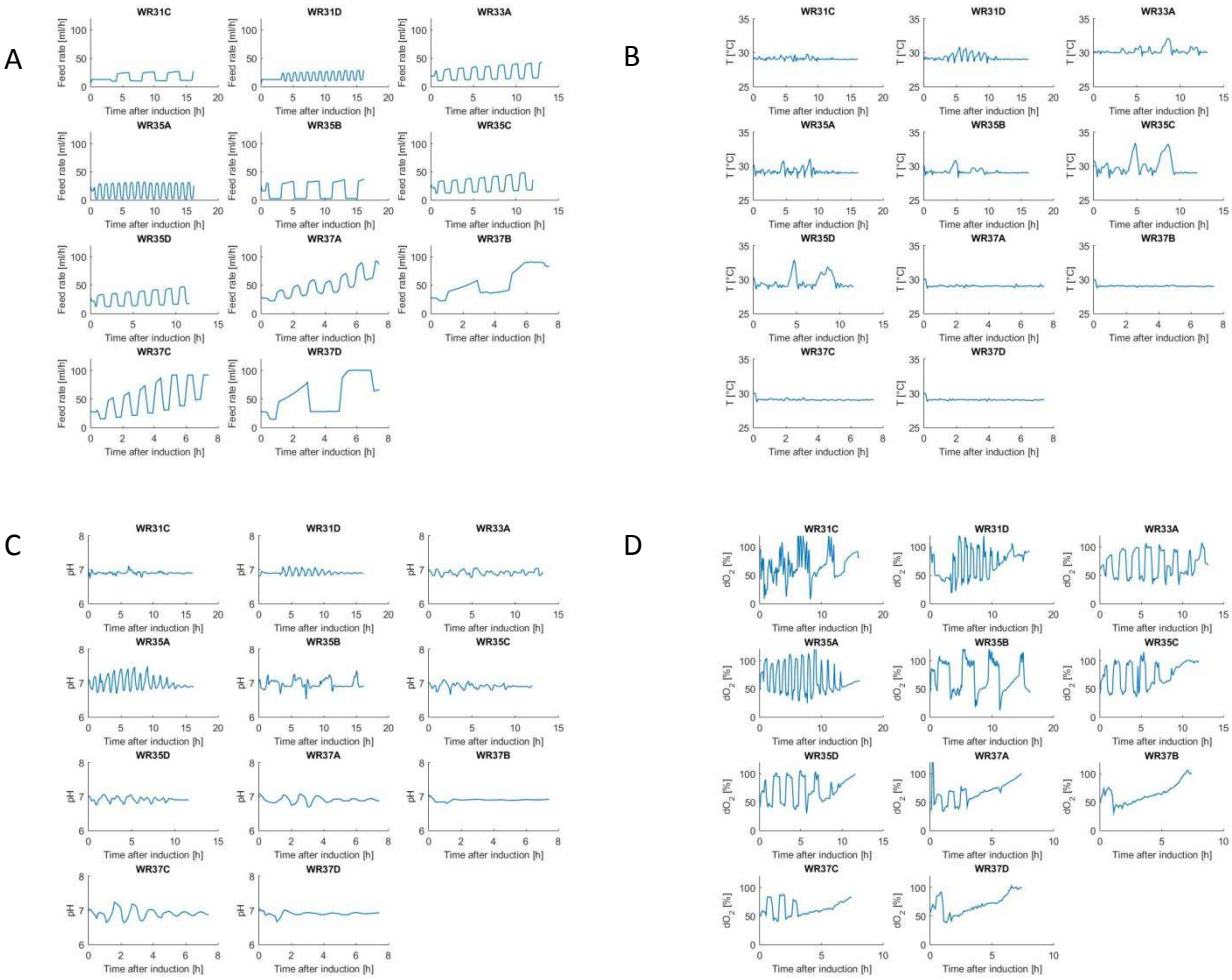
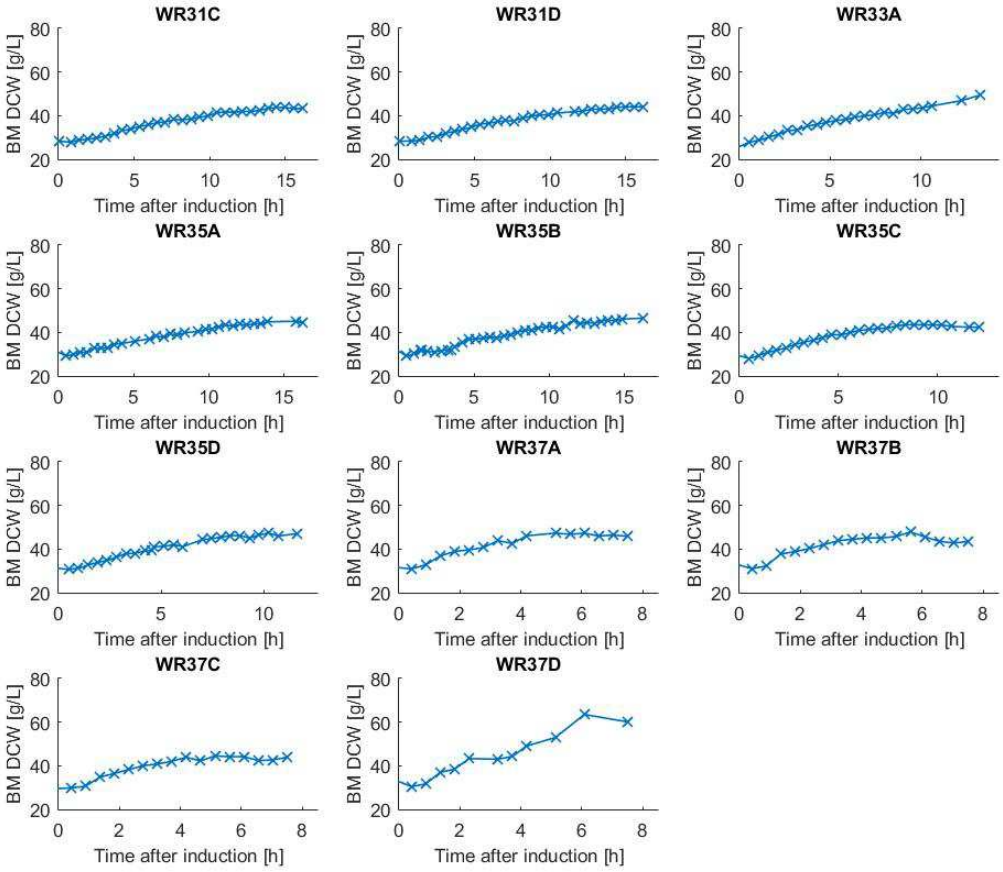


Figure 1: Complete data set of online data used for Quality control as well as the Feed rate. (A) Feed rate [ml/h]; (B) Temperature [°C]; (C) pH; (D) dO_2 [%]

A



B

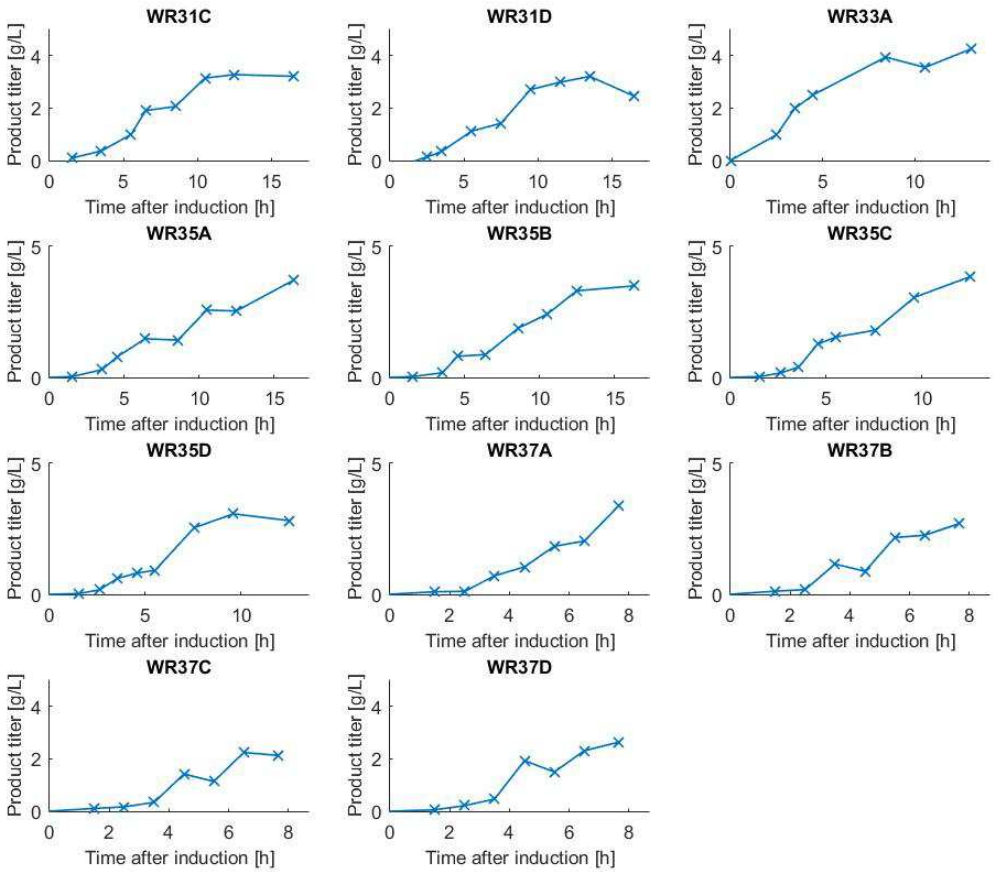


Figure 2: Offline BM and Product titer. **(A) Biomass dry cell weight [g/L]; (B) Product titer [g/L]**

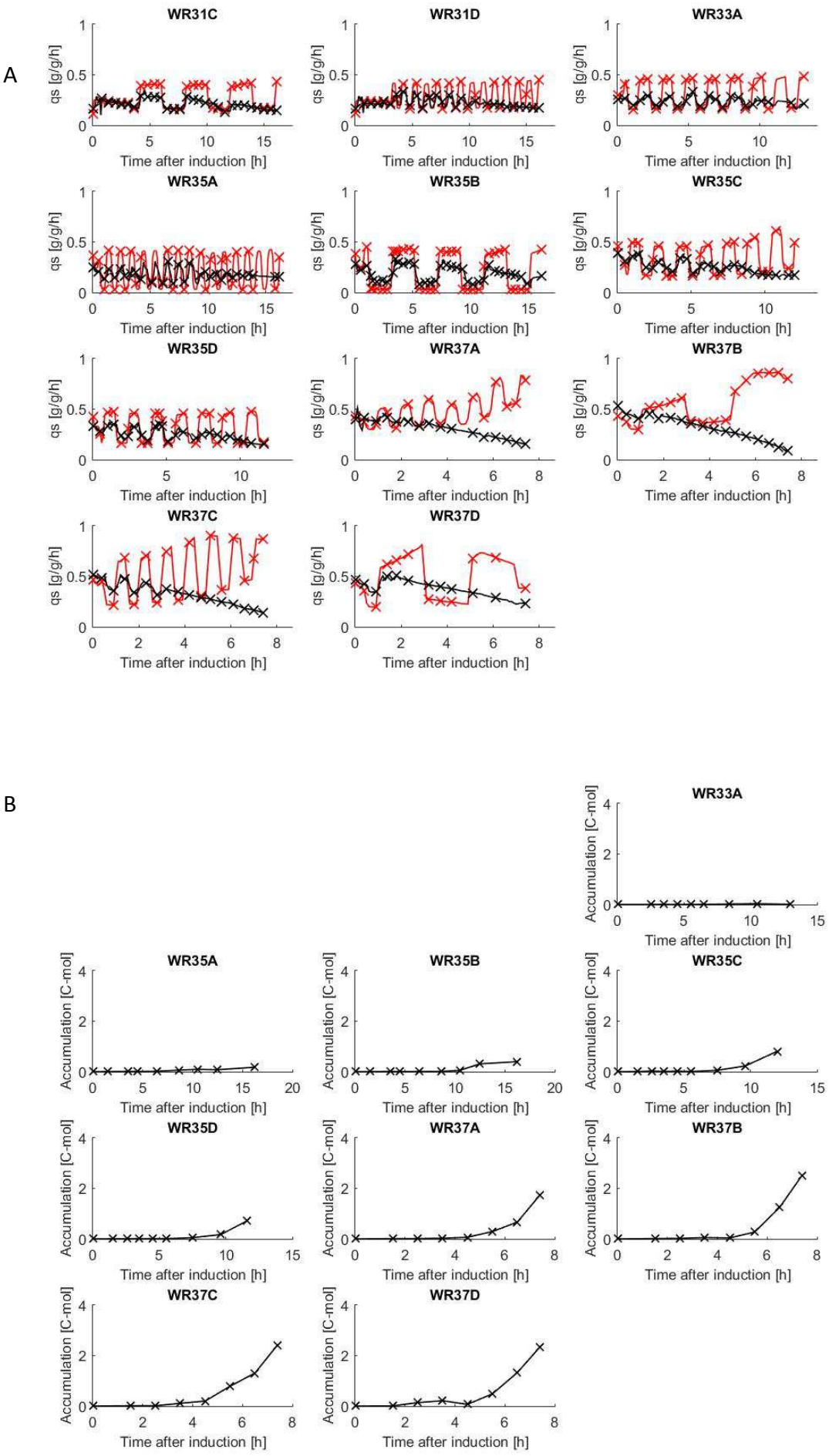


Figure 3: q_s trajectories and accumulation for all experiments. (A) q_s calculated with Feed rate and offline Biomass (red line) & q_s calculated with offgas signal and offline Biomass (yellow line), blue dots indicate the sampling points for offline Biomass; (B) Accumulation of Glucose and Acetate in the Supernatant (sum) [C-mol]

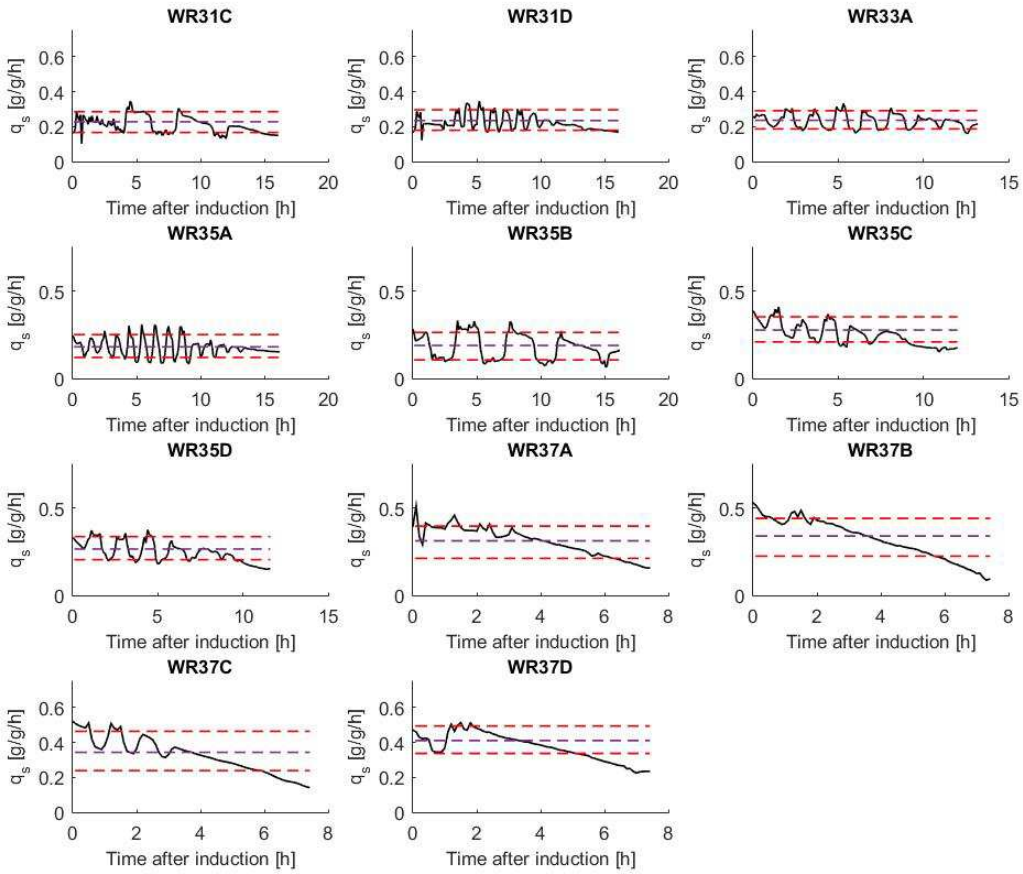


Figure 4 Calculation of the DoE factors q_s mean and q_s amplitude for all experiments. q_s calculated from offgas signal and offline Biomass (black line) over the time after induction, q_s mean (pink dashed line) and q_s amplitude levels (low/high) (red dashed line).

Data reconciliation for K2S1 softsensor

```
% reconcilation
for j=1:size(calc.(Fexn).time)
xm=[r_s; r_o2; r_co2];

E=[+1 0 +1 +1;gamma_s gamma_o2 gamma_x 0];
Em=[+1,0,+1;gamma_s,gamma_o2,0]; % First row C balance 2nd row DR balance
Ec =[1; gamma_x]; % Biomass Estimation

e_s=0.03;
e_o2=0.03;
e_co2=0.03;

Xi=[e_s 0 0;0 e_o2 0 ;0 0 e_co2];

Ec_star=(inv(Ec'*Ec))*Ec';

R=Em-Ec*Ec_star*Em;
[U,S,V]=svd(R);
Sconv=[1 0];
C=Sconv*S;
K=C*S'*U';
```

```
Rred=K*R;
eps = Rred * xm;
sai = diag(diag(xm * xm' * Xi * Xi'));
Phi = Rred * sai *Rred';
delta = (sai*Rred'*inv(Phi) * Rred)* xm;
xmbest=xm-delta;calc
xcbest = -Ec_star*Em*xmbest;
h = eps' * inv(Phi) * eps;

calc.(Fexn).rS_rec(j,2)=xmbest(1);
calc.(Fexn).OUR_rec(j,2)=xmbest(2);
calc.(Fexn).CER_rec(j,2)=xmbest(3);
calc.(Fexn).rX_rec(j,2)=xcbest;
calc.(Fexn).h_value(j,2)=h;

end
```

6. References

Ami, D. et al., 2006. Structural analysis of protein inclusion bodies by Fourier transform infrared microspectroscopy. *Biochimica et biophysica acta*, 1764(4), pp.793–799.

Calo-fern, B. & Mart, J.L., 2012. Biologics Market. , pp.1393–1408.

Caspeta, L. et al., 2013. Enhancing thermo-induced recombinant protein production in Escherichia coli by temperature oscillations and post-induction nutrient feeding strategies. *Journal of Biotechnology*, 167(1), pp.47–55. Available at: http://dx.doi.org/10.1016/j.jbiotec.2013.06.001.

Cheng, L. et al., 2003. Effect of specific growth rate on the production of a recombinant nuclease by Escherichia coli. , 14, pp.101–107.

Choi, J.H., Keum, K.C. & Lee, S.Y., 2006. Production of recombinant proteins by high cell density culture of Escherichia coli. *Chemical Engineering Science*, 61(3), pp.876–885.

Dietzsch, C., Spadiut, O. & Herwig, C., 2011. A dynamic method based on the specific substrate uptake rate to set up a feeding strategy for Pichia pastoris. *Microbial cell factories*, 10(1), p.14. Available at: http://www.microbialcellfactories.com/content/10/1/14.

Dürauer, A. et al., 2010. Autoprotease Fusion Technology: Development, Characteristics, and Influential Factors. *Separation Science and Technology*, 45(15), pp.2194–2209.

Fahnert, B., Lilie, H. & Neubauer, P., 2004. Inclusion Bodies : Formation and Utilisation. , pp.93–142.

FDA, 2004. Pharmaceutical CGMPs FOR THE 21 CENTURY — A risk based approach final report. , (September).

Gottschalk, U., 2008. Bioseparation in Antibody Manufacturing : The Good , The Bad and The Ugly. , pp.496–503.

Heyland, J., Blank, L.M. & Schmid, A., 2011. Quantification of metabolic limitations during recombinant protein production in Escherichia coli. *Journal of Biotechnology*, 155(2), pp.178–184. Available at: http://dx.doi.org/10.1016/j.jbiotec.2011.06.016.

Hollander, J.A. De, 1993. Kinetics of microbial product formation and its consequences for the optimization of fermentation processes. , pp.375–381.

ICH, 2009. ICH Harmonized Tripartite Guideline Pharmaceutical Development Q8(R2). , 8(August).

Jozefczuk, S. et al., 2010. Metabolomic and transcriptomic stress response of Escherichia coli. *Molecular systems biology*, 6(364), p.364. Available at: http://dx.doi.org/10.1038/msb.2010.18.

Kiviharju, K., Salonen, K. & Moilanen, U., 2008. Biomass measurement online : the performance of in situ measurements and software sensors. , pp.657–665.

Lin, H.Y. et al., 2001. Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of Escherichia coli. *Biotechnology and Bioengineering*, 73(5), pp.347–357.

Lionberger, R.A. et al., 2008. Review Article Quality by Design : Concepts for ANDAs. , 10(2), pp.268–276.

Lu, Æ.A., 2006. Estimation of biomass concentrations in fermentation processes for recombinant protein production. , pp.19–27.

Luo, J. et al., 2006. Size characterization of green fluorescent protein inclusion bodies in E. coli using asymmetrical flow field-flow fractionation-multi-angle light scattering. *Journal of Chromatography A*, 1120(1-2), pp.158–164.

Margreiter, G., Schwanninger, M., et al., 2008. Impact of different cultivation and induction regimes on the structure of cytosolic inclusion bodies of TEM1-??-lactamase. *Biotechnology Journal*, 3(9-10), pp.1245–1255.

Margreiter, G., Messner, P., et al., 2008. Size characterization of inclusion bodies by sedimentation field-flow fractionation. *Journal of Biotechnology*, 138(3-4), pp.67–73.

Neubauer, P. et al., 1995. Response of guanosine tetraphosphate to glucose fluctuations in fed-batch cultivations of Escherichia coli. *Journal of biotechnology*, 43(3), pp.195–204.

Norsyahida, A., Rahmah, N. & Ahmad, R.M.Y., 2009. Effects of feeding and induction strategy on the production of Bm R1 antigen in recombinant E . coli. , 49, pp.544–550.

Paalme, T. et al., 1995. The computer-controlled continuous culture of Escherichia coli with smooth change of dilution rate (A-stat). *Journal of Microbiological Methods*, 24(2), pp.145–153.

Pan, S. et al., 2014. Integrated continuous dissolution, refolding and tag removal of fusion proteins from inclusion bodies in a tubular reactor. *Journal of Biotechnology*, 185, pp.39–50. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0168165614002910.

Panda, A.K., 2003. Bioprocessing of therapeutic proteins from the inclusion bodies of Escherichia coli. *Advances in biochemical engineering/biotechnology*, 85, pp.43–93.

Sagmeister, P. et al., 2013. Soft sensor assisted dynamic bioprocess control: Efficient tools for bioprocess development. *Chemical Engineering Science*, 96, pp.190–198.

Schaepe, S. et al., 2014. Avoiding overfeeding in high cell density fed-batch cultures of E. coli during the production of heterologous proteins. *Journal of Biotechnology*, 192, pp.146–153. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0168165614008499.

Schlegel, S. et al., 2013. Optimizing heterologous protein production in the periplasm of E . coli by regulating gene expression levels. *Microbial Cell Factories*, 12(1), p.1. Available at: Microbial Cell Factories.

Soini, J. et al., 2005. Transient increase of ATP as a response to temperature up-shift in Escherichia coli. *Microbial cell factories*, 4, p.9.

Sørensen, H.P. & Mortensen, K.K., 2005. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. , 8(Figure 1), pp.1–8.

Sunya, S. et al., 2012. Comparison of the transient responses of Escherichia coli to a glucose pulse of various intensities. *Applied Microbiology and Biotechnology*, 95(4), pp.1021–1034.

Sunya, S. et al., 2013. Short-term dynamic behavior of Escherichia coli in response to successive glucose pulses on glucose-limited chemostat cultures. *Journal of Biotechnology*, 164(4), pp.531–542. Available at: http://dx.doi.org/10.1016/j.jbiotec.2013.01.014.

Swartz, J.R., 2001. Advances in Escherichia coli production of therapeutic proteins. *Current Opinion in Biotechnology*, 12(2), pp.195–201.

Upadhyay, A.K. et al., 2012. Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in escherichia coli. *PLoS ONE*, 7(3).

Waegeman, H. et al., 2012. Effect of iclR and arcA deletions on physiology and metabolic fluxes in Escherichia coli BL21 (DE3). *Biotechnology Letters*, 34(2), pp.329–337.

Walsh, G., 2010. Biopharmaceutical benchmarks 2010. , 28(9).

Walther, C. et al., 2014. Prediction of Inclusion Body Solubilization From Shaken to Stirred Reactors. , 111(1), pp.84–94.

Wechselberger, P. et al., 2012. Efficient feeding profile optimization for recombinant protein production using physiological information. *Bioprocess and Biosystems Engineering*, 35(9), pp.1637–1649.

Wechselberger, P. & Sagmeister, P., 2013. Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes. , pp.1205–1218.

Williams DC, Van Frank RM, Muth WL, Burnett JP (1982) Cytoplasmic inclusion bodies in Escherichia coli producing biosynthetic human insulin proteins. Science 215: 68–9.

Keil Peter 2014 Novel methodological approaches for fast and efficient physiological bioprocess characterization, Diploma Thesis.