Design and application of software sensors for monitoring of fed-batch cultivations containing recombinant *Escherichia coli*

CYRUS DJAHEDEI

Department of Physics, Chemistry and Biology
August 2010
LITH-IFM-A-EX-10/2366
MASTER THESIS

Design and application of software sensors for monitoring of fed-batch cultivations containing recombinant *Escherichia coli*

Cyrus Djahedi

Supervisor: Prof. Carl-Fredrik Mandenius, Linköping University

August 2010
LITH-IFM-A-EX-10/2366
Abstract

Bioprocesses are industrial processes that use living cells to produce desired products. These processes are complex and require advanced monitoring and control in order to reach their highest potential. Software sensors are a potent tool for real time monitoring and control of bioprocesses. In this project, three novel software sensors were designed and tested in three fed-batch cultivations containing recombinant *Escherichia coli* producing green fluorescence protein (GFP). Special attention was given to the sensors for reacting to any stress triggered when production of the recombinant protein was initiated by an inducer molecule.

The sensors calculated the oxygen consumption per substrate, carbon emission per substrate and the sum of the specific growth rates of various acids. Their mathematical algorithms were derived from already existing software sensors and programmed on to computer software designed for monitoring and control.

The results showed that all three sensors gave stable signals as expected. Unfortunately they did not provide any useful information regarding the recombinant protein expression in this particular study. Complications such as measurement errors and incorrect system configurations also created difficulties. It is clear however, that the sensors work well and should be highly considered for further studies.
Preface

This report describes my master thesis project, in the Engineering Biology program at Linköping University, about the design and application of new software sensors for monitoring of fed-batch cultivations with recombinant *Escherichia coli*.

The experimental work was carried out at the Division of Biotechnology, IFM, Linköping University, during Spring 2010.

The project was a continuation of a previous master thesis project carried out by Benedikt Warth (2008, LiTH-IFM-A-EX-08/2013-SE) where six other new software sensors were evaluated for their ability to provide information of what impact physiological stress may cause in recombinant *E. coli* fermentations as well as improving bioprocess monitoring and control.

In this project the use of the software sensor methods were further developed and refined.
Index

1. Thesis objective .............................................................................................................. 1

2. Introduction .................................................................................................................... 2
   2.1 Bioprocesses need efficient monitoring and control ......................................................... 2
   2.2 The software sensor concept .............................................................................................. 3
   2.3 Using software sensors for monitoring and control of bioprocesses ............................... 3
   2.4 Some obstacles that must be overcome when using software sensors ............................ 6
   2.5 The production microorganism - Escherichia coli ........................................................... 7
   2.6 The product from the bioprocess - Green Fluorescence Protein .................................... 7
   2.7 The expression mechanism in E. coli ................................................................................. 8

3. Materials and methods ................................................................................................. 11
   3.1 Strain .................................................................................................................................. 11
   3.2 Medium .............................................................................................................................. 11
   3.3 Analytical instrumentation ............................................................................................... 12
   3.4 Cultivation protocol .......................................................................................................... 13
   3.5 On-line hardware sensors ................................................................................................. 15
   3.6 Off-line measurements ...................................................................................................... 16

4. Modeling and theoretical aspects ................................................................................ 18
   4.1 Nomenclature and definition of parameters ................................................................... 18
   4.2 Estimation of biomass ....................................................................................................... 19
   4.3 Specific glucose uptake rate ............................................................................................. 19
   4.4 Specific acetate production rate ....................................................................................... 20
   4.5 Specific oxygen uptake rate .............................................................................................. 20
   4.6 Specific carbon dioxide production rate ......................................................................... 21
   4.7 Oxygen consumption per substrate ................................................................................. 21
   4.8 Carbon dioxide emission per substrate ........................................................................... 21
   4.9 Specific sum mixed acid production rate ........................................................................ 22

5. Results and Discussion ................................................................................................ 23
   5.1 Biomass growth estimation by a NIR-probe in a software sensor ................................. 24
   5.2 The software sensors for the ratios between the O₂ uptake or the CO₂ evolution rates and the substrate uptake rate ................................................................. 25
      5.2.1 Ratio of specific O₂ uptake rate and specific substrate uptake rate ........................... 31
      5.2.2 Ratio of the CO₂ evolution rate and substrate uptake rate ........................................ 36
   5.3 The software sensor for specific sum mixed acid production rate ................................ 39
1. Thesis objective

The objective of this master thesis project has been to investigate new software sensors for monitoring of fed-batch cultivations producing a recombinant protein. These sensors shall estimate the specific oxygen uptake rate, the specific carbon uptake rate, the ratios between these and specific glucose uptake rate, and specific production of organic acids from mixed acid fermentation.

The new software sensors shall be based on a previous project carried out by Benedikt Warth (2008, LiTH-IFM-A-EX-08/2013-SE) where six other software sensors were evaluated for their ability to provide information of what impact physiological stress may cause in recombinant *E. coli* fermentations as well as improving bioprocess monitoring and control.

The objective presupposed that the new software sensors were to be programmed and applied in operating software and tested in several test fed-batch fermentations prior to being evaluated in three validation fed-batch fermentations. The recombinant protein to be used was the green fluorescence protein (GFP) under control of the T7 vector. The sensor and measurement methods were typical standard bioreactor measurement devices (SBM), on-line high performance liquid chromatography (HPLC), near-infrared (NIR)-monitoring, gas analyzers for effluent oxygen and carbon dioxide, and off-line determination of GFP fluorescence and biomass dry weight.
2. Introduction

2.1 Bioprocesses need efficient monitoring and control

Reliable and accurate measuring, monitoring and control are prerequisites for efficient bioprocessing. When producing recombinant proteins it is essential to collect information about the course of the cultivation process, and the changing metabolic and physiological states of the production organism. When done correctly this can reduce expenses, save production time and resources, as well as increase yields and desired product quantities.

In this thesis project a recombinant *Escherichia coli* was used to produce Green Fluorescence Protein (GFP). Producing the protein GFP using recombinant *E. coli* bacteria cells is a well studied process (Li et al, 2002; Reischer et al, 2004; Nemecek et al, 2008; Kensy et al, 2010). It is clear however, that present GFP production processes are still far from being optimized to their full potential. Changes in the cell composition and specific interactions of the recombinant protein with the host cell proteins in *E. coli* are poorly understood. Several authors state that monitoring and control systems for bioreactors require further development and refinement in combination with better evaluation techniques, such as proposed for Process Analytical Technology in order to reach the necessary understanding of the organisms in the production process (Mandenius, 2004; Mandenius, 2006; Gnoth et al, 2008). In particular, this concerns existing estimation and control methods.

*E. coli* cells can be grown very easily in simple media and their genetic characteristics have been determined. Because of this they have been found to be effective hosts for the preparation of biological polymers, including polypeptide hormones, proteins and carbohydrates. By incorporating the genetic information required into the *E. coli* genome, it is simple to produce these products in large amounts. It should be noted however, that *E. coli* bacterium is very sensitive to changes or interferences during the production processes, which requires cultivations to be carried out under very stable conditions. This implies that the monitoring and control must be as good as possible since the loss of a batch could result in delivery delays and financial losses. In addition, only a few on-line measurements are available and these can be expensive to maintain as well as increase the contamination risk.

*Software sensors* provide a very simple and cost efficient way of generating new information regarding the metabolic and physiological state of cultivations producing bioproducts (Warth et al, 2010). What makes these sensors so practical is that they can provide additional information about a typical pharmaceutical production process without requiring any additional instruments and thus sparing additional expenses (Dochain, 2003). PAT and Good Manufacturing Practice (GMP) related issues such installation, qualification and validation of software sensors is facilitated by simple procedures that only require a computer with the given software. Furthermore, it becomes less expensive because of easier installations, calibrations, and service.
The repeatability of industrial bioprocesses is another important need for efficient manufacture (PAT, 2004, Mandenius et al, 2009). Especially the protein production processes lack this quality (Gnoth et al, 2008). Here, software sensors can play a significant role if they can cope with the demanding conditions in the large-scale cultures. The software sensors have perhaps a better chance to include the many influences that affect the course of a production process for food, pharmaceuticals and other biotechnology products.

### 2.2 The software sensor concept

A software sensor is a mathematical algorithm that uses on-line data to calculate an immeasurable variable or kinetic parameter that of significance to the process (Cheruy et al, 1997). The term is derived from combining a sensor (hardware) and an estimation algorithm (software). Software sensors can be extremely useful for real time monitoring, decision making and control. They are applied in many industrial areas such as reactor units in the chemical industry (Régnier et al, 1996), analytical instrumentations (Matsumura et al, 1998), environmental processes (Carstensen et al, 1996), and biotechnology processes (Kiviharju et al, 2008).

![Figure 1 General principle of a software sensor. Software sensor estimates can be used in estimation algorithms to estimate other software sensors.](image)

### 2.3 Using software sensors for monitoring and control of bioprocesses

In a software sensor, the sensor has to provide reliable measurements informative enough to deduce unmeasured process variables and parameters. It is also important that the estimation algorithm is designed in order to produce on-line estimations surely and quickly that are converging towards the real values of the unmeasured variables (Cheruy et al, 1997). Modern powerful software sensors utilize advanced mathematical models to establish these relations.
In biotechnology processes special attention is devoted to such software sensor solutions that allow estimation of biomass growth, effluent of growth dependent gases from the bioreactor and uptake of nutrients and other metabolites related to the cell mass formation.

**Software sensors for estimation of biomass growth**

The growth of biomass, volumetrically as well as specifically, is one of the most important parameters in biotechnological processes and therefore it is fundamental to have a reliable on-line estimation of it (Iversen-Lönsmann et al, 1994; Ödman et al, 2009, Warth et al, 2010). Since it is of such high priority, most of the presented software sensors have been applied to this field and the technologies that have been developed for measuring the biomass *in-situ* have evolved drastically over the years (Hoffmann et al, 2000; Sundström and Enfors, 2008; Kiviharju et al, 2008). A literature search shows that many methods that were designed over the years have ended up as mere prototypes or just one-time applications. It also presents more recent developments in the field which offer more practical and economic ways of estimating the biomass concentration. The article concludes with stating that modern biomass measurements that have been developed into probe technologies are mainly dielectric spectroscopy and various optical methods and that they have advantages and disadvantages over each other.

**Software sensors for biomass based on analyzing outlet oxygen and carbon dioxide**

In biomass estimation, the most popular correlation method in use today is the off-gas analysis. The biomass is related to the analyzed amounts of oxygen and carbon dioxide in the off-gas using a variety of model calculations as shown by several groups for *E. coli* (Simutis et al, 1993; Claes & Van Impe, 2000, Kviharju et al, 2008). The methodology is close to the software sensor concept as has been suggested by some authors.

In this thesis project on-line gas analysis of O₂ and CO₂ is also used connected to the software sensor concept. Indirectly is also relates to biomass as will be described herein.

**Software sensors for biomass based on pH**

Another way of estimating biomass concentration is through the chemical properties of the microorganisms. Measurements taken of the chemical properties of the medium such as pH, conductivity and the demand of pH control agents have been useful for that purpose. Acuna et al (1994) evaluated on-line estimations of biological variables during pH controlled lactic acid fermentations.

Also for other organisms similar methods have been applied. Wang, Lee and Chang (1998) describe on-line state estimations of biomass based on acid production in *Zymomonas mobilis* cultures.
Hoffman et al (2000) presents a similar method for simultaneous on-line estimation of biomass and acetate from base consumption and conductivity measurements, again in *E. coli* cultures. This is a proven and effective way of estimating the biomass provided that ammonia is the only nitrogen source available in the reactor.


**Software sensors for estimation of nutrient substrates and metabolites**

As cell cultures in bioprocesses grow, they consume nutrient substrates such as glucose and produce side-metabolites such as ethanol and acetate. Keeping track of these concentrations is critical for monitoring and eventually, if needed, for improving the conditions to ensure better growth. An example of this is demonstrated in a multi-analyzer monitoring and control system designed specifically for bioprocesses (Cimander et al, 2003). The system includes an on-line HPLC system with filtration probe and computer software. The HPLC monitors the medium components by automatically taking samples every 20 min and analyzing them using the expert system software.

In this thesis project the on-line HPLC software sensor approach was adapted to glucose, acetate and side-metabolites assumed to be formed as a result of mixed-acid fermentation in *E. coli* (figure 2), to combine these estimates with uptake rates and various ratios added new software sensor opportunities.

![Central metabolic network of E. coli metabolism](http://www.microbialcellfactories.com/content/8/1/54/figure/F102/08/10)

**Figure 2** The central metabolic network of *E. coli* metabolism (retrieved at [http://www.microbialcellfactories.com/content/8/1/54/figure/F1](http://www.microbialcellfactories.com/content/8/1/54/figure/F1) 02/08/10).
Software sensor estimates for recombinant protein production

More difficult is to set up software sensors for protein production. However, the physiological stress caused by the overproduction of heterologous proteins and plasmid vectors and which is referred to as metabolic burden, may provide an opportunity. Furthermore, adapting the expression system to the metabolic burden is an essential step for optimizing bioprocesses and achieving high productivity. To add complexity however, strong overproduction may actually stun cell growth, while moderate stress on the other hand may result in increased protein production (Bentley et al., 1990; Austin et al., 1998). This may indirectly indicate protein production. A study carried out by Bachinger & Mandenius (2001) showed that chemical gas sensor arrays, also known as electronic noses, can be use for monitoring the effects of metabolic burden in recombinant bioprocesses. A more recent article by Sundström & Enfors (2008) presents a new software sensor known as $R_{O/S}$. The sensor calculates the ratio between oxygen and energy substrate consumption, thereby providing useful information about the energy metabolism, which in turn can be correlated with stress influences.

2.4 Some obstacles that must be overcome when using software sensors

Software sensors provide additional opportunities for on-line monitoring owing to their ability to extract and refine data from hardware. However, there are important obstacles that must be overcome:

- It is necessary to access on-line sensors which can meet the requirements of the harsh conditions in a bioreactor. These are at shortage.
- The sensors may not be able to exhibit long term stability.
- The sensors may be too expensive to acquire.
- The sensors may be too labor-demanding to maintain. This becomes obvious when on-line sensors fail to pass operation and process qualifications (OQ/PQ) and GMP validation (Warth et al., 2010).
- The validity of the parameters used in the algorithm calculations decides the quality of the obtained estimates (as in all models).
- If too complex models are used the bigger the error amplification and uncertainty of the results. Especially at low cell densities, larger relative errors may occur due to small differences in measured values.
- Importantly, the sensors must be carefully cross-validated with reference data, something that not always are available.
2.5 The production microorganism - Escherichia coli

In this study *Escherichia coli* stain HMS 174 (DE3) transformed with plasmid pET30a containing GFPmut3.1 was used.

*E. coli* is one of the most used microorganisms in industrial biotechnology, especially for recombinant protein production (Madigan et al, 2000). It is fully sequenced and significant parts of the physiology are known on a proteomic and transcriptomic level.

However, there are many complications with the organism. There exist several pathogenic strains, and many strains produce endotoxins. Purification of recombinant proteins produced in *E. coli* is rather difficult.

The cultivation technique for *E. coli* is well established. Ideal composition of cultivation media is well known. It is possible to culture to high densities of bacteria with accompanying high protein expression.

2.6 The product from the bioprocess - Green Fluorescence Protein

The green fluorescent protein (GFP) was discovered by Shimomura already in 1962 (Shimomura et al (1962). GFP has a typical beta-barreled structure (Figure 3a). In its natural element the protein is expressed by a species of jellyfish known as *Aequorea victoria*, from which it has been isolated (Figure 3b). It emits green fluorescence when exposed to blue light. Wild type GFP has a major excitation peak at wavelength 395 nm and a minor at 475 nm. Its emission peak is 509 nm which is in the lower green region of the visible spectrum.

![Figure 3a](http://jj-thecavendish.blogspot.com/2009_04_01_archive.html 02/08/10)

![Figure 3b](http://other95.blogspot.com/2008/10/nobel-jelly-aequorea-victoria.html 02/08/10)
Different mutants of GFP have been engineered including variants that emit blue and yellow light (Phillips, 2001; Chalfie et al, 1994). GFP is more qualified than other potential reporter proteins since it does not need any cofactors, substrates or additional stabilization to give a detectable signal (Chalfie et al, 1994; Cha et al, 1999). Thus, the different variants of the protein have been possible to use for many purposes with reporting gene expressions.

Reischer et al (2004) evaluated the GFP signal and its aptitude for on-line monitoring strategies of recombinant fermentation processes. They investigated GFP and its blue fluorescent variant in fed-batch fermentations and concluded that fluorescent reporter proteins are very suitable for designing new strategies for on-line bioprocess monitoring.

In this study GFP was used as a model system that could elucidate the behavior of common recombinant protein production systems in E. coli. The lack of appropriate on-line in-situ methods for monitoring the metabolic burden and critical state variables is a major roadblock. However, the GFPmut3.1 variant used in the study has high fluorescence yield in relation to the amount of protein produced, a good signal-to-noise ratio as well as a low detection limit.

2.7 The expression mechanism in E. coli

The vector expression mechanism of the E. coli system used in the study is based on the lac operon. The lac operon is an operon (a set of genes) that handles transportation and metabolism of lactose in E. coli. It is very well characterized and has been used to develop and illustrate molecular genetic methods. It was the first proposed control of prokaryotic gene expression (Walker & Rapley, 2000). The lac operon consists of three structural genes, a promoter, a regulator, a terminator and an operator. The tree structural genes code for lacZ, lacY and lacA which are enzymes and transport proteins that break down lactose in E. coli.

The operon is regulated by a protein referred to as the lac repressor, which binds specifically to the lac operator site. The inducer molecule, isopropyl-thiogalactoside (IPTG), binds to the repressor molecule, and initiates the transcription of the operon (Figure 4). The induction system is optimized by combining the over-expression of the lac repressor with a lacI mutation and increasing the promoter activity with a lacUV5 promoter mutation (Miesfeld, 1999; Walker & Rapley, 2000).
**Figure 4** The expression system in *E. coli*. The repressor protein blocks transcription of the operon but are inactivated when IPTG is added.

In this project the expression plasmid pET30aGFPmut3.1 was used as designed by Nemecek et al (2008). The reporter gene in the vector, GFPmut3.1, was controlled by the T7/lac promoter, containing a 25bp lac operator sequence.

The use of strong promoter systems for recombinant protein production is an effective way of generating high product yields. This is a key process in the biotechnology industry and revenues in the billions have been achieved by producing therapeutic biomolecules. This does not come without complexity however as strong vector systems overburden the host cell metabolism and cause the cells to collapse.

Recombinant proteins in *E. coli* are normally produced through over-expression. Unfortunately this leads to cell collapse within a relatively short time period. This is connected to the balance between the cells need of the substrate for energy metabolism (protein production) and for the cell anabolism. The produced protein easily tends to overburden the cells which need energy resources both to the protein production of the foreign protein and to native proteins.
Because of this over-expression of recombinant protein which consequently happens in the GFPmut3.1 a heat shock-like response is elicited involving a lot of relevant proteins. This may also include genes involved in osmotic, pH, temperature and starvation stress which may change their expression levels when it comes to metabolic burden. The imbalance in the cell, due to the formation of GFP, and native cellular proteins, and ensuing interactions of GFP with the endogenous E. coli metabolism is most probably the cause of a global stress response mediated by guanosine tetraphosphate, ppGpp (Dürrschmid et al, 2007). This physiology behavior can therefore be assumed to interfere with the GFP model production system applied with the E. coli strain used and should be kept in mind when interpreting the software sensor results.
3. Materials and methods

3.1 Strain

The organism used during the course of the entire project was the *E. coli* strain HMS 174 (Novagen, Madison, WI, US) transformed with the expression plasmid pET30a (Novagen). The plasmid contained GFPmut3.1 (Clontech, US), which worked as the reporter gene. The plasmid was under control of the T7/lac promoter, which contained a 25 bp *lac* operator sequence. The strain was obtained from the Department of Biotechnology, University of Natural Resources and Applied Life Sciences, Vienna.

3.2 Medium

The semi-synthetic medium described by Nemecek et al (2008) was the same medium used for every fed-batch fermentation in this project. KH₂PO₄ and K₂HPO₄ were added in relation to the working volume to act as K- and P-sources as well as to provide buffer capacity. All other components (C₆H₅Na₃O₇ × 2 H₂O, MgSO₄ × 7 H₂O, CaCl₂ × 2 H₂O, (NH₄)₂SO₄, NH₄Cl, trace element solution, yeast extract and glucose) were calculated in relation to the target biomass according to Table 1. The composition of the trace element solution can be seen in Table 2. The chemicals were purchased from Merck KGaA, Darmstadt, Germany.

*Table 1* The chemicals described below were mixed to a medium with their concentrations measured in relation to the desired BDW (biomass dry weight). The final two components were measured in relation to the working volume instead.

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration (g/g BDW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C₆H₅Na₃O₇ x 2 H₂O</td>
<td>0.25</td>
</tr>
<tr>
<td>MgSO₄ x 7 H₂O</td>
<td>0.10</td>
</tr>
<tr>
<td>CaCl₂ x 2 H₂O</td>
<td>0.01</td>
</tr>
<tr>
<td>(NH₄)₂SO₄</td>
<td>0.45</td>
</tr>
<tr>
<td>NH₄Cl</td>
<td>0.37</td>
</tr>
<tr>
<td>Glucose</td>
<td>3.00</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>0.05</td>
</tr>
<tr>
<td>Trace element solution</td>
<td>50.0 (μL/g BDW)</td>
</tr>
<tr>
<td>KH₂PO₄</td>
<td>3.00 g/L</td>
</tr>
<tr>
<td>K₂HPO₄</td>
<td>6.00 g/L</td>
</tr>
</tbody>
</table>
Table 2 These components made up the trace element solution. The chemicals were dissolved in 5 M HCl.

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration (g/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FeSO₄ x 7 H₂O</td>
<td>40.0</td>
</tr>
<tr>
<td>MnSO₄ x H₂O</td>
<td>10.0</td>
</tr>
<tr>
<td>AlCl₃ x 6 H₂O</td>
<td>10.0</td>
</tr>
<tr>
<td>CoCl₂</td>
<td>4.0</td>
</tr>
<tr>
<td>ZnSO₄ x 7 H₂O</td>
<td>2.0</td>
</tr>
<tr>
<td>Na₂MoO₂ x 2 H₂O</td>
<td>2.0</td>
</tr>
<tr>
<td>CuCl₂ x 2 H₂O</td>
<td>1.0</td>
</tr>
<tr>
<td>H₃BO₃</td>
<td>0.5</td>
</tr>
</tbody>
</table>

3.3 Analytical instrumentation

Table 3 Analytical instrumentation and labwares used in off-line analysis of samples taken from cultivations

<table>
<thead>
<tr>
<th>Material</th>
<th>Function</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectrophotometer</td>
<td>Determination of optical density</td>
<td>Ultraspec 1000, Pharmacia Biotech, GB</td>
</tr>
<tr>
<td>Cuvette</td>
<td>Sampling of fermentation broth for determination of optical density</td>
<td>VWR international, Stockholm, Sweden</td>
</tr>
<tr>
<td>Microtube</td>
<td>Sampling of fermentation broth for fluorescence measurements</td>
<td>Sarstedt, Numbrecht, Germany</td>
</tr>
<tr>
<td>96-well microtiter plate</td>
<td>Sampling of fermentation broth for fluorescence measurements</td>
<td>Corning, US</td>
</tr>
<tr>
<td>Fluorometer</td>
<td>Determination of fluorescence</td>
<td>Fluostar Galaxy, MTX lab systems, VI, US</td>
</tr>
<tr>
<td>Plastic pipe</td>
<td>Sampling of fermentation broth for determination of BDW</td>
<td>Sarstedt, Numbrecht, Germany</td>
</tr>
<tr>
<td>Centrifuge</td>
<td>Determination of BDW</td>
<td>Eppendorf international</td>
</tr>
<tr>
<td>Oven</td>
<td>Determination of BDW</td>
<td>Memmert</td>
</tr>
<tr>
<td>Scale</td>
<td>Determination of BDW</td>
<td>Sartorious, Goettingen, Germany</td>
</tr>
<tr>
<td>BDW sampling filters</td>
<td>Determination of BDW</td>
<td>Schleicher &amp; Schnell, Dassel, Germany</td>
</tr>
</tbody>
</table>
3.4 Cultivation protocol

Throughout the course of the project cultivation conditions were altered. Initially several batch cultivations were carried out in order to test the functionality of the equipment. Following this, the equipment was further tested in fed-batch runs prior to the verification experiments.

Cell bank procedure

The *E. coli* strain was transported on dry ice and stored on a filter pad at 4 °C. For working cell bank preparation, a shake flash cultivation was performed in a 500 mL flask with 200 mL semi synthetic medium and 50 μg/L kanamycin at 37 °C overnight. Kanamycin is an amino glycoside antibiotic meant to prevent the growth of any bacteria not carrying the plasmid. The cultivation was carried out in a rotary incubator at 200 rpm. The aspired cell concentration during the shake flash cultivation was 2 g bacterial dry matter per liter. Cell bank aliquots of 1.2 mL cell suspension and 0.6 mL glycerol (60%) were prepared and stored at -70 °C as working cell bank (WCB) vials.

Fed-batch cultivation procedures

All batch and fed-batch cultivations were carried out in a 10 L computer regulated bioreactor (Model LMS 2002, Belach, Sweden). The control software regulating the reactor was BioPhantom (Belach AB). The configuration of the software was altered overtime in order to provide the best possible conditions for the novel software sensors. Before batch processing, an overnight culture was prepared through aseptic inoculation of one vial working cell bank into a 500 mL shake flask containing 200 mL media. In the morning, the overnight culture had reached an OD-value of around 8 and was inoculated into the reactor. The working volume during the batch cultivations was 5 L. The media was composed to reach a bacterial dry matter of 10 g/L, but after sterilization the volume decreased with about 550 mL. The pH range was set in between 6.95 and 7.05 and regulated by the addition of 1 M sulfuric acid and 20% ammonia. The aeration rate was set to 1 vvm (5 L per min) and the temperature was set to 37 °C. Dissolved oxygen (DO) was controlled to 20% with stirrer speeds of in between 500 and 1200 rpm. Foaming was hampered by addition of 10% anti-foam solution (Dow Corning, BDH, Germany).

0.119 g IPTG dissolved in 20 mL water was inoculated into the reactor at 6 h with a syringe and a filter membrane attached. The total time of the processes lasted to around 11 h when cells entered the stationary state and stopped growing.
As mentioned, batch cultivations utilized pre-cultures prior to the fermentation process. In the fed-batch cultivations, two vials of the working cell bank were injected directly into the reactor. The medium was calculated to reach a biomass concentration of 5 g/L at the end of the batch phase and 25 g/L after the feed phase. The initial volume consisting of the batch medium was 4 L. The feed added 2 L medium over the course of 21 h. The regulation of pH, aeration rate, temperature and anti-foam solution was initially the same as described for the batch cultivations, yet some alterations were made overtime. The DO control was set to 40% but stirrer speeds were kept in between 500 and 1200 rpm. While slower stirrer speeds were investigated during test runs it became clear that faster speeds were more effective for growth with this working volume. Also, a slight overpressure in the reactor set to 0.1 Bar proved to be effective as well. To avoid strong foaming formations 2 mL anti-foam solution (100%; 1 ml/L) was added to the feed medium.

After a batch phase lasting up to 14 h, exponential feeding was initiated when the culture entered the stationary phase where it had consumed all of the available substrate. The exponential feed profile software was connected to a peristaltic pump (P4 U1-MXV, Alitea, Sweden). The feed weight was detected by a scale (XL-3100, Denver Instrument) connected to the software where a pre-calculated exponential feed profile along with a recently installed PID-regulation system regulated the pump speeds over a 21 h period during which the feed was meant to be consumed. The specific growth rate was set to an exponential value of 0.1 h⁻¹.

The pulse induction with 0.119 g IPTG was launched 7 h after feed start through inoculation of the ultra filtrated inducer through a membrane. A sample for fluorescence measurement was taken immediately upon injection as a reference followed by additional samples throughout the remainder of the process. Samples were immediately stored in a freezer kept at -20 °C for off-line measurements. The entire process lasted to around 32 h.
3.5 On-line hardware sensors

Standard bioreactor devices (SBD)

The reactor was equipped with standard devices for measuring and controlling DO, pH, temperature, headspace pressure and foaming levels. DO and pH were measured through electrodes while temperature, pressure (dTrans p31, Jumo, Germany) and foaming were measured with probes. Signals from these devices were acquired every second.

On-line HPLC

An in-situ filtration membrane probe (PP-19, Advanced Biotechnology Corporation, Germany) taking continuous samples served as a link between the reactor and the analytical devices of the HPLC system. In this study the HPLC system was used to detect changes in the concentrations of glucose, acetate and various other organic acids throughout the process. The membrane of the probe was a hydrophobic polypropylene membrane (Advanced Biotechnology Corporation, Germany) with the dimensions 135x5.5x1.5 mm. The probe was ultrasonicated in water and hydrophilized in 2-propanol prior to usage. Before implementing it to the reactor, test runs with 2-propanol and distilled water were carried out to ensure the function of the filter. In this, as well as previous projects, some difficulties occurred when at times the filter appeared to be impermeable for the fermentation broth. What causes this is unclear but it may be prevented by purging it regularly as described and setting a slight overpressure of 0.01 MPa in the reactor to support the pumping.

The HPLC system in which the samples were analyzed was of the brand Shimadzu. It consisted of a pump unit LC-20AD, a column oven CTO-20A and a refractive index detector RID-10A. Measurements were taken every 20 minutes through sampling from the bioreactor with the membrane probe. A peristaltic pump (Alitea, C8/2-XV, Sweden) as well as a six-port two-position automatic injection valve (Rheodyne, US) with a 20 μL loop (TPMV, Rheodyne, US) were controlled by the software LC Solution (Shimadzu, Japan).

An ion-exclusion column (Aminex®HPX-87H, 300x7.8mm; BioRad, US) works on a resin-based principle and separates organic acids, alcohols and carbohydrates. A guard column (Guard Cartridge 125-0129, BioRad, US) protects the analytical column from impurities and carries out a preseparation. It has to be changed frequently since the filtration probe does not exclude proteins and other macromolecules, which may lead to a slight clogging in the injection valve as well as in the guard column.

The validity of the system was proved by standard solutions of glucose, acetate, formic acid, lactic acid and acetone. The column was applied at 60 °C with degassed 5 mM sulfuric acid as mobile phase and a flow rate of 0.8 ml/min. Detection of glucose, acetate and the other components were carried out by the refractive index detector after their set retention times.
**NIR-monitoring**

To measure the cell density on-line, a near infrared probe (Cell Growth Monitor Model 650; Wedgewood Technology, CA) with a path length of 5 mm was used. The sterilizable probe provides in-situ measurements of the OD-value/biomass at 560 nm. To achieve a linear correlation, a first order linearization of the instrument was realized by selecting the third linear power set position. The optical density ranging from 0 to 1 was set to 0.2 and the aeration suppression, which suspends noise, was switched on. The probe was implemented to the reactor from the side next to the stirrer. To get the correlation between the output reading and the current biomass concentration, a factor needs to be calculated and established through measuring known cell concentrations. Off-line BDW measurements taken at different stages of fed-batch cultivation were plotted against their corresponding output readings from the NIR-probe. The trend line acquired from the plot was used to calculate this factor. The qualification procedure is described in 4.1.

**Gas analysis**

The outlet gas generated from the bioreactor was measured by a gas analyzer (CP460-O2/CO2; Belach, Sweden) that analyzed oxygen- and carbon dioxide levels. The instrument was calibrated through a two step procedure. First it was calibrated with nitrogen gas, setting its O2- and CO2-values to 0. Afterwards it was calibrated with regular air, setting its values to O2=21% and CO2=0.04%.

**3.6 Off-line measurements**

Samples for the optical density and the dry matter of the biomass were taken every two hours or so from the fermentation broth through a steam sterilizeable pipe. Steam was passed though the pipe before and after sampling. For determination of recombinant protein production, fluorescence measurements were performed every hour in the hours following IPTG induction and less frequently throughout the remainder of the process.

**Optical density**

When determining of the optical density, the fermentation broth had to be diluted with deionizer water in order to reach an absorbance within the linear region (≤ 0.6). Samples taken throughout the earlier stages of the process were diluted 1:100 while later samples had to be diluted 1:1000. The samples were analyzed in a UV-spectrometer (Ultraspec 1000, Pharmacia Biotech, UK) with absorbance at λ=600 nm measured with a path length of 1 cm.
**Biomass dry weight (BDW)**

Determining the biomass dry weight was done using two different methods. In the principle method, 3 mL fermentation broth was centrifuged for 5 min at 4 °C in a pre-weighted tube. The supernatant was discarded, the cell pellet re-suspended and stirred in 3 mL ice cooled phosphate buffered saline (PBS) buffer and centrifuged again. The washed cell pellet was dried over night at 108 °C. All samples were weighed after the run and thus the biomass dry weight could be determined by calculating the weight difference.

The other method was based upon filtration. A pre-weighted and dried 0.2 µm membrane filter (Schleicher & Schnell, Dassel, Germany) was placed upon a vacuum filtration device. 3 mL fermentation broth was pipetted on to the filter followed by 3 mL PBS. The filter was stored at 108 °C overnight and weighted again in order to determine the biomass dry weight.

**Fluorescence**

Albano et al (1996) demonstrated that the fluorescence of GFP can be used as a quantitative reporter of the protein concentration in *E. coli* since the ratio of fluorescence signal to protein concentration correlates. To measure the fluorescence, 1 mL rough fermentation broth was injected into a sampling plastic pipe and frozen at -20 °C. Samples were taken regularly after IPTG induction. One sample was taken as blank directly following the induction to obtain the culture background fluorescence, which is mainly due to NAD(P)H. The samples were diluted 1:1000 in two steps and measured in a spectro-fluorometer (Fluostar Galaxy, MTX lab systems, VI, US). Fluorescence was measured with excitation/emission wavelengths of 470/515 nm.
4. Modeling and theoretical aspects

4.1 Nomenclature and definition of parameters

Table 4 The table below presents the model equations that were used in the study along with their corresponding symbols, descriptions and units.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Unit</th>
<th>Software range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{NIR}$</td>
<td>Signal from NIR probe converted to an OD value</td>
<td>%</td>
<td>0-100</td>
</tr>
<tr>
<td>$X_{NIR}$</td>
<td>Biomass concentration calculated from $A_{NIR}$</td>
<td>g/L</td>
<td>0-100</td>
</tr>
<tr>
<td>$K_{NIR}$</td>
<td>Factor for converting the OD signal to biomass concentration</td>
<td>g/L/%</td>
<td>0-100</td>
</tr>
<tr>
<td>$V$</td>
<td>Calculated culture volume</td>
<td>L</td>
<td>0-10</td>
</tr>
<tr>
<td>$c_{glu}$</td>
<td>Glucose concentration</td>
<td>g/L</td>
<td>0-100</td>
</tr>
<tr>
<td>$c_{glu, feed}$</td>
<td>Glucose concentration in feed</td>
<td>g/L</td>
<td>0-500</td>
</tr>
<tr>
<td>$r_{glu}$</td>
<td>Glucose uptake rate</td>
<td>g/L/h</td>
<td>0-100</td>
</tr>
<tr>
<td>$q_{glu}$</td>
<td>Specific glucose uptake rate</td>
<td>g/g/h</td>
<td>0-100</td>
</tr>
<tr>
<td>$q_{glu, feed}$</td>
<td>Specific glucose uptake rate during fed-batch phase</td>
<td>g/g/h</td>
<td>0-100</td>
</tr>
<tr>
<td>$c_{ac}$</td>
<td>Acetate concentration</td>
<td>g/L</td>
<td>0-100</td>
</tr>
<tr>
<td>$r_{ac}$</td>
<td>Acetate production rate</td>
<td>g/L/h</td>
<td>0-100</td>
</tr>
<tr>
<td>$q_{ac}$</td>
<td>Specific acetate production rate</td>
<td>g/g/h</td>
<td>0-100</td>
</tr>
<tr>
<td>$C_{O2}$</td>
<td>Oxygen concentration</td>
<td>%</td>
<td>0-100</td>
</tr>
<tr>
<td>OUR</td>
<td>Oxygen uptake rate</td>
<td>g/min</td>
<td>0-100</td>
</tr>
<tr>
<td>$q_{O2}$</td>
<td>Specific oxygen uptake rate</td>
<td>g/g/min</td>
<td>0-100</td>
</tr>
<tr>
<td>$C_{CO2}$</td>
<td>Carbon dioxide concentration</td>
<td>%</td>
<td>0-100</td>
</tr>
<tr>
<td>CPR</td>
<td>Carbon production rate</td>
<td>g/min</td>
<td>0-100</td>
</tr>
<tr>
<td>$q_{CO2}$</td>
<td>Specific carbon dioxide production rate</td>
<td>g/g/min</td>
<td>0-100</td>
</tr>
<tr>
<td>$R_{O2/S}$</td>
<td>Oxygen consumption per substrate</td>
<td>mol/mol</td>
<td>0-100</td>
</tr>
<tr>
<td>$R_{CO2/S}$</td>
<td>Carbon emission per substrate</td>
<td>mol/mol</td>
<td>0-100</td>
</tr>
<tr>
<td>$q_{SumMxc.acid}$</td>
<td>Specific sum mixed acid production rate</td>
<td>g/g/h</td>
<td>0-100</td>
</tr>
</tbody>
</table>
4.2 Estimation of biomass

While there are two different methods for estimating the biomass, one being through on-line OD readings and the other through base titration, on-line OD readings were the only method utilized in this project. This model is based on the well-known correlation between the optical density of a cell suspension and the biomass, described by Lambert Beers law. It is dependent on all optical properties of the medium that cause absorbance in the 900-1100 nm range.

The computer system reads the NIR probe and converts the signal to an $A_{NIR}$ value. The system then utilizes a separately calculated factor ($K_{NIR}$) to convert the $A_{NIR}$ signal [0-100%] to biomass ($X_{NIR}$, g/L/%). The adjusted factor was calculated as described in 4.1 and the on-line output was derived through the formula:

$$X_{NIR} = K_{NIR} * A_{NIR}$$

4.3 Specific glucose uptake rate

Glucose serves as the substrate for the recombinant $E. coli$ cells used in this project. With the data provided by the HPLC system described in chapter 2.2.2, the specific substrate uptake rate could be calculated with the following equation:

$$q_{glu} = (dc_{glu} / dt) * 1 / X_{NIR}$$

Upon entering the fed-batch phase, the feeding was included according to:

$$q_{glu} = [D * (c_{glu,feed} - c_{glu}) - (dc_{glu} / dt)] * 1 / X_{NIR}$$

The system was programmed to read data files generated by the HPLC software LC Solution through a shared folder network according to a certain time schedule. Files were generated every 20 min, transferred and read by BioPhantom. The files provided concentration values for the substrate, acetate and the organic acids. With these values, rates and specific rates were calculated compared to the “last” file, read 20 min earlier. Because of this, the results were “stepwise”. In the new software sensors provided here steps were set to the 20 min run time. The procedures behind the file transfer and rate calculations were carried out as following:

The files generated by LC Solution were named and stored in the following manner:

HPLC_XXXX_001.TXT
HPLC_XXXX_002.TXT
HPLC_XXXX_003.TXT

The generated files were stored in a shared folder an once the operator initiated the "import" function on the BioPhantom mainframe, files were automatically read in a
synchronized order, with the software checking for updates every 10 s until another file appeared 20 min later, upon which it was read. When read, the previous glucose concentration was stored as $c_{glu,n-1}$ while the more recent one was stored as $c_{glu,n}$. Simultaneously, the time of the old value ($t_{n-1}$) was stored while the updated time was placed in $t_n$. The rates were calculated through the following equations:

$$r_{glu} = (c_{glu,n} - c_{glu,n-1}) / (t_n - t_{n-1})$$  \hspace{1cm} (4)

$$q_{glu} = r_{glu} / X_{NIR}$$  \hspace{1cm} (5)

$$q_{glu,feedbatch} = [D * (c_{glu,feed} - c_{glu}) - r_{glu}] / X_{NIR}$$  \hspace{1cm} (6)

4.4 Specific acetate production rate

In *E. coli* cultivations the accumulation of acetate, a common by-product, is considered as a sign of either overflow metabolism or anaerobic conditions (Sharmaa et al, 2007). Overflow metabolism appears when a high glucose concentration results in an imbalanced carbon flux through the TCA-cycle in comparison to the glycolysis (Enfors & Häggström, 2000). The specific acetate formation rate can be estimated according to the equation below:

$$q_{ac} = (dc_{ac} / dt) * 1 / X_{NIR}$$  \hspace{1cm} (7)

As with glucose, the concentrations of acetate were analyzed every 20 min and the rates were calculated in similar fashion as well:

$$r_{ac} = (c_{ac,n} - c_{ac,n-1}) / (t_n - t_{n-1})$$  \hspace{1cm} (8)

$$q_{ac} = r_{ac} / X_{NIR}$$  \hspace{1cm} (9)

4.5 Specific oxygen uptake rate

The gas analyzer provided measurements of the oxygen concentration at any given time. Along with the airflow value, it is used to calculate the oxygen uptake rate (*OUR*) through the equation below. Airflow is a variable that measures the flow rate of air into the reactor in L/min:

$$OUR = FE2_{airflow} * [(0.209 – cO2) / 100] * 60$$  \hspace{1cm} (10)

With the rate established, the specific uptake rate could be calculated with the help of equation (10) as described below:

$$qO2 = OUR / (X_{NIR} * V)$$  \hspace{1cm} (11)
4.6 Specific carbon dioxide production rate

In a similar fashion, the gas analyzer provided values of the carbon dioxide concentrations as well. Along with the airflow, the carbon production rate ($CPR$) was calculated in the following manner:

$$CPR = \text{FE}_{\text{airflow}} \times \left[ \left( \frac{c_{CO2}}{100} \right) - 0.003 \right] \times 60 \quad (12)$$

With the rate established, the specific production rate could be calculated with the help of equation (12) as described below:

$$q_{CO2} = \frac{CPR}{(X_{NIR} \times V)} \quad (13)$$

4.7 Oxygen consumption per substrate

This new sensor provides the ratio between oxygen- and energy substrate consumption. This is done by using the previously described equations for oxygen specific uptake rate and glucose specific uptake rate and dividing them:

$$R_{O2/S} = \frac{q_{O2}}{q_{glu}} \quad (14)$$

4.8 Carbon dioxide emission per substrate

Also a new sensor, this variable describes the ratio between the specific carbon dioxide emission rate and energy substrate consumption. This is done by using the described equations for carbon dioxide specific production rate and glucose specific uptake rate:

$$R_{CO2/S} = \frac{q_{CO2}}{q_{glu}} \quad (15)$$

21
4.9 Specific sum mixed acid production rate

The sum mixed acid specific production rate is the third novel sensor reviewed in this study. Three acids known to accumulate during *E. coli* fermentations with recombinant GFP were calibrated and programmed in the LC Solution software. In BioPhantom these acids are referred to as Acid1 - Acid3. Concentration measurements were imported from the HPLC-system every 20 min along with the previously mentioned values. Each acid has its specific production rate calculated separately in the same way as acetate i.e. as is presented for $q_{acid,1}$ below:

$$r_{acid,1} = \frac{c_{acid,1,n} - c_{acid,1,n-1}}{(t_n - t_{n-1})}$$

(16)

$$q_{acid,1} = \frac{r_{acid,1}}{X_{NIR}}$$

(17)

The specific production rates of the three acids along with acetate are summed together and form the specific sum mixed acid production rate according to the following equation:

$$q_{SumMixAcid} = q_{Acid,1} + q_{Acid,2} + q_{Acid,3} + q_{ac}$$

(18)
5. Results and Discussion

In this section the main results of the project are presented and discussed. First, the important biomass software sensor results of a more accurate calibration are shown. Second, the software sensor data based on gas analysis are exhibited and critically interpreted. Third, the HPLC-based software sensor monitoring data are analyzed and compared. As mentioned above, several technical cultivations in batch and fed-batch mode preceded the three validation cultivations shown here. Thus, these shown results are founded on additional experiments.

In Figure 5 the principle outlines of the software sensors used below are shown. It is noteworthy how all of software sensor interacts with each other by providing secondary estimates. For example, the software sensor for ratio oxygen uptake rate versus substrate uptake rate combine data from two sensors, the HLPC-based glucose rate sensor and the gas analyzer rate sensor where also the biomass estimate is used.

Figure 5 Principle outline of the software sensors used in this study.
5.1 Biomass growth estimation by a NIR-probe in a software sensor

A NIR-probe for monitoring the biomass in the bioreactor was used in a previous project (Warth et al., 2010). The software sensor function was mainly a matter of converting the probe signal to (1) a biomass dry weight concentration, and (2) to transform this concentration to a specific growth rate value.

In order to convert the signal output from the NIR-probe ($A_{NIR}$) to a correct dry weight concentration value, $X_{NIR}$, according to Equation (1) above, a calibration factor, $\kappa_{NIR}$, had to be calculated. This was done by measuring dry weight concentrations during a test fed-batch run and plotting these versus their corresponding $A_{NIR}$-values as presented in the graph below (Figure 6):

![Graph showing biomass dry weight versus the $A_{NIR}$ output signal of the NIR-probe.](image)

Figure 6 Off-line measured biomass dry weight versus the $A_{NIR}$ output signal of the NIR-probe.

The calibration factor was determined from the k-value in the plot 0.194 g/L/%. Compared to previous estimates the factor appeared to be reasonable and was used in the validation fed-batch runs. It should be noted, however, that this is an estimate related to the cultivation procedure and that the cultures may grow differently from batch to batch. It should also be noted that the optical density of the NIR signal had a degree of exponential shape, which complicated the estimation of the factor. This was improved by the linearization function in the instrument, which corrected the exponential signal to linearity (in principle also a software sensor-like procedure).
5.2 The software sensors for the ratios between the O$_2$ uptake or the CO$_2$ evolution rates and the substrate uptake rate

The purpose with the ratio sensors was to compare the input of energy through nutrient and oxygen and the produced carbon dioxide as a result of the metabolic activity.

Cultivation data

The three validation fed-batch cultivations were run consecutively with all sensor equipment in action. Figures 7, 8 and 9 show the bioreactor interface from the process control software (BioPhantom). As can be seen, fed-batch cultivation no. 1 was deviating whereas cultivation no. 2 and cultivation no. 3 were rather similar.
Figure 7 Cultivation chart from the bioreactor process control interface of the first validation fed-batch cultivation. Part A shows the first eight hours of the cultivation and Part B the subsequent eight hours. The software sensors for the oxygen/substrate ratio, \( R_{O2S} \) (EFR\_O2, turquoise) and the carbon dioxide/substrate ratio, (EFR\_CO2, purple) are shown together with the DO, \( X_{OD} \), airflow, pressure, \( q_{O2} \) and \( q_{CO2} \) signals.
Figure 8  Cultivation chart from the bioreactor process control interface of the second validation fed-batch cultivation. Part A shows the first eight hours of the cultivation and Part B the subsequent eight hours. The software sensors for the oxygen/substrate ratio, \( R_{O2S} \) (EFR \_O₂, turquoise) and the carbon dioxide/substrate ratio, (EFR \_CO₂, purple) are shown together with the DO, X \_OD, airflow, pressure, \( q_{O2} \) and \( q_{CO2} \) signals.
Figure 9 Cultivation chart from the bioreactor process control interface of the third validation fed-batch cultivation. Part A shows the first eight hours of the cultivation and Part B the subsequent eight hours. The software sensors for the oxygen/substrate ratio, \(R_{O2S}\) (EFR\(_{O2}\), turquoise) and the carbon dioxide/substrate ratio, \(R_{CO2S}\) (EFR\(_{CO2}\), purple) are shown together with the DO, \(X_{OD}\), airflow, pressure, \(q_{O2}\) and \(q_{CO2}\) signals.
Analysis of the biomass dry weight estimates

The $X_{NIR}$ sensor has previously been used in several investigations and provided accurate on-line estimates of the fermentation broth biomass concentration. The sensor is relatively new however, and may still be in need of refinements before it replaces more conventional methods. Therefore off-line biomass dry weights and OD$_{600}$ measurements were taken throughout each fed-batch cultivation in order to verify the $X_{NIR}$ readings. The graphs in Figure 10 compare the biomass estimates for each cultivation versus OD$_{600}$.

(a)

(b)
Figure 10 (a, b and c) describe fed-batch cultivation no. 1, no. 2 and no. 3. The growth patterns for the graphs look similar for the two methods. The last OD_{600}-value in (a) may be a measurement error caused by the instrument.
5.2.1 Ratio of specific $O_2$ uptake rate and specific substrate uptake rate

This software sensor calculated the ratio between specific oxygen uptake rate and specific substrate uptake rate and was assumed to be able to provide useful information about the physiological stress caused on the cultivation by the IPTG induction. In order to properly assess the accuracy and validity of the sensor its signals need to be compared with signals that may have affected its trend as well as events that may have occurred during the cultivation.

Interpretation of software sensor signals

The graphs in Figures 11 to 13 compare the signal of the $R_{O2/S}$ software sensor with the $X_{NIR}$, $q_{O2}$, $c_{glu}$ and GFP-values recorded in the three fed-batch validation cultivations.

**Figure 11** Results acquired from fed-batch cultivation no. 1. This cultivation grew poorly as is evident by the low $X_{NIR}$-values and high substrate concentrations. Feed started after 10 h and stopped at 31 h. IPTG was added after 17 h.
Figure 12 Results acquired from fed-batch cultivation no. 2. Feed started after 14 h and stopped at 35 h. IPTG was added after 21 h.

Figure 13 Results acquired from fed-batch cultivation no. 3. Feed started after 14 h and stopped at 35 h. IPTG was added after 20 h.
As is shown in Figure 11 the first validation cultivation grew extremely poorly with biomass concentrations peaking at 3 g/L and GFP production only reaching about 2000 fluorescence units. The initial glucose present in the batch phase had not been consumed. During the feed phase the glucose concentrations remained high and only minor growth could be observed. After some trouble-shooting it was clear that the sterilization procedure of the bioreactor containing the batch medium caused too much evaporation of the medium. The sterilization hampered the calibration of the DO-electrode which could not be carried out due to low medium volume. Probably, this caused the level of oxygen to go down too low. Another reason for the low yield could have been that stirrer speed was too low. Despite this, the $R_{O2/S}$ sensor gave substantial amount of data in the batch-phase before fading out once the process entered the feed-phase.

The other two fed-batch cultivations, illustrated in Figures 12 and 13, grew to significantly higher biomass density much due to the experiences gained from the first cultivation. Both cultivations reached biomass levels of 19.4 g/L and consequently the GFP production was also much higher for these cultivations with end concentrations of close to 400,000 fluorescence units compared to 2300. The medium volumes lost in the sterilization procedure were compensated for so that the DO-electrode could be accurately calibrated. Also, stirring speeds were set higher so that the oxygen flow could be distributed much more effectively, especially in the larger culture volumes at the end of the feed-phase. Initially the feed was intended to begin at 10 h after the batch start for these cultivations as part of a standard procedure, but when it was discovered that glucose still remained in fed-batch cultivation no. 2 it was set to start when all substrate had been consumed and since fed-batch cultivation no. 3 had the same configurations this timing was used there as well. Although fed-batch cultivation no. 2 and no. 3 grew much more efficiently than fed-batch cultivation no. 1 the results obtained from the $R_{O2/S}$ sensor were almost the same. Also, the sensor gave steady readings in the batch phase. Fed-batch no. 2 gave the highest readings for the sensor with values between 70 – 80 mol/mol. Fed-batch cultivation no. 3 gave modest values for the sensor and extremely low values for the substrate concentration. This may have been caused by measurement errors occurring in the LC Solution™ software or clogging of the HPLC-probe. Cultivation no. 1 and no. 3 only obtained high $q_{O2}$-values in the batch phase while run 2 exhibited high levels throughout the entire process.

In a previous study, Sundström and Enfors (2008) described a similar software sensor for Ro/s where they used the sensor to provide information about physiological stress in an *E. coli* fed-batch culture (Sundström and Enfors, 2008). The study explained and demonstrated how a $R_{O2/S}$ software sensor also can be supplemented with a compensation for maintenance energy. The maintenance of a cell culture can be described as the amount of substrate consumed that does not result in new biomass. In a fed-batch culture with constant glucose feed, the specific substrate uptake rate declines gradually. At high specific substrate uptake rates in the earlier stages of the process the cells use both the oxygen and the substrate for growth (anabolism) and energy metabolism, from which only a small percentage goes to maintenance. As the specific substrate uptake rate declines the percentage of substrate and oxygen utilized for maintenance increases drastically if the maintenance coefficient is constant (Figure 14).
Figure 14 Schematic (adapted from Sundström and Enfors, 2008) illustrating that at high specific glucose uptake rates the flux to maintenance is of only a minor portion, while at declining rates as much as 40% at a constant maintenance coefficient. The figure also illustrates how the oxygen per glucose consumption increases at declining growth rates.

The study by Sundström and Enfors (2008) also demonstrated how the $R_{OS}$ sensor increased when the glucose uptake rate declined and the maintenance coefficient remained constant. The oxygen consumption rate is proportional to the part of the substrate flux that is used for energy metabolism. Because this flux takes a larger part of the total substrate uptake when $q_{S}$ declines, the oxygen consumption per substrate consumption ($R_{OS}$) increases when the feed rate is constant.

In this project study similar trends can be observed between the acquired results and the result from Sundström & Enfors (2008). All fed-batch cultivations produced values for the sensor in the batch phase, where $q_{glu}$ slowly declined prior to the exponential (0.1 h$^{-1}$) feeding phase. In the Sundström & Enfors report, the $R_{OS}$ sensor gave low values during a brief exponential feed phase with the exponent 0.3 h$^{-1}$ preceding the constant feed, resulting in rising levels of specific substrate uptake rate. Arguments can be made that since fed-batch cultivation no. 1 and no. 3 lacked $q_{O2}$-values in their corresponding feed-phases, it is only natural for the $R_{O2S}$ sensor to decline. However, the second fed-batch cultivation had high readings for $q_{O2}$ during the feed-phase as well, as is shown in Figure 12, yet still not providing high readings for the sensor.
Contrary to the study by Sundström and Enfors (2008) the feed in the three validation cultivations was exponential, making readings from the $R_{O2/S}$ sensor too low to properly assess it in the feed phase and thereby how IPTG induction may have affected it. Because of this, the sensor was not entirely tested for its purpose, making limitations and shortcomings difficult find as well. There are yet some conclusions that can be drawn about the sensor based on these experiments. Since it is highly dependent on results acquired from the HPLC-system through the $q_{glu}$-value, it is essential that the system works properly and continuously, which was not always the case. As had been reported in a previous project, the HPLC-probe as well as tubing was frequently clogged (Warth et al, 2010). This hindered the probe from taking samples of the culture medium and thus not being able to provide data for the sensor. Another shortcoming was discovered when the LC Solution™ software unexpectedly changed the retention time for the substrate when its concentrations were declining at the end of the batch phase. However, this error was noted and corrected manually before it could have affected the results. Nevertheless, these are two shortcomings that need to be addressed if the sensor is to be able to provide useful information about the physiological stress in a continuous and effective fashion.
5.2.2 Ratio of the CO₂ evolution rate and substrate uptake rate

The \( R_{CO2/S} \) sensor calculated the ratio between the specific carbon dioxide production rate and specific substrate uptake rate. As with the previous sensor, it was tested for its ability to provide information about the physiological stress by comparing it to the values which may have affected it.

Interpretation of the software sensor signals

Results from the first fed-batch run are discarded for this sensor as the values were too low to be observed. The following graphs compare the software sensor signal to the \( X_{NIR} \), \( q_{co2} \), \( c_{glu} \) and GFP-values acquired from the fed-batch cultivations:

![Fed-batch 2](image)

**Figure 15** Fed-batch cultivation no. 2. Feed started after 14 h and stopped at 35 h. IPTG was added after 21 h.
Figure 16 Fed-batch cultivation no. 3. Feed started after 14 h and stopped at 35 h. IPTG was added after 20 h.
As stated in Section 5.2.1, the second fed-batch cultivation (Figure 15) grew very well. The fermentation provided good measurements for $R_{CO2/S}$ as well. The highest peak value recorded was 35 mol/mol at 13 h, just before the feed phase started. The $q_{CO2}$-values remained high and stable throughout most of the fermentation as well. The results and patterns acquired from this sensor are very similar to those of the $R_{O2/S}$ sensor for obvious reasons. The difference here is that carbon dioxide is produced from the metabolism of *E. coli* as a by-product while oxygen is consumed. The carbon dioxide specific production rate has the same trend as the oxygen specific uptake rate but is throughout lower in absolute values, whereby the $R_{CO2/S}$ and $q_{CO2}$ sensors have the same trends as their corresponding O$_2$ ratios, but are lower. Fed-batch cultivation no. 3 gave only minor results for the $q_{CO2}$ sensor and no visible values for $R_{CO2/S}$.

The parallels drawn between the results and the experiment conducted by Sundström & Enfors (2008) in Section 5.2.1 can also be drawn here. The $R_{CO2/S}$-values rise as the substrate concentrations decline towards the end of the batch phase (Figure 15) and hence $R_{CO2/S}$ declines and fades out when the exponential feed causes $q_{glu}$ to rise.

The limitations hindering the performance of the $R_{O2/S}$ sensor (Section 5.2.1) are the same limitations present for this sensor. Since $q_{CO2}$-values were high throughout most of the feed phase including the IPTG induction, it can be assumed that $R_{CO2/S}$ would give readings here as well, provided that the feed is constant. This sensor, along with $R_{O2/S}$, has the potential to provide useful information about the physiological stress that IPTG may have on the metabolism of *E. coli* if properly evaluated. The two sensors can complement each other and will most likely provide a wider array of information if evaluated simultaneously rather than either one by itself.
5.3 The software sensor for specific sum mixed acid production rate

Cultivation data

The graphs presented in the following pages have been printed from bioreactor control software interface and describe the $q_{\text{SumMixAcid}}$ (green) signals for each of the three validations fed-batch runs.
Figure 17 Cultivation chart from the bioreactor process control program of the first validation fed-batch cultivation. Part A shows the first eight hours of the experiment and Part B the subsequent eight hours. The $q_{\text{SumMixAcid}}$ (green) software sensor signal is shown with the DO, $X_{\text{OD}}$, pumpspeed, pH, $q_{O_2}$ and $q_{CO_2}$ signals.
Figure 18 Cultivation chart from the bioreactor process control program of the second validation fed-batch cultivation. Part A shows the first eight hours of the experiment and Part B the subsequent eight hours. The $q_{\text{SumMixAcid}}$ (green) software sensor signal is shown with the DO, $X_{\text{OD}}$, pumpspeed, pH, $q_{O_2}$ and $q_{CO_2}$ signals.
Figure 19 Cultivation chart from the bioreactor process control program of the third validation fed-batch cultivation. Part A shows the first eight hours of the experiment and Part B the subsequent eight hours. The $q_{\text{SumMixAcid}}$ (green) software sensor signal is shown with the DO, $X_{\text{OD}}$, pumpspeed, pH, $q_{O_2}$ and $q_{CO_2}$ signals.
**Interpretation of signals**

The $q_{\text{SumMixAcid}}$ sensor sums together the specific growth rates of acetate and three other acids. In the bioreactor software the specific production rate for these acids was presented as $q_{\text{acid}}$ 1 to 3, leaving the operator to calibrate which retention time to assign to which sensor. In this study the $q_{\text{acid}}$ sensors 1 to 3 were assigned to read the specific production rates of formic acid, lactic acid and acetone. These acids are known to accumulate in fed-batch fermentations with recombinant *E. coli* as by-products of the cell metabolism. The acids are undesired and since they consume substrate their yields need to be minimized in order to achieve the highest possible productivity yield.

To properly evaluate the sensor its signals need to be compared with signals that may have affected it as well as events that may have occurred during the fermentation runs. The following graphs compare the software sensor signal to the $X_{\text{NIR}}$, $c_{\text{glu}}$ and GFP-signal acquired from the fed-batch cultivations.

**Figure 20** Fed-batch cultivation no. 1. Feed started after 10 h and stopped at 31 h. IPTG was added at 17 h.
**Figure 21** Fed-batch cultivation no. 2. Feed started after 14 h and stopped at 35 h. IPTG was added at 21 h.

**Figure 22** Fed-batch cultivation no. 3. Feed started after 14 h and stopped at 35 h. IPTG added at 20 h.
The third fed-batch cultivation provided the highest visible results while the first provided modest values. The sensor gave very high readings in the batch phases, with specific production rates reaching as high as 28 g/g/h before fading out as the exponential feed was initiated. Since no results could be observed in the feed phases, the sensors reaction to IPTG induction could not be observed. In other \textit{E. coli} strains, an IPTG induction had no effect on the acetate accumulation (Sharmaa et al, 2007).

The pattern shown by the software sensor trends in Figure 22 looks quite similar to the patterns obtained by the $q_{ac}$ sensor in Warth (2010). This is understandable since this sensor is included in the calculation. The sensor has very high potential in covering broader aspects of acid formation in recombinant \textit{E. coli} cultivations. This may reveal new opportunities for increasing productivity yields and lowering production costs for industrial fermentations.

When evaluating novel sensors, initial results may not live up to theorized expectations due to errors and complications not taken into consideration. Since there are many potential things that can cause problems in on-line measuring it is highly probable that errors may have hindered some of the readings. This is likely why trends differed somewhat between fed-batch cultivation no. 2 and 3.
6. Conclusions

The objective of this study was to develop and evaluate software sensors that cover important state parameters and provide useful information for improving monitoring and control of recombinant *E. coli* fed-batch processes. Three software sensor models, derived from already present software sensors, were successfully tested in fed-batch processes. Although there is much room for improvements, the sensors provided stable signals which could be explained in theory.

These software sensors were primarily developed to provide more information of what effects IPTG induction may have on cells expressing recombinant proteins, providing opportunities for the development of more cost efficient processes. While these effects could not be observed in this particular study, the sensors opened up a large variety of options for further studies and evaluations. In future experiments a lot of attention should be given to what configurations for the feed-regulation may be the most optimal when using these sensors. The recently installed PID-regulation system controlling the feed profile opens up a lot of interesting possibilities.

The study further proves the theory that new and useful information can be derived from already existing sensor algorithms. Software sensors are a cost efficient and easily applicable tool for monitoring and control of industrial fermentation processes and are likely play a large role in the future of industrial biotechnology.
7. Acknowledgements

I have had a great time working with this project and it has been an honor and a pleasure working with the staff of the Division of Biotechnology at Linköping University.

First and foremost I would like to thank my supervisor Professor Carl-Fredrik Mandenius for giving me this opportunity and providing me with hints and suggestions at times when I was confused and disappointed for not acquiring the results I had anticipated.

A big thank you to research engineers Gunnar Bergström and Maria Carlsson for helping me out with the many technical difficulties experienced during the equipment testing phase of the project. Thank you for your patience and support. Without you this would never have been possible and I would love to work with you again if time and opportunity allows me to do so.

My acknowledgements also go out to Dr. Michael Fritzsche and Benedict Warth who with their experience provided me with hints as well as György Rajkai from Belach Biotechnology who updated the BioPhantom software for us. Good luck to all of you with your future projects!

Last but not least I would like to thank all of my family and friends for supporting me as well as the people at IFM for being so friendly and social, creating a welcoming place for me to work in.
8. References


