Department of Physics, Chemistry and Biology

MASTER THESIS

A Physiology-based Soft Sensor for Monitoring and Control of Fed-batch Recombinant *E. coli* Culture Producing Green Fluorescent Protein

Cornelia Lukasser

February 2012

Linköpings universitet

Linköpings universitet
Institutionen för fysik, kemi och biologi
SE-581 83 Linköping

Abstract

The aim of process analytical technology (PAT) is to monitor and to control bioprocesses with powerful tools such as soft sensors. Soft sensors enable the use of unmeasurable variables for keeping a process on track.

In this work a soft sensor consisting of two hardware sensors and an estimation algorithm was used to control the feeding of an *Escherichia coli* cultivation expressing the recombinant green fluorescent protein.

The on-line NIR probe and carbon dioxide measurements were successfully combined to estimate the specific carbon dioxide rate which was used to control the feeding. Different settings for the specific carbon dioxide rate have been tested with the result that the used soft sensor was able to keep the acetate level low. In this study it is also shown that overflow metabolism did not affect the used strain and that too low feeding rates had an evident negative effect on cell growth and protein production.

Zusammenfassung

Das Ziel der Process Analytical Technology (PAT) ist es Bioprozesse mit leistungsfähigen Instrumenten wie zum Beispiel Softsensoren zu überwachen und zu steuern. Softsensoren ermöglichen den Einsatz von nicht messbaren Variablen um einen Prozess zu regeln.

Diese Arbeit beschreibt die Verwendung eines Softsensors der aus zwei Hardware-Sensoren und einem Algorithmus besteht. Der Softsensor wird für die Feed-Steuerung einer *Escherichia coli* Kultivierung, bei der das grün fluoreszierende Protein GFP exprimiert wird, verwendet.

Die on-line Biomassemessungen und die Kohlendioxidmessungen konnten erfolgreich zu einer spezifischen Kohlendioxidproduktionsrate zusammengefasst werden. Die berechnete Rate wurde verwendet um die Zufuhr von Feed-Medium zu steuern. Die Kultivierungen wurden mit verschiedenen Einstellungen durchgeführt mit dem Resultat, dass es möglich ist die Acetatkonzentration im Reaktor niedrig zu halten.

Weiters konnte mit den Ergebnissen gezeigt werden, dass der Overflow Metabolismus auf das Wachstum und die Proteinproduktion des verwendeten Stammes keinen Einfluss hatte und zu niedrige Feedraten zu deutlich schlechterem Ausbeuten führen.

Contents

List of Figures and Tables

1 Introduction

1.1 *Escherichia coli* as a host

The gram-negative bacteria *Escherichia coli* was identified in 1885 by the german pediatrician Theodor Escherich. It is distributed in the intestinal-tract of humans and warm-blooded animals. *E. coli* belongs to the family of *Enterobacteriaceae* which includes the genera of *Salmonella, Shigella and Yersinia* for example. Most *E. coli* strains can be seen as opportunistic pathogenic strains but there are also pathogenic strains, which can cause gastrointestinal illnesses in healthy humans (Feng et al., 2002).

One laboratory strain is the *E. coli* strain K12. It was isolated from the stool of a diphtheria patient in 1922 and used at the laboratories of the bacteriological department of Stanford University (Bachmann, 1972).

In 1997 the genome of the K12 strain was completely sequenced (Blattner et al., 1997). The detailed knowledge of the genome is one reason for the frequent use of *E. coli* in pharmaceutical production although it produces lipopolysaccharides, which are pyrogenic in humans and other mammals (Petsch and Anspach, 2000). Furthermore, it can be grown at predetermined growth rates to high cell densities on cheap carbon sources (e.g. glucose) (Riesenberg et al., 1991).

1.2 Green fluorescent protein

In 1962 the green fluorescent protein (GFP) was for the first time extracted from the luminous hydromedusan *Aequorea victoria* (Shimomura et al., 1962).

The structure of the wild-type protein was described by Yang et. al. in 1996. It forms a cylinder with a beta-structure on the outside and an alpha helix on the inside and is named the beta-can. The fluorophore is located in the center of the beta-can and protected from diffusible ligands (Yang et al., 1996).

Two proteins in *A. victoria* are responsible for the green light - the aequorin that provides excitation energy and the green fluorescent protein that absorbs blue

light and emits green light at a wavelength of 509 nm with a shoulder at 540 nm (Chalfie et al., 1994).

The cloning of a gfp10 cDNA which encodes for a 27 kDa protein (Prasher et al., 1992) and the demonstration that GFP could be used as a marker for gene expression in prokaryotes and eukaryotes (Chalfie et al., 1994) formed the basis for the application of GFP in molecular biology. Moreover, there are several mutants of GFP with different excitation properties and emission colors which can be used for e.g. a two-color assessment of differential gene expression (Heim et al., 1994).

One such mutant is GFPmut3.1, which is used in this experiment. It has improved folding properties, which make it a highly soluble protein, and an excitation spectrum at 488 nm for better fluorescing. Furthermore, it is detectable after a short lag phase, has a good signal to noise ratio and a low detection limit (Reischer et al., 2004).

1.3 Regulation of the protein production with the T7/lac promotor

The pET expression system is often used to regulate protein production in *E. coli* (see Figure 1).

pET plasmids contain a T7/lac promotor, a multiple cloning site and a T7 terminator. One important part of the regulation mechanism is the lactose operon (lac operon). It is coding for the three structural genes lacZ, lacY and lacA, whereby lacZ and lacY have an active role in the regulation of the lac operon. LacZ codes for beta galactosidase, which chops lactose into galactose and glucose. LacY codes for the membrane protein lac permease, which is needed for lactose uptake (Santillan and Mackey 2008).

The lac repressor protein is a product of lacI and inhibits the transcription of the T7 RNA polymerase gene, which is part of the *E. coli* genome and needed for the protein expression. It can be inactivated by the naturally occurring allolactose or by the synthetic inducer isopropylthiogalactoside (IPTG).

To induce the protein production IPTG is added to the bacterial growth medium. It binds to the repressor protein and the transcription of T7 RNA polymerase can start. The synthesized T7 RNA polymerase can now bind to the T7 promotor and transcribe the mRNA for the cloned protein (Chmiel, 2011; Clark and Pazdernik, 2009).

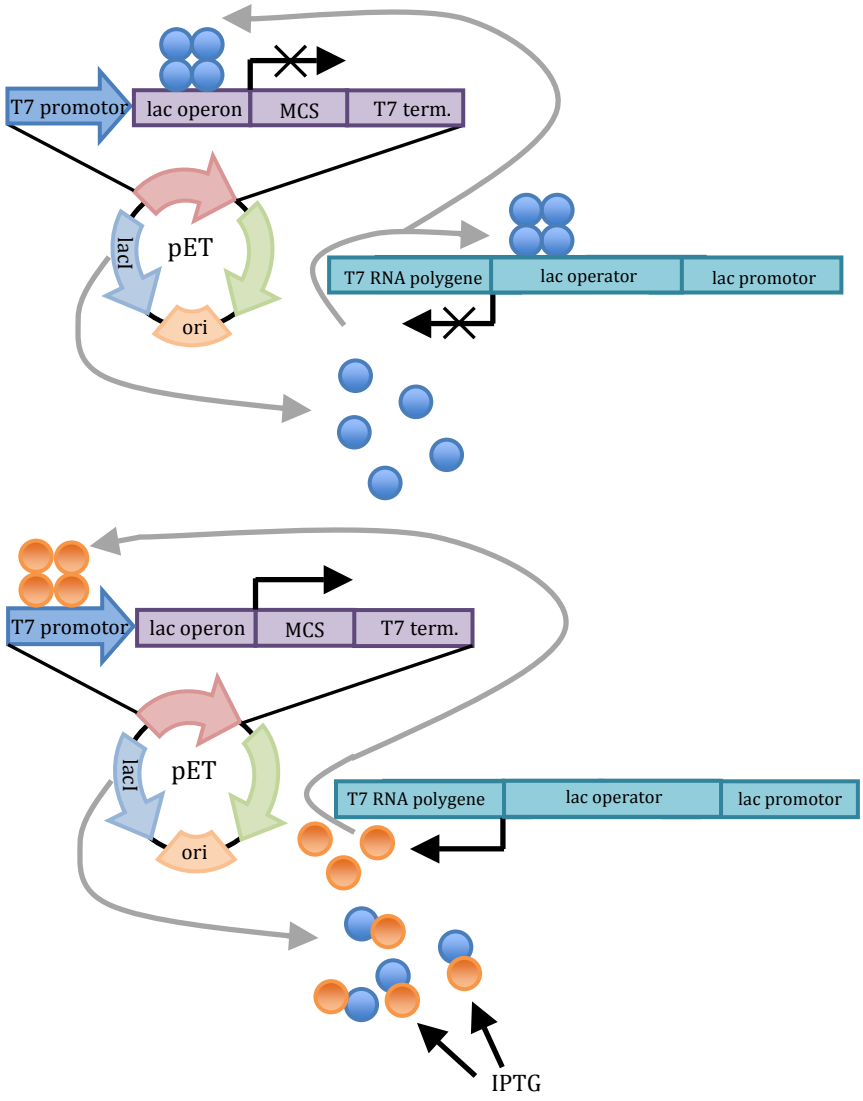

Figure 1 pET protein expression system compare with Clark and Pazdernik (2009)

1.4 Bacterial growth

The bacterial growth in batch cultivations can be described in five phases (see Figure 2). After inoculation the cells go through the lag phase (I) where they have to adapt to the media and which is characterized with nearly no growth. The second phase is the log phase (II) where the biomass grows exponentially until the stationary phase (III) where the biomass remains constant. In the end the cells enter the decline phase (IV) when the nutrients are consumed. In this phase no new cells are formed and the old cells die and lyse (Pommerville, 2010).

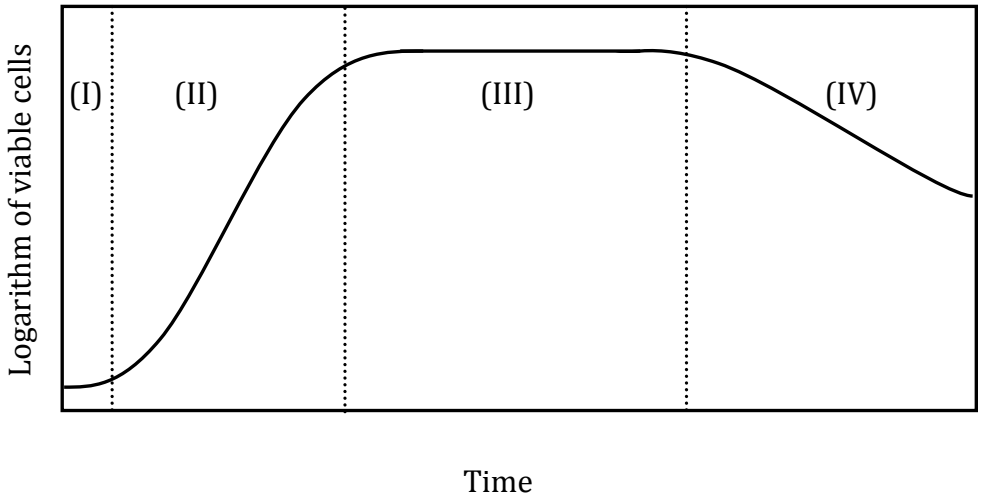

Figure 2 Bacterial growth compare with Pommerville (2010)

For batch cultivations it is not necessary to add further substrates during the growth phase. The substrates are added in the beginning of the cultivation and metabolized to biomass and product during bacterial growth. Substrate concentration and inhibition effects caused by accumulated metabolites limit the batch.

To overcome growth limitations it is possible to feed fresh media to the broth after the batch phase has ended. There are different strategies to feed, e.g. linear or exponential feed (Antranikian, 2006).

1.5 Bacterial Crabtree effect

When *E. coli* is growing at high growth rates and high glucose concentrations acidic by-products such as acetate are produced. This mechanism can be seen as a bacterial Crabtree effect due to an overflow metabolism at aerobe conditions (see Figure 3). Also mixed acid fermentation at anaerobe conditions plays an important role in fed-batch cultivations. Feeding a bigger amount of glucose leads to acetate accumulation in the cultivation medium and, due to that, to inhibition of cell growth and a worse protein production (Luli and Strohl, 1990; Xu et al., 1999).

In the overflow metabolism acetate occurs under aerobic conditions when too much acetyl CoA is produced but cannot enter the tricarboxylic acid cycle completely. Instead it is converted with help of pyruvate dehydrogenase to acetyl phosphate and later to acetate (Schlegel, 1992). Pyruvate oxidase, which is mostly build in the stationary growth phase, can directly transform pyruvate to acetate (Chang et al., 1994).

Under anaerobic conditions formate lyase converts pyruvate to formic acid and acetyl CoA. These intermediates are converted to H_2, CO_2, ethanol and acetate (Kessler and Knappe 1996).

Two types of *E. coli* strains produce acetate. Strains that produce acetate and consume it when no more glucose is left and strains that just produce acetate and cannot reassimilate it. In 1990 Luli and Strohl compared different *E. coli* strains and drew up a model that predicts inhibition of growth by acetate at concentrations of 5.3 to 9.8 g/L (Luli and Strohl, 1990).

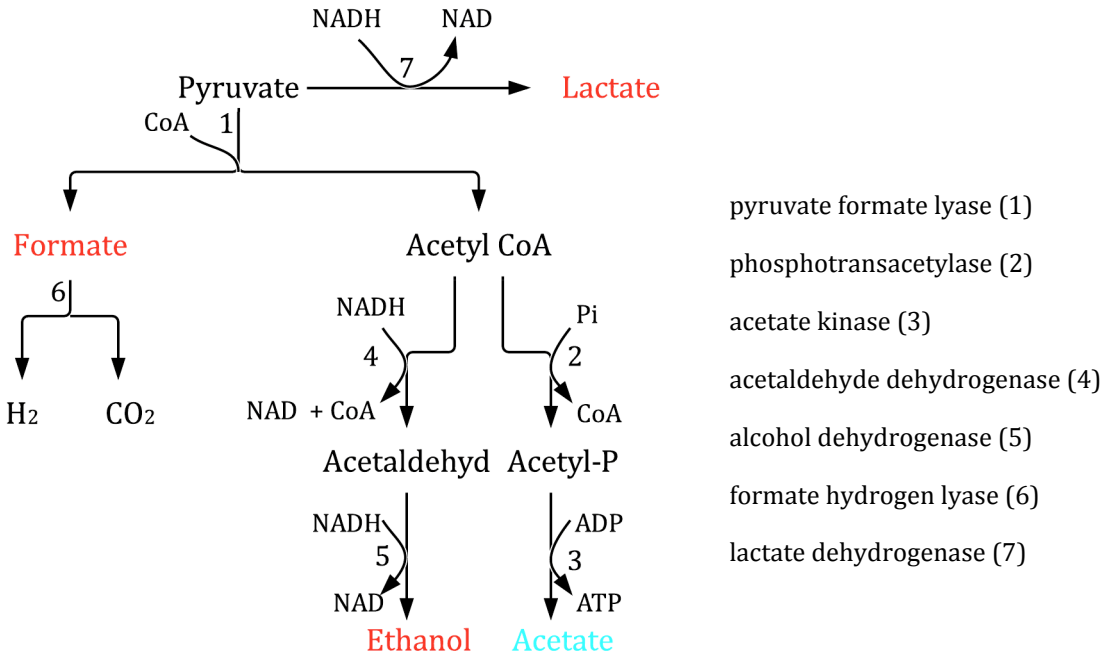

Figure 3 Overflow metabolism and mixed acid fermentation compare with Kessler and Knappe (1996)

The *in silico* model of Kotte et al. (2010) shows how *E. coli* grow on glucose, acetate or both, assuming that the selected strain is able to reassimilate acetate. *E. coli* growing on glucose at about 5 g/L forms acetate until acetate is the only carbon source left. Taking these bacteria to a new environment containing the same amount of acetate (5 g/L) results in a longer adaption phase, slower growth and less biomass. The third environment in the model contains 3 g/L of both. The *E. coli*, which are used to acetate now, adapt quite fast to glucose again. They are consuming and growing similar to the beginning (Kotte et al., 2010).

1.6 Soft sensors and process analytical technology

For Process Analytical Technology (PAT) it is important to use modern measurement techniques to get a better knowledge about a process and to use that knowledge for quality assurance and process control (Gnoth et al., 2007).

Soft sensors are tools which can be used to monitor and control bioprocesses. They consist of a sensor (hardware) and an estimation algorithm (software) to provide information about immeasurable variables and kinetic parameters (see Figure 4).

It is necessary to use hardware sensors which produce reliable measurement results and estimation algorithms which deliver on-line estimates approximate to the real values of the unmeasured variables (Chéruy, 1997).

To fulfill the requirements of PAT, it is of great value to use soft sensors to keep a process on track. In cases where the signal deviates from the setpoint value there should be a quick response to adjust the deviation. This type of feedback control leads to reproducible batches and has also a positive effect on downstream processing because of smaller variances of input materials (Gnoth et al., 2007).

Several soft sensors have already been tested for cultivations e.g. soft sensors for estimating biomass concentrations by NIR absorbance or ammonia titration, specific uptake rates and oxygen transfer capacity (Sundström and Enfors 2008; Warth et al., 2010).

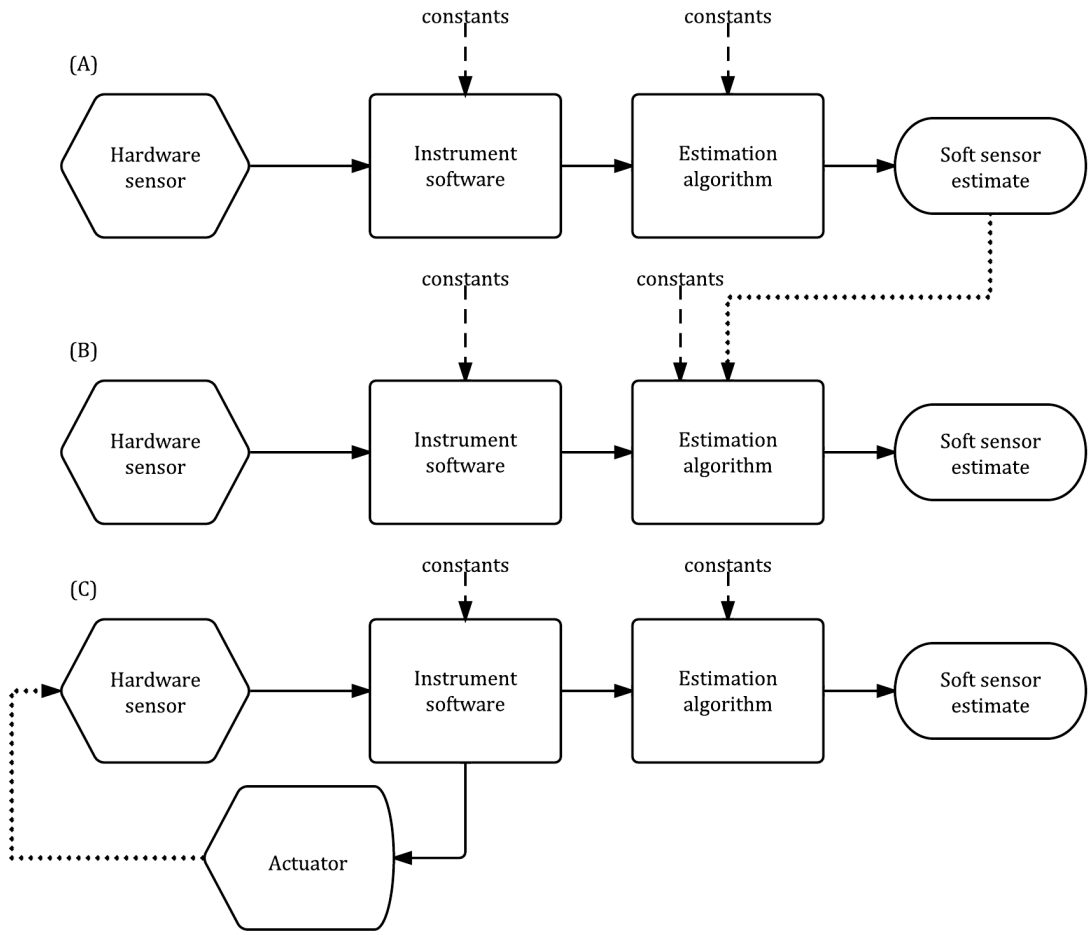

Figure 4 Soft sensor architectures compare with Warth, Rajkai and Mandenius (2010)

The choice of a soft sensor is product and process dependent. Every host-vector system and cultivation procedure has different characteristics. To involve those characteristics, soft sensors can be refined with appropriate statistical correction factors (Warth et al., 2010).

2 Aim

The aim of this study has been to control the growth of *E. coli* using carbon dioxide and to increase the yield of recombinant protein while the acetate concentration is kept as low as possible.

For this, two soft sensors were combined and used as a control unit.

3 Material and methods

3.1 Strain

For all cultivations the *E. coli* strain HMS 174(DE3) (Novagen, Madison, WI, USA) transformed with the plasmid pET30aGFPmut3.1 (Novagen and Clontech, USA), which is controlled by the T7/lac promotor, was used.

The strain was provided by the Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna.

3.2 Media

Three different types of media were used for the cultivations, a semi-synthetic medium (Nemecek, et al. 2007) for the pre-culture and two modified media for the batch and fed-batch cultivations.

The compositions of the media are listed in Table1 to Table 3. All components are from Merck.

Table 1 Composition of the pre-culture medium

Compound	Concentration [g/L]
KH_2PO_4	3.0
$K_2HPO_4 \cdot 3\ H_2O$	4.5
$C_6H_5Na_3O_7 \cdot 2\ H_2O$	2.5
$MgSO_4 \cdot 7\ H_2O$	1.0
$(NH_4)_2SO_4$	4.5
NH_4Cl	3.7
Yeast extract	1.5
Glucose	6,0
Trace element solution	0.5 [ml/L]

Table 2 Composition of batch and feed media

Compound	Batch medium [g/L]	Feed medium [g/L]
Glucose	5.0	100.0
KH_2PO_4	6.665	
$K_2HPO_4 \cdot 3\ H_2O$	0.25	
NaCl	1.2	
K_2SO_4	1.1	
Yeast extract	0.5	
$(NH_4)_2SO_4$	10.0	27.5
$MgSO_4 \cdot 7\ H_2O$	0.15	1.5
$CaCl_2 \cdot 2\ H_2O$	0.013	0.026
Trace element solution	0.125 [ml/L]	2.5 [ml/L]

The trace elements solution was added to the feed solution through a filter after autoclaving to avoid precipitation.

Table 3 Composition of trace elements solution

Compound	Concentration [g/L]
$FeSO_4 \cdot 7\ H_2O$	40.0
$MnSO_4 \cdot H_2O$	10.0
$AlCl_3 \cdot 6\ H_2O$	10.0
$CoCl_2$	4.0
$ZnSO_4 \cdot 7\ H_2O$	2.0
$Na_2MoO_2 \cdot 2\ H_2O$	2.0
$CuCl_2 \cdot 2\ H_2O$	1.0
H_3BO_3	0.5

3.3 Cultivation

For all cultivations a 10 L bioreactor (Model LMS 2002, Belach, Sweden) controlled by the software program BioPhantom was used.

The temperature was regulated at 37 °C and the pH was set to a range from 6.9 to 7.1. For keeping the pH in this range 1 M sulfuric acid or 20 % ammonia were added to the reactor. The regulation of the dissolved oxygen was controlled by the stirrer speed and set to 30 % whereby the minimum stirrer speed was 300 rpm. To reduce foaming during the cultivations, antifoam (Glanapon, Bussetti, Vienna) was added.

3.3.1 Pre-culture

For the pre-culture 200 ml medium was sterilized in a 500 ml shake flask. 16 hours before batch start 2 ml glucose (600 g/L) and one vial of the working cell bank were added and placed on a shaker at 200 rpm and 37 °C.

Before inoculation the OD value of the pre-culture was measured.

3.3.2 Batch cultivation

For every batch 3.7 L batch medium was sterilized in situ at 121 °C for 20 minutes. After sterilization 500 ml glucose and 200 ml pre-culture were added. Because of the water loss of about 10% during sterilization, the start volume for the batch was about 4 L.

The batch phases lasted in between 4 and 5 hours depending on the OD values of the pre-cultures.

3.3.3 Fed batch cultivation

After the glucose from the batch was consumed the feed could be started. Instead of waiting for the HPLC on-line result of glucose, which was 25 minutes delayed, a sudden drop in the carbon dioxide level was used as a starting point.

2 L of feed medium was added in about 8 to 10 hours using a peristaltic pump (P4 U1-MXV, Alitea, Sweden) controlled by a feed profile with an exponential growth rate of 0.3 h^{-1}. To control the amount of feed added, the feed medium was placed on a balance (XL-3100, Denver Instrument) that was connected to the BioPhantom software.

The induction with 0.12 g IPTG was carried out by a filter with pore size 0.2 µm after the cell concentration of the feed start has doubled, which was about OD 12 to 15. In the end the concentration of IPTG was about 0.02 g/L.

3.4 Analysis

3.4.1 Near infrared monitoring

For the on-line measurement of biomass an in situ sterilizable near infrared probe (Cell Growth Monitor Model 650; Wedgewood Technology, CA) with a pathlength of 5 mm was used. Different settings for the cell growth monitor were tested to get a linear relationship between measured NIR values and cell concentration. The chosen measurement range was 0-2 and the use of the second power function proved to be the most reliable for this kind of cultivations.

To calculate the current biomass from the growth monitor signal correlation, a factor was established by comparing biomass dry weight with the recordings from the NIR probe at different biomass concentrations.

3.4.2 Gas analysis of CO_2 and O_2

The changes in oxygen and carbon dioxide were measured with a gas analyzer (CP-460, Belach). Oxygen was measured with an oxygen sensor based on galvanic cell principles, while carbon dioxide was measured by infrared absorption.

Both sensors were calibrated with air and 100 % carbon dioxide. The 100 % carbon dioxide signal was read into the program as 10 % carbon dioxide and 0 % oxygen (*c1*). As second point the air was taken as 20.9 % oxygen and 0.04 % carbon dioxide into the program (*c2*). The slopes (*m*) of the two measured points were calculated and the raw values (*RAW*) were converted to calibrated values (*CALIBRATED*) (see Figure 5).

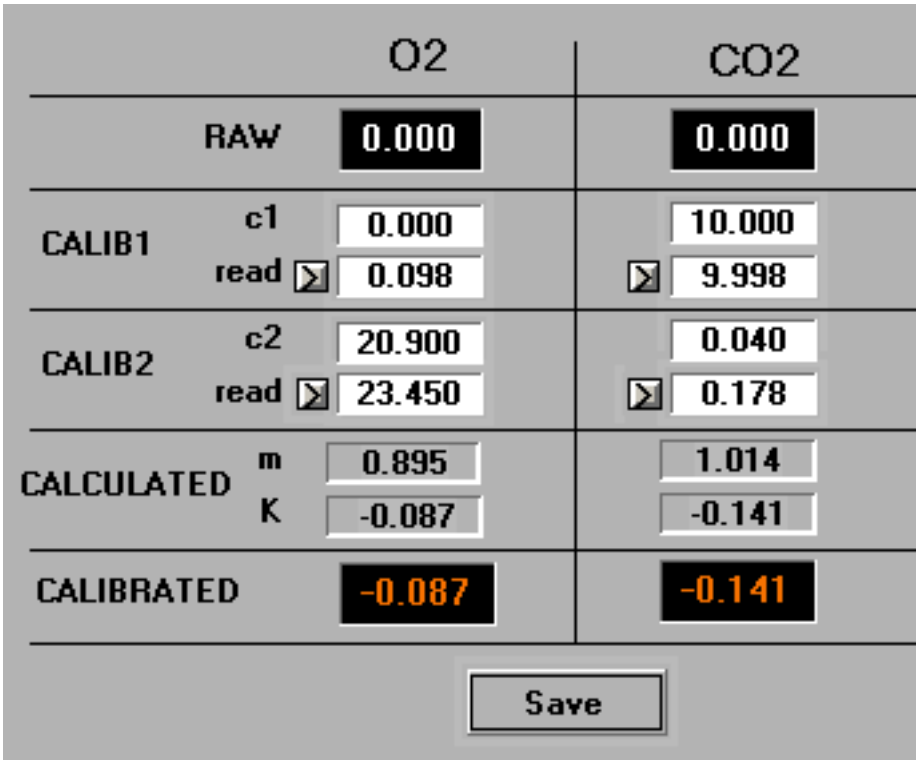

Figure 5 Screenshot BioPhantom: Calibration window of gas analyzer

3.4.3 High performance liquid chromatography (HPLC)

For every cultivation the concentrations of glucose, acetate, formate, lactate and ethanol were measured on-line by HPLC.

The bioreactor was equipped with an in situ ceramic membrane probe with pore size of 0.2 µm (FISP-sampling probe, Flownamics, Madison, US). A peristaltic pump took continuously culture medium from the reactor to a six-port two-position automatic injection valve (Rheodyne, US) using a 20 µL loop (TPMV, Rheodyne, US). Every 25 minutes a sample was injected into the HPLC system.

After injection the samples passed a guard column at a flow rate of 0.6 ml/min (Guard Cartridge 125-0129, BioRad, US). This protected the analytical column from impurities and an ion exclusion column (Aminex®HPX-87H, 300x7.8mm;BioRad, US) heated to 60°C. The column was packed with a polymer-based matrix to separate organic acids, alcohols and carbohydrates.

The detection was carried out with a refractive index detector (13 -detector RID-10A, Shimadzu, Japan) and the results could be seen with the software LC solution (Shimadzu).

To calibrate the system, standard solutions of glucose, organic acids and ethanol were used.

3.4.4 Optical density (OD)

During all cultivations the cell density was measured with a UV-spectrometer (Ultraspec 1000, Pharmacia Biotech, UK) at a wavelength of 600 nm and a pathlength of 1 cm. To make the results comparable with the results from the NIR-probe a commonly described factor of 1 g/L DW is equivalent to OD of 3 was used (Welschof and Krauss 2003).

3.4.5 Biomass dry weight (BDW)

8 ml of cell suspension were centrifuged for 10 min at 5000 rpm and 4 °C. The supernatant was discarded after centrifugation and the pellet was resuspended in 8 ml phosphate buffer and centrifuged again. The remaining pellet was resolved in RO water and transferred into dried and pre-weighed beakers. After 24 hours at 105 °C in the dry box, the beakers were put in a desiccator to cool down. The beakers were weighed and the biomass concentration calculated.

3.4.6 Fluorescence measurement

Fluorescence was measured with a spectrofluorometer (Fluostar Galaxy, MTX lab systems, VI, US) at excitation/emission wavelengths of 470/515 nm. Samples of 1 ml culture broth were taken before and after induction on a regular base. The samples were stored at -20°C until used for the measurement. After unfreezing, the samples were diluted 1:100 with phosphate buffer in two steps. The results from the measurement are given in relative fluorescence units (rfu).

3.5 Feedback control with soft sensors

Two hardware sensor signals were combined to control the feeding of the cultivations (see Figure 6). The measured concentrations of biomass and carbon dioxide were transferred to the BioPhantom program. With these two hardware sensor signals the specific carbon dioxide rate was calculated. A proportional controller used qCO_2 as control variable and together with the predetermined feed profile a feed rate was calculated. A PID controller, connected to a pump and a scale, ensured that the calculated amount of feed medium was added.

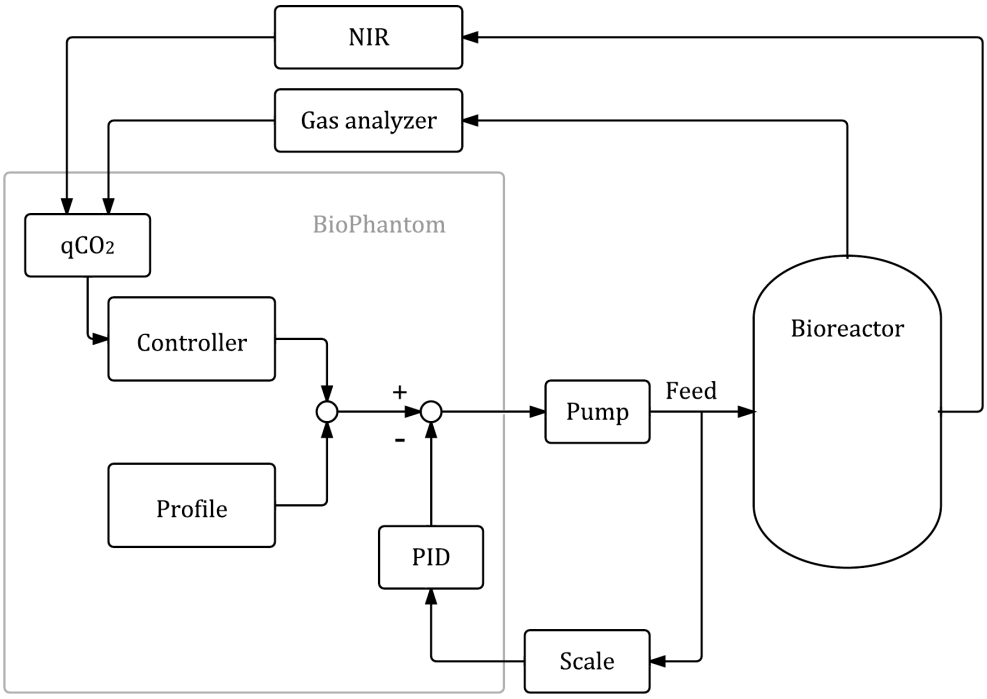

Figure 6 Feedback control using soft sensors

3.5.1 Biomass estimation with near infrared spectroscopy

To calculate the biomass concentration following equation was used:

$$x_{NIR} = A_{NIR} \cdot K_{NIR}$$

x_{NIR}	Biomass concentration [g/L]
A_{NIR}	Signal output from NIR measurement [%]
K_{NIR}	Empiric factor [g/L/%]

The A_{NIR} value is the signal from the NIR probe appearing in BioPhantom with values between 0 and 100 % (see Figure 7). To convert this signal to a biomass concentration in g/L, A_{NIR} is multiplied with a predetermined factor K_{NIR} (see 4.1).

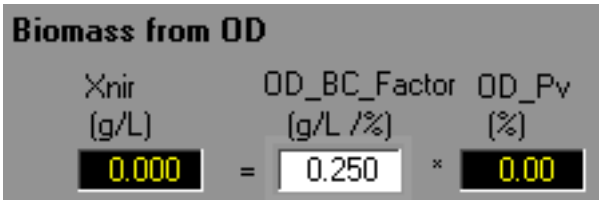

Figure 7 Screenshot BioPhantom: Biomass estimation

3.5.2 Carbon dioxide evolution rate

The carbon dioxide evolution rate was calculated from the measured carbon dioxide concentrations.

$$CER = Q \cdot 60 \cdot \left(\frac{CO_{2out} - CO_{2in}}{100} \right) \cdot \left(\frac{M}{V} \right)$$

CER	Carbon dioxide evolution rate [g/h]
Q	Volumetric airflow rate [L/min]
M	Molar mass [g/mol]
V	Molar volume of carbon dioxide [L/mol]
X	Biomass [g]

3.5.3 Specific carbon dioxide evolution rate

The specific carbon dioxide evolution rate was calculated with the CER and the biomass concentration as follows:

$$qCO_2 = \frac{CER}{X} \qquad qCO_2 = \frac{CER}{(x \cdot (V_0 + Cw \cdot \rho))}$$

CER	Carbon dioxide evolution rate [g/h]
V_0	Start volume [L]
Cw	Consumed weight [g]
ρ	Density feed medium [g/L]
x	Biomass [g/L]
X	Biomass [g]

For this calculation it is important to note that the fed batch volume changes. The term 'consumed weight multiplied with feed density' completes the calculation.

3.6 Feeding

A predetermined profile was used for the feeding phase (Gustavsson, 2011). The feed rate increased with time according to the exponential growth of the cells. In all experiments the profile started with theoretical feed rates of 100 g/h and increased to 400 g/h after 8 hours with a chosen μ of 0.3 h^{-1} (see Figure 8).

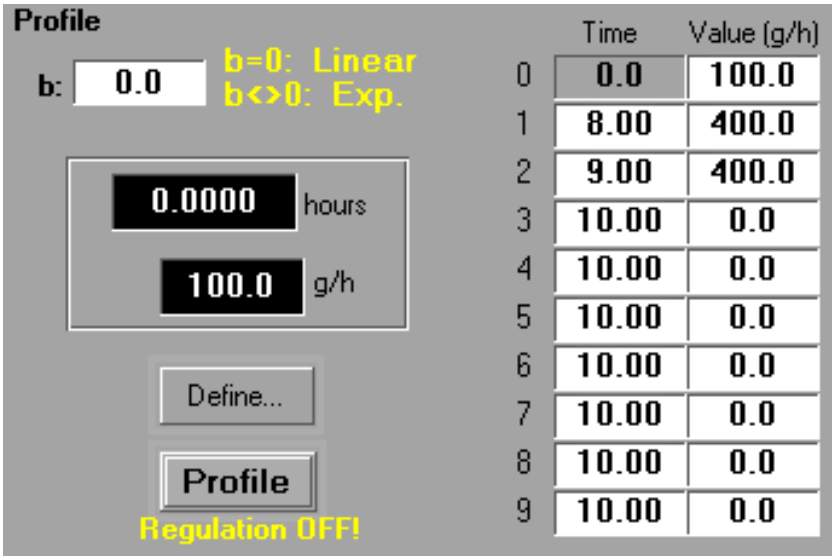

Figure 8 Screenshot BioPhantom: Used profile

Depending on the controller the actual values differed from the theoretical.

$$y(t) = (y_2 - y_1) \cdot \left(\frac{e^{(\mu \cdot t_1)} - 1}{e^{(\mu \cdot t_2)} - 1} \right) + y_1$$

t_1	Starting time (0)
t_2	End time
y_1	Initial feed rate
y_2	Final feed rate
μ	Specific growth rate

3.7 Feedback control with a proportional controller

A proportional controller was used to control the process value (*Pv*). Setpoint (*Sp*) and proportional band (*Kc*) could be set manually (see Figure 9). In this experiments the specific carbon dioxide evolution rate was applied as process value to control the feed rate.

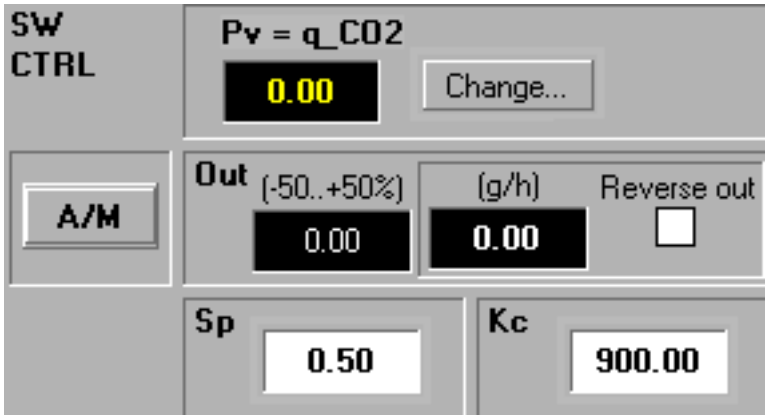

Figure 9 Screenshot BioPhantom: Control window

$e(t) = (Sp - Pv)/300 \cdot 100\ \%$

$u(t) = Kc \cdot e(t)$

$Y_2(s) = u(t) \cdot Fr$

$Y(s) = Y_1(s) + Y_2(s)$

Pv	Process value, qCO_2 [g/g/h]
Sp	Setpoint value for the controller
Kc	Amplification factor, increases the controller output
e(t)	Err%, error between Sp and Pv
u(t)	Out%, output [%]
Fr	HiRange-LoRange [g/h]
$Y_1(s)$	Profile output
$Y_2(s)$	Controller output
Y(s)	Feed output

After the scaled error (*e(t)*) between process value and setpoint was calculated, it was multiplied with the proportional band (*Kc*) to get the output (*u(t)*) in percent. For the controller output ($Y_2(s)$) the range of the pump (+/- 400 g/h) had to be included in the calculation. Controller output ($Y_2(s)$) and profile output ($Y_1(s)$) were summed up to result in the feed output (*Y(s)*).

4 Results and discussion

This section presents data of five fed-batch cultivations and the data for the determination of the calibration factor for the NIR probe.

For the first cultivation the predetermined feed profile was used shown in 3.6. Those results were compared to the following cultivation results where the feeding was controlled by the implemented proportional controller. In order to find out which setpoint for the specific carbon dioxide rate would result in a low acetate level and most GFP, setpoints between 0.2 and 0.5 g/g/h have been tested.

4.1 Calibration of biomass estimation

To estimate the factor K_{NIR} described in 3.5.1 biomass dry weight results were plotted against the NIR probe output (see Figure 10). The slope is 4.017 and hence $K_{NIR} = 0.25$.

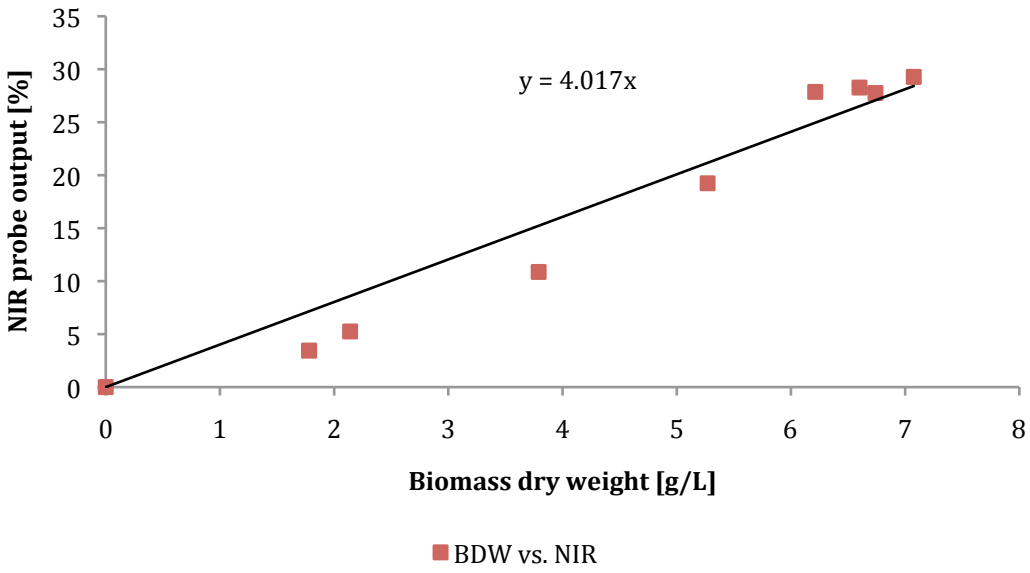

Figure 10 Linearity of BDW and NIR probe output

4.2 Cultivation A

The first cultivation was performed with the predetermined profile without using the controller system. The purpose of this cultivation was to compare the results to the ones carried out with the controller system.

The profile starts at a feed rate of 100 g/h and increases to 400 g/h during eight hours with a growth rate of 0.3 h^{-1}. The feed was started after the batch phase ended at hour 4.75 and at an OD value of approximately 6. The culture was induced after the first doubling of the biomass at an OD value of 12 and at hour 6.5.

Figure 11 shows the measured NIR probe signal, OD/3 values and GFP results. The results of the NIR probe are higher than the OD/3 values until hour 8. After induction the OD measurement results increased with the produced amount of GFP. This happened in every cultivation. As the energy of the light at 600 nm is not able to excite GFP it can be assumed that fluorescence was not present during these measurements (Zanzotto et al., 2005). Although the used GFPmut3.1 protein is highly soluble, it is possible that the cells produce inclusion bodies which higher the measured optical density. No inclusion body testing has been done to prove this assumption. During cultivation A GFP increased to a plateau of about 4100 rfu.

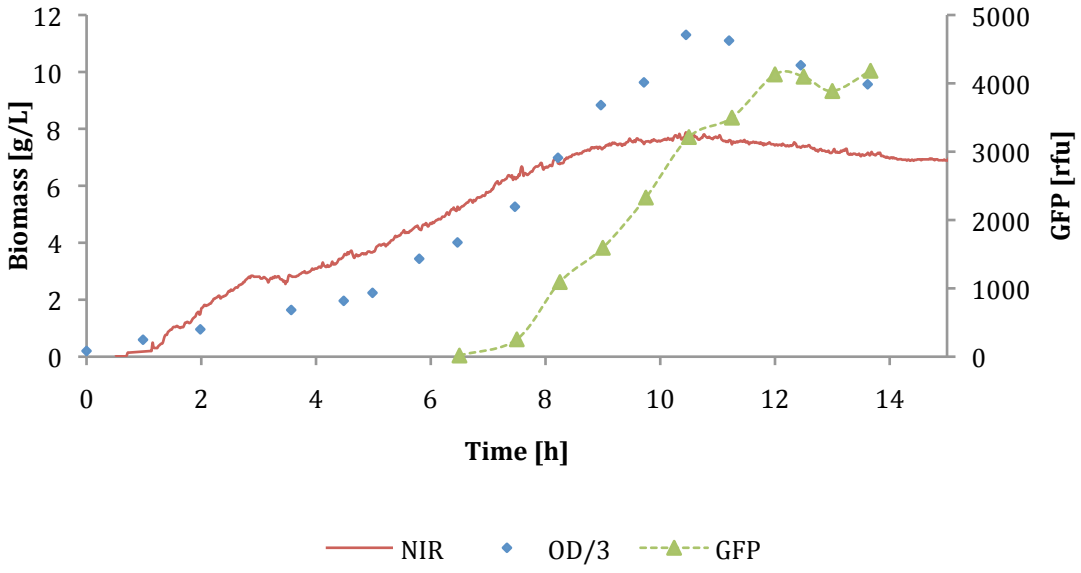

Figure 11 Cultivation A: Biomass and GFP

The results of the HPLC were interrupted due to technical problems, from hour 4.5 to 6.6. The results shown in Figure 12 demonstrate the consumption of glucose in the batch phase and the overfeeding with glucose after 10 hours. Acetate is formed up to a concentration of 1.8 g/L and is then consumed slowly. After 10 hours of feeding, glucose starts to accumulate. The bacteria are growing slower than calculated in the exponential profile due to the protein expression. The protein production, which is induced after 6.5 hours, is a burden for the cells and the used expression system pET30 with the T7/lac promotor and the synthetic inducer IPTG forces them to produce a high amount of GFP.

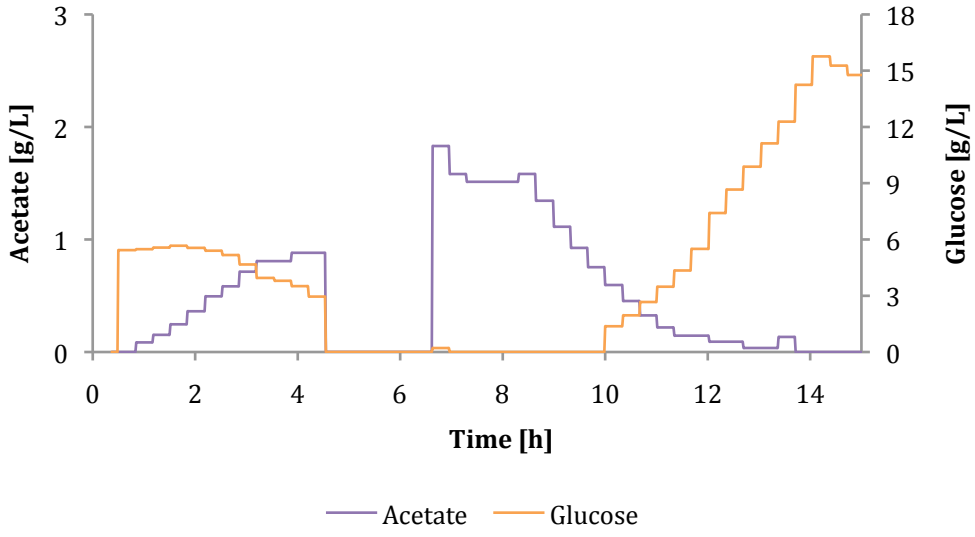

Figure 12 Cultivation A: Online data from the HPLC

The feed profile shown in Figure 13 works well until induction happens. The cells consume less glucose and it would be necessary to have another feed profile for the phase after induction than the one below.

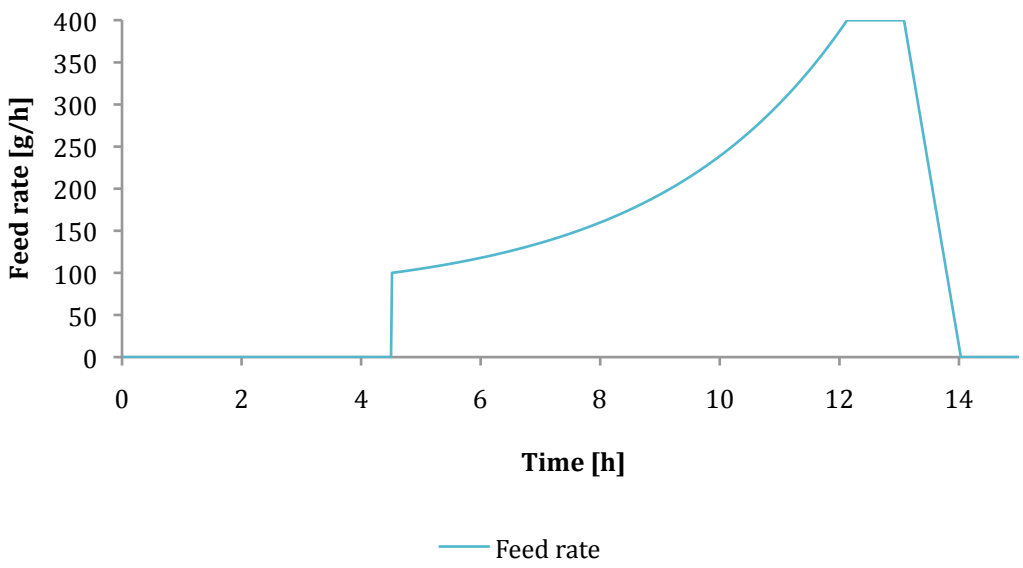

Figure 13 Cultivation A: Used feed rate

4.3 Cultivations B and C

The cultivations B and C were carried out with a low setpoint for the specific carbon dioxide rate to test the controller system and to see if the acetate level can be kept lower than in cultivation A. The chosen setpoint for both cultivations was 0.2 g/g/h.

The measured OD values of the pre-cultures prior to inoculation differ quite a lot. Cultivation B was inoculated with a pre-culture with a measured OD of 7.24 while the OD value for cultivation C was just 2.74 because a shake flask without baffles was used. Because of that, the feed of cultivation B starts an hour earlier than in cultivation C.

The biomass concentrations in both cultivations reached about 7 g/L as highest concentration (see Figures 14 and 15). The produced amount of GFP in these two cultivations is just half the amount of produced GFP in cultivation A (see Table 4).

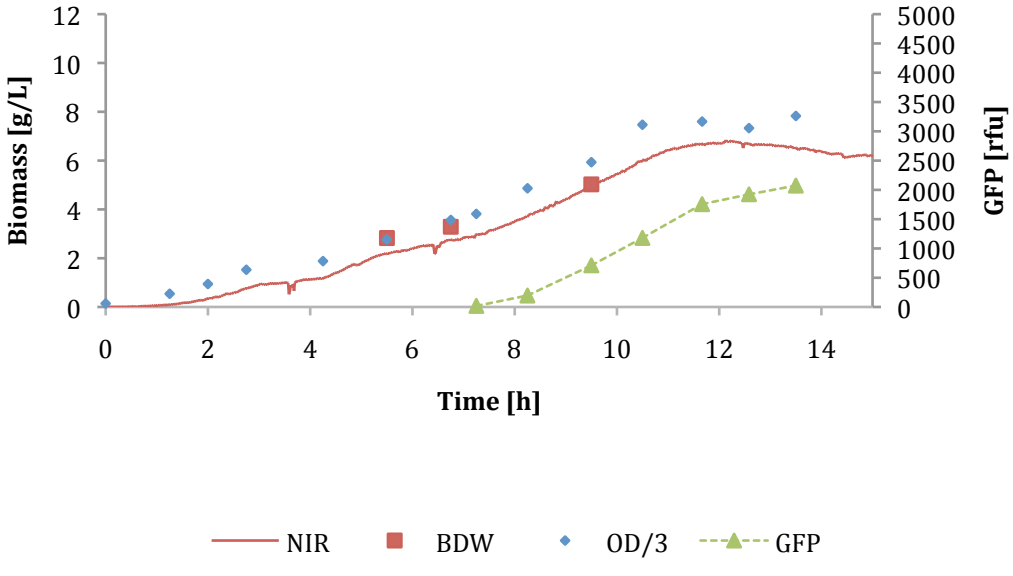

Figure 14 Cultivation B: Biomass measured with OD and NIR, biomass dry weight and GFP

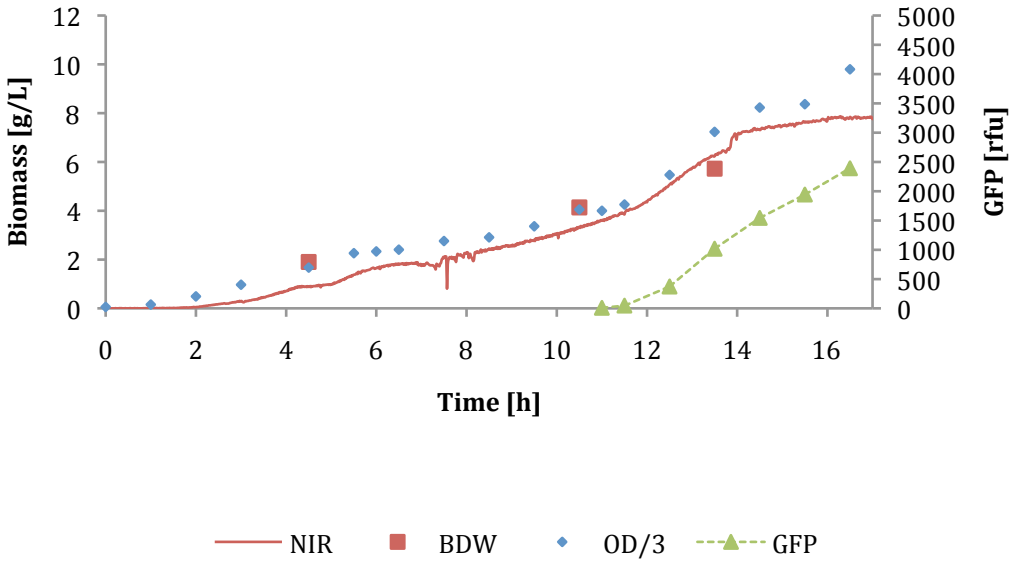

Figure 15 Cultivation C: Biomass measured with OD and NIR, biomass dry weight and GFP

The curves of glucose and acetate shown in Figures 16 and 17 are quite similar. During batch the cells consume glucose and produce acetate. After the batch glucose is consumed, the cells start to consume acetate and the small amount of glucose, which is fed. The two graphs show that there is no glucose in the broth for about 6 to 7 hours. The fed glucose is consumed right away and probably below the amount, which the cells would need. Feeding so little can be a problem

for the cell metabolism and growth. Bacterial cells need some time to adapt to the carbon source. In this case they get little glucose and consume acetate beside, which leads to delayed growth.

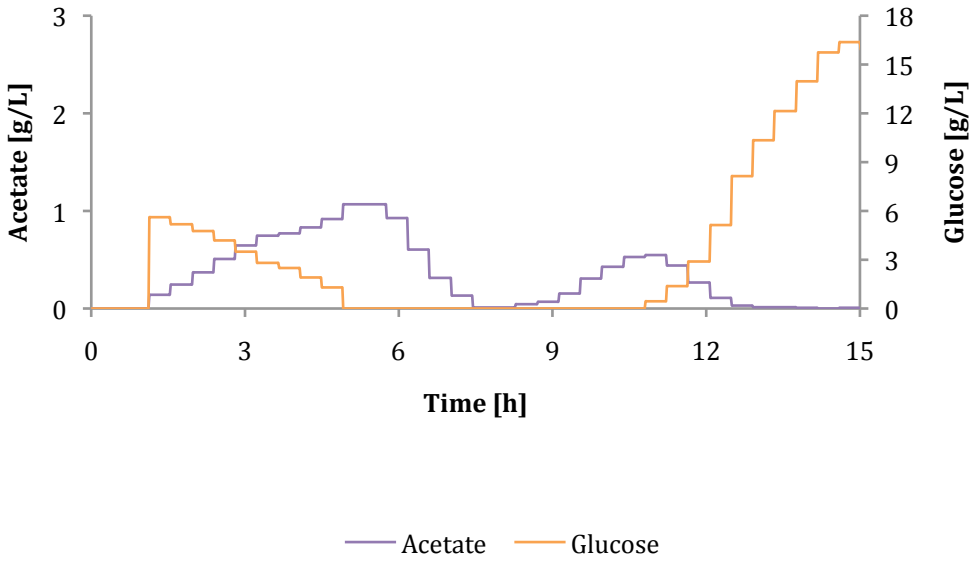

Figure 16 Cultivation B: HPLC data for acetate and glucose

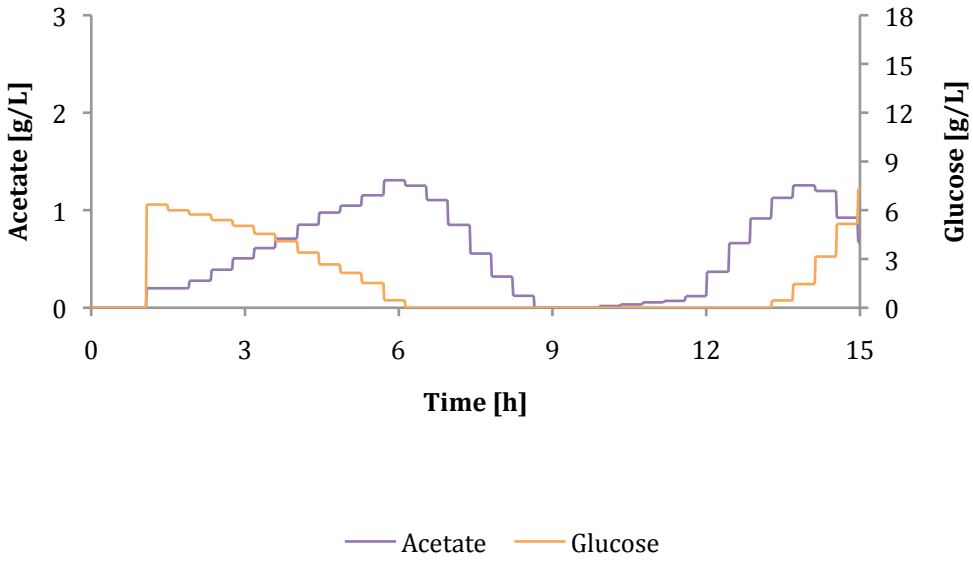

Figure 17 Cultivation C: HPLC data for acetate and glucose

For cultivation B, the value for the proportional band within the first 1.5 feeding hours was increased from 200 to 900 while the feeding of cultivation C had a proportional band of 900 from the beginning.

Comparing the two feeding rates shown in Figures 18 and 19 to each other, one can see that cultivation B starts with a low feed rate while in cultivation C the feed rate is zero for the first 1.5 hours. The reason for that is the different proportional band in the beginning and that the used strain can grow on acetate and produce carbon dioxide without consuming glucose.

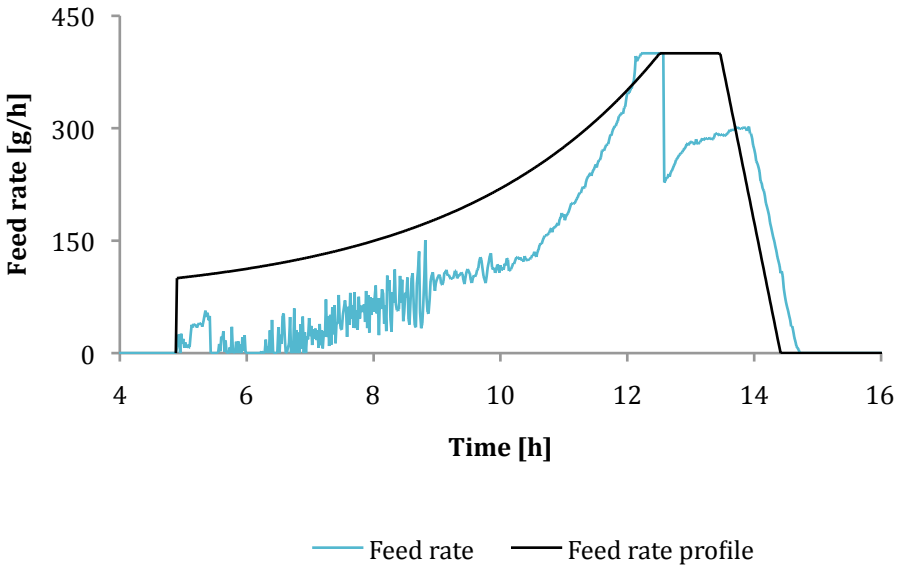

Figure 18 Cultivation B: Feed rate used and feed rate from profile

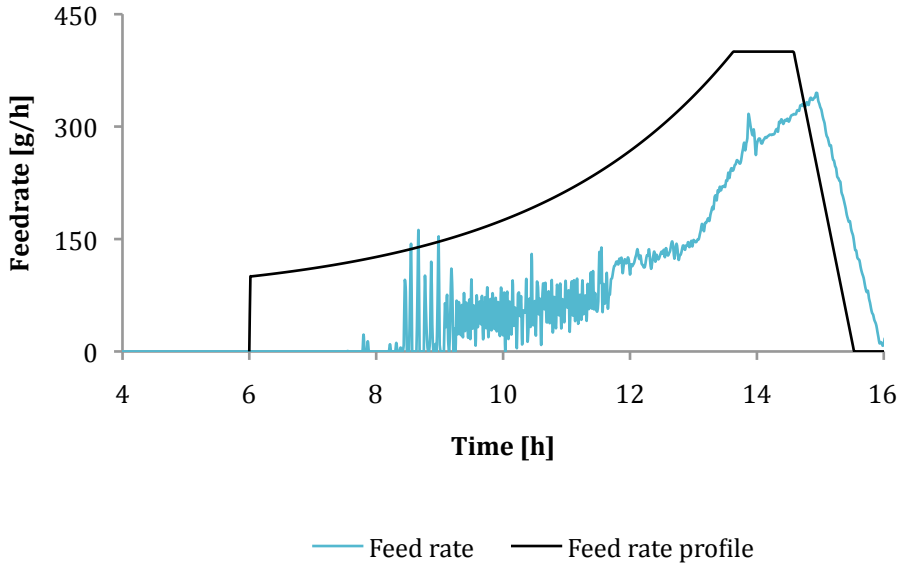

Figure 19 Cultivation C: Feed rate used and feed rate from profile

In both cultivations it needed some time until the set value of 0.2 g/g/h specific carbon dioxide evolution rate was nearly reached. As shown in Figures 20 and 21 the proportional controller was not able to reach exactly 0.2 g/g/h but stayed constant at 0.25.

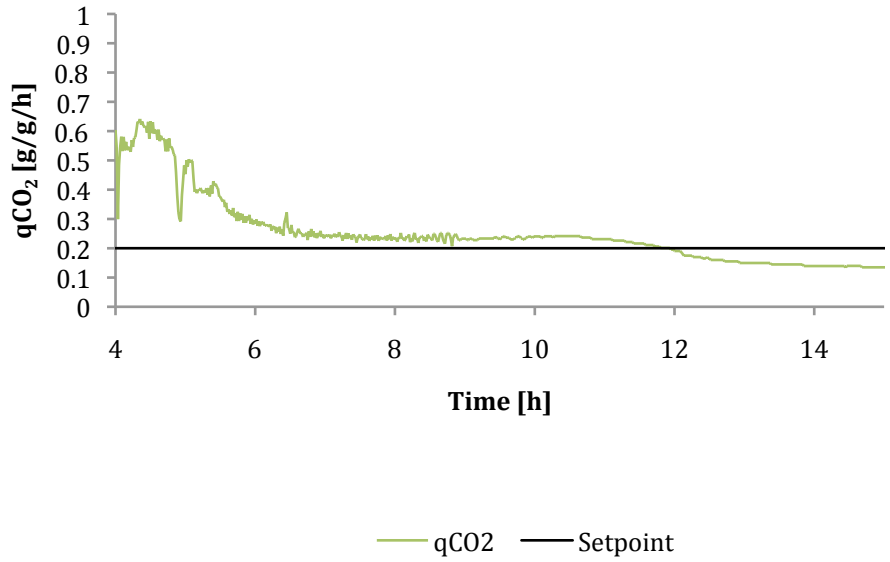

Figure 20 Cultivation B: qCO_2 calculated from the soft sensors and setpoint

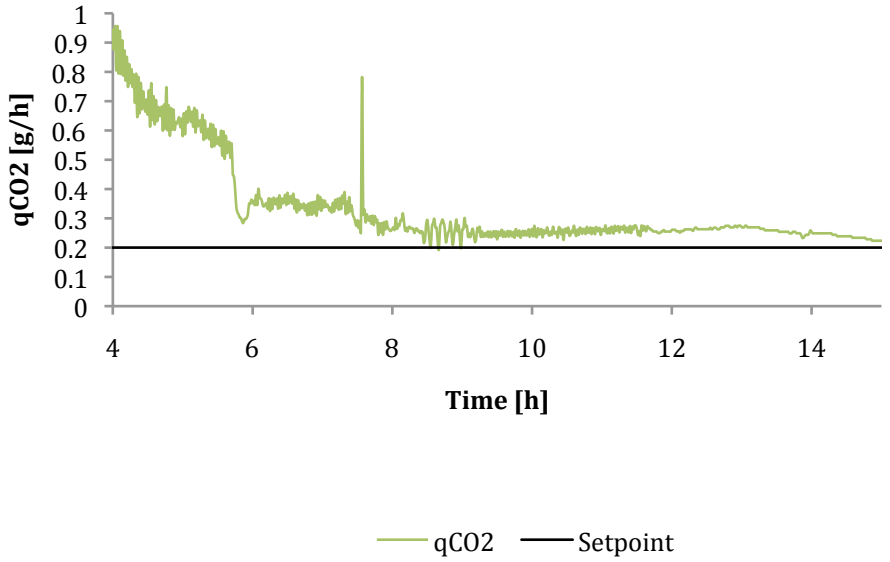

Figure 21 Cultivation C: qCO_2 calculated from the soft sensor and setpoint

Because of the low setpoint and the low feeding rates, the bacterial growth was slower than in cultivation A and led to a later induction and overall to less GFP in comparable time intervals.

4.4 Cultivation D

In order to find out if a higher setpoint had a negative influence on cell growth and acetate concentration, a setpoint of 0.5 g/g/h was tested.

The pre-culture had an OD of 4.6 and the feed was started after 5.5 hours. Due to some changes in the system, the NIR probe signal was slightly higher than in cultivation B and C. It corresponded more to the results of the OD measurements. In addition to that the gas analyzer was changed and this had also an influence on the results.

As shown in Figure 22, the biomass reached a concentration of 9.1 g/L. The culture was induced after 7.5 hours and the amount of GFP increased to nearly 4700 rfu. Compared to the amount of GFP produced in cultivation A it is an increase of about 15 %.

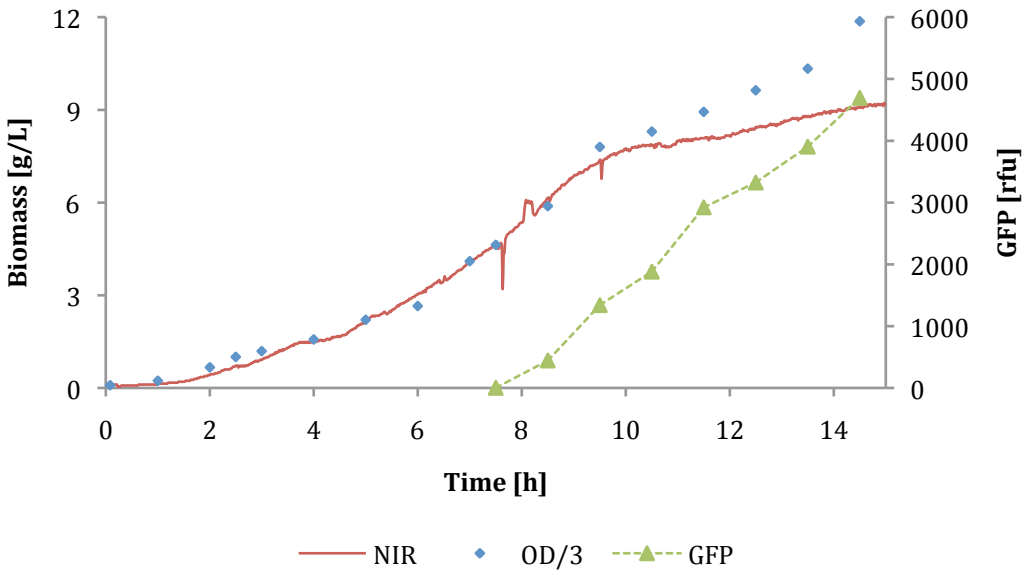

Figure 22 Cultivation D: Biomass measured with OD and NIR and GFP

The HPLC showed quite interesting results (see Figure 23). Because of the high feed rate glucose started to accumulate already after 7.5 hours which is 2.5 hours earlier than in cultivation A.

The high glucose concentration of 19 g/L started to decline after the 2 L of feed media were pumped into the reactor and no more fresh glucose was provided. The acetate increased over the whole cultivation up to 8 g/L.

Despite the high concentration of acetate more biomass and GFP was produced than in cultivation A. This means that the used strain can grow and produce recombinant proteins, although acetate concentration in the culture medium is high.

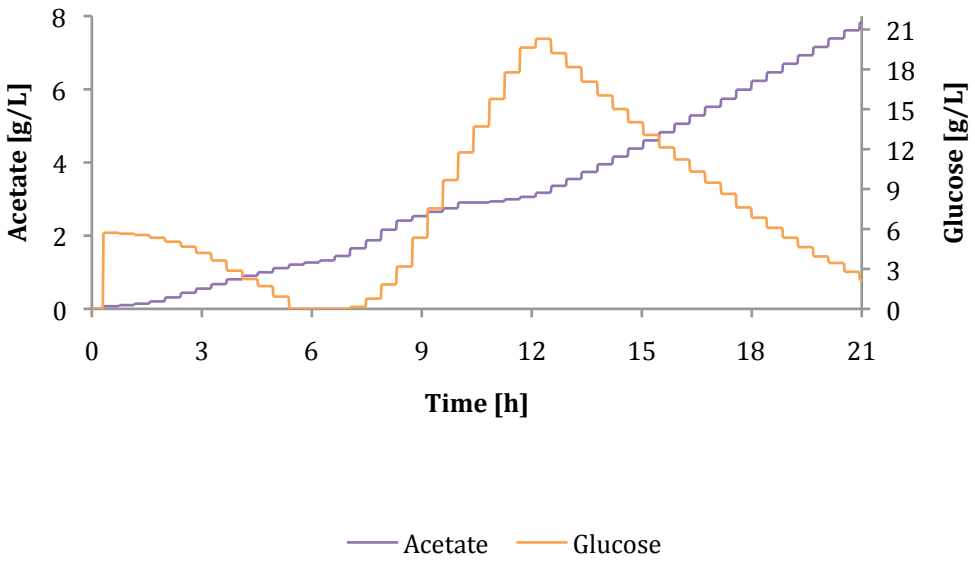

Figure 23 Cultivation D: HPLC data for acetate and glucose

As shown in Figure 24 the feed rate increased for a period of 2.5 hours to its maximum and continued with 400 g/h until the feed media was finished.

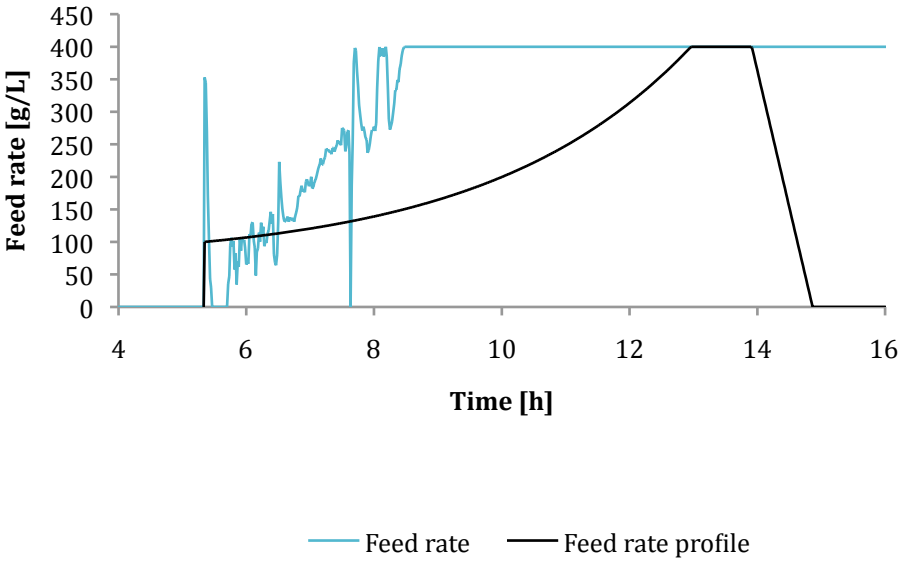

Figure 24 Cultivation D: Feed rate used and feed rate from profile

Even though the feed rate was so high and glucose was already accumulating, it was impossible to keep the qCO_2 level at 0.5 g/g/h after induction at hour 7.5 (see Figure 25).

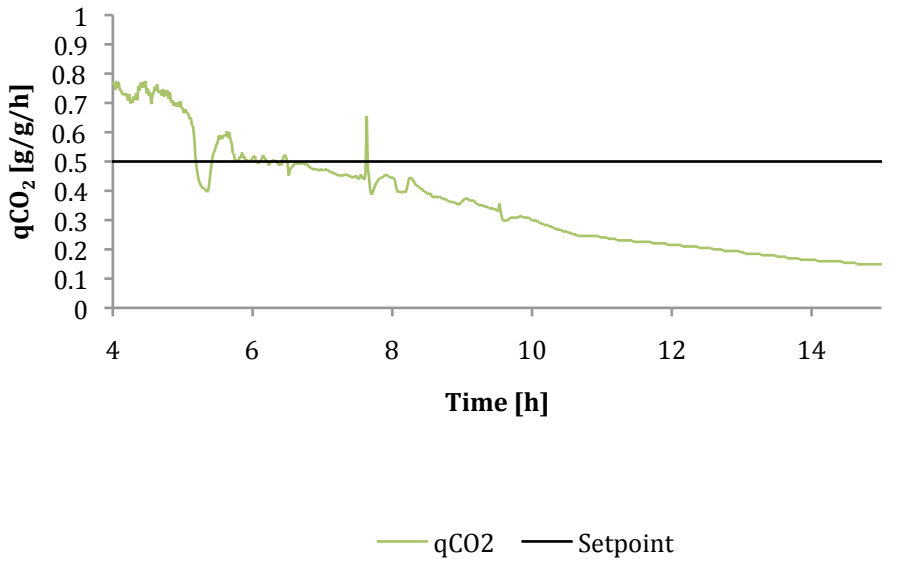

Figure 25 Cultivation D: qCO_2 calculated from the soft sensor and setpoint

This cultivation showed that the used strain is not sensitive to acetate concentrations of about 8 g/L and that the induction had a big influence on the carbon dioxide production.

4.5 Cultivation E

In a final cultivation it was tested if a specific carbon dioxide rate of 0.35 could increase biomass and GFP production. The last cultivation was started with a pre-culture of OD 8.82. The results in Figure 26 show the highest biomass concentration at 9.5 g/L, which is similar to cultivation D. The amount of produced GFP is as high as in cultivation A where just the predetermined feed profile was used.

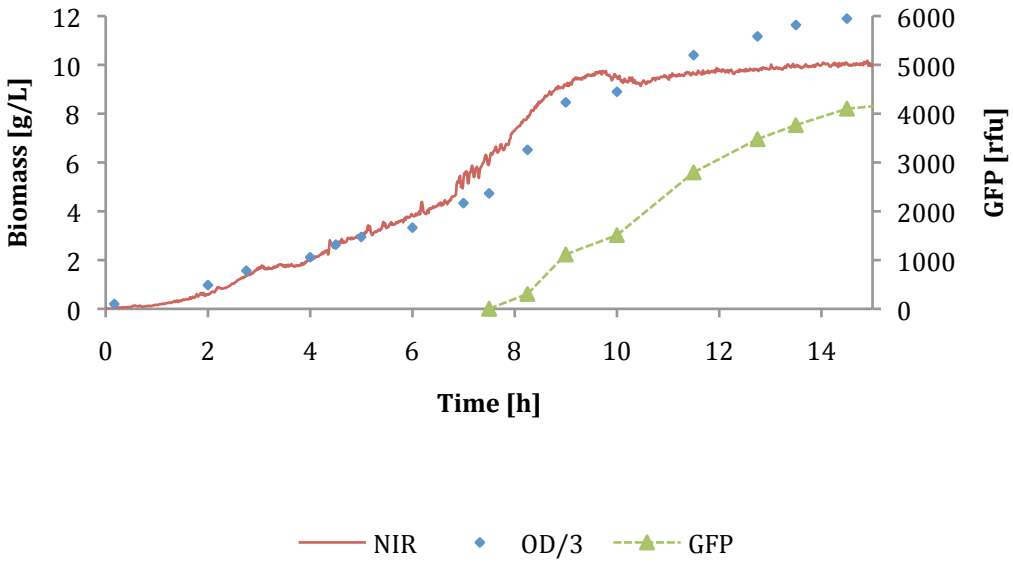

Figure 26 Cultivation E: Biomass measured with OD and NIR and GFP

The feed was started after 4 hours. The glucose concentration in Figure 27 was zero until hour 9 while the acetate concentration decreased after the batch phase. That means, that the cells did not get enough glucose and consumed acetate beside. At hour 7.5 IPTG was added. The bacteria grew slower and the

acetate did not decrease anymore. It seems that the fed glucose from hour 7.5 to hour 8.5 is exactly the amount the cells needed.

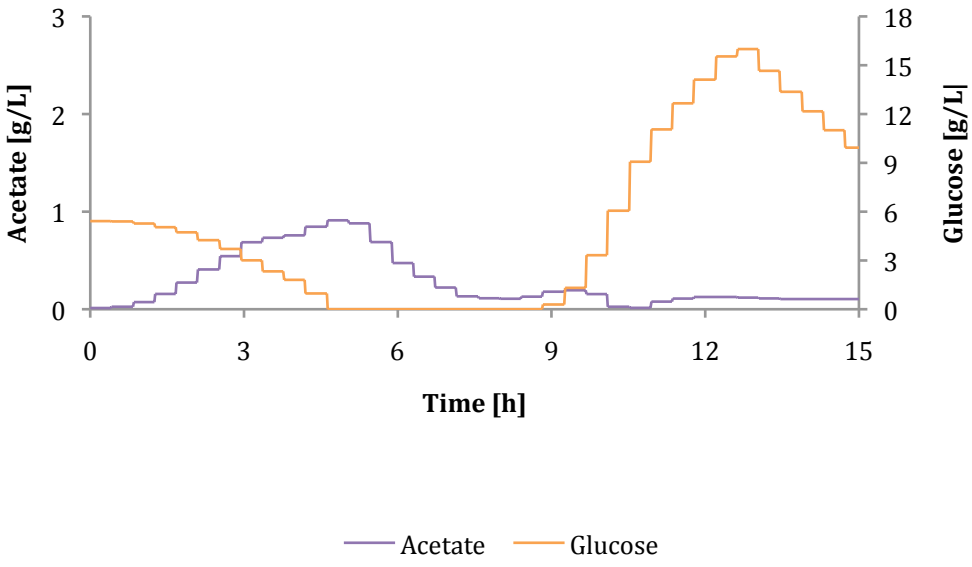

Figure 27 Cultivation E: HPLC data for acetate and glucose

Figure 28 shows that the feed rate for the first half feeding hour was zero. There was no need to feed glucose because the cells were consuming acetate and hence producing enough carbon dioxide.

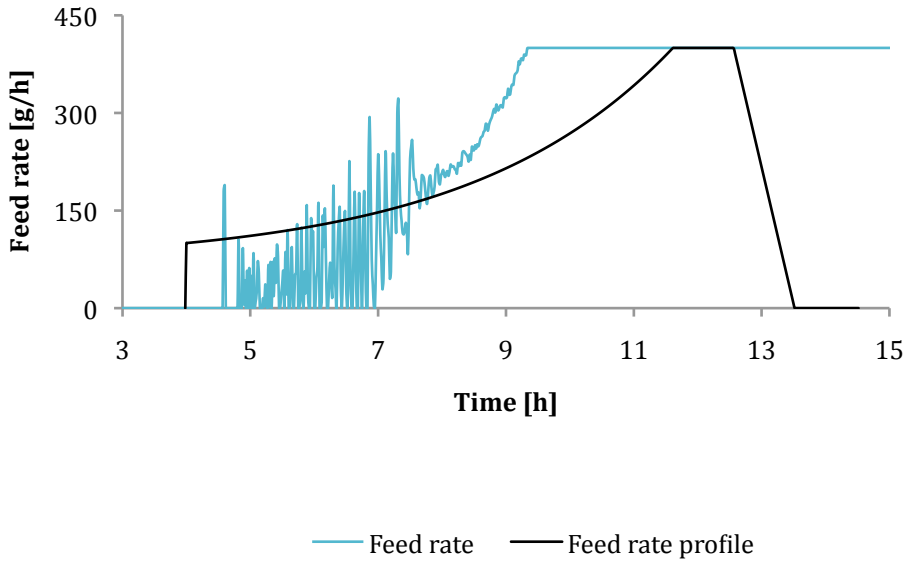

Figure 28 Cultivation E: Feed rate used and feed rate from profile

The results of the measured specific carbon dioxide rate shown in Figure 29 are oscillating. The reason is as follows:

The solubility of gases in stirred vessels depends on the gas flow and the stirrer speed. The higher the stirrer speed, the smaller are the bubbles and the slower is the bubble rise velocity. If the speed is too low and the gas flow too high, the distribution of the gas will be poor (Doran 1995).

Figure 30 shows the changes in the stirrer speed during cultivation E. This changes affected beside the DOT also the carbon dioxide measurements with the gas analyzer. As the stirrer speed controls the DOT, the settings for this PID regulation could be improved to get smaller responses.

The stirring had an influence on the results of carbon dioxide measurements in all cultivations. Usually it became more stable with higher biomass concentrations.

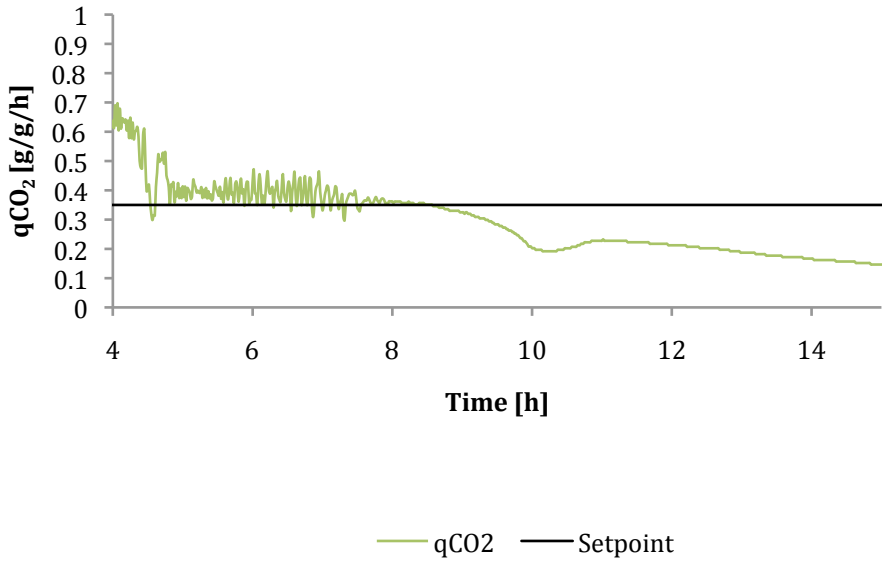

Figure 29 Cultivation E: qCO_2 calculated from the soft sensor and setpoint

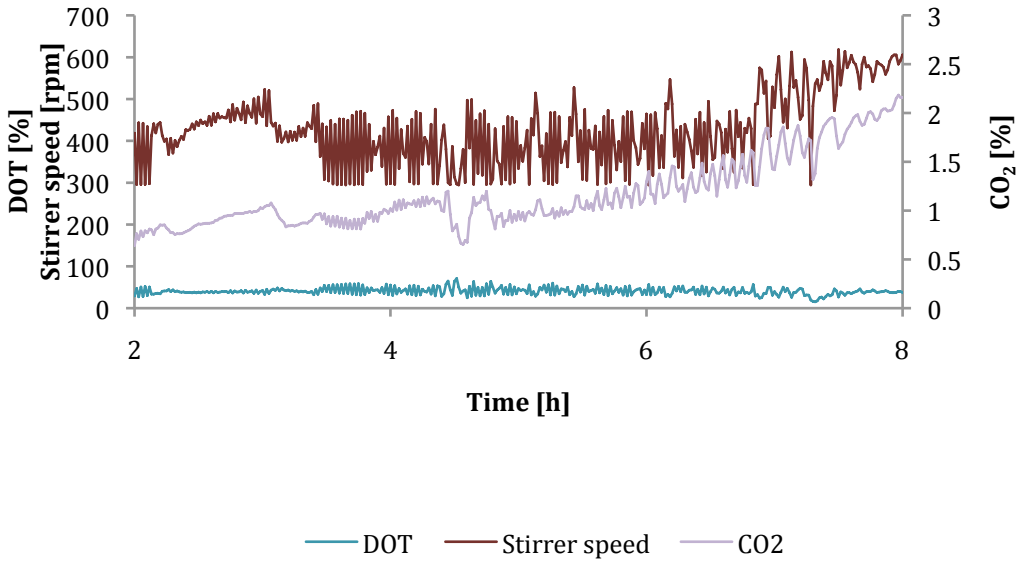

Figure 30 Cultivation E: Stirrer noise

4.6 Summary of results

In Table 4 the results of this study are summarized. An important condition for a time efficient cultivation was the OD value of the pre-culture. The lower OD values of cultivation C and D resulted in the two latest feed starts. As slow feeding resulted in slow growth, cultivation B and C were not very economic. Because of the low setpoint they produced less biomass and GFP than the others did in 14 hours.

The cultivation with the highest acetate concentration (D) produced the highest rfu of GFP after 14 hours. But anyways it was not possible to increase the amount of produced GFP significantly more compared to the feed profile (A).

Table 4 Summary of results

Key parameters	Cultivation				
	A	B	C	D	E
qCO_2 setpoint	-	0.2	0.2	0.5	0.35
OD pre-culture	11.68	7.24	2.74	4.6	8.82
Highest amount of GFP [rfu]	4183	2073	2393	4694	4103
Feed start hour	4.75	5	6	5.5	4
Induction hour	6.5	7.25	11	7.5	7.5
Highest acetate concentration [g/L]	1.83	1.06	1.25	7.83	0.88

5 Conclusion

5.1 General conclusions of the work

The experiments have shown that the influence of acetate concentration on growth and protein production was quite low for the used strain. Nevertheless it is possible to use the specific carbon dioxide evolution rate to control feeding and to keep the acetate level low. One advantage in using carbon dioxide results is that they can be produced quickly and continuously. Furthermore it is cheap to use infrared spectroscopy because most sensors have a long lifetime and a high stability.

The results also show that a predetermined feed profile can keep the whole process very simple but still result in a high biomass concentration and a high amount of GFP.

One thing that should be kept in mind is that using more than one hardware sensor to estimate one soft sensor makes the application more difficult. The hardware probes have to give reproducible and reliable results and if just one of them works in a proper way, the result of the soft sensor will be useless.

5.2 Why not to use carbon dioxide to control the feeding?

Using just the produced amount of carbon dioxide for the feed control would result in over- or underfeeding.

A high setpoint for the produced carbon dioxide would result in a very high feed rate in the beginning of the feed. As the cell concentration at this time of the process would not be high enough to produce a very high amount of carbon dioxide the culture would be overfed.

Another option would be, to set the carbon dioxide level low. But in this case the feeding would start slow, because the entered setpoint could be reached easily. After a while there would be a higher biomass concentration so that the fed glucose would be consumed immediately. Due to the low setpoint the amount of fed glucose would be too low to feed all cells (underfeeding).

5.3 Suggestions for future work

To refine the application of this soft sensor it would be an option to use it on different strains. Strains that do not reassimilate acetate and that are affected in their growth by higher acetate concentrations.

If a strain does not grow on acetate, the feeding starts immediately after the batch phase. The delay, which occurs because of acetate consumption, could be avoided. It could also be shown with strains that are inhibited in their growth by acetate that the controller can improve cell growth and protein production with keeping acetate concentration low.

Other objectives would be to implement this feed forward control for larger scale recombinant protein productions or to use it as an alternative to a more cost intensive on-line HPLC controlled feeding.

6 Acknowledgements

First of all I would like to thank my supervisor Carl-Fredrik Mandenius for giving me the opportunity to do my master thesis in Sweden, for his help and his suggestions on my thesis.

I would also like to thank the work group technical biology at University of Linköping for their hospitality, their support and for making my stay so pleasant.

Der meiste Dank gebührt meiner Familie - meinen Eltern, meinem Mann und meinen Brüdern - für die grenzenlose Unterstützung vor und während meines Studiums und dafür, dass ihr immer für mich da seid.

7 References

Antranikian G, *Angewandte Mikrobiologie.* Springer Verlag, 2006.

Bachmann B J, "Pedigrees of some mutant strains of Escherichia coli K-12." *Bacteriology Reviews* 36 (1972): 525-557.

Blattner F R, Plunkett G 3rd, Bloch C A, Perna N T, Burland V, Riley M, Collado-Vides J, Glasner J D, Rode C K, Mayhew G F, Gregor J, Davis N W, Kirkpatrick H A, Goeden M A, Rose D J, Mau B, Shao Y, "The Complete Genome Sequence of Escherichia coli K-12." *Science* 277 (1997): 1453-1462.

Chalfie M, Tu Y, Euskirchen G., Ward W W, and Prasher D C. "Green fluorescent protein as a marker for gene expression." *Science* 263 (1994): 802-805.

Chang Y Y, Wang A Y, and Cronan Jr J E, "Expression of Escherichia coli pyruvate oxidase (Poxb) depends on the sigma factor encoded by the rpoS (hatF) gene." *Molecular Mikrobiology*, 1994: 1019-1028.

Chéruy A, "Software sensors in bioprocess engineering." *Journal of Biotechnology* 52 (1997): 193-199.

Chmiel H, *Bioprozesstechnik.* Vol. 3. Heidelberg: Springer, 2011.

Clark D P, Pazdernik N J, *Molekulare Biotechnologie.* Heidelberg: Spektrum, 2009.

Doran P M, *Bioprocess Engineering Principles.* London: Academic Press Limited, 1995.

Feng P, Weagent S D, Grant M A. "Enumeration of Escherichia coli and the Coliform Bacteria." *Bacteriological Analytical Manual*, September 2002: Chapter 4.

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

Gustavsson R, "On-line control of glucose feeding in an Escherichia coli fed-batch cultivation expressing a recombinant protein." Master thesis, Linköpings universitet (2011)

Heim R, Prasher D C, Tsien R Y, "Wavelength mutations and posttranslational autoxidation of green fluorescent protein." *Biochemistry* 91 (1994): 12501-12504.

Kessler D, and Knappe J, "Anaerobic dissimilation of pyruvate." *Cellular and Molecular Biology*, 1996: 199–205.

Kotte O, Zaugg J B, Heinemann M, "Bacterial adaption through distributed sensing of metabolic fluxes." *Molecular Systems Biology 6*, 2010: 355.

Luli G W, Strohl W R, "Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed batch fermentations." *Applied and environmental microbiology* 56 (1990): 1004-1011.

Nemecek S, Marisch K, Juric R, Bayer K, "Design of transcriptional fusions of stress sensitive promoters and GFP to monitor the overburden of Escherichia coli hosts during recombinant protein production." *Bioprocess and Biosystems Engineering* 31 (2007): 47-53.

Petsch D, Anspach F B, "Endotoxin removal from protein solutions." *Journal of Biotechnology* 76 (2000): 97-119.

Pommerville J C, *Alcamo's Fundamentals of Microbiology.* 9th Edition. Jones & Bartlett Publishers, 2010.

Prasher D C, Eckenrode V K, Ward W W, Prendergast F G, and Cormier M J. "Primary structure of the Aequorea victoria green-fluorescent protein." *Gene* 111 (1992): 229-233.

Reischer H, Schotola I, Striedner G, Pötschacher F, Bayer K, "Evaluation of the GFP signal and its aptitude for novel on-line monitoring strategies of recombinant fermentation processes." *Journal of Biotechnology* 108 (2004): 115-125.

Riesenberg D, Schulz V, Knorre W A, Pohl H D, Korz D, Sanders E A, Roß A, Deckwer W D, "High cell density cultivation of Escherichia coli at controlled specific growth rate." *Journal of biotechnology* 20 (1991): 17-27.

Santillan M, Mackey M C, "Quantitative approaches to the study of bistability in the lac operon of Escherichia coli." *J. R. Soc. Interface* 5 (2008): 29-39.

Schlegel H G, *Allgemeine Mikrobiologie.* Stuttgart: Thieme Verlag, 1992.

Shimomura, O., F. Johnson, and Y. Saiga. "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan Aequorea." *Journal of Cellular and Comparative Physiology* 59 (1962): 223-239.

Sundström H, Enfors S O, "Software sensors for fermentation process." *Bioprocess and Biosystems Engineering* 31 (2008): 145-152.

Warth B, Rajkai G, Mandenius C-F. "Evaluation of software sensors for on-line estimation of culture conditions in an Escherichia coli cultivation expressing a recombinant protein." *Journal of Biotechnology* 147 (2010): 37-45.

Welschof M, Krauss J, "Recombinant Antibodies for Cancer Therapy." *Methods in Molecular Biology* 207 (2003): 438.

Xu B, Jahic M, Blomsten G, Enfors S-O, "Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with Escherichia coli." *Applied Microbiology and Biotechnology* 51 (1999): 564-571.

Yang F, Moss L G, Phillips Jr G N, "The molecular structure of green fluorescent protein." *Nature Biotechnology* 14 (1996): 1246-1251.

Zanzotto A, Boccazzi P, Gorret N, Van Dyk T K, Sinskey A J, Jensen K F, "In Situ Measurement of Bioluminescence and Fluorescence in an Integrated Microbioreactor." *Biotechnology and Bioengineering* 93 (2005): 40-47.