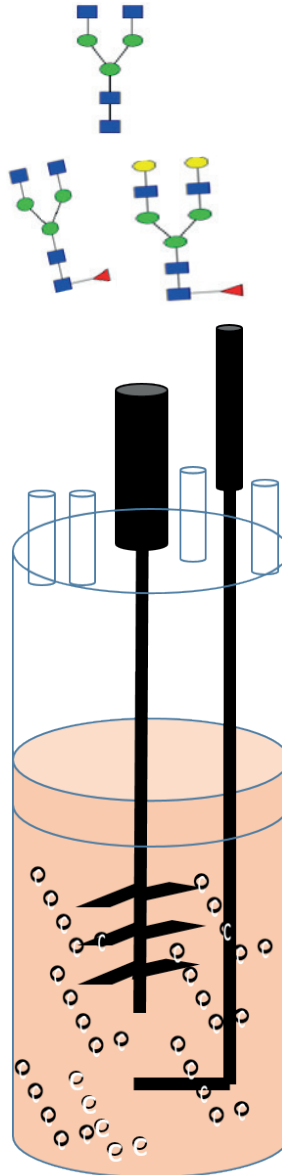


Transcriptomic profiling as a tool to characterize CHO cell processes



Abdulaziz Abdullah Al Sayyari

Transcriptomic profiling as a tool to characterize CHO cell processes

Abdulaziz Abdullah Alsayyari

Thesis committee

Promotor

Prof. Dr R.H. Wijffels,
Professor of Bioprocess Engineering
Wageningen University & Research

Co-promotors

Dr D.E. Martens, Associate professor, Bioprocess Engineering Group
Wageningen University & Research
Dr J.A. Hageman, Assistant professor, Biometris Group Wageningen University &
Research

Other members

Prof. Dr A.H. Kersten, Wageningen University & Research
Dr C. Clarke, The National Institute For Bioprocessing Research and Training
(NIBRT), Dublin, Ireland
Dr M. Streefland, Kiadis Pharma, Amsterdam, Netherlands
Dr Y.E. Thomassen, Intravacc, Utrecht, Netherlands

This research was conducted under the auspices of the Graduate School VLAG
(Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health
Sciences).

Transcriptomic profiling as a tool to characterize CHO cell processes

Abdulaziz Abdullah Al Sayyari

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 17 October 2018
at 4 p.m. in the Aula.

Abdulaziz Abdullah Al Sayyari

Transcriptomic profiling as a tool to characterize CHO cell processes,
163 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)

With references, with summary in English

ISBN: 978-94-6343-359-4

DOI: <https://doi.org/10.18174/460722>

Dedicated to my Father and Mother

Table of contents

Chapter 1

Introduction and thesis outline	9
--	----------

Chapter 2

Differential gene expression in CHO cells cultivated in shake flasks and bioreactors	21
---	-----------

Chapter 3

Shake flask as scale down model for fed-batch controlled bioreactor: a transcriptome comparison	45
--	-----------

Chapter 4

Transcriptome analysis for the scale-down of a CHO cell fed-batch process	71
--	-----------

Chapter 5

Transcriptome analysis as a tool for characterizing a CHO cell perfusion culture	107
---	------------

Chapter 6

General discussion	129
---------------------------	------------

Summary	156
----------------	------------

Chapter 1

Introduction and thesis outline

In recent years, increasing attention has been directed towards the development of biological medicines. For example, biopharmaceutical proteins are a class of medicines used to treat life threatening diseases such as cancer and auto immune diseases [1]. A special class of these biopharmaceutical proteins is formed by the monoclonal antibodies. Monoclonal antibodies can bind specific antigens and thus specifically target certain cells like cancer cells [2], [3]. With the increased knowledge on disease progression, better antibodies are designed that more effectively target the diseases. As a consequence monoclonal antibodies have become a major class of biopharmaceutical proteins and the number of antibodies on the market and in clinical trials is rapidly growing. Biopharmaceutical proteins are complex proteins that have specific post translational modifications, among which glycosylation is the most important one [4]. Correct glycostructures are important for the efficacy of the protein as well as for the side effects. Human like glycosylation can only be attained by production in mammalian cells. Thus, the main production host for biopharmaceutical proteins is the Chinese Hamster Ovary cell line [5]. A problem with these biopharmaceutical medicines is their high price, which imposes an increasing financial pressure on the health care system. The high price can at least in part be explained by the high development costs of these products, the long development time and the high failure rate during development. However, currently a number of biopharmaceutical products are going off patent [6]. This opens up the possibility for other companies to produce a similar copy of this protein, which is called a biosimilar. If the biosimilar has a similar structure as the originator molecule the development paths can be substantially shorter and less expensive and thus will result in cheaper products.

1.1 Antibody biosimilar development

The regulatory pathway for biosimilar medicines is designed to ensure that the proposed biosimilar medicine is highly similar to the off-patent approved biological medicine (reference product) in terms of structure, efficacy, safety, and quality [7]. Thus approval of biosimilars relies on well-designed comparative studies to demonstrate the high similarity between proposed biosimilar product and reference product. Similarity is based on intensive characterisation such as structural analyses, functional assays, functional, non-clinical, immunogenicity assessments, and comparative clinical assessments [8], [9].

Biosimilar medicines have the potential to reduce the high cost of original biological medicine up to 30% [10]. Nevertheless, the costs of biosimilar development are still high with a range of 100-250 million dollars per product [10]. Costs show variation depending on the biological medicine class. For example, the costs of biosimilar development for antibodies is much higher than for human growth factor biosimilars, due to, amongst others, the higher complexity of the product and of the biomanufacturing process [10]. As stated, antibody-based drugs are among the major classes of therapeutics against cancer, and immunological disorders [11], and more than 70 antibody-based drugs have been approved by the European medicines agency EMA [12]. Moreover, more than 600 therapeutic antibodies are currently in different phases of clinical trials [13].

As stated pharmaceutical proteins such as antibody-based drugs are complex proteins with extensive post translational modifications like glycosylation. Therefore, production of these proteins is done in mammalian cells like CHO cells. These cells require a complex chemically defined medium containing in general more than 80 components. Furthermore, to optimize production more complex processes like fed-batch or perfusion are used, increasing further the number of parameters that can influence the quality of the product like the composition of feeds and the feed addition strategy. To design a robust and good process for the production of these high-value therapeutic proteins, in general, a large number of experiments are required, which is time and resource consuming. Traditionally, process development for biopharmaceuticals was quite empirical and trial and error based. Once a good enough process was obtained, all parameters were fixed and deviations were not allowed. Quality of the product was tested after it was manufactured. This traditional process development approach is associated with an high failure rate during development as during scale-up unexpected changes occurred in the process and product quality. Furthermore, once a process was approved it also led to high failure rates during manufacturing as process deviations often resulted in rejection of the production batch. Thus, there was a need for a more pro-active scientific approach, such as quality by design (QbD), for the development of new biopharmaceutical products as well as of biosimilar products to increase the production efficiency and reduce the failure risk.

1.2 Quality by Design

Quality by design is an approach that focusses on a better scientific understanding of the product and the manufacturing process including the associated risks and how to mitigate these risks. This should lead to a process where the quality of the end product is an integral part of the manufacturing process. For the product this means that there is a good understanding of the relation between the structure of the product and its efficacy and side effects. This understanding results in the definition of the so called critical quality attributes (CQAs), which are the structural aspects of the product important for quality. Next the quality can be built into the process by a proper understanding of the relation between all the process parameters, including their interactions, and the CQAs. Central to this and to the QbD approach are the concepts of knowledge space, design space and control space. The knowledge space is the range of parameter values that is studied and typically also includes parameter values resulting in insufficient product quality. The design space is the range of parameter values within which a proper quality is attained. The control space is the range of parameter values within which the process should be controlled, which falls usually well within the design space. Given the complexity of the process and the large number of parameters, identification of the design space requires a large number of experiments. Usually first a set of screening experiments are done to select a proper clone and identify parameters, the so called critical process parameters (CPPs), that are important, as well as an initial range for these parameters. Next, using high throughput experimentation the design space is identified for the CPPs. **Figure 1** illustrates how critical process parameters (CPPs) influence cell physiology and through that also the quality of the product during the cultivation process. These experiments are performed in small-scale systems such as shake flasks or 15-100ml minibioreactor systems (e.g. ambr) (see **Figure 2**). It is very important that these small scale systems are a proper scale-down of the production reactor, meaning results obtained can be translated to the production reactor without changes occurring.

The glycosylation pattern is the main indicator of quality throughout the life cycle of production and for process development, including cell line development, feed and process optimisation and cultivation [14]. Glycosylation is a post-translational

modification (PTM) that is considered one of the most important CQAs [15], [16]. All approved glycoprotein therapeutics contain N-linked glycans. The patterns of these N-linked carbohydrates have been widely reported to have a direct impact on the safety and efficacy of IgG therapeutics administered to patients. For example, glycoforms can negatively affect the protein half-life, tissue targeting and modulate biological activity [15].

The concept of establishing a design space is based on a statistical knowledge on the relation between the critical process parameters and the critical quality attributes of the product like glycosylation structure. As stated this requires extensive experimentation requiring considerable resources but more importantly considerable time. In addition still a limited number of parameters is studied. This problem can be reduced by obtaining a better biological understanding on how the CPPs affect the CQAs. If the biological mechanisms are known critical process parameters and possible interaction effects can be predicted, meaning a reduction in the amount of parameters and interaction effects that need to be studied. Furthermore, results can easier be translated to other products using other clones. Finally, possibly markers or combinations of markers can be found that can directly predict product quality and can be measured on-line, allowing for on line control of product quality. A possible tool to obtain this biological understanding is by measuring global gene expression or transcriptome analysis. With the availability of the CHO genome and a CHO cell specific microarray measuring gene expression in CHO cells became relatively easy either using RNAseq or micro-arrays. However, not everything is regulated on the gene expression level, as there is also regulation on the translational level and on protein activity. Thus transcriptome analysis can clarify whether certain processes are at least in part regulated on the gene-expression level and these changes in gene expression can link changes in process parameters to changes in for example glycosylation. Apart from giving mechanistic insight, the transcriptome can also be used as a fingerprint of the physiological state of the cells. As such it can be used to compare certain conditions and study whether they affect the state of the cells on the gene expression level. If yes one can do a functional analysis to find the cause of these differences.

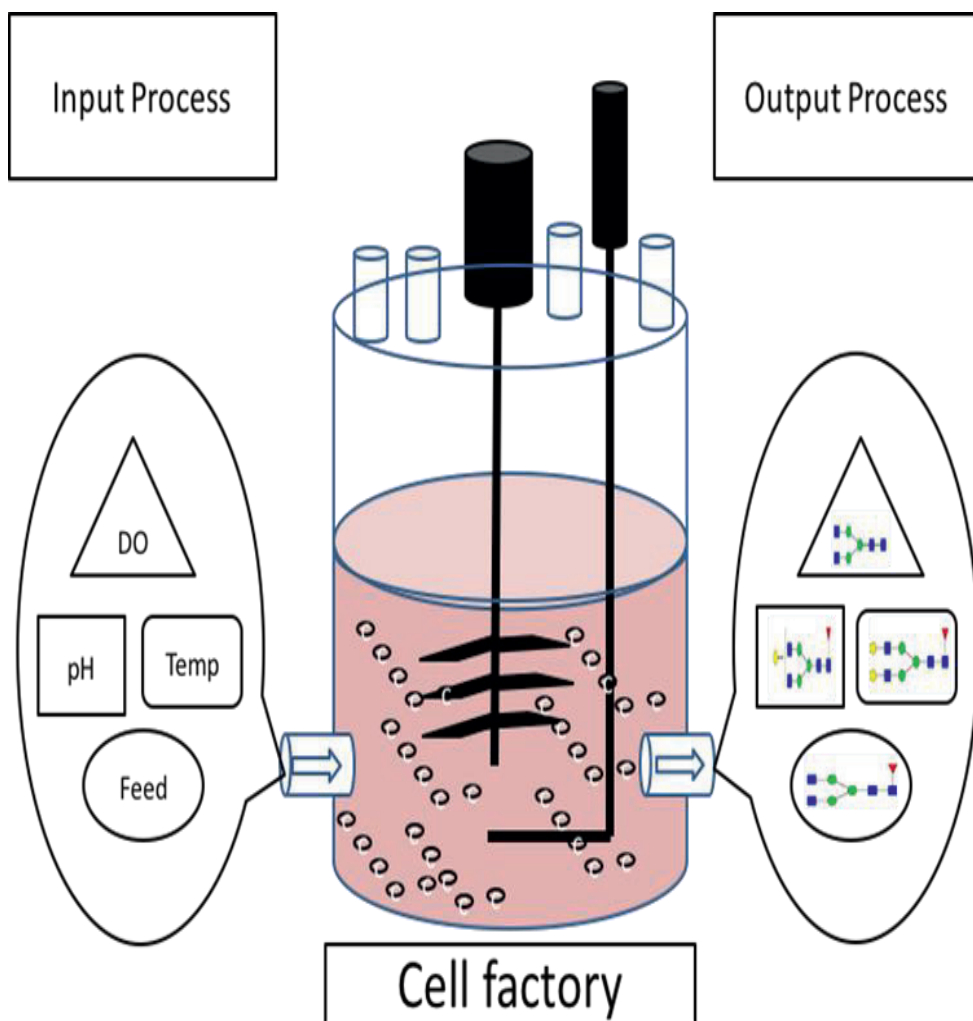


Figure 1: The influence of process parameters on product formulation in a cell factory.

1.3 Aim and outline of the thesis

The aim of this thesis is to study the value of transcriptome analysis for process development for the upstream cultivation process. In the experiments presented in this thesis, a commercial Affymetrix® Chinese Hamster Ovary (CHO) Gene 2.1 ST Array Plate was used for transcriptomic profiling of CHO cells producing a potential antibody biosimilar in batch, fed-batch and perfusion cultivation modes. Transcriptomic profiling

is used in the two ways as described before: (1) to carry out process fingerprinting using an unsupervised methodology such as principal component analysis (PCA); and (2) to gain understanding of biological processes in the cell by identifying correlations between gene expression, CQAs and CPPs, as well as between gene expression and key performance indicators (KPIs). As such the results of this thesis can contribute to applying transcriptomic profiling in a QbD approach.

Different cultivation systems and scales are used in this thesis and the associated cultivation system are illustrated in **Figure 2**. In Chapter 2 to 4 transcriptome analysis is used to assess different small-scale reactor systems. As stated for process development high throughput experiments are required, which need to be done at small scale. These small scale systems must be representative for the larger scale systems. Two small scale systems are studied being shake flasks and the ambr® 15 bioreactor system.






	A	B	C	D	E
					
Cultivation system	Shake Flask (250ml)	ambr® 15ml Miniature bioreactor system (Sartorius)	1 L Desgip Bioreactor system (Eppendorf)	3L bioreactor vessel and 10L BioSep system cell retentions (Applikon biotechnology)	10 L Biostat STR stir tank bioreactor (Sartorius)
Cultivation mode	Batch Fed-batch	Fed-batch	Batch Fed-batch	Perfusion	Fed-batch

Figure 2 several types of cultivation system used in this thesis. (A) shake flask, and (C) 1L bioreactors used in Chapter 2 (batch mode) and Chapter 3 (Fed-batch). (B) ambr®, and (E) 10 L biostat STR used in Chapter 4 (fed-batch). (D) perfusion cultivation mode used in Chapter 5. Figure 2. B and E are adopted from Sartorius with their permission to be published in this thesis. Figure 2 D adopted from Applikon biotechnology. Figure 2. C adopted from Eppendorf.

In Chapter 2 the shake flask system is compared to 1l bioreactors for a batch cultivation process on the gene expression level with the aim to study whether shake flasks are representative for the bioreactors. Shake flasks are in principle not a proper scale-down from controlled reactors. They lack control of the pH and of the dissolved oxygen tension as opposed to bioreactor systems where these are controlled. Furthermore mixing is done by shaking and not by an impeller as in the bioreactors and gas transfer occurs through the surface only whereas in bioreactors sparging is used. Nevertheless, the shake flask system is still used for early process development and consequently it is important to know to what extent they are representative for the bioreactor systems. Gene expression analysis is done on 3 time points to capture the dynamics of the cultures, being early and late exponential phase and the stationary phase. Apart from gene expression also the glycosylation is measured together with other important parameters like cell concentrations, and glucose and lactate concentrations.

In Chapter 3 a similar study is done for a fed-batch process. For a fed batch process it is more likely to find differences between the uncontrolled shake flasks and controlled bioreactors due to the more complex process (addition of feed) and higher cell densities reached.

In Chapter 4, a fed-batch process in a 10 litre bioreactor is scaled down to the ambr® 15 system which contains 48 miniature bioreactors with a maximum working volume of 15 mL. Two different scale-down criteria are compared with one based on matching the agitation tip-speed with the 10 L process and the other based on matching the agitation power input per volume of liquid. The ambr® 15 results are compared to the 10 L process in terms of cell growth, metabolism, mAb production, mAb glycosylation and transcriptome. The main difference with the shake flasks is that in the reactors of the ambr® 15 system the DO and pH can be controlled and that as compared to shake flasks the agitation and gassing system are more comparable to that in the larger scale bioreactors. Thus we can assess whether having this more comparable system is beneficial for the validity of the scale-down and whether the transcriptome analysis is useful to assess this. In Chapter 5 transcriptome analysis is used for a different purpose namely to characterize a perfusion process in a 1.5 litre bioreactor using CHO

cells. Continuous manufacturing draws a lot of attention currently because it allows a more efficient use of equipment and thus a reduction in costs. For the upstream part of the manufacturing process continuous perfusion is the preferred cultivation mode to fit in this approach. In such a system a steady state can be reached at high cell densities allowing for high volumetric productivities under constant conditions. Here we study the use of transcriptome analysis to identify the different states of a perfusion process and the variation in gene expression during the steady state as defined based on other culture parameters. For this samples are analysed for gene expression throughout the culture period. Again next to transcriptome analysis also cell growth, metabolism, mAb production, and mAb glycosylation are measured. Finally, in Chapter 6 the application of transcriptome analysis in process development and quality by design is discussed.

1.4 References

- [1] P. Zhang *et al.*, “Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs,” *Drug Discov. Today*, vol. 21, no. 5, pp. 740–765, May 2016.
- [2] M. Spearman, B. Dionne, and M. Butler, *Antibody Expression and Production*, vol. 7. 2011.
- [3] P. J. Carter, “Potent antibody therapeutics by design,” *Nat. Rev. Immunol.*, vol. 6, no. 5, pp. 343–357, 2006.
- [4] Walsh G, “Biopharmaceutical benchmarks 2006,” *Nat. Biotechnol.*, vol. 24, no. 7, 2006.
- [5] M. Matasci, D. L. Hacker, L. Baldi, and F. M. Wurm, “Recombinant therapeutic protein production in cultivated mammalian cells: current status and future prospects,” *Drug Discov. Today Technol.*, vol. 5, no. 2–3, pp. e37–e42, Sep. 2008.
- [6] F. Braido, S. Holgate, and G. W. W. Canonica, “From ‘blockbusters’ to ‘biosimilars’: an opportunity for patients, medical specialists and health care providers.,” *Pulm. Pharmacol. Ther.*, vol. 25, no. 6, pp. 483–6, Dec. 2012.
- [7] FDA-Biosimilar-2015-draft, “Additional Questions and Answers Regarding Implementation of the Biologics Price Competition and Innovation Act of 2009 Guidance for Industry,” *Required*, no. May 2015, 2009.
- [8] European Medicines Agency, “Guideline on similar biological medicinal products containing monoclonal antibodies – non-clinical and clinical issues Guideline on Similar Biological Medicinal Products Containing Monoclonal Antibodies – Non-clinical and Clinical Issues Table of contents,” vol. 44, no. May, pp. 1–16, 2014.
- [9] CDER/CBER and FDA, “Scientific Considerations in Demonstrating Biosimilarity to a Reference Product Guidance for Industry Scientific Considerations in Demonstrating Biosimilarity to a Reference Product,” *Guid. Ind.*, vol. April, pp. 1–24, 2015.
- [10] J. P. F. Erwin A. Blackstone, “The Economics of Biosimilars,” *Am. Heal. Drug Benefits*, vol. 6, no. 8, pp. 469–478, Nov. 2013.
- [11] X. Geng *et al.*, “Research and development of therapeutic mAbs : An analysis based on pipeline projects,” no. December, pp. 2769–2776, 2015.
- [12] Animal cell technology industrial platform, “Monoclonal Antibodies Approved by the EMA and FDA for Therapeutic Use (status 2017),” *biopharma.com*. [Online].

Available: <http://www.actip.org/products/monoclonal-antibodies-approved-by-the-ema-and-fda-for-therapeutic-use/>.

[13] J. M. Reichert, "Metrics for antibody therapeutics development," no. December, pp. 695–700, 2010.

[14] J. R. Birch and A. J. Racher, "Antibody production," *Adv. Drug Deliv. Rev.*, vol. 58, no. 5–6, pp. 671–685, 2006.

[15] N. Sethuraman and T. A. Stadheim, "Challenges in therapeutic glycoprotein production," 2003.

[16] S. Sha, C. Agarabi, K. Brorson, D. Lee, and S. Yoon, "N-Glycosylation Design and Control of Therapeutic Monoclonal Antibodies," *Trends Biotechnol.*, vol. 34, no. 10, pp. 835–846, Oct. 2016.

Chapter 2

Differential gene expression in CHO cells cultivated in shake flasks and bioreactors

This chapter is submitted as

Abdulaziz A. Alsayyari, Jos A. Hageman, Guido J. Hooiveld, Rene H. Wijffels, Dirk E. Martens, "Differential gene expression in CHO cells cultivated in shake flasks and bioreactors"

Abstract

Chinese Hamster Ovary cell lines are currently the primary host for production of therapeutic glycoproteins. Fast process development resulting in robust and scalable processes is a critical success factor in the highly competitive market for biosimilars. In process development screening of hundreds of clones and selection of process conditions are routinely performed in uncontrolled cultivation systems like shake flasks. A handful of potential candidate clones is nominated to be evaluated more intensively in well controlled small-scale bioreactors. Cell performance in the uncontrolled systems and to a lesser extent in the small-scale bioreactors may, however, be different from that in the final production reactor, which may result in failures during scale-up and thus extra development time. In this work, the focus is on better understanding the differences in cell performance between controlled and uncontrolled systems, which can be used to make process development faster and more robust in terms of scale-up. For this, we evaluated differences in gene expression profiles between shake flask and bioreactor cultures at three different time points during the exponential and stationary phase of a batch culture using commercially available Affymetrix CHO Gene 2.1 ST arrays and multivariate data analysis on the outcomes. The outcomes were correlated with differences in glycosylation patterns and other culture parameters. Results showed large differences in gene expression over time and much smaller differences between the two cultivation systems. Furthermore, our study identified differentially expressed genes and corresponding metabolic and mechanistic pathways that can be related to degree of the control of the system.

2.1 Introduction

Chinese hamster ovary (CHO) cells are currently the main host for production of biopharmaceutical proteins [1]. In the last 10 years, a large increase in volumetric productivity and product titers has been achieved [2]. Nevertheless, process development is still time-consuming, taking an average of 9-12 months [3], [4]. Due to patent expiration of major glycoprotein therapeutics, biosimilars are now being developed. For these potential biosimilars it is important to accelerate process development and achieve robust and scalable processes in a shorter time to meet the highly competitive and regulated markets. Early process development involves clone selection and finding the proper process conditions like pH, dissolved oxygen (DO), medium composition and for fed-batch processes feed composition as well as a feeding strategy. Part of the early process development is usually done in a high-throughput way in small-scale systems without pH and DO control, which results in a potential risk of failures due to scalability issues [2], [5]–[8]. For example, screening of clones can involve hundreds of clones and is for practical and logistic reasons often performed in batch cultivations in well plates or shake flasks without proper control of pH and DO. A few selected clones are then studied more intensively in well controlled lab-scale bioreactors in which, if applicable, also fed-batch cultivation is studied. In principle, these bioreactor studies should be done in such a way that no issues occur anymore during scale-up.

The conditions in the uncontrolled high-throughput systems used in early process development are different from the conditions in the controlled bioreactors in the final production process. Consequently, clones that give good results in the uncontrolled systems may fail in the small-scale bioreactor studies or in the final production process due to an unexpected drop in productivity and titer or changes in the glycosylation pattern. Furthermore, clones that would have done well at the manufacturing scale may not be selected.

A proper understanding of how culture conditions and culture systems affect cell physiology and how this will affect the performance of the clone in the final production system will reduce the process development time, result in selection of better clones

for the final production process and decrease the risk of failure during scale-up. One way to obtain more insight into the physiology of CHO cells is by measuring gene expression profiles and correlate this to cell performance and culture conditions. Analysis of gene expression profiles can help to mechanistically link changes in process conditions to important process outputs like product titer and quality. It can also be used as a fingerprinting tool where certain expression patterns are linked to a good or bad clone or process. Finally, it is possible that this fingerprint can be reduced to a few marker genes, which would make analysis simpler and faster. There are some examples of previous studies that integrate transcriptomic profiling in CHO cell lines into process development, such as identifying genes that are related to the growth rate, or other process characteristics [9]. However, the differences in gene expression between uncontrolled shake flasks and controlled bioreactors during the time course of cultivation process has not been thoroughly studied, since all of the previous studies used only a limited CHO cDNA microarray either customized or in-house [9], [10]. Most of these studies are either performed in shake flasks or in bioreactors. Only a few of these studies are performed in both systems, however without direct comparison of the systems. For example Clarke et al. (2011) did large-scale microarray profiling experiments associated with growth and productivity in 123 shake flasks, and 172 bioreactors using proprietary CHO-specific WyeHamster2a oligonucleotide arrays on which only 10-15 % of the CHO transcriptome is represented [11].

The aim of this work is to study how representative uncontrolled batch cultures in shake flasks are for controlled bioreactors on the level of gene expression and cell culture characteristics like cell number, viability, glucose consumption, lactate metabolism, product titer and product glycosylation. For this, gene expression profiling was performed in an IgG-producing CHO-K1 cell line cultivated in batch mode in shake flasks and bioreactors using the commercially available Affymetrix GeneChip CHO Gene 2.1 ST arrays. The expression design of this array provides the most comprehensive coverage of the transcribed genome of CHO cells currently available.

2.2 Materials and Methods

2.2.1 Cell line and culture medium

CHO-k1 IgG-secreting cell lines BC® were provided by Bioceros Holding BV, The Netherlands. All the cultures used chemically defined CD FortiCHO™ Medium (Gibco®, Life Technologies), supplemented with 4 mM L-glutamine and 0.5% (v/v) anti-clumping agent (both from Gibco®, Life Technologies). Pre-cultures were maintained in a 125 mL un-baffled shake flask (VWR, USA) with 20 ml working volume supplemented with selection reagents (200 µg/mL Zeocin™ and 5 µg/mL Blasticidin, both from Life technologies) for five passages prior to the experiments.

2.2.2 Bioreactor and incubation conditions

Bioreactor batch culture conditions were kept at 37°C with a working volume of 600 ml in a 1 litre DasGip parallel bioreactor systems (DasGip, an Eppendorf Company, Germany). pH was kept constant at 7.2 using CO₂ sparging combined with 0.5 M NaOH. Agitation was kept at 100 rpm. Sparging with a mixture of oxygen and nitrogen was used to maintain the DO at 50 % air saturation. Total of 10 ppm concentration of Antifoam (Antifoam C Emulsion, Sigma-Aldrich®) was added to the bioreactors. Addition was done when needed. For the cultivation in a batch shake flask, the working volume was 60 ml in 250 ml total volume (VWR, USA) in an incubator (Multitron CO₂ incubator; Infors HT) operated with 90% humidity, 8% CO₂, and 37 °C at 100 rpm rotational speed and 50 mm orbital shaking diameter. Inocula for the batch shake flasks and the bioreactors were prepared from one pool of cells (shake flask). The starting cell density in both systems was 0.3×10^6 cells.ml⁻¹.

2.2.3 Analytical methods

2.2.3.1 Growth rate and metabolic nutrients

Total cell density, viable cell density and cell diameter were measured daily using an automated cell counter (TC20™; BIO-RAD) using the trypan blue dye (Sigma-Aldrich) exclusion method. Glucose and lactate were measured using an YSI analyser (YSI

2700, YSI Life Sciences). All samples were filtered through a 0.2 µm filter (Minisart, Sartorius™) and the supernatant was stored at -20 °C for further analysis.

2.2.3.2 Quantitative measurement for IgG titrations

The concentration of the IgG1 was determined by Ultra High-Performance Liquid Chromatography UHPLC with a bio-monolith protein A column from Agilent technologies according to biopharmaceutical application note from Agilent (Urbas et al., 2008).

2.2.3.3 Glycosylation structure analysis

For determining the composition of the N-glycan structures from samples taken from both shake flasks and bioreactors, the method developed by Waters® [13] was used. In short, filtered supernatant was desalted by PD-10 columns (GE Healthcare, USA) and then purified using 1 mL HiTrap MabSelect SuRe column (GE Healthcare, USA) in combination with an Äkta pure system Unicorn 6.3 (GE Healthcare). Next, the IgG was concentrated with Amicon Ultra 0.5 mL centrifugal filters (Sigma-Aldrich, USA) and N-glycans were released using peptide N-glycosidase F, called PNGase F solution (Sigma-Aldrich, USA). Glycans were labelled with fluorescence active 2-Aminobenzamide (2-AB), followed by post-labelling clean-up of glycan released by hydrazinolysis with module cartridges GlycoPrep® (Prozyme, USA) and elution of N-glycans. The labeled glycans were analyzed using HILIC UPLC with a column from Waters® [13].

The overall galactosylation and fucosylation were calculated as follows [14].

$$Galactosylation = \frac{G1 + G1F + 2 \times (G2 + G2F)}{[2(G0 + G0F + G1 + G1F + G2 + G2F)]}$$

$$Fucosylation = \frac{G0F + G1F + G2F}{(G0 + G0F + G1 + G1F + G2 + G2F)}$$

2.2.3.4 RNA extraction and sampling for gene expression profiling by array

Samples (3 ml) were taken from both the bioreactors and the shake flasks, centrifuged at 300 g for 10 minutes, after which the cell pellet was snap-frozen in liquid nitrogen

and stored at -80 °C for later analysis. Total RNA was extracted using TRIzol® Reagent (Life Technologies, USA), and further purified using the RNeasy Mini kit (Qiagen, Valencia, CA). RNA concentration was measured on a Nanodrop 2000 (Thermo Scientific, Wilmington, DE), and RNA quality was assessed on a 2100 Bioanalyzer (Agilent, Santa Clara, CA) using the RNA 6000 Nano Kit. RNA samples were only included for microarray analysis if the RNA integrity number (RIN) was > 8.

2.2.4 Gene expression profiling and analysis

Affymetrix CHO Gene 2.1 ST arrays were used for transcriptome profiling (Affymetrix, Santa Clara, USA). Briefly, 100 ng of total RNA was labelled by the Whole-Transcript Sense Target Assay (Affymetrix) and hybridized to Affymetrix CHO Gene 2.1 ST arrays, as per the manufacturer's recommendations. Quality control and data analysis pipeline have been described in detail previously [15]. Briefly, normalized expression estimates of probe sets were computed by the robust multiarray analysis average (RMA) algorithm [16], as implemented in the Bioconductor library AffyPLM. Probe sets were redefined according to Dai et al. [17] using well-annotated reference sequences based on the CriGri_1.0 genome assembly (NCBI Reference Sequence Project (RefSeq) Release 72), which resulted in the profiling of 60626 annotated sequences (transcripts) (custom CDF v20). After averaging the expression levels of probe sets targeting the same gene, expression data for 20859 unique genes was obtained, which was used for all subsequent analyses. Principle component analysis (PCA) was used to visually inspect the biological significance and the characteristics of the data set of gene expression changes [18]. Differentially expressed genes were identified by using linear models (library limma) and an intensity-based moderated t-statistic [19], [20]. Correlation due to the repeated sampling from the same bioreactors or shakers flasks was incorporated in the linear model fit, and hence into all tests for differential gene expression, by applying a strategy similar to fitting a random effects model, with the difference that all genes were constrained to share the same intra-block correlation. This consensus correlation was calculated by limma's duplicateCorrelation() function [19], [21]. P values were corrected for multiple testing using a false discovery rate (FDR) method [22]. Genes that satisfied the criterion of $FDR < 0.05$ and absolute fold change > 1.4 were considered to be significantly regulated. All genes that were significantly differentially expressed between day 4 and day 2 (d4-d2), day 5 and day

4 (d5-d4), and day 5 and day 2 (d5-d2) were selected. Genes selected from the comparison of d4-d2 and d5-d4, are those that are strongly up or down regulated in one of these two comparisons or they are strongly transiently regulated, meaning first up and then down or vice versa. The d5-d2 comparison includes the genes that are slowly up or down regulated in time, such that they are not significant in both of the other comparisons. The selected genes were next subdivided in three groups of genes being 1. genes uniquely regulated in shake flask, 2. genes uniquely regulated in bioreactor and 3. genes regulated in both systems at the same time but in none of the three comparisons regulated uniquely. All these significant regulated genes groups were subjected to pathway overrepresentation analysis using Fisher's exact test and CHO-specific pathway information derived from the Kyoto Encyclopaedia of Genes and Genomes (KEGG).

2.3 Results and Discussions

2.3.1 Cell culture characteristics

IgG producing CHO cells were cultivated in shake flasks and bench top bioreactors in batch mode in triplicate. The three shake flasks and three bioreactors were all inoculated from the same preculture at an initial concentration of 3.0×10^5 cell/ml. In Figure 1 the average values for the total and viable cell density, glucose, and lactate concentration, product concentration and the integral of the viable cell density as a function of time for the shake flask and bioreactor are shown. The integral of the viable cell density is shown because many parameters like total product formed or total essential amino acid consumed are proportional to this characteristic. As can be seen in Figure 1 there are no differences between these parameters for the shake flasks and the bioreactors, despite the fact that in the shake flasks, as opposed to the bioreactors, pH and DO are not controlled and, as a consequence, the cells may experience larger fluctuations with respect to these parameters in the shake flasks.

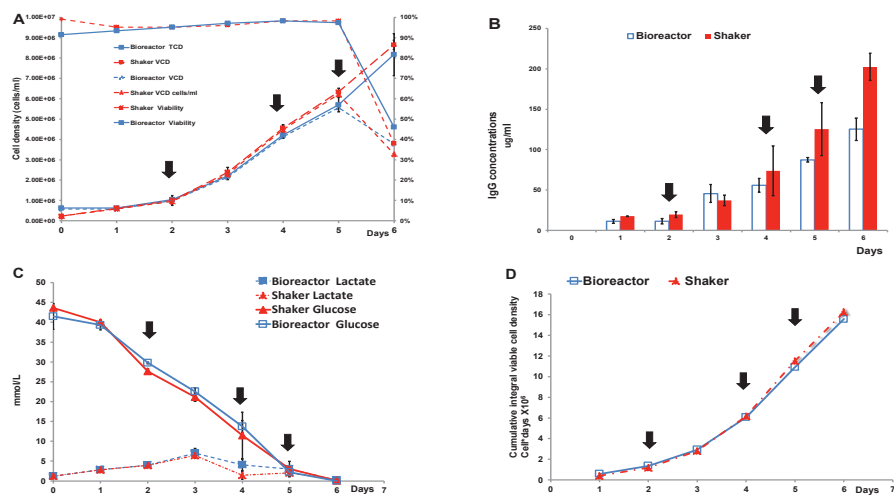


Figure 1 Comparison of cell performance between shake flask and bioreactor; A: Total and viable cell concentration, B. Antibody concentration, C. Lactate, and glucose concentration, D. Cumulative integral viable cell density. Data points are the average of 3 replicate shake flask and bioreactor runs. Error bars show the standard deviation.

Glycosylation patterns are considered critical quality attributes in therapeutic IgG products [23], [24]. Any alterations or changes in the structure of the glycans can lead to adverse reactions in the patient and impact therapeutic efficacy [25]. Therefore, in upstream process optimisation, glycosylation patterns are a major concern [24]. The glycosylation patterns in both bioreactor and shake flask are summarized in Figure 2. It shows a significantly higher percentage of G0F for the shake flask, while the G1F and G2F glycoforms are significantly lower than in the bioreactor. As a consequence, also the overall galactosylation level was significantly lower in the shake flask (19%) than in the bioreactor (27%). The fucosylation level was identical (81%) for both systems. These differences in glycosylation patterns may be either directly or indirectly caused by the absence (shake flasks) or presence (reactors) of DO and pH control. Other studies have shown that glycosylation patterns are affected by process conditions such as pH, DO, temperature, and others [26].

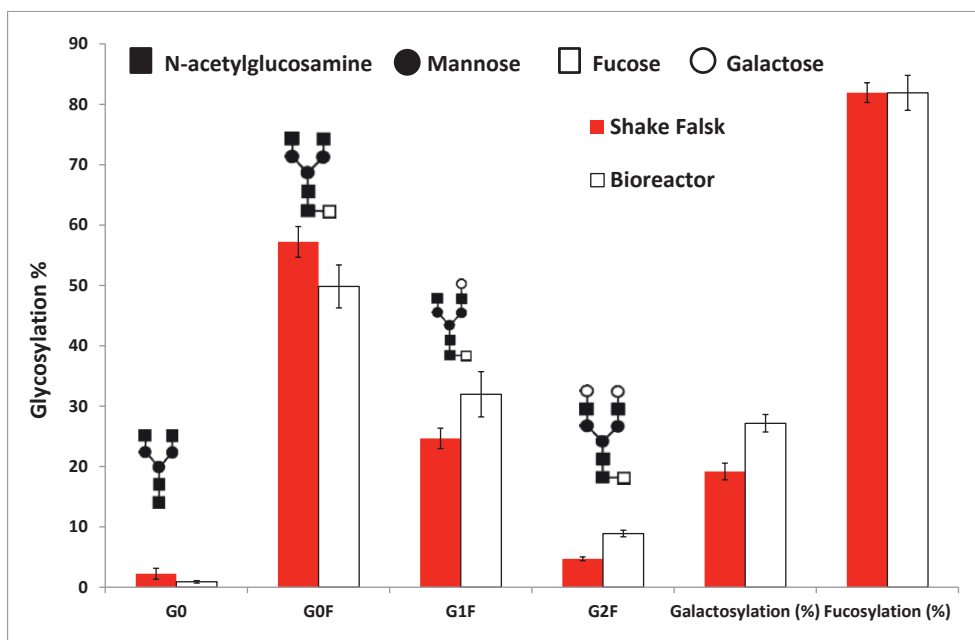


Figure 2 Average glycosylation patterns for the shake flask and the bioreactor. Error bars show the standard deviation of 3 replicate shake flask and bioreactor runs.

2.3.2 Differential gene expression between both systems

In order to study differences in cell physiology between cells cultivated in batch mode in uncontrolled shake flasks and in controlled bioreactors, we studied differential gene expression. RNA samples were taken during the exponential growth phase (day 2), just after the end of the exponential phase (day 4) and at the peak cell density just before the death phase (day 5) for both systems in triplicate and analyzed using a full genome CHO microarray. The complete dataset from the microarray experiment is available at gene expression omnibus (GEO) under accession number **GSE104787**. Figure 3 shows the result of principal component analysis (PCA) on the obtained expression data. The first two principal components explain 42% of the total variation. Principal component 1 captured the changes in gene expression as a function of culture time, i.e., progression of the culture from the exponential phase to the non-exponential phase and finally to the death phase. Principal component 2 seemed to be associated with genes that were transiently regulated to deal with the changes in the culture conditions resulting in the stop of cell growth. On day 2 both culture systems

are grouped together meaning gene expression profiles were highly similar. This is to be expected as medium conditions, like pH, and nutrient concentrations, will still be comparable in both systems and differences in agitation apparently have no big influence on gene expression during the first two days. As culture time progressed to day 4 and 5, shaker and bioreactor cultures are more separated although the distance between the groups remained relatively small. The separation could be due directly to differences between the systems or indirectly due to a small difference in the speed with which the cultures develop. However, based on viable and total cell concentrations cells seem to grow equally fast in both systems and also have a comparable glucose metabolism (figure 2). To obtain more insight in the small difference between the two systems, the expression data are studied in more detail.

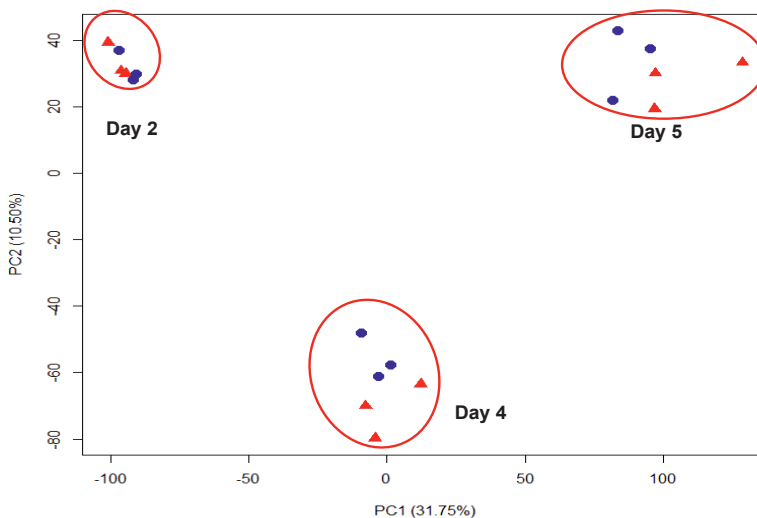


Figure 3 PCA plot of gene expression for the first two principal components. Different culture days are shown with circles for day 2, day 4, and day 5. For 3 shake flask (triangle red) and 3 bioreactor (circle blue) cultures.

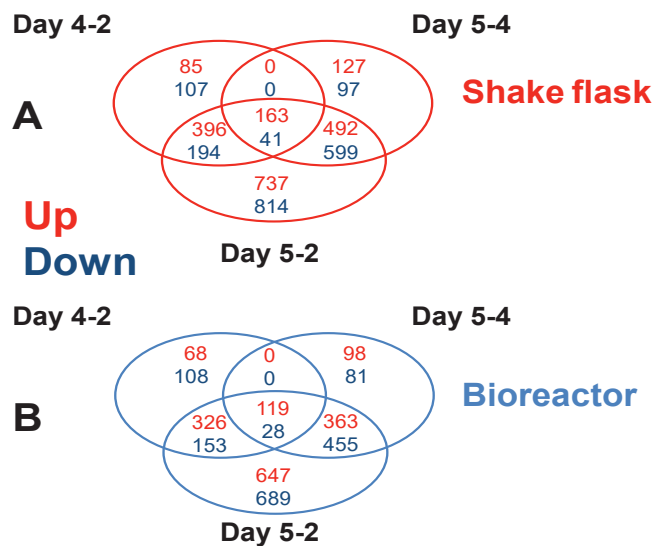


Figure 4 Venn diagrams showing common and uniquely differentially expressed genes between different days for the shake flask (A) and the bioreactor (B). Applying ($\text{FDR-BH} \leq 0.05$ and absolute fold change > 1.4).

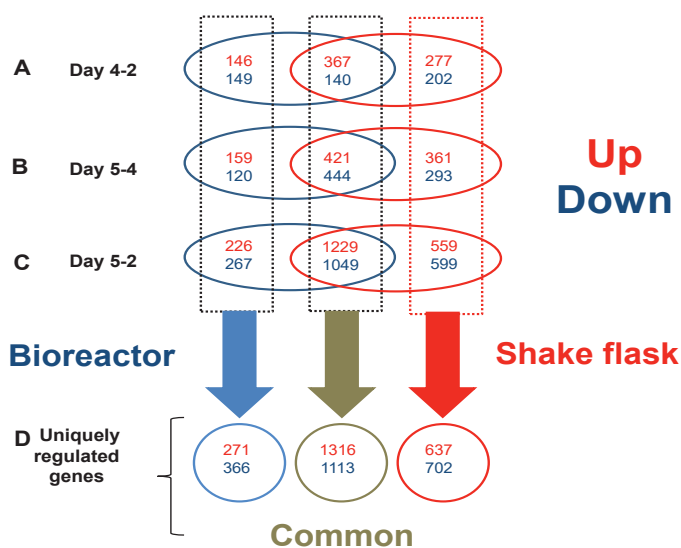


Figure 5 Venn diagrams showing uniquely and overlapping differentially expressed genes (FDR-BH ≤ 0.05 and absolute fold change > 1.4 for both systems) for day 4-2 (A), day 5-4 (B), and day 5-2 (C) respectively. (D) Genes that are differentially expressed only in the bioreactor and only in the shaker at any time and genes that are only differentially expressed in both systems at any time (common).

2.3.3 Differential pathways expressions

Figure 4 shows for both the bioreactors and shake flasks the number of differential expressed genes for the three time comparisons. The number of genes regulated are comparable for day 4-2 and day 5-4. However, there is a relatively large number of slowly changing genes as can be seen in the day 5-2 comparison. To understand the differences in gene expression between the systems all the regulated genes in both systems and time points were divided into three categories as shown in Figure 5. Bioreactor, common and shake flask. Genes unique to the bioreactor or shaker are differentially expressed in the respective system in at least one of the time comparisons but never in the other system. Genes in the common category are differentially expressed in both systems for one of the time comparisons. The genes present in the three categories were next mapped onto CHO-specific pathways using the KEGG pathway database. The shake flasks had more uniquely regulated genes and these genes are distributed into a larger number of pathways as compared to the bioreactors.

This could be a reflection of the more variable conditions in the shake flasks. The unique genes expressed in the shake flasks could not be linked to one or a few specific pathways and thus cannot directly linked to differences in process conditions between both systems. Apparently, the absence of pH and DO control in the shake flasks is affecting many pathways. Complete pathways data can be found in the supplementary data file.

Table 1 shows an overview of pathways involved in the basic processes related to growth being DNA replication and repair, transcription, translation and turnover of RNA, and the folding, sorting and degradation of proteins. These are the pathways in which a change in gene expression is expected because they are involved in cellular processes that change when cells pass from the exponential into the stationary phase. As can be seen in Table 1 indeed most regulation occurs in the common category meaning regulation occurs in both the bioreactor and the shake flask. Furthermore, mostly it concerns down regulation of genes, which agrees with the stop of growth that occurs upon the phase transition from the exponential to the stationary phase [27]. Only for protein processing also significant up regulation is observed. For the bioreactor system, there is almost no additional regulation of genes compared to the common category. For the shake flask, however there is clearly down regulation of additional genes in, RNA transport, mRNA surveillance pathway, ribosome biogenesis in eukaryotes and RNA degradation.

Table 2 shows pathways involved in glycan biosynthesis and metabolism. Generally, in both systems N-glycan biosynthesis is down regulated and degradation is upregulated. Remarkably, in the shake flasks additional genes are upregulated in the N-glycan biosynthesis, which is opposite the common trend and also opposite the bioreactor, which shows additional genes being down-regulated in this pathway. However, no clear link with the decreased galactosylation in the shake flask cultures can be made.

Table 1 Processing transcriptional, translation, folding, sorting, and degradation pathways that are significantly uniquely regulated in shake flask, bioreactor, and common regulated as showing in figure 5.D. (FDR-BH < 0.05 and fold change > ± 1.4 for significant genes). Asterisk is equal to a p-value < 0.05 of over- representation of the KEGG pathways.

Biological processes and corresponding pathways		Number of genes in pathway	Number of genes changed in pathway Figure 5.D										
Genetic information processing			Common			Bioreactor			Shake flask				
			All	Up	Down	All	Up	Down	All	Up	Down		
Transcriptional													
RNA polymerase		31	0	0	0	0	0	0	2	0	2		
Basal transcription factors		44	4	2	2	2	0	*2	0	0	0		
Translation													
Ribosome		236	7	1	*6	1	0	1	7	2	5		
RNA transport		185	*33	2	*31	3	1	2	*21	3	*18		
mRNA surveillance pathway		99	*10	2	*8	2	1	1	*14	1	*13		
Ribosome biogenesis in eukaryotes		90	*12	2	*10	*7	0	*7	*13	0	*13		
Folding, Sorting, and Degradation													
Protein export		27	2	1	1	0	0	0	1	0	1		
Proteasome		48	3	0	*3	1	0	1	*4	1	*3		
Sulfur relay system		8	0	0	0	0	0	0	0	0	0		
SNARE interactions in vesicular transport		33	2	2	0	*3	*3	0	2	1	1		
Protein processing in endoplasmic reticulum		172	*21	*14	*8	*3	0	3	5	3	2		
RNA degradation		79	*18	3	*15	0	0	0	*8	1	*7		
Replication and repair													
DNA replication		38	*29	2	*27	0	0	0	*4	0	*4		
Base excision repair		41	*13	2	*11	2	0	*2	*5	*3	2		
Nucleotide excision repair		45	*17	3	*14	0	0	0	*4	0	*4		
Mismatch repair		25	*17	*3	*14	0	0	0	*3	0	*3		
Homologous recombination		42	*23	2	*21	1	0	1	*3	0	*3		

Table 2 Glycan biosynthesis and metabolism pathways that are significantly uniquely regulated in shake flask, bioreactor, and common regulated as showing in figure 5.D. (FDR-BH < 0.05 and fold change > \pm 1.4 for significant genes). Asterisk is equal to a p-value < 0.05 of over-representation of the KEGG pathways.

	Number of genes in pathway	Number of genes changed in pathway Figure 5.D									
		Common				Bioreactor				Shake flask	
		All	Up	Down	All	Up	Down	All	Up	Down	All
Glycan biosynthesis and metabolism											
N-Glycan biosynthesis	51	*7	2	*5	2	0	*2	*4	*3	1	
Other glycan degradation	23	*8	*8	0	1	1	0	0	0	0	
Mucin type O-Glycan biosynthesis	31	2	1	1	1	1	0	1	0	1	
Other types of O-glycan biosynthesis	21	2	1	1	0	0	0	2	*2	0	
Glycosaminoglycan degradation	25	*8	*8	0	0	0	0	*3	*2	1	
Glycosaminoglycan biosynthesis											
Chondroitin sulfate / dermatan sulfate	21	*4	2	2	1	1	0	2	1	1	
Keratan sulfate	15	2	*2	0	0	0	0	1	1	0	
Heparan sulfate / heparin	23	*4	*4	0	0	0	0	1	1	0	
Glycosylphosphatidylinositol (GPI)	26	1	1	0	1	1	0	2	1	1	
Glycosphingolipid biosynthesis											
Lacto and neolacto series	32	0	0	0	1	0	1	*4	*2	*2	
Globo and isoglobos series	19	*6	*6	0	0	0	0	0	0	0	
Ganglio series	17	*5	*5	0	0	0	0	0	0	0	

Table 3 General metabolism pathways that are significantly regulated in uniquely regulated in shake flask, bioreactor, and common regulated as showing in figure 5.D. (FDR-BH < 0.05 fold change > ± 1.4 for significant genes). Asterisk s equal to a p-value < 0.05 of over- representation of the KEGG pathways.

	Number of genes in pathway	Number of genes changed in pathway Figure 5.D									
		Common				Bioreactor					
		All	Up	Down	All	Up	Down	All	Up	Down	Shake flask
Global and overview maps											
Fatty acid metabolism	57	*6	1	*5	1	1	0	3	*3	0	
Biosynthesis of amino acids	81	*12	*10	2	*4	1	*3	3	1	2	
Carbohydrate metabolism											
Glycolysis / Gluconeogenesis	68	5	*5	0	1	0	1	2	1	1	
Citrate cycle (TCA cycle)	33	1	1	0	0	0	0	2	0	*2	
Pentose and glucuronate interconversions	31	2	1	1	*2	1	1	2	1	1	
Pentose phosphate pathway	34	*5	*4	1	*4	1	*3	0	0	0	
Fructose and mannose metabolism	35	*5	2	3	1	1	0	2	1	1	
Galactose metabolism	33	*4	*4	0	1	0	1	2	0	2	
Ascorbate and aldarate metabolism	25	1	1	0	1	0	1	1	0	1	
Starch and sucrose metabolism	32	*4	*3	1	1	1	0	1	0	1	
Amino sugar and nucleotide sugar	54	*7	*6	1	*4	*2	*2	1	1	0	
Pyruvate metabolism	41	5	3	2	1	0	1	1	1	0	
Glyoxylate and dicarboxylate metabolism	32	6	1	5	0	0	0	2	*2	0	
Propanoate metabolism	34	3	1	2	1	0	1	2	*2	0	
Butanoate metabolism	29	*7	*3	*4	*3	0	*3	*3	*3	0	
Inositol phosphate metabolism	71	*10	*7	3	0	0	0	*6	*4	2	
Energy metabolism											
Oxidative phosphorylation	131	3	1	2	*5	2	*3	3	2	1	
Nitrogen metabolism	19	1	1	0	0	0	0	1	1	0	

Table 3 (continued).	Number of genes in pathway	Number of genes changed in pathway Figure 5.D									
		Common			Bioreactor			Shake flask			
		All	Up	Down	All	Up	Down	All	Up	Down	
Lipid metabolism											
Fatty acid biosynthesis	15	*4	*2	*2	0	0	0	1	1	0	0
Fatty acid elongation	30	3	1	2	*2	*2	0	0	0	0	0
Fatty acid degradation	50	*5	2	3	1	0	1	*4	*4	0	0
Synthesis and degradation of ketone bodies	14	*5	*2	*3	*2	0	*2	*3	*3	0	0
Steroid biosynthesis	21	*11	1	*10	0	0	0	1	0	1	1
Primary bile acid biosynthesis	17	*3	*3	0	0	0	0	2	0	2	2
Steroid hormone biosynthesis	73	4	3	1	*3	1	2	*5	2	*3	*3
Glycerophospholipid metabolism	105	*12	*7	*5	2	2	0	*6	*5	*1	*1
Glycerolipid metabolism	67	*6	*6	0	0	0	0	*5	*3	2	2
Ether lipid metabolism	55	*7	*5	2	1	1	0	*3	*3	0	0
Sphingolipid metabolism	53	*9	*9	0	1	1	0	2	1	1	1
Arachidonic acid metabolism	99	*15	*11	4	1	1	0	*9	*4	*5	*5
alpha-Linolenic acid metabolism	35	3	2	1	1	1	0	1	1	0	0
Biosynthesis of unsaturated fatty acids	34	1	0	1	1	1	0	1	0	1	1
Nucleotide metabolism											
Purine metabolism	184	*21	7	*14	*5	1	*4	*10	4	*6	*6
Pyrimidine metabolism	104	*24	*6	*18	2	0	2	*7	2	*5	*5

Genes regulated in metabolism of carbohydrate, energy, lipid and nucleotides are listed in table 3. Biosynthesis of amino acids pathway has a significant number of up-regulated genes in both the bioreactor and the shake flask (common), which may point to a shortage of one or more amino acids upon the transition to the stationary phase. In carbohydrate metabolism only a small number of genes are significantly up-regulated in both systems. Furthermore, there are almost no genes additionally regulated only in the bioreactor or only in the shake flask. Likely the regulation of these pathways is more on a metabolite level and not transcriptional. Inositol phosphate metabolism is significantly enriched in upregulated genes in both systems and also additional genes are upregulated in this pathway for the shake flasks. Inositol phosphate is involved in signalling pathways. Pathways in lipid metabolism are significantly enriched in regulated genes in both systems, with genes being mostly upregulated. An exception is the steroid biosynthesis pathway, which is enriched with genes that are down regulated. When entering the stationary phase in principle no additional membranes have to be synthesized anymore and also no sterols are required explaining the down-regulation of this pathway. The fact that other pathways are upregulated may be related to remodelling of membranes and lipids upon entering the stationary phase. In the shake flask regulation of additional genes in these pathways can be seen, which is not the case for the bioreactor. As expected Purine and Pyrimidine metabolism pathways are down-regulated in both systems, which is in agreement with the down regulation of DNA replication and the stop of growth upon entering the stationary phase.

In general, most pathways are more enriched in up or down regulated genes for the shake flasks as compared to the bioreactor. The biggest difference between bioreactor and shake flask is found for RNA transport, mRNA surveillance pathway, ribosome biogenesis in eukaryotes and RNA degradation, which contains significantly more down regulated genes for the shake flask. A smaller difference is seen in lipid metabolism, which contains slightly more upregulated genes for the shake flask.

2.4. Conclusion

Currently bioprocess development is hampered by a lack of understanding on the cellular mechanism that determining process performance. Measuring and comparing gene expression can help to develop a scientific understanding of the process and though this contribute to a more rational, data driven process development for biopharmaceuticals in the near future. In this paper, a time series gene expression analysis is done to compare a batch culture of an IgG-secreting CHO cell line performed in uncontrolled shake flasks and controlled bioreactors. Cell growth, glucose and lactate metabolism, and product formation were comparable for both systems. For the product, the galactosylation level was slightly lower in the shake flask as compared to the bioreactor. For gene expression analysis, the Affymetrix GeneChip CHO Gene 2.1 ST arrays was used, which provides the most comprehensive coverage of the transcribed genome of CHO cells currently available. Overall the differences in gene expression between the shake flasks and bioreactors were small compared to the variation in gene expression over time. Nevertheless, more genes were regulated in shake flasks compared to bioreactors, which agrees with the absence of pH and DO control in the shake flask. The uniquely regulated genes in shake flask were, however, distributed into a large number of pathways and could not be linked directly to the differences in control between the systems.

2.5 References

- [1] S. Kyriakopoulos and C. Kontoravdi, "Analysis of the landscape of biologically-derived pharmaceuticals in Europe: dominant production systems, molecule types on the rise and approval trends.," *Eur. J. Pharm. Sci.*, vol. 48, no. 3, pp. 428–41, Feb. 2013.
- [2] F. Li, N. Vijayasankaran, A. (Yijuan) Shen, R. Kiss, and A. Amanullah, "Cell culture processes for monoclonal antibody production," *MAbs*, vol. 2, no. 5, pp. 466–479, Sep. 2010.
- [3] J. Schaub *et al.*, "CHO gene expression profiling in biopharmaceutical process analysis and design," *Biotechnol. Bioeng.*, vol. 105, no. 2, pp. 431–438, 2010.
- [4] S. M. Browne and M. Al-Rubeai, "Selection methods for high-producing mammalian cell lines," *Trends Biotechnol.*, vol. 25, no. 9, pp. 425–32, Sep. 2007.
- [5] C. S. Alves *et al.*, "Integration of cell line and process development to overcome the challenge of a difficult to express protein," *Biotechnol. Prog.*, vol. 31, no. 5, pp. 1201–1211, 2015.
- [6] J. R. Birch and A. J. Racher, "Antibody production.," *Adv. Drug Deliv. Rev.*, vol. 58, no. 5–6, pp. 671–85, Aug. 2006.
- [7] R. Bareither and D. Pollard, "A review of advanced small-scale parallel bioreactor technology for accelerated process development: Current state and future need," *Biotechnol. Prog.*, vol. 27, no. 1, pp. 2–14, 2011.
- [8] S. Moses and M. Manahan, "Assessment of AMBRTM as a model for high-throughput cell culture process development strategy," *Adv. Biosci. ...*, vol. 2012, no. November, pp. 918–927, 2012.
- [9] A. M. Lewis, N. R. Abu-Absi, M. C. Borys, and Z. J. Li, "The use of 'Omics technology to rationally improve industrial mammalian cell line performance.," *Biotechnol. Bioeng.*, vol. 113, no. 1, pp. 26–38, Jan. 2016.
- [10] J. Becker *et al.*, "Next-generation sequencing of the CHO cell transcriptome.," *BMC Proc.*, vol. 5 Suppl 8, no. Suppl 8, p. P6, 2011.
- [11] C. Clarke *et al.*, "Large scale microarray profiling and coexpression network analysis of CHO cells identifies transcriptional modules associated with growth and productivity," *J. Biotechnol.*, vol. 155, no. 3, pp. 350–359, 2011.

- [12] B. M. Urbas L, Brne P, Gabor B, Plevcak S, "Rapid Human Polyclonal IgG Quantification Using the Agilent Bio-Monolith Protein A HPLC Column Application Note," *Agil. Technol. Inc.*, 2008.
- [13] M. Hilliard *et al.*, "Development of a Glycan Database for Waters ACQUITY UPLC Systems," *Waters Appl. Note 720004202en*, pp. 1–7, 2012.
- [14] M. J. Gramer *et al.*, "Modulation of antibody galactosylation through feeding of uridine, manganese chloride, and galactose," *Biotechnol. Bioeng.*, vol. 108, no. 7, pp. 1591–1602, Jul. 2011.
- [15] K. Lin *et al.*, "MADMAX - Management and analysis database for multiple -omics experiments.," *J. Integr. Bioinform.*, vol. 8, no. 2, p. 160, Jul. 2011.
- [16] R. A. Irizarry *et al.*, "Exploration, normalization, and summaries of high density oligonucleotide array probe level data," *Biostatistics*, vol. 4, pp. 249–264, 2003.
- [17] M. Dai *et al.*, "Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data," *Nucleic Acids Res.*, vol. 33, no. 20, pp. 1–9, 2005.
- [18] M. Ringnér, "What is principal component analysis?," *Nat. Biotechnol.*, vol. 26, no. 3, pp. 303–4, Mar. 2008.
- [19] M. E. Ritchie *et al.*, "limma powers differential expression analyses for RNA-sequencing and microarray studies.," *Nucleic Acids Res.*, vol. 43, no. 7, p. e47, Apr. 2015.
- [20] M. A. Sartor, C. R. Tomlinson, S. C. Wesselkamper, S. Sivaganesan, G. D. Leikauf, and M. Medvedovic, "Intensity-based hierarchical Bayes method improves testing for differentially expressed genes in microarray experiments.," *BMC Bioinformatics*, vol. 7, p. 538, Dec. 2006.
- [21] G. K. Smyth, J. Michaud, and H. S. Scott, "Use of within-array replicate spots for assessing differential expression in microarray experiments," *Bioinformatics*, vol. 21, no. 9, pp. 2067–2075, 2005.
- [22] Y. Benjamini and Y. Hochberg, "Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. R. Stat. Soc. B*, vol. 57, no. 1, pp. 289–300, 1995.
- [23] D. Reusch and M. L. Tejada, "Fc glycans of therapeutic antibodies as critical quality attributes.," *Glycobiology*, vol. 25, no. 12, pp. 1325–34, Dec. 2015.

- [24] Y. Fan *et al.*, "Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation.," *Biotechnol. Bioeng.*, vol. 112, no. 3, pp. 521–35, Mar. 2015.
- [25] R. Jefferis, "Recombinant antibody therapeutics: the impact of glycosylation on mechanisms of action," *Trends Pharmacol. Sci.*, vol. 30, no. 7, pp. 356–362, Jul. 2009.
- [26] P. Hossler, S. F. Khattak, and Z. J. Li, "Optimal and consistent protein glycosylation in mammalian cell culture," *Glycobiology*, vol. 19, no. 9, pp. 936–949, Sep. 2009.
- [27] J. A. H. Bort *et al.*, "Dynamic mRNA and miRNA profiling of CHO-K1 suspension cell cultures," *Biotechnol. J.*, vol. 7, no. 4, pp. 500–515, 2012.

Chapter 3

Shake flask as scale down model for a fed-batch controlled bioreactor: a transcriptome comparison

This chapter is submitted as

Abdulaziz A. Alsayyari, Jos A. Hageman, Guido J. Hooiveld, Rene H. Wijffels, Dirk E. Martens, “Shake flask as scale down model for a fed-batch controlled bioreactor: a transcriptome comparison”.

Abstract

Fed-batch is the main used process mode for the production of glycoprotein therapeutics, due to the high cell density, volumetric productivity and the fact that it is well-accepted from a regulation perspective. Screening clones and process parameters, like for example feed composition and addition for fed-batch process development is often done in disposable shake flasks. However, critical process parameters like pH and DO are not controlled in shake flasks, which may lead to changes in process performance at scale-up. Here we study the differences in gene expression for a CHO cell fed-batch process performed in either a shake flask or a controlled bioreactor. Overall the cells in the shake flask grew faster than in the reactor and entered the stationary and death phase earlier. Moreover, also the glycosylation pattern was different for both systems. With respect to gene expression PCA was able to discriminate between two types of differences being that growth was faster in the shake flask as compared to bioreactors (represented by PC1, 28.50%) and a difference between both systems that was independent from time (represented by PC2, 10.95%). Apart from being a sensitive tool to detect differences in cell physiology, transcriptome data can also identify differentially expressed pathways, which may lead to the root cause of the differences. Differences related to regulation of pathways were especially related to the fact that shake flask cultures entered the stationary and death phase earlier, with down regulation of growth related processes and the up regulation linked to arrest of growth and stress in shake flasks.

3.1 Introduction

Monoclonal antibodies form an important and rapidly growing class of therapeutic glycoproteins [1]. They are used to treat complex diseases, such as cancer and autoimmune disease [2]. The majority of the regulatory approved therapeutic monoclonal antibody products are produced in fed-batch with the Chinese Hamster Ovary (CHO) cell line as the primary host cell line. Fed-batch cultivation allows for high cell densities and product titres and at the same time gives a well-defined batch period, which is an advantage from the regulation perspective [3].

The development of fed-batch processes relies on small scale, high throughput experiments to define the process design space. In the first step generally screening experiments are done for the selection of a few suitable clones, media and feed combinations and the determination of critical process parameters. These experiments are often done in shake flasks. Shake flasks are easy to use, relatively cheap, allow for high throughput experimentation and are flexible with respect to the number of experiments. At a certain moment during development, depending on the number of experiments that need to be done, a switch is made to controlled bench-scale bioreactors to more precisely define the design space and confirm the outcome of the shake flask experiments. A major issue in process development is that the conditions in the shake flasks are not representative of those in the controlled bench-scale bioreactor and the conditions in the production scale reactor. This leads to a difference in process performance between the shake flasks, the bench scale bioreactors and the final production scale, which in turn may lead to a delay in process development and/or suboptimal processes.

The major difference between shake flasks and controlled bioreactors is in the absence of pH and DO control in the shake flask. Thus, cells in shake flasks are exposed to varying oxygen concentrations ranging from a dissolved oxygen tension of 100% air saturation to ultimately oxygen limitation depending on the cell concentration reached and volume used in the flask. Likewise, the pH will vary during the process depending on the level of lactate production and or consumption and the CO₂ levels. Addition of feed with, for example, a relatively high pH will cause additional pH fluctuations during the run. A second important difference is in the absence of sparging in the shake flasks.

This results in differences in gas transfer and may result in oxygen limitation as mentioned before, but will also cause differences in CO₂ levels and in cell death related to sparging. Finally, there is a difference in agitation, which may cause differences in mixing and in shear forces. Differences between shake flasks and controlled bioreactors with respect to the effect they have on cell physiology can be assessed in different ways as for example on a metabolic level or using transcriptome analysis.

There are a number of studies where transcriptome analysis was done for CHO cell lines producing a high-value protein therapeutic. Vishwanathan et al. (2014) studied the transcriptome with the aim to find targets for genetic engineering, to design superior production cell lines. A limited number of transcriptome studies were done to obtain more insight into the effect of changes in cultivation conditions, such as temperature, on cell physiology [4]–[6]. However, these comparisons were done on a single time point and thus did not take into account the culture dynamics [7], [8]. For example, if one culture develops faster than another comparison on one-time point is not valid anymore. For dynamic processes like batch and fed-batch transcriptome analysis has to be done on a series of time points spread over the culture period to obtain true insight in cell physiology and its dynamics. With the sequencing of the CHO cell genome and the availability of a CHO cell specific microarray transcriptome analysis in CHO cells could be used to study biological processes in the cell in response to process conditions [9], [10].

In our previous study, we showed that for a batch cultivation process only minor differences in gene expression were present between shake flasks and controlled bioreactors (Chapter 2). However, this may be different for fed-batch processes as cell concentrations reached are much higher, which for example may easily result in larger differences in DO, CO₂ concentration and pH between both systems. In addition, a feed is added which may result in nutrient and pH fluctuations that affect cell physiology. Despite the importance of shake flask experiments as part of the current process development practice, very little is known about the effect of the differences between shake flasks and controlled bioreactors on cell physiology. This limited knowledge increases the risk of unexpected cell performance at a later stage of process development requiring additional experiments or in case a bad clone was

selected starting at the clone selection again, which both mean a serious delay in process development. Therefore, the aim of this study is to obtain insight in the effect of the differences between shake flasks and controlled bioreactors on cell physiology for a fed-batch process. This is done by studying gene expression and link this to other measured parameters like cell metabolism. For this, gene expression profiling was performed using the commercially available Affymetrix GeneChip CHO Gene 2.1 ST array in an IgG-producing CHO-K1 cell line cultivated in fed-batch mode in shake flasks and bioreactors.

3.2 Materials and Methods

3.2.1 Cell line and fed-batch culture

A CHO-k1 BC[®] IgG-producing cell line was provided by Bioceros (Bioceros Holding BV, The Netherlands). The chemically defined CD FortiCHO[™] Medium (Gibco[®], Life Technologies) complemented with 4 mM L-glutamine and 0.5%(v/v) anti-clumping agent (both from Gibco[®], Life Technologies) was used for all cultures. Pre-cultures were maintained in a 125 mL un-baffled shake flask (VWR,USA) with 20 mL working volume supplemented with selection reagents (200 µg/mL Zeocin[™] and 5 µg/mL Blasticidin[™], both from Life technologies) for five passages prior to inoculation into 3 replicate 250 mL un-baffled shake flask (VWR,USA), (initial working volume 30 mL) and 3 replicate 1 litre DasGip Parallel bioreactors (DasGip, Eppendorf Company, Germany), (initial working volume 500 ml). An inoculation density of $0.3\text{--}0.4 \times 10^6$ cells mL^{-1} was used for the three shake flasks and the three bioreactors. For both systems identical feeding regimes were applied, which started from day 3 using a ratio of 1:1 (v/v) chemically defined Efficient Feed[™] A and B (EFA/B) (Gibco[®], Life Technologies). In short, the daily feeding amount was calculated in such a way as to keep the glucose concentration above 5mM until the next day. This calculation was based on the specific glucose consumption rate and the expected cell concentration the next day. Detailed calculations are described in Pan et al., 2017, [11].

3.2.2 Bioreactor and incubation conditions

The bioreactors were controlled at temperature of 37 °C. The pH was controlled at 7.2 by either sparging CO₂ or adding 0.5 M NaOH. The DO was controlled at 50 % by a cascade regulation changing the oxygen fraction in the gas phase and the agitation rate in the range 100-300 rpm. Antifoam (Antifoam C Emulsion, Sigma-Aldrich®) was added on a daily basis to only the bioreactors. Shake flasks were placed in an incubator (Multitron CO₂ incubator; Infors HT), operated at 90 % humidity, 8% CO₂, and 37 °C at 100 rpm rotational speed and 50 mm orbital shaking diameter.

3.2.3 Analytical methods

3.2.3.1 Growth rate and metabolite concentrations

Dead and viable cell density were measured by a TC20™ automated cell counter (BIO-RAD) using trypan blue (Sigma-Aldrich) exclusion method. Glucose and lactate were measured daily using the YSI analyzer (YSI 2700, YSI Life Sciences).

3.2.3.2 IgG concentration

The concentration of the IgG1 was measured by Ultra High-Performance Liquid Chromatography (UHPLC) with a bio-monomer protein A column from Agilent technologies according to the instruction of biopharmaceutical application note from Agilent (Urbas et al., 2008)[12].

3.2.3.3 Glycosylation structure analysis

For determining the composition of the N-glycan structures, samples were taken in triplicate on day 7 from both shake flasks and bioreactors were analyzed using the method developed by Waters® [13]. In short, filtered supernatant was desalted by PD-10 columns (GE Healthcare, USA) and then purified using 1 mL HiTrap MabSelect SuRe column (GE Healthcare, USA) in combination with an Äkta pure system Unicorn 6.3 (GE Healthcare). Next, the IgG was concentrated with Amicon Ultra 0.5 mL centrifugal filters (Sigma-Aldrich, USA) and N-glycans were released using peptide N-

glycosidase F, (PNGase F solution, Sigma-Aldrich, USA). Glycans were labelled with the fluorescent 2-Aminobenzamide (2-AB), followed by post-labelling clean-up of glycan released by hydrazinolysis with module cartridges GlycoPrep® (Prozyme, USA) and elution of N-glycans. The labeled glycans were analyzed using HILIC UPLC with a column from Waters® [13].

3.2.3.4 RNA Sampling and quality

3mL cell suspension was sampled from both bioreactors and shake flasks and mixed with TRIzol® Reagent (Life Technologies, USA), on day3, 5, and 7. Samples were next snap-frozen in liquid nitrogen and stored at -80 °C for later analysis. Total RNA was extracted and purified using the RNeasy Mini-kit (Qiagen, Valencia, CA). RNA quality was assessed as described in Alsayyari, et al [14].

3.2.4 Transcriptomic analysis

Transcriptomic analysis on Affymetrix CHO Gene 2.1 ST arrays was carried out as described [14], [15]. Briefly, normalized expression estimates of probe sets were computed by the robust multiarray analysis (RMA) [16], as implemented in the Bioconductor library *AffyPLM*. Probe sets were redefined according to Dai et al. [17], based on the *CriGri_1.0* genome assembly (NCBI Reference Sequence Project (RefSeq) Release 72), which resulted in the profiling of 60626 transcripts (custom CDF v20). After averaging the expression levels of probe sets targeting the same gene, expression data for 20859 unique genes were obtained, which were used for all subsequent analyses.

After scaling and centering, principle component analysis (PCA) was performed in R using the *prcomp* function [18] to visually inspect the characteristics of the data set [19]. Differentially expressed genes were identified by using linear models (library *limma*) and an intensity-based moderated t-statistic [20], [21]. Correlation due to the repeated sampling from the same bioreactors or shake flasks was taken into account in the linear model fit by incorporation of the consensus intra-block correlation calculated by *limma*'s duplicate correlation function [20], [22]. P values were corrected for multiple testing using a false discovery rate (FDR) method [23]. Genes that satisfied the criterion of $FDR < 0.05$ and absolute fold change > 1.4 were considered to be

significantly regulated. Regulated genes were subjected to KEGG pathway overrepresentation analysis using Fisher's exact test, and visualized using the library *pathview* [24].

3.3 Results and discussion

3.3.1 Cell culture characteristics

Both shake flask and bioreactor cultures were done in fed-batch mode in triplicate. Nutrient feeding started from day 3 as described in material and methods. The glucose concentration was maintained between 2-6 g/L (11.2-35.2 mmol/L). The starting cell concentrations are slightly different for the shake flask (0.3×10^6 cells.mL⁻¹) and the bioreactor (0.45×10^6 cells.mL⁻¹). **Figure 1A** shows the growth curve for both cultivation systems. Starting from day 2 the growth in the bioreactor and shake flask started to be different. The growth in the bioreactor was slower than in the shake flask and reached the stationary phase about 1-2 days later than the shake flasks. The cell concentrations reached in the bioreactors were higher than in the shake flasks. For both systems the viable cell concentration dropped after the maximum cell concentration was reached. However, for the bioreactor the total cell concentration remained constant, while in the shake flasks the total cell concentration dropped in the same way as the viable cell concentration. Consequently, the viability as measured by trypan blue exclusion decreased in the bioreactors after reaching the maximum viable cell concentration, while in the shake flasks it remained constant except for the last two days. During the fed-batch the cultures are diluted due to addition of feed to the bioreactors and the shake flasks and the addition of base and antifoam to the bioreactors. The dilution due to antifoam and base addition is negligible compared to the dilution caused by addition of the feed. Furthermore, the sample sizes for both systems are the same, meaning that relatively more sample is taken from the shake flask. The chosen feeding strategy resulted in feed additions that were comparable relative to the culture volume and thus dilution of both systems was also comparable. On day 8 the total dilution factor for the shake flask and the bioreactor is respectively 1.9 and 1.7. At the day 10 the differences have become slightly higher at 2.4 and 1.9 for the shake flasks and bioreactors, respectively. The growth curves corrected for dilution are shown in appendix Figure A.1. As expected the correction does not change the observed differences between

the two systems. The decrease in viable cells at a constant high viability as observed in the shake flask means that substantial cell lysis occurs. Why this happens in the shake flask and not in the bioreactors is not clear, but could be due to different limitations occurring in both systems upon reaching the maximum cell density, which in turn may be due to the differences in conditions. Also until day 6 the viability in the bioreactor was slightly lower than in the shake flask, which could be due to the higher shear forces due to sparging and agitation in the bioreactors.

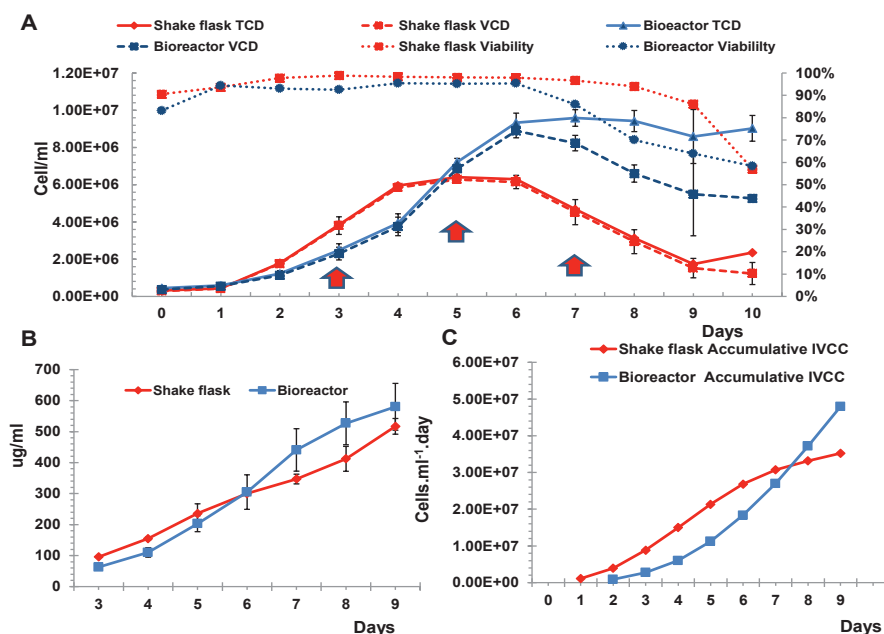


Figure 1. Comparison of cell performance between fed-batch shake flask and bioreactor; A: Total and viable cell concentration, B: IgG concentration, C: Cumulative integral viable cell concentration (IVCC). Data points are the average of three replicate shake flask and bioreactor cultivations.

Until day 4 the glucose concentration (Appendix figure A.2) was comparable between both systems, while from day 4 on the glucose concentration was consistently lower in the bioreactor system, which could be explained from the higher cell concentrations in the bioreactor. The lactate concentration (Appendix figure A.1) was slightly higher in the bioreactors for the first 4 days, after which it became comparably low for both systems. The IgG concentrations are lower in the bioreactor during the first 5 days and become higher from day 7 on, which can be explained from the cell concentrations as the integral of the viable cell density (see figure 1.C) becomes higher for the bioreactor at around the same time. This means that the specific productivity was comparable for both systems. For glycosylation three replicate samples were analysed taken from the shake flasks and bioreactors at day 7, which represents the harvest point as defined as the latest time point where the viability is still high (above 80 %) for both cultivations. The glycosylation profile (Figure 2) shows that major differences are present in the

glycan structure in G0, G0F and G1F between shake flask and bioreactor at day 7. Clearly the overall galactosylation is lower in the shake flask system with an increase in the G0F and G0 and a major decrease in G1F and minor decrease in G1 and G2F. These results are in agreement with our previous study comparing shake flasks to the bioreactors in batch cultivation mode [14].

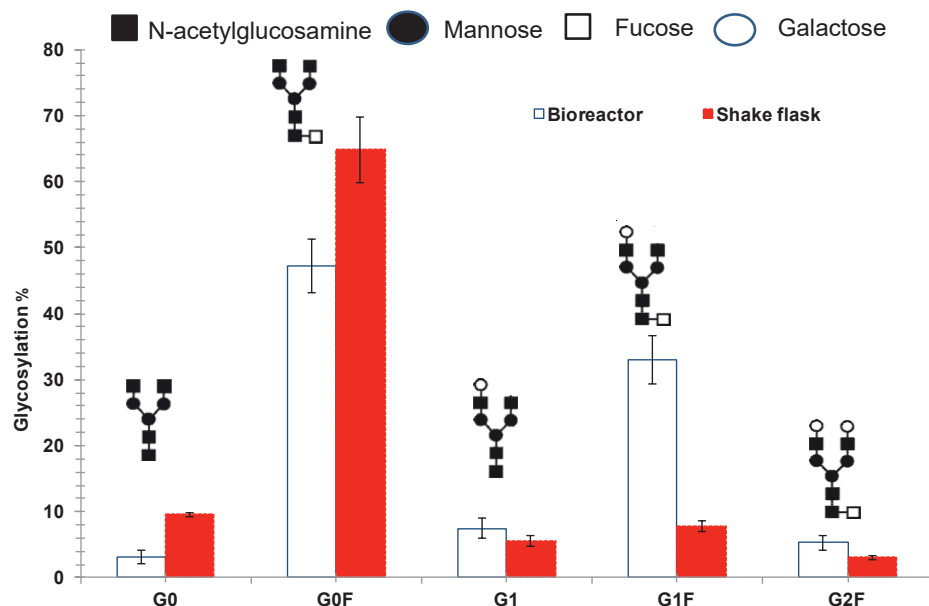


Figure 2. Average glycosylation pattern for the shake flask (red) and the bioreactor (white) at day 7. Error bars show the standard deviation of 3 replicate shake flask and bioreactor runs.

3.3.2 Differential gene expression

In **Figure 3** principal component analysis (PCA) for gene expression on three time points in the shake flask and bioreactor is shown. The three time points represent the dynamics of the cultures being; early exponential phase (day3) for both shake flasks and bioreactors, late exponential phase for bioreactors (day5) early stationary phase in shake flasks (day5) and bioreactors (day7), and late stationary phase in shake flasks (day7). All samples were taken in biological triplicates (i.e. three bioreactors and three shake flasks) except the shake flasks on day 3, where one shake flask sample was excluded due to technical failure of the DNA microarray chip. PC1, contains 28.5% of the variation and represents the development of the cultures in time. On the PC1 axis

day 5 of the shake flasks is close to day 7 of the bioreactors and day 5 of the bioreactors falls in between day 3 and day 5 of the shake flasks, confirming the faster growth of the cells in the shake flasks and the fact that the development of the shake flask cultures is ahead of that in the bioreactors by 1-2 days as observed in Figure 1. PC2 covering 10.95% of the variation represents a difference between both systems that is independent from time.

In **Figure 4** the number of genes that are differentially expressed between day 3-5 (A), 5-7 (B) and 3-7 (C) are shown and divided into genes that change only in the shake flask, only in the bioreactor and in both systems. Note that the day 3-7 comparison may contain genes that change expression slowly and are thus absent from the day 3-5 and day 5-7 comparison. Furthermore, genes that have an opposite regulation on day 3-5 as compared to day 5-7 may be absent in the day 3-7 comparison. From these three comparisons, we next select genes that are uniquely regulated in one system, meaning that for none of the comparisons they are regulated in the other system. Furthermore, we also select genes which are only regulated in common and never uniquely. Clearly always more regulated genes are found in the shake flasks than in the bioreactors. Also, the number of uniquely regulated genes is higher for the shake flasks than for the bioreactors. This may reflect the more dynamic conditions in the shake flask caused by the lack of pH and DO control. However, it can also be caused by the difference in development of the cultures over time with the shake-flask being ahead of the bioreactor. Since day 7 of the bioreactors is more or less comparable to day 5 of the shake flasks also the comparison day 3-7 bioreactor to day 3-5 shake flask was made. For this comparison, the bioreactor is still slightly ahead of the shake flask resulting now in more genes differentially expressed for the bioreactor and much less for the shake flasks. This shows that the period around day 5 is a very dynamic period and that differences in differential gene expression are highly dependent on the developmental status of the culture and the moment of sampling. Finally, it should be noted that since the shake flasks are ahead of the bioreactors the unique genes for the bioreactor in figure 4D are probably really unique for this system, whereas the unique genes for the shake flask may contain many genes that are differentially expressed during late stationary phase and thus not changed yet in the bioreactors. Likewise, for the comparison in Figure 4E the bioreactor is slightly further in development than the

shake flask, and here the unique genes of the shake flasks are probably related to the shake flask, while those of the bioreactor may now be due to the time effect.

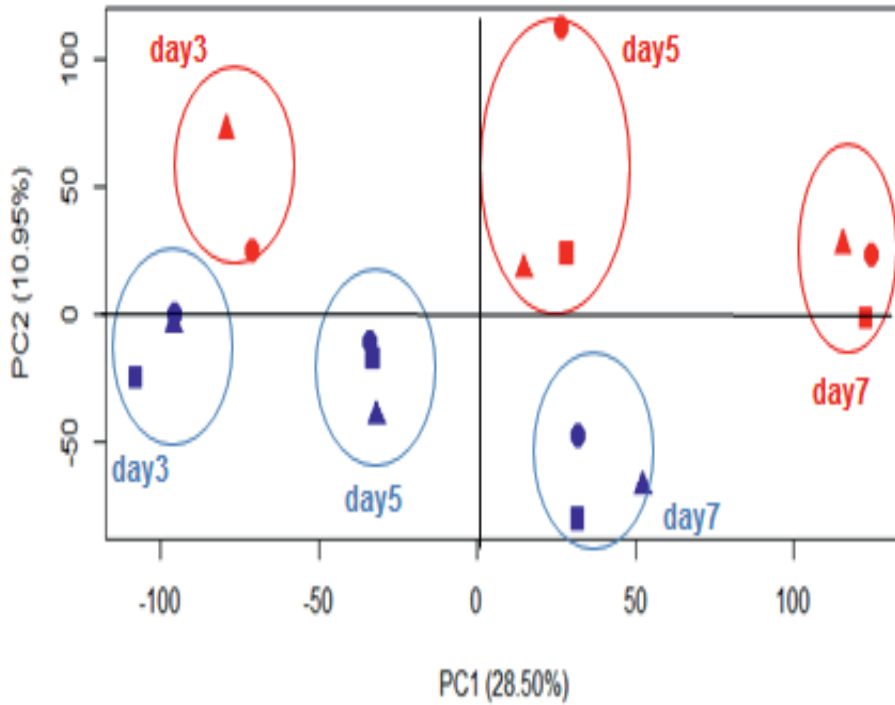


Figure 3 Principle component analysis shake flask (red), and bioreactor (blue). Different culture days are shown day 3, day 5 and day 7.

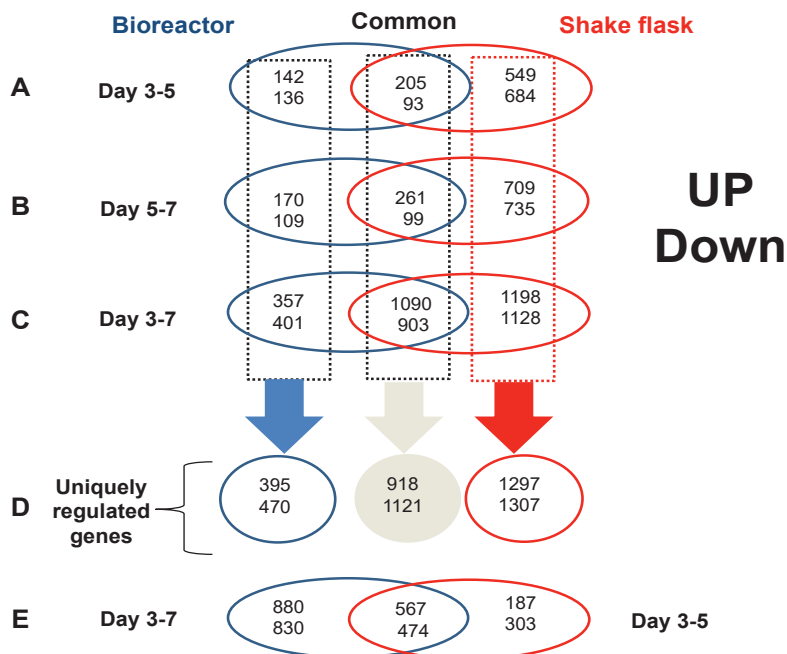


Figure 4 Venn diagram showing uniquely and common differentially expressed genes ($\text{FDR-BH} \leq 0.05$ and fold changed $> \pm 1.4$) for three time points **A**: day3-5, **B**: day 5-7, and **C**: day 3-7. **(D)** represent all the significant uniquely regulated genes in the shake flasks and the bioreactors, and the common regulated genes in both systems, after eliminating all duplicated and overlapped genes within the three time point comparisons. **(E)** comparison of bioreactor day 3-7 and shake flask day 3-5.

3.3.3 Differential Pathway expression

To obtain more insight in the effect of the differences in conditions between shake flasks and bioreactors the pathway enrichment of the genes selected in Figure 4 is studied. The pathways and biological functions are necessary to understand the dynamic changes occurring in both cultivation systems as well as the differences between them. The Venn diagram in **Figure 4** shows that the number of genes up regulated is comparable to the number of down regulated genes for all three time points for both systems. Pathway analysis will show whether genes in certain pathways are consistently up or down regulated or contain a mix of up and down regulated genes. Complete pathways data can be found in the supplementary data file A. From the top

significantly up and down regulated pathways of comparison 4A,B,C,D a selection is made of pathways that could be relevant for this study and these are shown in Table 1.

As can be seen most pathways represent the fact that the shake flask culture grows faster than the bioreactor. Thus, at day 5 the shake flask is already close to the stationary phase and the bioreactor is still in the exponential phase. At day 7 the shake flask is in the death phase while the bioreactor is in the stationary phase. Thus the genes unique to the shake flask are involved in processes that are related to down regulation of growth like the down regulation of DNA replication, and up regulation of processes that are related to stress and cell death like apoptosis and autophagy. Furthermore, in the oxidative phosphorylation pathway, 14 genes are uniquely up regulated in the shake flasks. As shown in Figure 5, six genes (Atp6v1a: Atp6v1c2: Atp6v0d2: Atp6v1d, Atp6v1e1, Atp6v1h) out of these 14 encode components of vacuolar ATPase (V-ATPase) which is responsible for transporting Hydrogen ions (H⁺) over cell membranes at the cost of ATP. Vacuolar ATPases have a wide variety of functions among which protein turnover, nutrient sensing, and a role in autophagy. Considering the late change in expression, it may be related to these functions here. Furthermore, the higher fluctuations of the conditions in the shake flask environment can lead to increasing energy demands for the cells to adapt to these changes, which may alter many genes and pathways related to stress adaption [25]. These genes are also involved in other pathways such as phagosome and mTOR signaling pathway. Additionally, two genes (Ndufa7 and Ndufb7) coding for NADH ubiquinone oxidoreductase subunit A7 and B7 of the translocating NADH dehydrogenase are uniquely up regulated in shake flasks. These genes are related to cellular respiration in the mitochondria. The upregulation of genes involved in oxidative phosphorylation may indicate oxygen limitation in the shake flasks, which can occur at the high cell densities reached in shake flasks [26]. The Oxidative phosphorylation pathway is important for ATP production process and lower flux through this pathway due to for example oxygen limitation can result in arresting the cell cycle and progression of the cells into the death phase [27].

Another example of an uniquely up regulated gene in the shake flask is the *Itpr3* gene which is involved in many pathways among which apoptosis. *Itpr3* codes for an inositol 1,4,5-trisphosphate receptor type 3 with functions as and intracellular signalling mechanism sensitive to oxidative stress [28]. *Itpr3* was up regulated uniquely in the shake flask during time points (days: 3-5, 5-7, and 3-7) with fold changes of 1.4, 1.5 and 2.21 respectively. This is another indication that varying and limiting oxygen concentrations may be a problem in the shake flasks. As stated limitations in oxygen can trigger cell cycle arrest and apoptosis [27]. **Table 1.B** illustrates a list of pathways regulated uniquely in the shake flask (Figure 4D). This more or less confirms the time comparisons discussed before. Cell cycle pathways, although not present in the top 10, consisting of 129 genes show significant down regulation in the shake flask as compared to the bioreactors, which agrees with the earlier arrest of growth in the shake flasks (see supplementary data table A). Similar results were obtained when comparing shake flasks and bioreactors in batch mode (**Chapter 2**). Furthermore, DNA replication, Fatty acid elongation, histidine metabolism, mismatch repair autophagy, phagosome, and apoptosis are all down regulated in shaker only. The down regulation of all these processes agrees with the earlier arrest of cell growth except for apoptosis. When cells are earlier exposed to for example nutrient limitation an upregulation of apoptosis would be expected. The reason for the down regulation is not clear but could be related to the fact that despite the decrease in viable cell concentration the cell viability remains high in the shake flasks indicating fast lysis of dead cells.

Many pathways that are regulated in Table 1 are thus the result of a difference in growth rate in both systems, which makes it difficult to study other differences between both systems that are not related to development time and possible form the root cause for the difference in time development. Information on this could be still extracted from Figure 4D and 4E. In **Figure 4.D** the number of significant differential expressed genes unique for each of the three groups (shake flask, common, and bioreactor) are given. Whereas, as stated, the genes unique to the shake flask may still be there because of the fact that the shake flasks are further in their development, the bioreactor genes are probably unique to the bioreactor conditions. Likewise, the genes in Figure 4E the bioreactor is further in development than the shake flask and thus the unique genes of

the shake flask are probably due to the specific conditions in the shake flask. However, no pathways that could be linked to system differences could be derived from this, which like stated before could be linked to the very strong time effect and the highly dynamic behaviour of gene expression around day 5, masking other effects. Thus, In Table 1B panel E shows that none of the pathways from the list of pathways selected for Table1A are significantly regulated. This again confirms that the different regulation of the pathways selected from figure 4,A,B,C and D is mainly related to the difference in development in time.

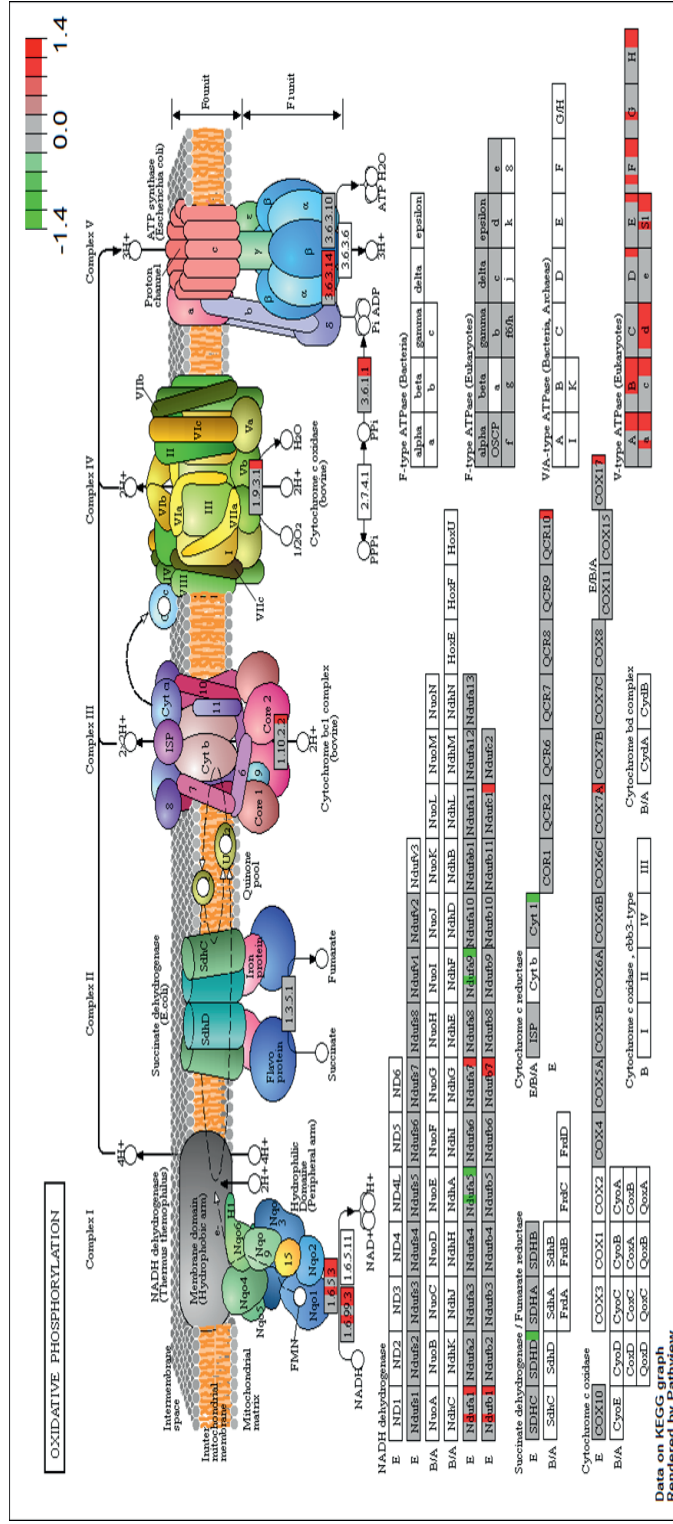


Figure 5 Pathview for oxidative phosphorylation pathways (FDR-BH >0.05, Fold changed $\leq \pm 1.4$).

3.4 Conclusion

In process development a major part of screening clones, media and feeds is performed in high throughput experiments in shake flasks. A good understanding of the differences of cell behaviour in these uncontrolled shake flasks and controlled bioreactors is needed to avoid unexpected changes upon transfer to the more controlled production runs. We have used commercial Affymetrix CHO plates to identify the transcriptomic changes between uncontrolled shake flask and the controlled bioreactor systems and to gain an understanding differences between them on a pathway level. Principal component analysis showed a clear difference in gene expression for both systems and is able to discriminate differences that are due to a different speed of development in time (i.e. differences in growth rate) from differences between the systems that seem independent from time. On a pathway level the differences in time development are clearly seen back in the down regulation of growth related processes and the up regulation of processes related to arrest of growth and stress in the shake flasks. It is more difficult to find time independent differences on a pathway level, which is at least in part due to the highly dynamic nature of the gene expression in the fed-batch cultures. Nevertheless, pathway analysis does indicate potential oxygen limitation in the shake flasks.

3.5 References

- [1] G. Walsh, "Biopharmaceutical benchmarks 2014," *Nat. Biotechnol.*, vol. 32, no. 10, p. 992, 2014.
- [2] R. Jefferis, "Recombinant antibody therapeutics: the impact of glycosylation on mechanisms of action," *Trends Pharmacol. Sci.*, vol. 30, no. 7, pp. 356–362, Jul. 2009.
- [3] D. M. Dinnis and D. C. James, "Engineering mammalian cell factories for improved recombinant monoclonal antibody production: Lessons from nature?," *Biotechnol. Bioeng.*, vol. 91, no. 2, pp. 180–189, 2005.
- [4] E. Trummer *et al.*, "Transcriptional profiling of phenotypically different Epo-Fc expressing CHO clones by cross-species microarray analysis," *Biotechnol. J.*, vol. 3, no. 7, pp. 924–937, 2008.
- [5] A. Kantardjieff *et al.*, "Transcriptome and proteome analysis of Chinese hamster ovary cells under low temperature and butyrate treatment," *J. Biotechnol.*, vol. 145, no. 2, pp. 143–159, Jan. 2010.
- [6] J. Y. Baik *et al.*, "Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin," *Biotechnol. Bioeng.*, vol. 93, no. 2, pp. 361–371, 2006.
- [7] M. Castro-Melchor, H. Le, and W.-S. Hu, "Transcriptome Data Analysis for Cell Culture Processes," in *Genomics and Systems Biology of Mammalian Cell Culture*, vol. 123, no. July 2015, Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 27–70.
- [8] N. Vishwanathan, H. Le, T. Le, and W. S. Hu, "Advancing biopharmaceutical process science through transcriptome analysis," *Curr. Opin. Biotechnol.*, vol. 30, pp. 113–119, 2014.
- [9] N. Vishwanathan *et al.*, "Global insights into the Chinese hamster and CHO cell transcriptomes," *Biotechnol. Bioeng.*, vol. 112, no. 5, pp. 965–976, 2015.
- [10] N. Carinhas, R. Oliveira, P. M. Alves, M. J. T. Carrondo, and A. P. Teixeira, "Systems biotechnology of animal cells: the road to prediction," *Trends Biotechnol.*, vol. 30, no. 7, pp. 377–385, Jul. 2012.
- [11] X. Pan, M. Streefland, C. Dalm, R. H. Wijffels, and D. E. Martens, "Selection of chemically defined media for CHO cell fed-batch culture processes," *Cytotechnology*, vol. 69, no. 1, pp. 39–56, 2017.

- [12] B. M. Urbas L, Brne P, Gabor B, Plevcak S, "Rapid Human Polyclonal IgG Quantification Using the Agilent Bio-Monolith Protein A HPLC Column Application Note," *Agil. Technol. Inc*, 2008.
- [13] M. Hilliard *et al.*, "Development of a Glycan Database for Waters ACQUITY UPLC Systems," *Waters Appl. Note 720004202en*, 2012.
- [14] A. A. Alsayyari *et al.*, "Transcriptome analysis for the scale-down of a CHO cell fed-batch process," *J. Biotechnol.*, vol. 279, no. May, pp. 61–72, Aug. 2018.
- [15] K. Lin *et al.*, "MADMAX - Management and analysis database for multiple ~omics experiments.," *J. Integr. Bioinform.*, vol. 8, no. 2, p. 160, Jul. 2011.
- [16] R. A. Irizarry *et al.*, "Exploration, normalization, and summaries of high density oligonucleotide array probe level data.," *Biostatistics*, vol. 4, no. 2, pp. 249–64, Apr. 2003.
- [17] M. Dai *et al.*, "Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data," *Nucleic Acids Res.*, vol. 33, no. 20, pp. 1–9, 2005.
- [18] K. Hornik, "Principal Component Analysis using R," 2009.
- [19] M. Ringnér, "What is principal component analysis?," *Nat. Biotechnol.*, vol. 26, no. 3, pp. 303–4, Mar. 2008.
- [20] M. E. Ritchie *et al.*, "limma powers differential expression analyses for RNA-sequencing and microarray studies.," *Nucleic Acids Res.*, vol. 43, no. 7, p. e47, Apr. 2015.
- [21] M. A. Sartor, C. R. Tomlinson, S. C. Wesselkamper, S. Sivaganesan, G. D. Leikauf, and M. Medvedovic, "Intensity-based hierarchical Bayes method improves testing for differentially expressed genes in microarray experiments.," *BMC Bioinformatics*, vol. 7, p. 538, Dec. 2006.
- [22] G. K. Smyth, J. Michaud, and H. S. Scott, "Use of within-array replicate spots for assessing differential expression in microarray experiments," *Bioinformatics*, vol. 21, no. 9, pp. 2067–2075, 2005.
- [23] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. R. Stat. Soc. B*, vol. 57, no. 1, pp. 289–300, 1995.
- [24] W. Luo, "Pathview: pathway based data integration and visualization," no. Figure 3, pp. 1–16, 2015.

- [25] J. B. Sieck *et al.*, “Development of a Scale-Down Model of hydrodynamic stress to study the performance of an industrial CHO cell line under simulated production scale bioreactor conditions,” *J. Biotechnol.*, vol. 164, no. 1, pp. 41–49, Mar. 2013.
- [26] F. M. Wurm, “Production of recombinant protein therapeutics in cultivated mammalian cells,” *Nat Biotechnol*, vol. 22, no. 11, pp. 1393–1398, 2004.
- [27] H. Swiderek, A. Logan, and M. Al-Rubeai, “Cellular and transcriptomic analysis of NS0 cell response during exposure to hypoxia,” *J. Biotechnol.*, vol. 134, no. 1–2, pp. 103–111, 2008.
- [28] N. Factor *et al.*, “Nuclear Factor, Erythroid 2-Like 2 Regulates Expression of Type 3 Inositol 1,4,5-Trisphosphate Receptor and Calcium Signaling in Cholangiocytes,” *Gastroenterology*, vol. 149, no. 1, p. 211–222.e10, 2015.

Appendix

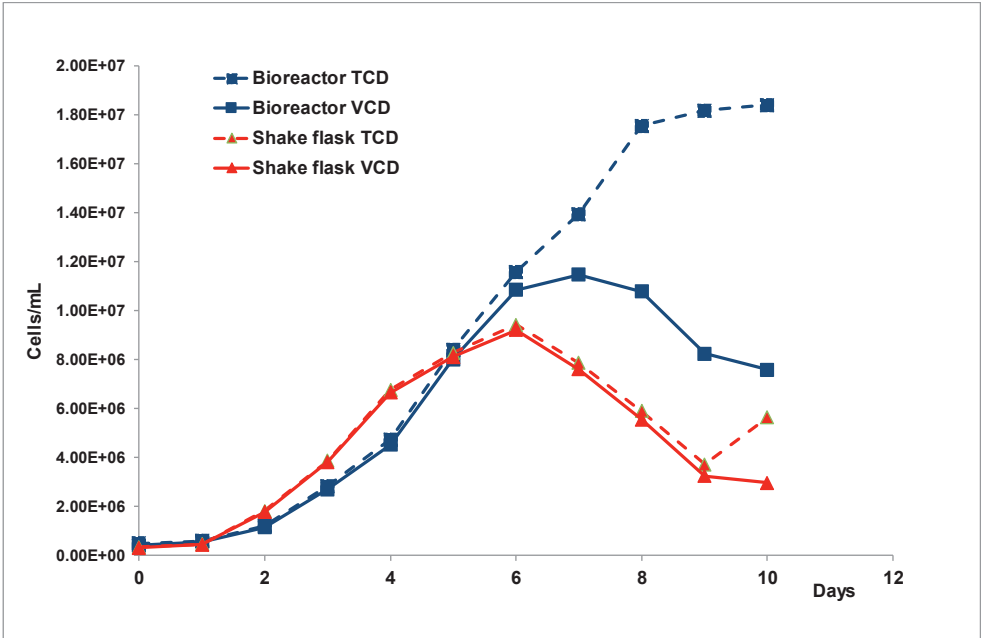


Figure A.1: Viable cell density (VCD) and total cell density (TCD) corrected for dilution for shake flasks and bioreactors as functional of the time. Each curve is the average of three replicate cultures.

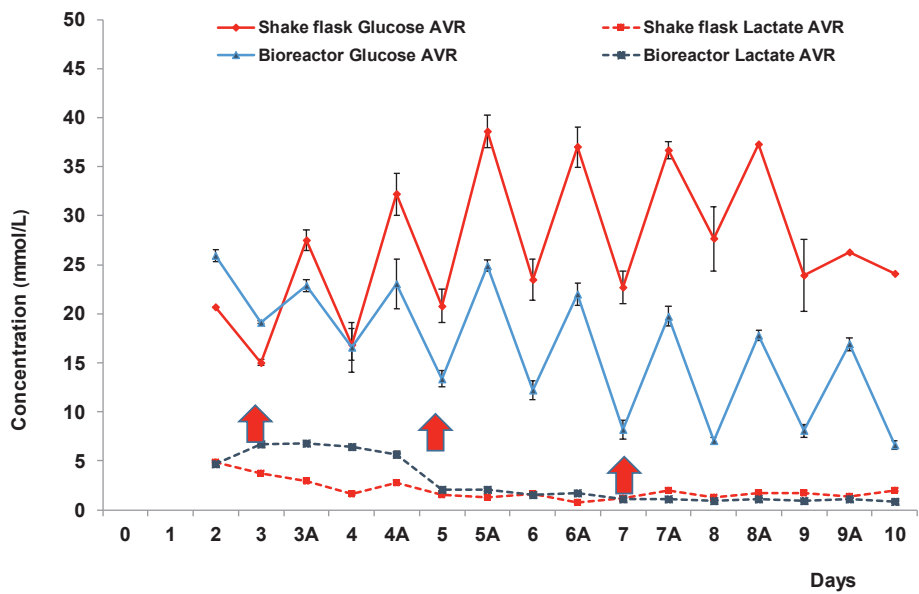


Figure A.2: Lactate and glucose concentration for shake flask and bioreactors.

Chapter 4

Transcriptome analysis for the scale-down of a CHO cell fed-batch process

This chapter is published as

Abdulaziz A. Alsayyari^a, Xiao Pan^a, Ciska Dalm, Jochem W. van der Veen, Nienke Vriezen, Jos A. Hageman, René H. Wijffels, Dirk E. Martens, 2018, “Transcriptome analysis for the scale-down of a CHO cell fed-batch process”, <https://doi.org/10.1016/j.jbiotec.2018.05.012>

^a. Both authors contributed equally to this publication

Abstract

Transcriptome and metabolism analysis were performed to evaluate the scale-down of a CHO cell fed-batch process from a 10 L bioreactor to an ambr 15[®] (ambr) system. Two different agitation scale-down principles were applied, resulting in two different agitation rates in the ambr system: 1300 RPM based on the agitator tip speed, and 800 rpm based on the volumetric power input (P/V). Culture performance including cell growth, product titer, glycosylation, and specific consumption/production rates of metabolites was the same for both agitation rates in the ambr and was comparable to that of the 10 L system. The initial variation in gene expression between the inocula for the ambr and 10 L system was no longer present after three days of culture, indicating comparable culture conditions in both systems. Based on principal component analysis, changes in gene expression over time were similar between both scales with less than 6% variation. 2455 genes were uniquely regulated in the ambr system compared to 1604 genes in the 10 L system. Functional analysis of these genes did not reveal their relations with scale or cellular function. This study further strengthens that the ambr system gives representative culture performance for the 10 L bench-scale bioreactor.

4.1 Introduction

The development of a Chinese Hamster Ovary (CHO) cell production process within the biopharmaceutical industry aims at obtaining a consistent and robust cell culture process with a high product volumetric productivity, as well as a consistent and desired product quality. To reach this, it is important to identify the parameters that influence the culture performance and the range within which these parameters should stay to ensure a reproducible process with a consistent product quality. This requires high throughput experimentation for which small-scale bioreactors are needed. The small scale bioreactors should be representative of the large-scale so that the process developed at small-scale, including the design space for the critical process parameters, can be translated to large-scale with the same process performance. Different scale-down criteria can be used to develop a representative small-scale process. These criteria include, for example, the overall volumetric gas-liquid mass transfer coefficient (k_La), the volumetric power input (P/V), and the agitator tip speed [1], [2]. Production scale reactors can be properly scaled down to bench-scale reactors (1-10 L) which are traditionally used as the main platform for process development [3], [4]. However, the number of experiments that can be done in a certain time frame at bench-scale is still too small for studying all the parameters that can be relevant to a culture process (pH, DO, temperature, nutrient concentrations, pre-culture age, seeding density, etc.). In order to increase the experimental throughput, several miniature bioreactor systems at mL scales have been developed during the past decade [5], [6]. Like in large-scale reactors, DO, pH, agitation, and gas sparging can be monitored and controlled in these small-scale systems as well. The higher throughput allows for a shorter development time and the use of design of experiment (DoE) with a higher resolution, resulting in a more accurate definition of the design space.

The ambr® 15 (from hereon called the ambr) system has advantages compared to other miniature systems due to the advanced automated operation which reduces handling errors and increases the experimental throughput. A characterization of both the ambr and large-scale stirred tank bioreactors (STRs) was conducted by Nienow et al. (2013a). It was shown that the physical environment between the ambr and conventional STRs differs in some aspects including the flow regime, gas superficial velocity, and volumetric power input. Several studies have been performed to compare

the culture performance of the ambr system to STRs ranging from 2 L up to 1000 L for CHO cell fed-batch processes [8]–[11]. Overall, similar process performances were reported across all scales. In a recent study done by Rouiller et al (2016), it was shown that the ranking of 12 CHO cell clones based on product titer reached in batch cultures was more similar between 3.5 L bioreactors and ambr (Pearson correlation coefficient=0.97) than between 3.5 L bioreactors and deep well plates and shake tubes. In the study of Janakiraman et al. (2015), for several studied process parameters, the ranges of these parameters which resulted in acceptable product quality attributes were similar between ambr and 5 L bioreactors. These studies show that the result of screening experiments as well as determined process parameters ranges are transferable from ambr to bench-scale systems. In contrast, Siva et al. (2015) showed that for a fed-batch process using a CHO DHFR- cell line the ambr gave different results compared to 5-L and 250-L STRs. Lower cell density and final titer together with an earlier decrease of cell viability were observed in the ambr system. They hypothesized that a foaming problem along with a higher ammonium concentration, a higher osmolality, and a lower pH in the ambr cultures were the causes of these differences. Summarizing, representative culture performances have been reported in ambr as compared to the bench-scale systems in all studies except for one and the ambr system has gained acceptance for use in process development in biopharmaceutical industries. A better understanding of the effects of the different physical environments present in the small-scale ambr system as compared to bench-scale systems on cell physiology could be helpful for further validating the ambr system, specifically for developing better small-scale reactors in general.

A better understanding of cell physiology and gene regulation in relation to process conditions can be obtained using transcriptomics [15]. Transcriptome analysis can be used in different ways for process understanding. (i) The expression levels of genes can be used as a fingerprint to compare two process conditions, *e.g.* small- and large-scale. For example, Jayapal and Goudar (2013) used transcriptome analysis to confirm the scalability of a BHK cell perfusion culture process. (ii) The genes that are differentially expressed can be studied on a functional level. This approach can give mechanistic information on how differences in process conditions (*e.g.* differences between scales) influence cell physiology and process performance. For example, Sieck et al. (2013; 2014) found that transcripts related to DNA damage and repair

mechanisms were up-regulated when elevated shear forces from agitation and sparging were applied to a CHO cell culture.

The aim of this study is to evaluate whether the physiology and performance of the cells in the ambr systems is comparable to that in a 10 L system for a fed-batch process. The CHO cells are compared on the metabolic and gene expression level. A detailed analysis of the 10 L process, including cell metabolism has been presented in a previous study [19] and was the starting point for the comparison with the ambr system in this work. CHO gene microarrays were used to measure the transcriptome. In addition, in order to investigate the impact of different scale-down criteria for agitation on cell physiology and culture performance, two different impeller agitation rates (800 and 1300 RPM) were included in the ambr experiment. The 800 RPM in ambr is equivalent to the impeller volumetric power input (P/V) value that was employed at the 10 L scale, whereas the 1300 RPM is equivalent to the maximum impeller tip-speed at the 10 L scale.

4.2 Materials and methods

4.2.1 Cell line and pre-culture

A suspension CHO^{BC®} cell clone (BC-P, provided by Bioceros Holding BV) producing a recombinant immunoglobulin G1 (IgG1) was used in this study. Cells were thawed from a working cell bank and maintained in ActiCHO-P medium (GE Healthcare) supplemented with 4 mM glutamine (Gibco, Life Technologies) and 0.5% Anti-clumping agent (Gibco®, Life Technologies). 200 µg×mL⁻¹ Zeocin[™] and 5 µg×mL⁻¹ Blasticidin (both from Life Technologies) were added as two selection reagents during the pre-cultures. Sub-culture was conducted in repeated batch cultures in shake flasks grown under 37°C and 8% CO₂ conditions. The inoculum for the 3×10 L bioreactors was prepared in a single culture bag on a rocking bioreactor system. For each 10 L bioreactor, a 500 mL inoculum at a viable cell concentration of 3×10⁶ cells×mL⁻¹ was directly inoculated in 4500 mL of pre-heated fresh medium, resulting in a starting density of 3×10⁵ viable cells×mL⁻¹ at a starting volume of 5 L. The inoculum for the ambr bioreactors was expanded in shake flasks. To prepare the inoculum for the ambr cultures, the cells from the flasks were pooled together and spun down at 300×g. Next, the used medium was discarded and cells were re-suspended in fresh medium at a

viable cell concentration of 3×10^6 cells \times mL⁻¹. 1.4 mL inoculum was inoculated in 12.6 mL of pre-heated fresh medium in each ambr bioreactor, resulting in a starting density of 3×10^5 viable cells \times mL⁻¹ at a starting volume of 14 mL.

4.2.2 10 L culture experiment

Triplicate fed-batch cultures were conducted in 10 L Sartorius bioreactors (Sartorius Stedim Biotech) controlled by BIOSTAT® B-DCU II. The geometry of the BIOSTAT® B-DCU II 10 L bioreactor is shown in Table I. Culture temperature was controlled at 37°C, dissolved oxygen (DO) was controlled at 40% by pure O₂ flow through a micro sparger, pH was controlled at 7.2, by using base, and CO₂ through a macro sparger. An air flow through the macro sparger was used to strip out the accumulating dissolved CO₂ (pCO₂). Mixing was done at 200 RPM. From day 3 on, feeds were added to each bioreactor daily. First, the glucose concentration was measured. If the concentration was lower than 18 mM, a 45% (w/w) glucose solution was added as one bolus to reach a glucose concentration of 28 mM in the reactor. Next, 4.5% (v/v) ActiCHO feed A, also containing about 500 mM glucose amongst other nutrients, and 0.45% (v/v) ActiCHO feed B (both from GE Healthcare, USA) per culture volume per day were fed to each reactor. FoamAway™ antifoam (ThermoFisher Scientific) solution was added to each bioreactor prior to inoculation and during the culture when needed.

For each 10 L bioreactor, a 15 mL sample was taken daily before and after feed addition. Total cell density, viable cell density, and cell diameter were measured using a CedexHiRes® analyzer (Innovatis). 3 mL of the sample was spun down at 300 \times g for 5 min (Heraeus Multifuge X3R, Thermo Scientific). The cell pellet was collected and the total RNA was extracted using 3 mL TRIzol reagent (Invitrogen™) and stored at -80°C for later transcriptome analysis. The remaining sample was spun down at 3345 \times g for 15 min. 1 mL supernatant from the sample was measured for off-line pH, pCO₂, glucose, lactate, ammonium concentrations and osmolality by a nova BioProfile FLEX analyzer (nova biomedical). The remaining supernatant was stored at -20 °C for later analysis. On culture day 4, 7 and 10 biomass samples were taken. The preparation of the biomass samples was performed in the same way as in [19].

Table I. Geometry of the BIOSTAT® B-DCU II (10 L) and the ambr® 15 (ambr) bioreactors, based on Sartorius Stedim Biotech

	10 L	ambr
Overall reactor shape	Cylinder	Rectangular parallelepiped
Internal cross-section area (mm²)	$\pi \times (80)^2$	28 × 15
Reactor height (mm)	470	60
Maximum working volume (mL)	10,000	15
Number of impeller	1 or 2	1
Impeller type	3-blade segment impeller	2-blade segment impeller
Impeller diameter (mm)	78	11.4
Impeller shaft location	Centroid	3 mm off the center
Number of baffles	Not baffled	Not baffled
Number of spargers	2	1
Sparger types	1× ring sparger 1× micro sparger	1× 1 mm open pipe sparger

4.2.3 ambr culture experiment

The ambr system (ambr® 15, from Sartorius Stedim Biotech) used in this study consisted of 4 culture stations (CS), each CS contained 12 bioreactors. In total 48 bioreactors were used in this study. The geometry of the ambr bioreactor is shown in **Table I**. The reactor volume was kept between 13 and 15 mL throughout the culture. For all 4 CSs, the temperature was controlled at 37°C, dissolved oxygen (DO) was controlled at 40% by oxygen enrichment in gas supply through an open pipe sparger, and pH was controlled by 0.5 M NaOH and CO₂ enrichment in gas supply. pH was measured off-line on a different analyzer as the 10 L cultures, because of the smaller sample volume. Measurement differences in pH due to the different analyzer used were accounted for creating matching bands of control. The daily feeding strategy was similar as in the 10 L culture experiment with a difference that the feed can only be added at one time as a bolus. 4×diluted FoamAway™ antifoam (ThermoFisher Scientific) solution was added to each ambr bioreactor when needed. Two different

agitation rates were tested in the ambr culture experiment being: CS-1 and CS-2 at 800 RPM which is equivalent to the volumetric power input (P/V) value of the 10 L culture experiment, and CS-3 and CS-4 at 1300 RPM which is equivalent to the agitator's tip-speed of 10 L culture experiment (**Table II**). The P/V value for the 10 L reactor is obtained from:

$$\frac{P}{V} = \frac{P_0 \rho N^3 D^5}{V} \quad \text{Equation 1}$$

where P is the power input from impeller (W), V is the culture volume (m³), P₀ is the power number of the 10 L bioreactor impeller and was determined as 2.1 based on [20], ρ is the liquid density (kg×m⁻³), N is the impeller speed (s⁻¹), D is the impeller diameter (m). The Reynolds number (Re) and the Kolmogorov length (λ_k, μm) scale in a turbulent flow are calculated based on [20]:

$$Re = \frac{\rho N D^2}{\eta} \quad \text{Equation 2}$$

$$\lambda_k = \left(\frac{P}{\rho V \vartheta^3} \right)^{-\frac{1}{4}} \quad \text{Equation 3}$$

where η is the dynamic viscosity (N·s·m⁻²), and ϑ is the kinematic viscosity (m²·s⁻¹). Maximum 800 μL daily sample volume was taken before feed addition from each individual ambr bioreactor. Individual ambr bioreactors were sampled every other day to measure off-line pH, pCO₂, glucose concentration, and lactate concentration on a blood gas analyzer (ABL90 FLEX). Larger samples were obtained daily from paired cultures run under identical conditions which were analyzed for metabolites, cell density and culture viability, IgG concentration, spent media, biomass composition, and RNA preparation by using the same methods as for the 10 L cultures.

Table II. Physical comparison between ambr and 10 L reactors. The 300 RPM of the 10 L experiment is not used in this study. It is shown for comparison only.

	10 L		ambr	
Agitation rate (RPM)	200	300	800	1300
Tip-speed ($\text{m}\times\text{s}^{-1}$)	0.82	1.23	0.48	0.78
P/V ($\text{W}\times\text{m}^{-3}$)	40	135	65*	280*
k_La (h^{-1})	2-40	2-40	4-6*	5-8*
Reynolds number	29000	43000	2500	4000
Flow type	Turbulent	Turbulent	Laminar	Transient
Kolmogorov (μm)	55	40	-	27

4.2.4 Biomass, spent medium, and mAb analysis

The analysis of biomass composition including cell dry weight, soluble cellular protein, fatty acids, and carbohydrate was performed using the same method as described in [19]. Compositions of the spent medium including extracellular amino acids and organic acids were quantified using NMR (Spinovation Biologics BV). mAb (IgG1) titer was quantified by using Protein-A chromatography (Agilent, 5069-3639). mAb N-glycan composition was analyzed by a Hydrophilic Interaction Chromatography (HILIC UPLC). The methods for mAb quantification and N-glycan analysis were described in Pan et al. (2017).

4.2.5 Average specific metabolic rates

The calculation of the specific production/consumption rate of a compound is the same as described in Pan et al. (2017). In brief, the following equation is used, assuming a constant q_x

$$M_x(t) - M_x(0) - V_f \times C_f = q_x \times \int_0^t X_{VC} dt \quad \text{Equation 4}$$

where M_x (mg; mmol) is the total amount of compound x in a culture at time 0 and t , X_{VC} (cells) is the number of viable cells in a culture, V_f (mm^3) is the total volume of feed added, C_f (mM) is the concentration of compound x in the feed, and q_x ($\text{mg}\cdot\text{cell}^{-1}\cdot\text{day}^{-1}$; $\text{mmol}\cdot\text{cell}^{-1}\cdot\text{day}^{-1}$) is the cell-specific production rate of compound x . When the rates are calculated based on cell volume, X_{VC} (mm^3) presents the volume of viable cells in a culture and q_x ($\text{mg}\cdot\text{mm}^{-3}\cdot\text{day}^{-1}$ or $\text{mmol}\cdot\text{mm}^{-3}\cdot\text{day}^{-1}$) is the cell volume-specific production rate of compound x . The average specific production rates (q) were obtained from the slope of a plot of accumulated consumed/produced compound mass against the accumulated integral of viable cell number (or the integral of viable cell volume) using linear regression.

4.2.6 Transcriptome analysis

For both scales, the transcriptome analysis was carried out on: cells from the inoculum just before inoculation, on day 3 which represents the exponential phase, on day 5 which represents the cell size increase phase, and on day 9 which represents the stationary phase. For the inoculum of the 10 L experiment, three samples were taken from the culture bag before inoculation. For the inoculum of the ambr experiment, three samples were taken from the pooled shake flasks just before centrifugation. During the culture, for the 10 L experiment, three samples were taken from the three individual bioreactors, whereas for the ambr, also three samples were taken with each sample obtained by pooling the samples from two bioreactors of the 800 RPM condition. The total RNA was purified using the RNeasy Mini kit (Qiagen, Valencia, CA). RNA concentration was measured on a Nanodrop 2000 (Thermo Scientific, Wilmington, DE). RNA quality was assessed on a 2100 Bioanalyzer using the RNA 6000 Nano Kit (Agilent, Santa Clara, CA), and all samples had a greater than 8 in RNA integrity number (RIN). Affymetrix GeneChip™ CHO Gene 2.1 ST Arrays (Affymetrix, Santa Clara, USA) were used for transcriptome expression profiling. In short, the same amount of RNA was labeled by the Whole-transcript Sense Target Assay (Affymetrix, Santa Clara, USA) and hybridized according to the manufacturer's instructions. Quality control and data analysis were done as described in [21]. Normalized expression estimates of probe sets were computed by the robust multi-array analysis (RMA) [22], using R/Bioconductor package AffyPLM [23]. The well-annotated reference sequences are based on the CriGri_1.0 genome assembly NCBI reference sequence

project (RefSeq) Release 72, which resulted in 60626 annotated sequences (transcripts profiles) (custom CDF v20). After averaging the expression levels of probe sets targeting the same gene, expression data for 20858 unique genes were obtained, which was used for all subsequent analysis. Linear models (Limma package) were used to identify the differential expressed genes taking into account correlation due to the repeated sampling from the same bioreactors followed by an intensity-based moderated t-statistic [24], [25]. P-values were adjusted for multiple testing using the Benjamin and Hochberg (FDR.BH) method [26]. Genes regulation that satisfied the criterion of $FDR.BH < 0.05$ and absolute fold-change (FC) > 1.4 were defined to be significant and were subjected to KEGG pathways over-representation analysis using fisher's exact test and Venn diagram. These cut-off values are commonly applied to reduce the number of false-positive genes and the background noises that are picked up by the differential gene analysis [27]–[29]. The raw data on transcriptome analysis is available in **Supplementary A**.

4.3 Results and discussion

4.3.1 DO and pH control

Proper control of dissolved oxygen (DO) and pH is critical for the culture performance. The online DO and pH profiles of a representative ambr reactor and a representative 10 L reactor are shown in **Figure 1**. For the ambr cultures, the DO and the pH were well controlled around the set-point. However, compared to the 10 L cultures, the ambr cultures showed higher spikes of DO and pH. In addition, slightly higher pH values from day 5 to 8 can be seen in the ambr cultures. The higher spikes of DO were caused by the opening of the vessels which was needed for feed addition, base addition, and sampling. The spikes in pH in both the ambr and the 10 L cultures were caused by the daily addition of ActiCHO feed B, which has a high pH value of 11. The spikes were higher in the ambr cultures, due to the fact that for the ambr the feed was added as a bolus whereas for the 10 L it was added over a period of 2 hours.

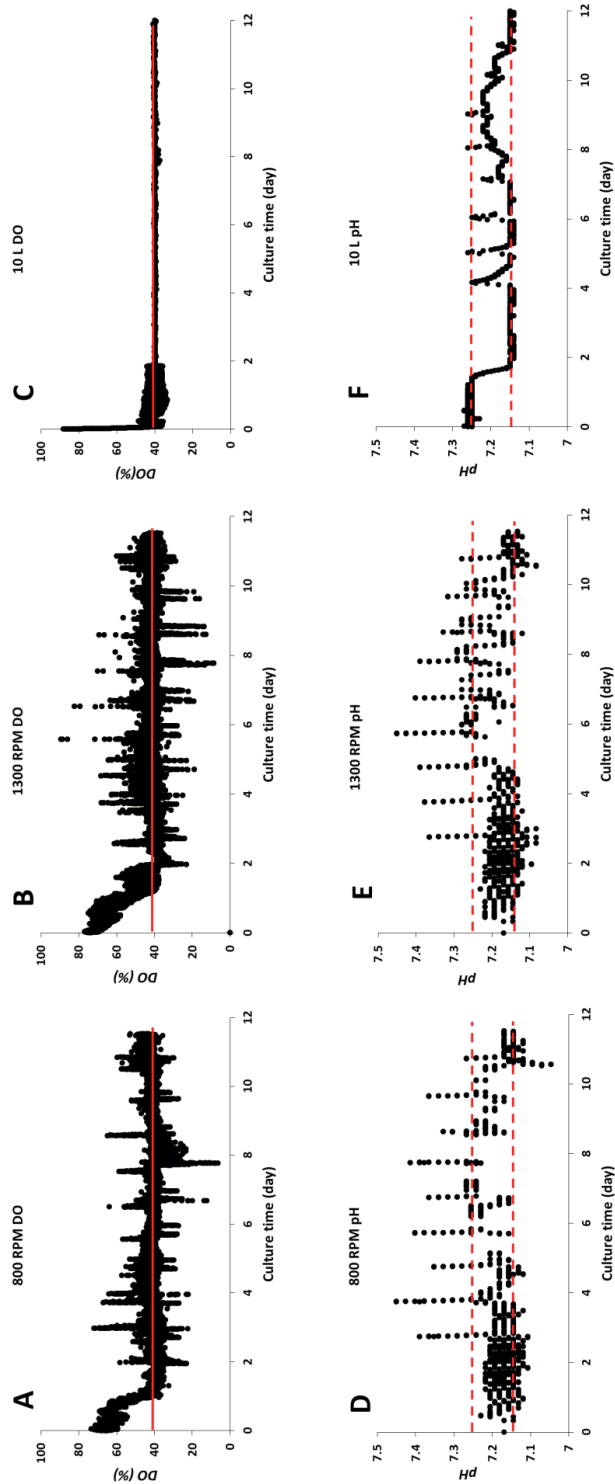


Figure 1. Examples of the online measured pH and DO profiles. Graph A, B, and C show DO profiles of the ambr 800 RPM, the ambr 1300 RPM, and the 10 L bioreactor, respectively. Graph D, E, and F show pH profiles of ambr 800 RPM agitation, ambr 1300 RPM agitation, and 10 L, respectively. For DO the set-points are shown by red solid lines, and for pH the dead-bands are shown in red dashed lines. The differences in pH measurement due to the different analysers used between the scales were corrected for as mentioned in materials and methods 4.2.2 and 4.2

4.3.2 Cell growth

The cell growth characteristics including viable cell density (VCD), viability, and average cell diameter are shown in **Figure 2** for the two agitation rates of the ambr experiment, and the 10 L experiment. As described in our previous study [19], the fed-batch culture can be divided into a cell number increase phase (NI phase, day 0-4), a cell size increase phase (SI phase, day 4-8), and a stationary phase followed rapidly by a death phase (day 8-12). Similar cell growth and viability profiles were observed between the two agitation conditions in the ambr cultures. Compared between the two scales, the ambr cultures resulted in a slightly higher peak VCD ($\sim 8 \times 10^6$ cells \times mL $^{-1}$) than the 10 L cultures ($\sim 7 \times 10^6$ cells \times mL $^{-1}$). Due to a difference in inoculation method (see section 2.1), the ambr cultures are started with 100% fresh medium whereas the 10 L cultures are started with medium that contains 10% spent medium from the inoculum. This led to lower starting nutrient concentrations in the 10 L cultures and is a possible cause for a lower maximum VCD in the 10 L cultures. In addition, the two experiments were conducted from different pre-culture trains which may also affect the first few days of culture. The difference between the inocula of the two scales is further investigated in the following transcriptome analysis (section 4.3.6.1).

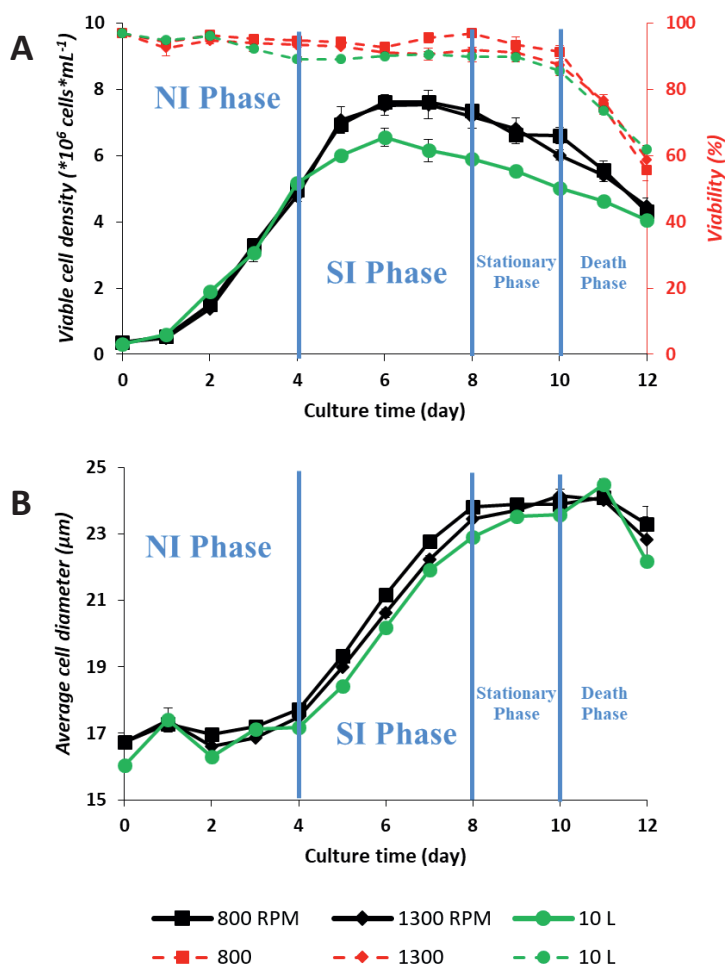


Figure 2. A. Viable cell density (VCD, 10^6 cells \times mL $^{-1}$) and viability (%), and B. average cell diameter (μ m) of ambr 800 RPM (closed square, black lines), ambr 1300 RPM (closed diamond, black lines), and 10 L (close circle, green lines) cultures. The error bars represent the standard deviation for ambr (n=5 for each condition) and 10 L (n=3) cultures.

4.3.3 mAb production and glycan distribution

Product characteristics including product titer, cell-volume based specific productivity, and N-glycan distribution are presented in **Figure 3** for the two agitation conditions of the ambr and 10 L cultures. In a previous study [19], it was shown that the specific productivity increases linearly with the cell volume. The cell-volume specific

productivity is constant during the whole process and therefore for comparison of the specific productivity the cell volume-based specific productivity is used instead of the cell-based value. First of all, the two agitation conditions in the ambr gave similar product characteristics. Compared between scales, the ambr cultures yielded a slightly higher mAb titer ($750 \text{ mg} \times \text{L}^{-1}$) than the 10 L cultures ($650 \text{ mg} \times \text{L}^{-1}$) (**Figure 3A**). However, the cell volume-based specific mAb productivity (**Figure 3B**) was the same between the two scales, meaning that the difference in mAb titer was caused by the difference in viable cell density, since the volume per cell is comparable between scales (**Figure 2**). Thus, the difference in mAb titer was most likely a result of the different inoculation methods (as discussed in section 3.2) and was not related to the difference between the two bioreactor systems. A difference in the N-glycan distribution is observed at culture day 12 between the ambr and 10 L cultures, represented by a higher fraction of G0 and G1, and a lower fraction of G0F for the ambr cultures compared to the 10 L cultures (**Figure 3C**). A survey from the historical data using the same cell line and culture media system (not published data from the Bioprocess Engineering group of Wageningen University) showed a 10-20% variation of the mAb N-glycan species, which could be caused by the difference in the pre-culture batches and in the sampling time. Considering this variation, the difference in mAb glycosylation between the ambr and 10 L experiment in this study (**Figure 3C**) is minor.

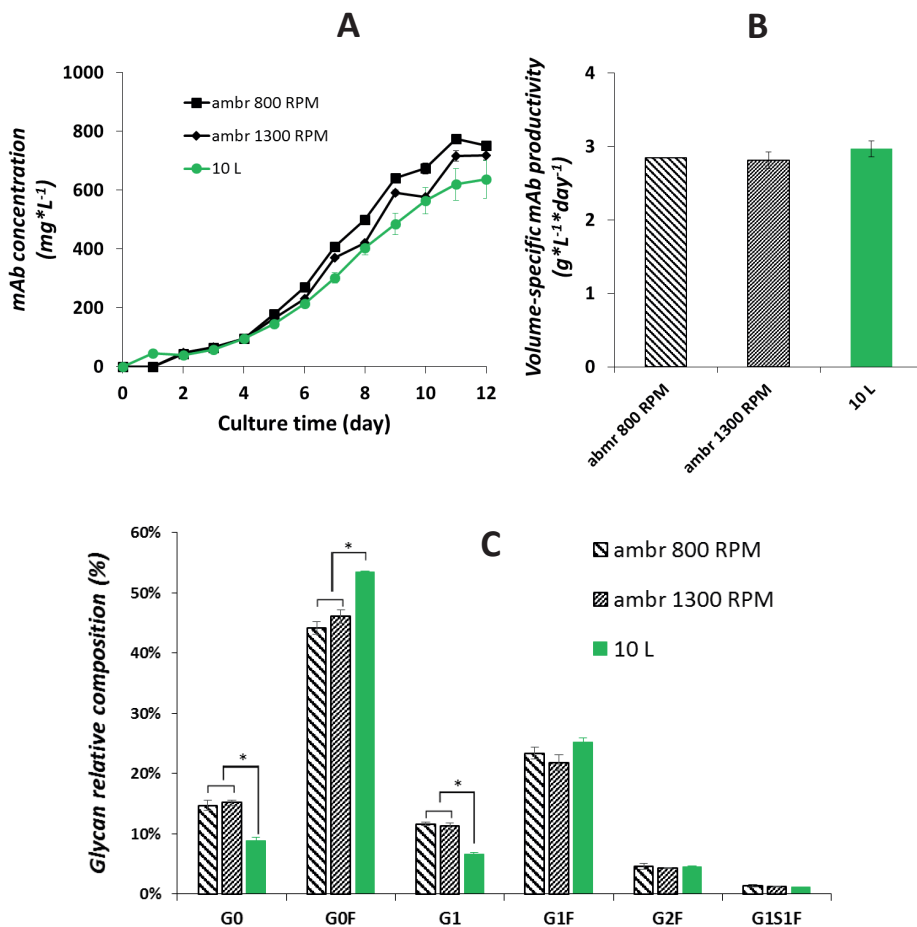


Figure 3. A. mAb concentration ($\text{mg} \times \text{L}^{-1}$) of ambr 800 RPM (closed square, blue line), ambr 1300 RPM (closed diamond, blue line), and 10 L (close circle, green line) cultures. **B.** Cell volume-specific mAb productivity ($\text{g} \times \text{L}^{-1} \times \text{day}^{-1}$). **C.** Relative mAb N-glycan composition (%) of ambr 800 RPM, ambr 1300 RPM, and 10 L cultures of the culture day 12. N-glycans with different numbers of terminal residuals (G: galactose, F: fucose, S: sialic acid) are shown. *: $P < 0.05$, T-test. The error bars represent the standard deviation ($n=3$) of biological triplicates.

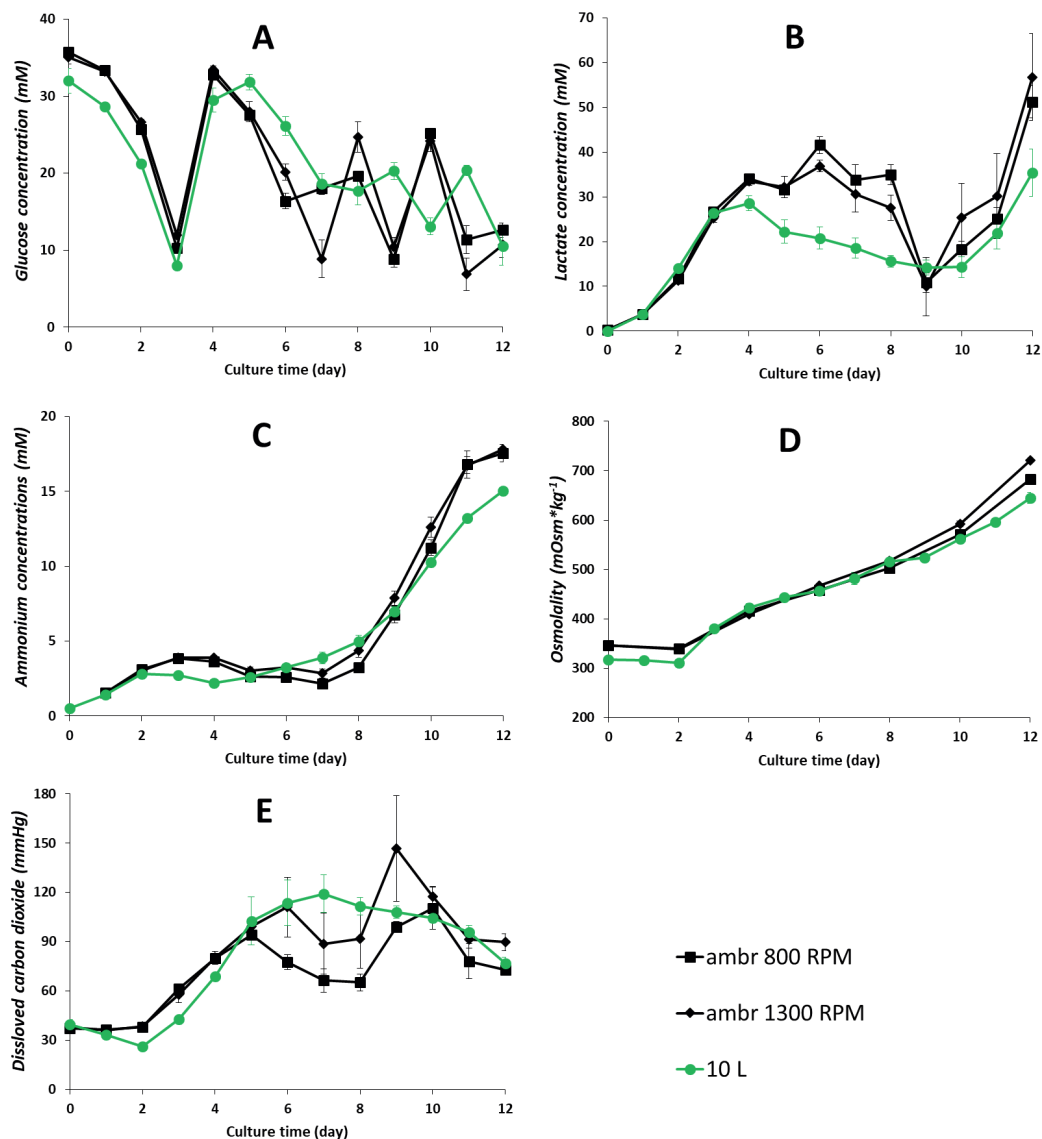


Figure 4. Profiles of the extracellular glucose concentration (mM) (A), lactate concentration (mM) (B), ammonium concentration (mM) (C), osmolality (mOsm \times kg $^{-1}$) (D), and dissolved CO₂ (mmHg) (E) of ambr 800 RPM (closed square), ambr 1300 RPM (closed diamond), and 10 L (closed circle) cultures. All measurements that were done using different methods were corrected between the two scales. The error bars show standard deviation for ambr (n=5 for each agitation condition) and 10 L (n=3) cultures.

4.3.4 Metabolite profiles

The primary metabolite concentrations and osmolality profiles are shown in **Figure 4** for the two agitation conditions tested in the ambr, and in the 10 L cultures. The other measured metabolite concentrations are shown in **Supplementary B**. The two agitation conditions in the ambr experiment gave similar results, only the pCO₂ profile was lower at 800 RPM from day 5 onwards. This was probably the result of increased CO₂ stripping due to the fact that the 800 RPM condition required a higher gas flow rate compared to the 1300 RPM condition to maintain the DO at the set-point. The higher gas flow rate was needed because of the lower oxygen transfer from the headspace due to the lower agitation rate.

The glucose concentration profiles showed a similar pattern between the two scales. Note that the slightly lower glucose concentration and osmolality on the first days in the 10 L bioreactor were caused by the different treatments in inoculum media as explained before. The lactate concentration for both scales showed a similar increase until day 3. From day 4 to day 8, however, the lactate concentration was up to 20 mM higher in the ambr as compared to in the 10 L cultures. This was a consequence of both the higher cell density (**Figure 2A**) and the higher specific lactate productivity (**Figure 5a**) in the ambr as compared to in the 10 L culture. After day 9, all cultures showed a sharp increase in the lactate concentration again, which coincided with the decrease in viability (**Figure 2**). The profiles of the ammonium concentration and osmolality were similar between the two scales. The culture osmolality increased from day 3 on for all conditions due to the feed addition. pCO₂ levels were comparable until day 5. From day 5 to day 9, a difference in the pCO₂ level was observed between the two scales. The pCO₂ in the ambr 1300 RPM cultures is more comparable to the 10 L cultures and they are both higher than in the ambr 800 RPM cultures.

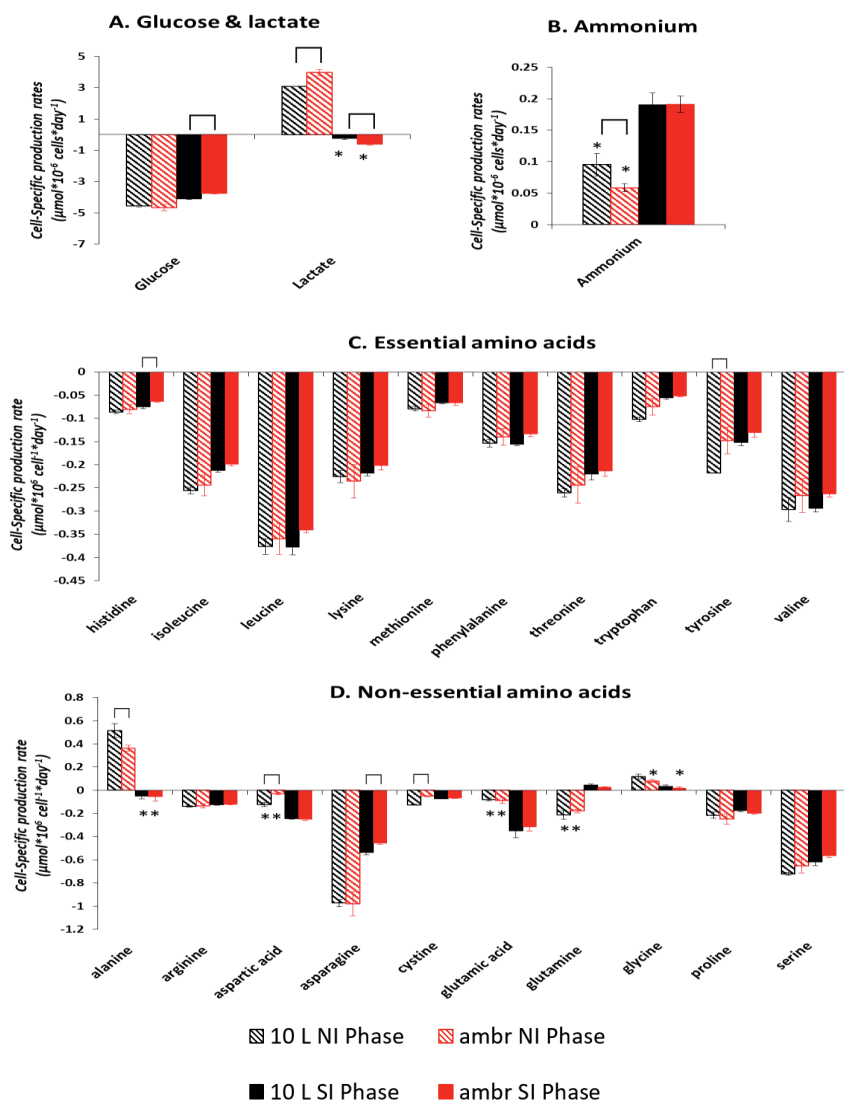


Figure 5. Average cell-specific production rates ($\mu\text{mol} \times 10^6 \text{ cell}^{-1} \times \text{day}^{-1}$) of glucose and lactate (A), ammonium (B), essential amino acids (C), and non-essential amino acids (D) of the 10 L (black) and the ambr (red) cultures during the number increase (NI, stripes), and the size increase (SI, filled) phase. Positive values indicate production, negative values indicate consumption. The error bars show the standard deviation ($n=3$) of biological replicates. * shows the rate was not constant over the measurement points in time ($R^2 < 0.9$). “ \square ” shows a significant difference ($P < 0.05$, T-test) between the values calculated from the two compared scales.

4.3.5 Cell metabolism

Small differences can be seen in the extracellular product (**Figure 3A**), lactate (**Figure 4B**), and nutrient (**Supplementary B**) concentrations between the two scales, which might be caused by the small difference in viable cell density. In order to compare the two scales at the cellular level we investigated the metabolism, by looking at specific metabolite consumption and production rates, and the biomass composition. The specific rates are the same for the two agitation rates in the ambr (data not shown). This demonstrates that the two agitation rates calculated based on two scale-down criteria did not cause a difference in cell metabolism. Therefore, in this section, only the ambr 800 RPM condition is shown and compared with the 10 L bioreactor.

4.3.5.1 Specific consumption/production rates

The cell specific rates of the primary metabolites and amino acids are compared between the ambr and 10 L cultures for the NI phase SI phase in **Figure 5**. In a previous study [19], it was shown that the cell specific rates are more or less constant despite the increase in volume. Slightly higher specific lactate production and consumption rates are observed in the ambr cultures compared to the 10 L cultures (**Figure 5A**). A possible cause was the different pH and DO profiles in these two systems (**Figure 1**). First of all, slightly higher pH values from day 5 to 8 and pH spikes towards higher values were observed in the ambr cultures (**Figure 1D&E**), which may favor the transport of lactate out of cells [30]. Furthermore, it was reported by Serrato et al. (2004) that compared to a constant DO, oscillating DO values resulted in an increased glycolytic metabolism and an increased lactate yield. Next to the specific rate of lactate production, minor differences in specific rates between the two scales can be observed for ammonium during the NI phase, and for a few amino acids in both the NI and the SI phase (**Figure 4B**). Although the values for these amino acids are statically different between the two scales, the differences are so small that they are not of biological relevance. Overall, the specific metabolic rates of these primary metabolites are very comparable between the two scales.

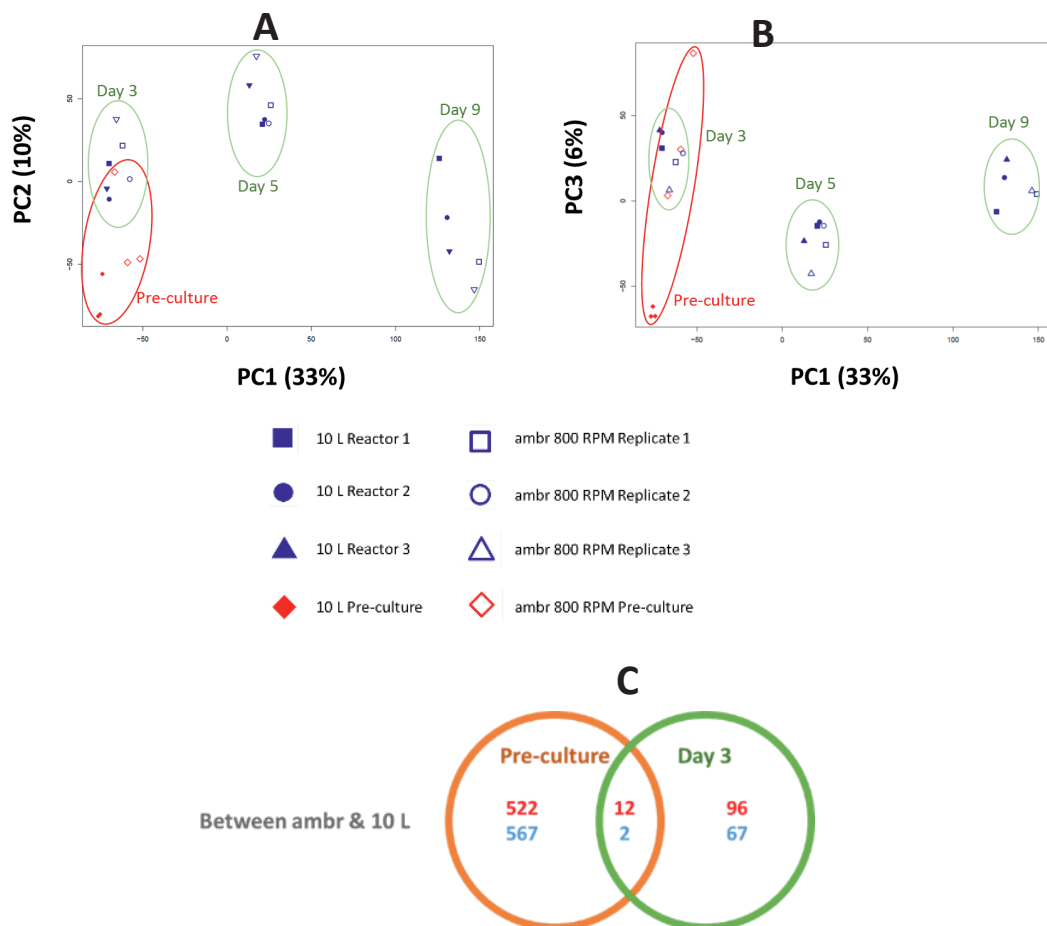


Figure 6. (A&B) Score plots generated from the transcriptome results of the 10 L (filled markers), ambr experiments with 800 RPM agitation rate (open markers), and pre-cultures at the moment of inoculation (filled diamond: pre-culture 10 L reactor, open diamonds: pre-culture ambr reactor). Samples for pre-cultures are grouped by red circles. Samples on the same culture day are grouped by green circles. On day 9, one replicate of the ambr culture was excluded due to a technical error in the microarray. **(C)** Venn diagram of the number of genes that are differentially regulated between the ambr and 10 L cultures for the pre-culture samples and the samples on day 3.

4.3.5.2 Biomass composition

The cell size increased during the fed-batch process (**Figure 2B**). In our previous study it was shown that the biomass composition changes when the cells increase in size [19]. The biomass composition between two scales is compared on day 7 that represents the cell size increase phase. In addition to biomass dry weight (DW), the relative composition (w/w) of several main biomass components were measured including protein, fatty acids, and carbohydrates (**Table III**). The results show identical biomass composition between the 10 L and the ambr cultures on culture day 7. Together with the similar cell diameter profiles (**Figure 2B**), this indicates highly comparable biomass dynamics between 10 L and ambr cultures.

It can be concluded that the process performance in terms of cell growth, product titer, glycosylation, and specific consumption/production rates of metabolites in the two bioreactors with different scales showed very comparable results. Thus, for the specific process used in this study, agitation seems to have no impact on the process performance and the scale-down based on equal k_{La} seems to be the best choice. However, this may be different for cells that are more sensitive to shear by agitation for which differences between both systems may occur and the P/V value has to be taken into account as well. Finally, also the pCO_2 level and cell death through sparging may still influence the process performance and have to be further studied.

Table III. Biomass composition of the CHO cells from the 10 L and ambr cultures on day 7 of the fed-batch cultures.

	10 L		ambr	
Dry weight (pg×cell⁻¹)	1530	±77	1483	±43
Protein (w/w %)	70.0	±1.2	71.3	±1.3
Fatty acids (w/w %)	4.5	±0.1	4.4	±0.2
Carbohydrates (w/w %)	3.8	±0.2	3.9	±0.1

4.3.6 Transcriptome analysis

Gene expression was measured using CHO Gene microarrays of the cells in the inocula just before inoculation and on day 3, 5, and 7 in both systems in triplicate. For the transcriptome study, only the 800 RPM condition of the ambr experiment is used to compare to the 10 L experiment. First, principal component analysis (PCA) is used to study the global variation in gene expression between the two scales. Next, functional analysis is performed for the genes that are uniquely regulated in one of the systems in order to study their relationship with the differences between the two bioreactor systems and with cell behavior.

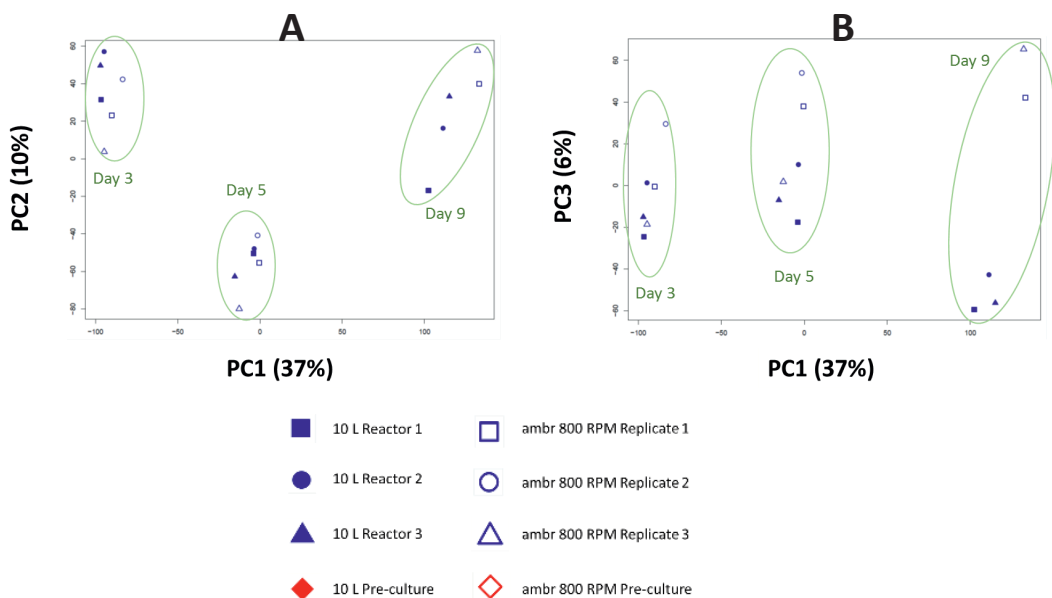


Figure 7. Score plots for the transcriptome results of the 10 L (filled markers) and ambr experiments with 800 RPM agitation rate (open markers) on day 3, 5, and 9. On day 9, one replicate of the ambr culture was excluded due to a technical error in the microarray.

4.3.6.1 Global variation analysis in gene expression

The principal component analysis (PCA) result is presented in **Figure 6A&B**, which shows the overall variation in gene expression for all analyzed samples including the inocula. The inocula are included at first because the difference in the inocula may cause differences in cell behavior between both systems. The principal components (PC) PC1, PC2, and PC3 represent 33%, 10%, and 6% of the total variation, respectively. The other individual PCs have only a minor contribution (less than 5% per individual component) to the total variation and are not considered. The variance explained by PC1 correlates with the culture development over time. For the variance explained by PC2, no clear unique correlation to the culture performance can be found. PC3, which represents 6% of the total variance, mainly explains the variance between the inocula. After three days of culture, both systems group together on all three

principle components and the variance in gene expression observed in the inocula has disappeared.

The genes that are significantly differentially expressed ($\text{FDR.BH} < 0.05$ and absolute $\text{FC} > 1.4$, see section 2.6) between the two pre-cultures are compared to those that are significantly differentially expressed between both systems on day 3 (**Figure 6C**). Between the two inocula, 1089 genes (total number of genes in the pre-culture circle) were differentially expressed, while between the two systems on day 3, only 167 genes (total number of genes in the day 3 circle) were differentially expressed. Moreover, of the genes that were differentially expressed between the inocula, only 14 were also differentially expressed between both systems on day 3. Thus the differences in gene expression on day 3 involve different genes than those that caused differences between the inocula. In conclusion, the difference in gene expression between the inocula probably did not cause differences in gene expression between the two scales later on and thus did not influence the scale comparison. As mentioned in section 3.2, the inocula for the two scales were treated in different ways. This does complicate the interpretation of the results a bit for the difference that is observed in the cell and product concentration between the two scales, which may now be caused by a difference in inoculation or a difference between scales. However, the specific metabolic and production rates and the profiles for viability, cell growth, and cell diameter are comparable between the two scales. Furthermore, here it was shown that on a transcriptome level the effect of different inocula is not present anymore on day 3. Thus, it is most likely that the small differences in biomass and product concentration are due to the difference in inoculation method and is not associated with the bioreactor system. In the future, this can easily be solved by also adding the pre-culture directly to the ambr reactor without replacing the medium.

Next, the differences in gene expression between the ambr and 10 L system are studied by comparing the gene expression data from day 3, 5, and 9, excluding the inocula samples. The results of PCA are shown in **Figure 7**. PC1, PC2, and PC3 now represent 37%, 10%, and 6% of the total variation, respectively. The other individual PCs each have only a minor contribution (less than 5% per individual component) to the total gene expression variance and are not considered. Samples taken on the same culture day of the two scales are closely grouped on PC1 and PC2 (**Figure 7A**). PC1 again correlates with the culture development over time. PC2 shows a transient

pattern that correlates with the cell size increase, since the cell size increased on day 5 but not on day 3 and 9. Based on PC1 and PC2 there is a slight separation of the ambr and 10 L cultures on day 3 and day 9. When looking at the development in time as represented by PC1 moving to the right and PC2 moving to a minimum on day 5, it seems that the ambr system is slightly ahead of the 10 L cultures in its development. In agreement with this, slightly larger cell diameters during the SI and the stationary phase are observed in the ambr cultures (**Figure 2B**). PC3 relates to the difference between the two scales. As can be seen the difference becomes more prominent for the later stage of the cultures (day 5 and 9). PC3 describes 6% of the total variance, which is much less than the variance that is due to the changes in gene expression over time (47%, PC1+PC2). To analyze whether this global difference in gene expression between the scales is relevant for the evaluation of the scale-down, the function of the genes that are differentially expressed between both systems were subsequently studied in more detail.

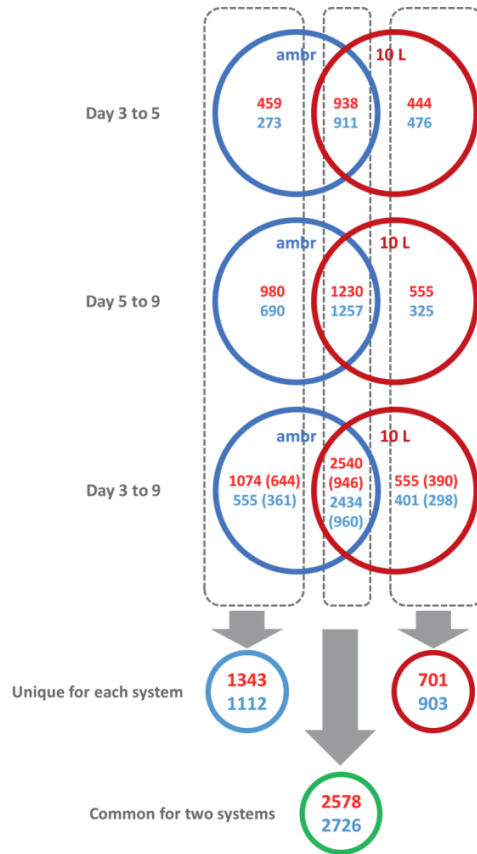


Figure 8. Venn diagram of the number of genes that are differentially regulated from day 3 to 5, 5 to 9, and 3 to 9 for the ambr and 10 L cultures. Red numbers indicate up-regulation, blue numbers indicate down-regulation. Numbers in brackets from day 3 to 9 show the number of genes significantly regulated from day 3 to 9, but not from day 3 to 5 and day 5 to 9. The number of regulated genes that are unique to each system, and are common for two systems (dash boxes) are also shown. A gene is unique to a system means it is differentially expressed in at least one of the three time comparisons for that system and never differentially expressed in one of the three time comparisons for the other system

4.3.6.2 Gene functional analysis

For the functional analysis, first, for each system the genes that are significantly differentially (as shown in section 2.6) expressed from day 3 to 5, day 5 to 9, and day 3 to 9 are selected. Genes are considered differentially expressed between two days if the FDR.BH is smaller than 0.05 and the absolute fold-change (FC) is larger than 1.4 (section 2.6). Next, the differentially expressed genes over time are compared between the ambr and 10L system and the result is shown in **Figure 8**. Note that the day 3-9 comparison picks up genes for which the expression level goes up or down slowly in time such that they were still insignificant in the 3-5 and 5-9 time frames. However, the day 3-9 time frame misses genes that are first up-regulated and then down-regulated or *vice versa*. These genes are picked up by the day 3-5 and 5-9 time frames. The number of the genes that slowly change over time are shown between brackets in the day 3-9 time frame in **Figure 8**. Their number is roughly half of the total regulated genes from day 3-9. The Venn diagrams show for each time frame the number of genes that change in the same way for both systems, and the number of genes that change uniquely in a specific system and not in the other. From these comparisons, genes are selected that are only regulated in time in one system and never in the other system (unique for the system) as well as the genes that are always regulated in the same way in both systems (common for two systems). The numbers are given at the bottom of the figure. In this way differences in gene expression between the two systems due to time delays are filtered out. It can be seen from **Figure 8** that: (i) The amount of commonly regulated genes is about twice the uniquely regulated genes for each system except for the up-regulated genes in ambr from day 5-9 where it is more or less comparable. (ii) From day 3-5 the amount of uniquely regulated genes is the same between the two systems whereas from day 5-9 it is twice as much in the ambr as compared to the 10 L system. This is in agreement with the observation in the PCA analysis that the difference between both systems becomes progressively more with time. This may be related to the larger fluctuations in conditions, like the pH, DO, and nutrient concentrations in the ambr system. (iii) The total amount of uniquely regulated genes in two systems is 4059 (total number of genes in “unique for each system” in **Figure 8**), which is in the same range as the total amount of common genes in two systems (5304, “common for two systems” in **Figure 8**). The total number of uniquely regulated genes is higher in the ambr than in the 10L

reactor, which again may be due to the higher fluctuations in conditions like pH, DO, and nutrient concentrations.

The biological function of the identified genes that were uniquely regulated in one of the systems (the genes in “unique for each system” in **Figure 8**) are next categorized based on the KEGG pathway database and their up/down-regulations are presented in **Supplementary C**. The aim is to study whether these genes specifically belong to certain pathways that can be linked to the differences that exist between the scales.

The genes unique to a system are distributed over almost all the functional pathways, including global metabolism of carbohydrate, protein, lipid, nucleotide, energy, signal transduction, etc. In the ambr cultures, overexpression of the hypoxia-inducible factor 1 (HIF-1) signaling pathways (**Supplementary C**) is observed. HIF-1 is known to respond to extracellular oxygen levels and to play a role in anaerobic respiration and lactate formation [32]. The higher regulation of the HIF-1 signaling pathways in ambr may be caused by the fluctuations in DO (**Figure 1**) and may be related to the higher lactate production (**Figure 5A**). Moreover, as mentioned before, higher pH levels from day 4 to 8 and higher pH spikes towards higher values are observed in the ambr cultures (**Figure 1D&E**), which would favor the transport of lactate out of cells. The transport is done by H⁺-monocarboxylate cotransporters (MCTs) [33]. Among several MCTs (MCT1-MCT4) that are known for lactate transport, MCT2 (encoded by *SLC16A7* gene) has a higher affinity for substrates compared to the other MCTs [34] and it showed a higher up-regulation from day 3-9 in the 10 L cultures (3.6 FC, FDR.BH=0.00) compared to that of in the ambr cultures (2.2 FC, FDR.BH=0.00) (see the data in **Supplementary A**). This might be associated with the differences in lactate production/consumption between the two systems (**Figure 5A**). Regulations of MCT1, MCT3, and MCT4 are however similar between the two systems. Furthermore, for the N-glycan biosynthesis pathway (**Supplementary C**), the one up-regulated gene unique to the 10 L system is *Fut8* (1.5 FC from day 3-9, FDR.BH=0.00, see the data in **Supplementary A**) which gene encodes the fucosyltransferase that is responsible for the α -1,6-fucosylation. This agrees with the observation of the higher percentage of fucosylated mAb in the 10 L cultures (**Figure 3C**).

In summary, only for a few genes a possible relationship between their expression and differences that exist between both scales can be found, which could be related to the fact that the differences between both systems in terms of process performance are

small. Given this fact, the fraction of uniquely regulated genes of 20% is rather high (4059 uniquely regulated genes out of a total of 20858, **Figure 8**). Possibly the difference in gene expression of these genes between both scales represents an adaptive response to differences that exist between scales, leading to a comparable process performance. Overall, this study shows that transcriptomics is a sensitive tool to measure differences in the physiological response of cells to different bioreactor environments. However, these differences in gene expression could not be related to the process performance, which was comparable between scales or differences in the bioreactor systems. To be able to rate the importance of the observed differences, comparison is needed on gene expression data of other scale comparisons from both successful and failed scale-up/scale down experiments. These data are currently lacking in literature.

4.4 Conclusion

For the specific cell line and the fed-batch process tested in this study, the results obtained in the ambr at two different agitation rates are comparable to that obtained in a 10 L bioreactor in terms of cell growth, metabolism, productivity, and product quality. Transcriptome analysis showed differences in gene expression between both systems. The number of genes that were significantly regulated over culture time in only one of the systems and never in the other was higher in the ambr (2455) than in the 10 L system (1604), which may be related to the higher fluctuation of conditions like DO, pH and nutrient concentrations in the ambr. In general, these genes were spread over all KEGG pathways and could not be linked to differences in bioreactor system or differences in cell physiology. For a few genes possible functional relationships with differences in DO and pH patterns, lactate metabolism, and glycosylation pattern could be identified. Furthermore, transcriptome analysis showed that a difference in gene expression between the inocula was not present anymore after three days of culture, and thus did not influence the system comparison. In addition to existing literature, this study further strengthens the opinion that the ambr system gives representative culture performances for the 10 L bench-scale bioreactor. Differences in gene expression between scales could be observed by using transcriptome analysis. However, the differences could not be linked to specific

process conditions. More studies are needed to determine the sensitivity of the transcriptome analysis on scale comparison.

4.5 References

- [1] P. Xu *et al.*, "Characterization of TAP Ambr 250 disposable bioreactors, as a reliable scale-down model for biologics process development," *Biotechnol. Prog.*, 2016.
- [2] Z. Xing, B. M. Kenty, Z. J. Li, and S. S. Lee, "Scale-up analysis for a CHO cell culture process in large-scale bioreactors," *Biotechnol. Bioeng.*, vol. 103, no. 4, pp. 733–746, 2009.
- [3] J. Yang *et al.*, "Fed-batch bioreactor process scale-up from 3-L to 2,500-L scale for monoclonal antibody production from cell culture," *Biotechnol. Bioeng.*, vol. 98, no. 1, pp. 141–154, 2007.
- [4] F. Li, Y. Hashimura, R. Pendleton, J. Harms, E. Collins, and B. Lee, "A systematic approach for scale-down model development and characterization of commercial cell culture processes.," *Biotechnol. Prog.*, vol. 22, no. 3, pp. 696–703, 2006.
- [5] M. H. Sani, "Evaluation of Microwell based Systems and Miniature Bioreactors for Rapid Cell Culture Bioprocess Development and Scale-up," pp. 1–214, 2016.
- [6] J. I. Betts and F. Baganz, "Miniature bioreactors: current practices and future opportunities.," *Microb. Cell Fact.*, vol. 5, p. 21, 2006.
- [7] A. W. Nienow *et al.*, "The physical characterisation of a microscale parallel bioreactor platform with an industrial CHO cell line expressing an IgG4," *Biochem. Eng. J.*, vol. 76, pp. 25–36, 2013.
- [8] W. T. Hsu, R. P. S. Aulakh, D. L. Traul, and I. H. Yuk, "Advanced microscale bioreactor system: A representative scale-down model for bench-top bioreactors," *Cytotechnology*, vol. 64, no. 6, pp. 667–678, 2012.
- [9] S. Moses and M. Manahan, "Assessment of AMBRTM as a model for high-throughput cell culture process development strategy," *Adv. Biosci. ...*, vol. 2012, no. November, pp. 918–927, 2012.
- [10] A. W. Nienow *et al.*, "Scale-down studies for assessing the impact of different stress parameters on growth and product quality during animal cell culture," *Chem. Eng. Res. Des.*, vol. 91, no. 11, pp. 2265–2274, 2013.
- [11] S. Rameez, S. S. Mostafa, C. Miller, and A. A. Shukla, "High-throughput miniaturized bioreactors for cell culture process development: Reproducibility,

- scalability, and control," *Biotechnol. Prog.*, vol. 30, no. 3, pp. 718–727, 2014.
- [12] Y. Rouiller, J. M. Bielser, D. Brühlmann, M. Jordan, H. Broly, and M. Stettler, "Screening and assessment of performance and molecule quality attributes of industrial cell lines across different fed-batch systems," *Biotechnol. Prog.*, vol. 32, no. 1, pp. 160–170, 2016.
- [13] V. Janakiraman, C. Kwiatkowski, R. Kshirsagar, T. Ryll, and Y.-M. Huang, "Application of high-throughput mini-bioreactor system for systematic scale-down modeling, process characterization, and control strategy development," *Biotechnol. Prog.*, 2015.
- [14] S. Siva *et al.*, "Leveraging high-throughput technology to accelerate the time to clinic: A case study of a mAb," *Eng. Life Sci.*, pp. 143–151, 2015.
- [15] A. Farrell, N. McLoughlin, J. J. Milne, I. W. Marison, and J. Bones, "Application of multi-omics techniques for bioprocess design and optimization in Chinese hamster ovary cells," *J. Proteome Res.*, vol. 13, no. 7, pp. 3144–3159, 2014.
- [16] K. P. Jayapal and C. T. Goudar, "Transcriptomics as a tool for assessing the scalability of mammalian cell perfusion systems," *Mamm. Cell Cult. Biol. Manuf.*, pp. 227–243, 2013.
- [17] J. B. Sieck *et al.*, "Development of a Scale-Down Model of hydrodynamic stress to study the performance of an industrial CHO cell line under simulated production scale bioreactor conditions," *J. Biotechnol.*, vol. 164, no. 1, pp. 41–49, 2013.
- [18] J. B. Sieck *et al.*, "Adaptation for survival: Phenotype and transcriptome response of CHO cells to elevated stress induced by agitation and sparging," *J. Biotechnol.*, vol. 189, pp. 94–103, 2014.
- [19] X. Pan, C. Dalm, R. H. Wijffels, and D. E. Martens, "Metabolic characterization of a CHO cell size increase phase in fed-batch cultures," *Appl. Microbiol. Biotechnol.*, vol. 101, no. 22, pp. 8101–8113, 2017.
- [20] K. Van't Riet and J. Tramper, *Basic bioreactor design*. CRC Press, 1991.
- [21] K. Lin *et al.*, "MADMAX - Management and analysis database for multiple ~omics experiments," *J. Integr. Bioinform.*, vol. 8, no. 2, p. 160, Jul. 2011.
- [22] R. A. Irizarry *et al.*, "Exploration, normalization, and summaries of high density oligonucleotide array probe level data," *Biostatistics*, vol. 4, no. 2, pp. 249–264, 2003.

- [23] M. Dai *et al.*, “Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data,” *Nucleic Acids Res.*, vol. 33, no. 20, pp. 1–9, 2005.
- [24] M. E. Ritchie *et al.*, “limma powers differential expression analyses for RNA-sequencing and microarray studies.,” *Nucleic Acids Res.*, vol. 43, no. 7, p. e47, Apr. 2015.
- [25] M. A. Sartor, C. R. Tomlinson, S. C. Wesselkamper, S. Sivaganesan, G. D. Leikauf, and M. Medvedovic, “Intensity-based hierarchical Bayes method improves testing for differentially expressed genes in microarray experiments.,” *BMC Bioinformatics*, vol. 7, p. 538, Dec. 2006.
- [26] Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” *J. R. Stat. Soc. Ser. B*, vol. 57, no. 1, pp. 289–300, 1995.
- [27] B. C. Mulukutla, M. Gramer, and W. S. Hu, “On metabolic shift to lactate consumption in fed-batch culture of mammalian cells,” *Metab. Eng.*, vol. 14, no. 2, pp. 138–149, 2012.
- [28] D. B. Allison, X. Cui, G. P. Page, and M. Sabripour, “Microarray data analysis: From disarray to consolidation and consensus,” *Nat. Rev. Genet.*, vol. 7, no. 1, pp. 55–65, 2006.
- [29] I. H. Yuk *et al.*, “Effects of copper on CHO cells: Cellular requirements and product quality considerations.,” *Biotechnol. Prog.*, pp. 226–238, 2014.
- [30] E. Trummer *et al.*, “Process parameter shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors,” *Biotechnol. Bioeng.*, vol. 94, no. 6, pp. 1033–1044, 2006.
- [31] J. A. Serrato, L. A. Palomares, A. Meneses-Acosta, and O. T. Ramírez, “Heterogeneous conditions in dissolved oxygen affect N-glycosylation but not productivity of a monoclonal antibody in hybridoma cultures,” *Biotechnol. Bioeng.*, vol. 88, no. 2, pp. 176–188, 2004.
- [32] T. N. Seagroves *et al.*, “Transcription factor HIF-1 is a necessary mediator of the pasteur effect in mammalian cells.,” *Mol. Cell. Biol.*, vol. 21, no. 10, pp. 3436–44, 2001.
- [33] R. C. Poole and A. P. Halestrap, “Transport of lactate and other

- monocarboxylates across mammalian plasma membranes.," *Am. J. Physiol.*, vol. 264, no. 4 Pt 1, pp. C761-82, 1993.
- [34] A. P. Halestrap and N. T. Price, "The proton-linked monocarboxylate transporter (MCT) family: structure, function and regulation," *Biochem J*, vol. 343 Pt 2, pp. 281–299, 1999.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jbiotec.2018.05.0>

Chapter 5

Transcriptome analysis as a tool for characterizing a CHO cell perfusion culture

This chapter is submitted as

Abdulaziz A. Alsayyari, Jort Altenburg, Jos A. Hageman, Guido J. Hooiveld, Rene H. Wijffels, Dirk E. Martens, “Transcriptome analysis as a tool for characterizing a CHO cell perfusion culture”.

Abstract

Perfusion cultivation technology provides a practical alternative for cost-effective and efficient glycoprotein production for continuous biomanufacturing. In this study, we explore the use of transcriptomic profiling for characterisation of CHO cell perfusion cultures. This study shows that transcriptome data can be linked to specific phases of a perfusion process like a period characterized by a cell size increase and a period with a specific nutrient limitation. More importantly, quantitative transcriptome data visualized using Principle Component Analysis (PCA) show a clear separation between the steady state data points and the other data points. During the steady state a high consistency in the glycosylation patterns is observed while for the non-steady state period variation in glycosylation patterns are seen. This paper shows the potential of integrating gene expression data into process development and process characterisation in mammalian cell perfusion culture.

5.1 Introduction

There is an increasing industrial interest to implement continuous biomanufacturing for the production of therapeutic glycoproteins in order to reduce the production costs [1]. Furthermore, both regulatory authorities and industry support continuous manufacturing as a way to reduce variability in product quality [2]. With respect to the upstream cultivation process, perfusion technology forms the fundamental cornerstone for continuous biomanufacturing.

In perfusion cultures cell concentrations can be reached that are substantially higher than in fed-batch processes and consequently higher volumetric productivities are achieved [3]. Furthermore, a steady state can be reached that in theory can be maintained for longer periods of time, resulting in less down time and thus contributing to a further increase in volumetric productivity. In this steady state, the conditions are constant in time, which in theory should result in a more constant product quality over time and a reduction in release testing of product quality. With a proper characterisation and control of the steady state, release testing may not be necessary anymore and product release can be done based on measurement of the critical steady state parameters. Finally, the higher volumetric productivity means that smaller bioreactor systems can be used meaning scale-up from clinical to commercial scale is easier and can be done faster.

In the past robustness and scale-up of the cell separation systems were a problem, which was the main cause preventing the use of perfusion systems [4]. Thus, in the last 30 years, perfusion cultivation was used mainly for the production of unstable therapeutic proteins, because the residence time of the protein in the reactor can be relatively short when using high perfusion rates [5]. For the more stable pharmaceutical proteins, such as monoclonal antibodies, fed-batch was and is still the preferred cultivation strategy [6], despite the fact that in principle it is less cost-effective than perfusion cultivation.

With the development of more robust and single use separation devices like the ATF system [7], perfusion technology has become more robust, scalable and flexible. Perfusion processes have a feed rate of usually 1 to 4 reactor volumes per day supplying fresh nutrients, a perfusion flow through the filter of spent medium and a certain bleed rate of cells and spent medium directly out of the bioreactor. The bleed rate usually has a value between 0 and 20% of the reactor volume per day, which is

needed to maintain an acceptable viability. Because the separation efficiency of the filter is often not 100% also the perfusion flow contributes to the bleed. The ideal perfusion process consists out of a short growth phase followed by a long steady state production phase. In the current practice, however, the period before steady state is reached may still be quite long and dynamic and the product quality may vary and be different from that in the steady state phase. Consequently, the volumetric productivity during part of the process will be suboptimal. In addition, part of the product produced may not be of sufficient quality and has to be discarded, which further lowers the volumetric productivity.

In order to reach steady state earlier and achieve higher volumetric productivities during the steady state a good understanding of the relation between the different process parameters, like feed rate, feed composition and bleed rate, and cell performance is needed. Moreover, perfusion cultivation in biopharmaceutical practices are associated with extra validations compared to fed-batch cultivation in terms of cellular performance, definition of the steady state phase, and quality compliance. Next to other analytical methods, transcriptome analysis is possibly a good tool to characterise the physiological state of the cells and find the relation between the physiological state and the different process parameters.

The aim of this study is to evaluate transcriptomics as a tool to characterize the dynamics of CHO-cell perfusion cultures as well as the steady state production phase. Quantitative transcriptome data visualized using principle component analysis (PCA) show a clear separation between pre-steady state and steady state. In addition principal components could be linked to culture parameters like, cell size, high cell density and metabolic activities.

5.2 Materials and Methods

5.2.1 Cell line and Culture medium

A recombinant CHO cell line producing an IgG was used (provided by Bioceros Holding BV, The Netherlands). The cells were grown in ActiCHO P Base (GE Lifesciences, United Kingdom), supplemented with 4mM glutamine and 0.5% Anti-Clumping Agent (both from Invitrogen, United State).

Two selection markers (200 µg/mL Zeocin™ and 5 µg/mL Blasticidin both obtained from life Technologies, USA) were applied in the pre-cultures for four passages. During scale-up for inoculation as well as in the perfusion run no selection markers were added. Perfusion medium was sterilized by 0.2 µm microfiltration using Sartobran® P 300 filters (Sartorius, Germany), then stored directly at 8°C for later use within 5 days. The pre-culture was incubated in un-baffled shake flasks at a shaking speed of 100 rpm, a humidified atmosphere with 5% CO₂ at 37°C (Multitron CO₂, Infors HT).

5.2.2 Bioreactor Operation

The perfusion run was carried out in a 3L bioreactor with a 1.5L working volume (Applikon Biotechnology, Netherlands) connected to a DASGIP® control system (Eppendorf, Germany). The cell retention device utilized was a 10L BioSep system, and an APS 990 controller (Applikon Biotechnology, Netherlands). The controller was set at run / stop cycle times of 10 min / 3s and a power level of 2W. The perfusion rate was 1 RV/day using a Watson-Marlow 205U (Watson-Marlow, England) peristaltic pump connected to the APS 990 controller. The recirculation tube had a length of 0.5 m and an internal diameter of 5 mm. The recirculation rate was maintained at 2.5 times the perfusion rate. A cell bleed of 0.1 RV/day was applied by using a dip tube. To ensure a constant reactor volume the reactor was placed on a balance, which was used to control the feed rate. The bioreactor was inoculated at initial density of 3.0×10^5 viable cells/mL.

5.2.3 Bioanalytical methods

Daily cell counts were done using the trypan blue exclusion method in combination with a TC20™ Automated Cell Counter (Bio-Rad Laboratories, United States), as well as with manual cell counts using a disposable counting chamber (INCYTO, Korea). Glucose and lactate levels of the culture were measured daily by an YSI 2700 Biochemistry Analyser (YSI Life Sciences, United States). Samples for product quality analysis and amino acids analysis were centrifuged at 10000 x g for 15 minutes and the supernatant was stored at -20°C until further analysis. The spent medium composition was analysed using NMR (Spinnoation biologics, Netherlands). The

product concentration was measured using an Agilent 1290 Infinity Binary LC System UHPLC and Bio-Monolith Protein A Column (Agilent Technologies, United States).

5.2.4 N-glycosylation analysis

3ml samples were desalted and purified with the Äkta Pure System (GE Lifesciences, United Kingdom) using a HiTrab MabSelect SuRe column (GE healthcare, United States). Samples were then concentrated and N-glycans enzymatically detached using PNGase F (Roche, Switzerland). In a next step, 2-AB reagent (Roche, Switzerland) was used to label oligosaccharides. Finally, N-glycans were analysed with a HILIC UPLC System on an ACQUITY UPLC BEH Glycan Amide column [8].

5.2.5 Gene expression data analysis

Gene expression samples were prepared by centrifuging 1.5 mL cell suspension at 300 x g for 15 minutes. Subsequently the supernatant was removed and the cell pellet was resuspended in 2 mL TRIzol® reagent (Life Technologies, United States) and immediately stored at -80°C. Samples stored at -80°C were thawed and next purified using the RNeasy Mini kit (Qiagen, USA), according to the manufacturer instructions. The total RNA concentration was quantified by a Nanodrop 200 (Thermo Scientific, Wilmington, DE). The RNA integrity number (RIN) was determined by RNA 600 Nano kit using 2100 Bioanalyzer (Agilent, Santa Clara, CA). Only RNA with RIN > 8 was used in transcriptome profiling analysis.

Affymetrix CHO Gene 2.1 ST arrays were used for transcriptome profiling (Affymetrix, Santa Clara, USA). In short, 100 ng of total RNA was labelled by the Whole-Transcript Sense Target Assay (Affymetrix) and hybridized to Affymetrix CHO Gene 2.1 ST arrays, according to the manufacturer instructions. The quality control and the data analysis pipeline, including normalization, have been described in detail previously [9] [10]. Normalized expression estimates of probe sets were computed by the robust multiarray analysis average (RMA) algorithm [11], as implemented in the Bioconductor library AffyPLM. Probe sets were redefined according to Dai et al. [12] using well-annotated reference sequences based on the CriGri_1.0 genome assembly (NCBI Reference Sequence Project (RefSeq) Release 72), which resulted in the profiling of 60626 annotated sequences (transcripts) (custom CDF v20). After averaging the

expression levels of probe sets targeting the same gene expression data for 20859 unique genes was obtained, which was used for all subsequent analyses.

5.2.6 Calculations

The separation efficiency is defined as:

$$SE = 100 \cdot \frac{C_{xvr} - C_{xvp}}{C_{xvr}} \quad (1)$$

Where SE is the separation efficiency (%), and C_{xvr} , C_{xvp} are respectively the viable cell concentration in the bioreactor and in the perfusion flow ($\text{cells} \cdot \text{m}^{-3}$).

The specific growth rate is calculated from a balance over the viable cells:

$$\frac{dC_{xvr}}{dt} = -BC_{xvr} - P(1 - SE/100)C_{xvr} + \mu C_{xvr} - \mu_d C_{xvr} \quad (2)$$

Where B and P are respectively the specific bleed rate (day^{-1}) and specific perfusion rate (day^{-1}), and μ and μ_d are respectively the specific growth and death rate (day^{-1}).

The specific death rate is calculated from the dead cell balance assuming that the separation efficiency is the same as for viable cells and there is no cell lysis:

$$\frac{dC_{xdr}}{dt} = -BC_{xdr}\mu_d - P(1 - SE/100)C_{xdr} + \mu_d C_{xvr} \quad (3)$$

Where C_{xdr} is the dead cell concentration in the reactor.

The specific production of the metabolites including product is calculated from their balance according to:

$$\frac{dC_i}{dt} = F(C_i^{in} - C_i) + q_i C_{xvr} \quad (4)$$

Where F is the specific feed rate (day^{-1}), which is the sum of the specific bleed and perfusion flow ($F=B+P$), C_i^{in} , C_i are respectively the concentration of compound i in the incoming feed flow and the bioreactor ($\text{mol} \cdot \text{m}^{-3}$, $\text{g} \cdot \text{m}^{-3}$) and q is the specific production rate of compound i ($\text{mol} \cdot \text{cell}^{-1} \cdot \text{day}^{-1}$, $\text{g} \cdot \text{cell}^{-1} \cdot \text{day}^{-1}$), where a negative value indicates consumption.

The accumulation term at the left side of equation 2,3 and 4 is not always zero and is estimated by:

$$\frac{dC_i}{dt}(t_2) = \frac{\frac{C_i(t_2) - C_i(t_1)}{t_2 - t_1} + \frac{C_i(t_3) - C_i(t_2)}{t_3 - t_2}}{2} \quad (5)$$

Where t_1, t_2, t_3 are three consecutive time points (day).

5.3 Results and discussions

A perfusion culture was run for a period of 39 days. The traditional key performance indicators (KPIs) like cell density and viability and the concentrations of various metabolites were measured daily and will be presented first. Next gene expression data will be presented and related to the KPIs.

5.3.1 Cell growth

Figure 1A&B show the cell concentration, viability, total biomass concentration as represented by total cell volume, and cell diameter. The culture was operated as a batch culture during the first three days. At day 3 perfusion and bleed were started at a rate of respectively one reactor volume per day and 0.1 reactor volume per day, which was maintained for the rest of the culture. From the start of the culture the cell concentration increased until day 8, during which the cell diameter was constant. After day 8 a sharp drop occurred in the cell concentration until about day 15 and at the same time the cell diameter increased from about 16 μm to 27 μm , which is equivalent to an about 5 times increase in cell volume. Notably, the biomass concentration as represented by the total cell volume per ml remained constant from day 8 until day 15. These changes can be explained by the fact that cells stop dividing, while biomass growth continues. The stop of cell proliferation can clearly be seen from the specific proliferation rate shown in Figure 1C. During this period from day 8 to day 15 a slight drop in viability occurred, which is related directly to the negative proliferation rates in Figure 1C. The stop of cell proliferation and concomitant increase of cell size was also observed earlier in fed-batch cultures using the same cell line and basal medium. In that study, it was shown that cells were arrested mainly in the G1/G0 phase and to a minor extent also in the G2/M phase [13].

From day 15 on cells started dividing again at a slow rate and at the same time the cell diameter dropped to the original values on day 20. During this period, the cell concentration remained more or less constant and the biomass concentration decreased with a factor 3. From day 20 on the cell diameter remained constant and

the cells grew with a relatively high growth rate to a cell density of about 14 million per ml on day 26, after which the cell concentration total cell volume and cell diameter remained constant and a steady state is reached.

To confirm that the observed dynamics is reproducible, a validation run was done. The results in comparison to the first run are shown in Appendix figure A2, A3, and A4. The validation run was terminated due to a contaminations on day 14. During these first 14 days, containing the main fluctuation, the results are very similar, showing that at least the main variation in cell concentration, glucose and lactate concentration and cell size is reproducible.

For CHO perfusion processes with a fixed perfusion and bleed rate the time to reach steady state is usually around 10-15 days [5,7]. This is somewhat shorter than the length of the dynamic period here, which is probably due to the growth arrest and cell size increase, which is a characteristic of the used medium cell line combination. For other cell lines substantial longer dynamic periods have been observed [14] . The dynamic period can be substantially shortened if the cell concentration is controlled at a certain level by changing the bleed rate using an on-line biomass measurement tool [15]. The cell concentrations reached are lower than those obtained in literature, which are usually between 40-60 10^6 /ml for comparable perfusion and bleed rates. Possibly this is due to glucose limitation as during steady state both glucose and lactate concentration are below the detection limit.

In summary, it took about 26 days for the system to reach steady state. The relatively long dynamic period in this specific situation seems related to the changes in cell size that occur in this period.

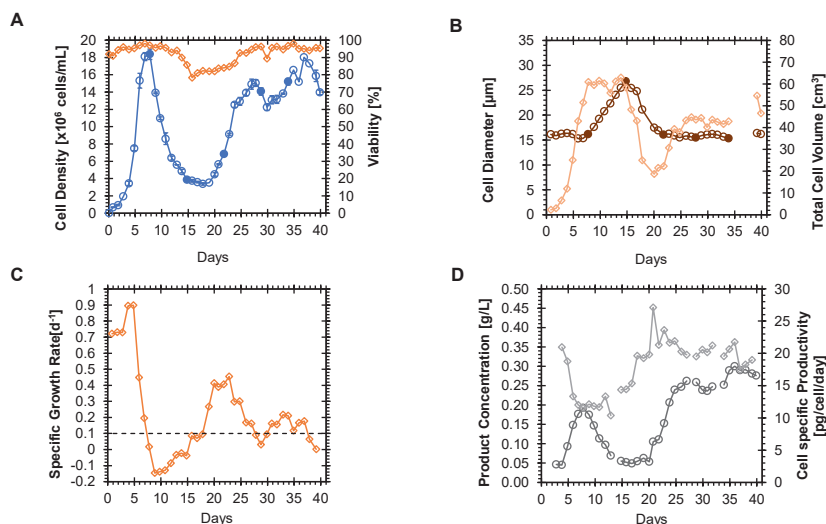


Figure 1 Cell performance data of a perfusion culture. Perfusion and bleed were started from day 3 on: **A:** Cell density (diamonds) and viable cell concentration (circles), **B:** Cell size (circles) and total cell volume (diamonds), **C:** Specific cellular proliferation rate. The dashed line indicates the specific bleed rate, **D:** Product concentration (circles) and cell specific productivity (diamonds).

5.3.2 Product formation and glycosylation patterns

The product concentration shows the same fluctuation as the cell density and became constant at day 26 (figure 1A). The specific productivity dropped from the beginning of the culture until day 8 and next stayed at a constant low value of 12 pg/cell/day until about day 15. From day 15 on it increased to the original value and became constant from day 26 on with a value of about 20 pg/cell/day. Thus, the period of low specific productivity coincided with the period where proliferation is arrested and the increase in cell size. This is in contrast to the work of Pan et al. where the cell specific productivity increased from about 7.5 pg/cell/day to 15 pg/cell/day during the cell size increase [16].

The glycosylation pattern is considered as a critical quality attribute for the glycoprotein therapeutics due to its clinical impact in terms of efficacy, pharmacokinetics, and side effects [17]–[20]. The glycosylation pattern consists of several glycan attributes such as fucosylation, galactosylation, and sialylation [21]. Therefore, the analysis of

glycosylation patterns is essential and part of the list of KPIs during the upstream phase [22].

The glycosylation profiles are plotted in Figure 2 and show the four major glycoforms (G0, G0F, G1, G1F, and G2F). The sample points were chosen in such a way that they represent various growth phases in the culture. The first five samples (day 7, 12, 13, 14 and 16) represent the pre-steady state, whereas the last four samples (day 27, 31, 34 and 40) represent the steady state. The steady state samples as expected show high consistency in the glycosylation pattern as was also observed by others [23]. In the pre-steady state the glycosylation pattern was clearly not consistent, which can be attributed to the varying conditions in the culture. Nevertheless, for some days (day 7 and 14), the profiles were still comparable to steady state glycosylation profile.

The main change occurring in the glycosylation profile was a drop in G0F at day 12 and a concomitant increase in G1. From day 12 to day 14 both glycoforms returned to the steady state level. This change occurred during the cell size increase part and thus also during the period of low specific productivity. Previous studies have shown the correlation between variations in glycosylation pattern and some of the process parameters such as dissolved oxygen tension (DOT), pH, temperature, and nutrient addition [24], [25]. Therefore, these changes in glycosylation pattern may be related to the changes in the nutrient and waste product concentrations during this period from day 5-18 (supplementary file A).

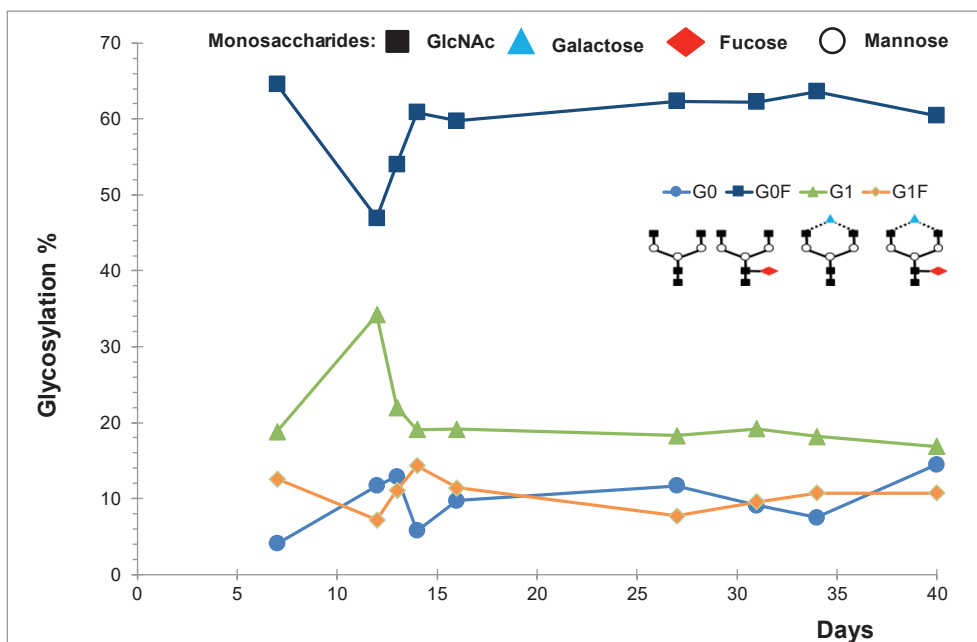


Figure 2: Glycosylation patterns: N-glycan structure distributions over time of the perfusion cultivation.

5.3.3 Metabolites: glucose and lactate concentration

During the first days, the glucose was rapidly consumed and lactate was produced as shown in Figure 3. The glucose concentration dropped below the detection level at day 6. From day 5 on lactate was consumed and the concentration was below the detection limit on day 7 which correlates with the stop of cell division. The lactate concentration started increasing slowly again from day 11. Likewise, from day 15 the glucose concentration increased again, which can be explained from the drop in the biomass concentration as represented by the total cell volume. This was followed by a more rapid increase in lactate concentration from day 16 till about day 21, which agrees with the more rapid proliferation and high glucose concentrations during this period. Glucose reached a peak at day 19 and with the increase in cell density dropped to zero at day 22. As soon as the glucose concentration became zero, the cells switched again from lactate production to consumption. The lactate concentration became zero at day 26 and from day 26 on both glucose and lactate were below the detection limit. Both at the start of the culture as well as during the period from day 16 to 21 rapid growth

correlates with glucose concentrations higher than zero and the production of lactate. As soon as the glucose concentration dropped to zero lactate was consumed again. Glucose and lactate concentrations of zero correlate with the cell size increase at day 7. However, also at day 21 at the start of the steady state phase their concentration became zero and no cell size increase was observed. This indicates that the fact that glucose and lactate are limiting is at least not the sole reason for the cell size increase.

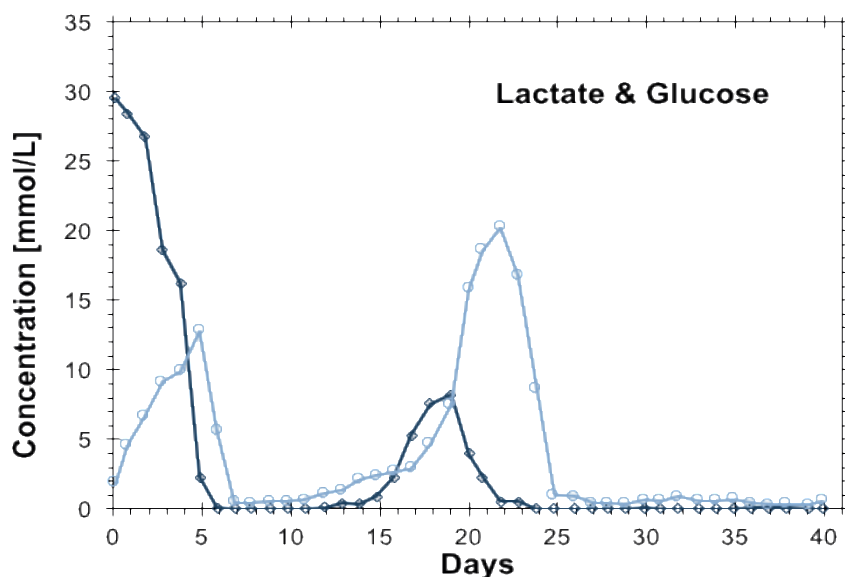


Figure 3 Glucose (open diamonds) and lactate (open circle) concentration as a function of time.

5.3.4 Other Organic acids

Next to lactic acid also citric acid, fumaric acid, formic acid, acetic acid, isovaleric acid and butyric acid were produced. Formic acid reached a peak at day 10 of 6 mM and then decreased to a more or less steady state value between 2 and 3 mM at day 15. Fumaric acid, butyric, 2-hydroxy butyric acid and isovaleric acid had a comparable concentration pattern with an increase until day 10 followed by a decrease and an increase to a more or less constant value. Concentrations were a bit lower than for formic acid being about 0.5 mM for fumaric acid, 0.35 for butyric, 2-hydroxy butyric acids and 1.2 mM for isovaleric acid. Acetic acid concentration varied between 0 and 0.7 mM, but seemed to level of around 0.3 mM. Citric acid showed a constant

concentration between 0.8-1.0 mM. Succinic acid and pyruvic acid were both consumed, where succinic acid became constant at 5 mM around day 15 and the pyruvic acid concentration became 0 mM after day 7. Moreover, the concentrations of all acids were constant from day 26 till the end of the culture, which confirms that a steady state was reached.

All amino acid concentrations showed a very comparable pattern with a minimum at day 7 and a peak at day 18 followed by a decrease to the steady state value at day 24. After day 24 the concentrations remained more or less constant, demonstrating again that indeed a steady state was reached. Glycine is the only amino acid with a different pattern with a peak at day 10 after which the concentration dropped to a more or less constant value from day 18 on. The concentration of most amino acids stayed above the detection limit. Only tyrosine (day 7) arginine (day 7) and cysteine (day 7 and 27, 35 and 37) became zero. However, amino acid concentrations are not analysed every day and it cannot be excluded that one of the other amino acids became zero at one of the other days. For example, glutamine and asparagine concentrations were quite low at day 7, and may have been zero at day 8 or 9. The cell size increase is possibly related to the limitation of tyrosine and arginine at day 7. Note that during the steady state (from day 24) these amino acids are not limiting. Other candidates triggering the cell size increase may be cystine, glutamine and asparagine. However, the concentrations of glutamine and cystine were also quite low during the steady state phase where no size increase occurred, which seems to exclude these amino acids. The complete amino acids analysis can be found in supplementary file-A.

5.3.5 Transcriptomic analysis

To study whether transcriptome analysis can further support characterisation of the culture, samples were analysed for almost all days (except day 0-3, 12, 35, and 38-40). In the PCA of all data points, day 16 and 30 were for the first 4 PCs always in a very different position than the surrounding days as shown in appendix figure A.1. In addition, PC2 was based almost entirely on these two points. Based on this it was assumed that something went wrong with the transcriptome analysis of these points and they were removed from the PCA analysis. Figure 4 shows the result of principal component analysis for PC1 to PC3, which together represent 45% of the variation.

PC1 separates days 9 to 18 from the rest with day 8 and 19 being in the transition zone. This correlates very well with the increase in cell size and the stop of cell proliferation. PC2 separates day 4,5,6, and 13-22 from the other points, which correlates very well with the presence of glucose. Since the lactate concentration correlates with the glucose concentration it of course also correlates with the lactate concentration. Finally, there is also some correlation with the pattern seen for the concentration of most amino acids. PC3 clearly separates culture time with days 4-11 separated from the later days with the exception of day 19 and 36. This correlates with the proliferation rate, which is higher than 0.5 day^{-1} until day 7 and stays below this value after that. It also correlates very well with the concentrations of some organic acids. Especially fumaric acid, which increases to a constant high value around day 11 and succinic acid which declines to a constant low value around day 11. Butyric acid and isovaleric acid also increase from day 0 to a high value, however they reach high values earlier on day 7. For both plots (PC1, PC2 and PC3) the samples from day 26 to the end of culture always closely group together, with the exception of day 36 (PC3) and 39 (PC2) supporting that a steady state was present during these days. In summary, transcriptome data clearly identify the different phases in the culture. Although some parameters become constant 1-2 days before day 26, the transcriptome data clearly show that the steady state period ranges from day 26 to 39, since these points are the only points that group together for all three principal components and do not show a clear order in time.

Furthermore, it is possible to discriminate specific events in the culture like a change in cell size and the presence of glucose. Therefore, the transcriptomic analysis is a powerful tool to indicate the physiological state of the cells during perfusion cultivation. Also, gene expression data can be used to identify gene expression patterns that are considered optimal for harvesting time points with a good product quality.

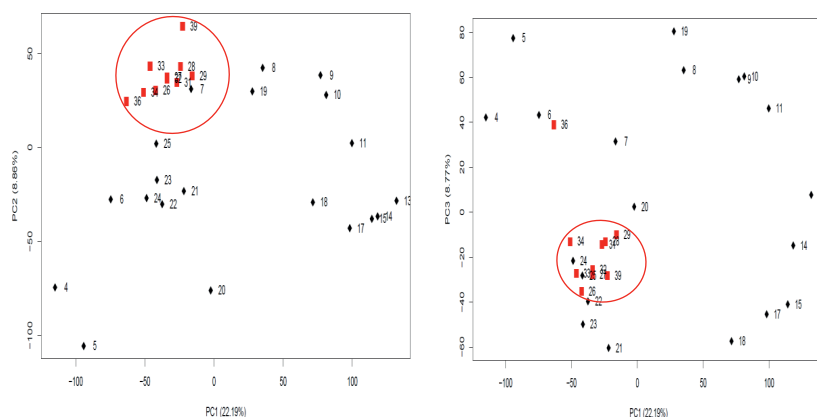


Figure 4: Principal Component Analysis score plots for PC1 vs PC2 and PC1 vs PC3: Day 4-25 is the pre steady state (black), day 26-39 is the steady state phase (red).

5.4 Conclusions

In this study, we study the use of transcriptomics to further characterize perfusion cultures. The dynamic period before reaching steady state was quite long, which is probably related to the stop of cell proliferation and accompanying change in cell size during this period. Quantitative transcriptome data were used as a key performance indicator (KPIs). Thus, transcriptome analysis was able to identify the steady state and the transcriptome data correlated well with events occurring in the perfusion cultivation such as the stop of cell proliferation, the increase in cell size, and nutrient concentrations. Furthermore, studying the transcriptome could be an useful approach to improve understanding of the perfusion process and CHO cell biology. Moreover, in combination with other analytical data, transcriptome data have potential to become a KPI for the batch definition and the release of batches.

5.5 References

- [1] K. B. Konstantinov and C. L. Cooney, "White Paper on Continuous Bioprocessing May 20–21 2014 Continuous Manufacturing Symposium," *J. Pharm. Sci.*, vol. 104, no. 3, pp. 813–820, Mar. 2015.
- [2] S. Chatterjee, "FDA Perspective on Continuous Manufacturing," *IFPAC Annual Meeting*, 2012. [Online]. Available: <https://www.fda.gov/downloads/aboutfda/centersoffices/officeofmedicalproductsandtobacco/cder/ucm341197.pdf>.
- [3] M. Pohlscheidt *et al.*, "Optimizing capacity utilization by large scale 3000 L perfusion in seed train bioreactors.," *Biotechnol. Prog.*, vol. 29, no. 1, pp. 222–9, 2013.
- [4] D. Voisard, F. Meuwly, P.-A. Ruffieux, G. Baer, and A. Kadouri, "Potential of cell retention techniques for large-scale high-density perfusion culture of suspended mammalian cells," *Biotechnol. Bioeng.*, vol. 82, no. 7, pp. 751–765, Jun. 2003.
- [5] C. Zheng *et al.*, "Improved process robustness, product quality and biological efficacy of an anti-CD52 monoclonal antibody upon pH shift in Chinese hamster ovary cell perfusion culture," *Process Biochem.*, vol. 65, no. October 2017, pp. 123–129, Feb. 2018.
- [6] Y. Sun, L. Zhao, Z. Ye, L. Fan, X. Liu, and W. Tan, "Development of a fed-batch cultivation for antibody-producing cells based on combined feeding strategy of glucose and galactose," *Biochem. Eng. J.*, vol. 81, pp. 126–135, Dec. 2013.
- [7] M.-F. Clincke *et al.*, "Very high density of Chinese hamster ovary cells in perfusion by alternating tangential flow or tangential flow filtration in WAVE bioreactor™-part II: Applications for antibody production and cryopreservation," *Biotechnol. Prog.*, vol. 29, no. 3, pp. 768–777, May 2013.
- [8] M. Hilliard *et al.*, "Development of a Glycan Database for Waters ACQUITY UPLC Systems," *Waters Appl. Note 720004202en*, 2012.
- [9] A. A. Alsayyari *et al.*, "Transcriptome analysis for the scale-down of a CHO cell fed-batch process," *J. Biotechnol.*, vol. 279, no. May, pp. 61–72, Aug. 2018.
- [10] K. Lin *et al.*, "MADMAX - Management and analysis database for multiple ~omics experiments.," *J. Integr. Bioinform.*, vol. 8, no. 2, p. 160, Jul. 2011.

- [11] R. A. Irizarry, "Exploration, normalization, and summaries of high density oligonucleotide array probe level data," *Biostatistics*, vol. 4, no. 2, pp. 249–264, Apr. 2003.
- [12] M. Dai, "Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data," *Nucleic Acids Res.*, vol. 33, no. 20, pp. e175–e175, Nov. 2005.
- [13] D. Martens, "The relevance of cell size in a CHO fed batch process: Metabolic and transcriptomic characterization," in *Cell Culture Engineering XVI, Engineering Conferences International*, 2018, p. 198.
- [14] M. C. F. Dalm, S. M. R. Cuijten, W. M. J. Van Grunsven, J. Tramper, and D. E. Martens, "Effect of Feed and Bleed Rate on Hybridoma Cells in an Acoustic Perfusion Bioreactor : Part I . Cell Density , Viability , and Cell-Cycle Distribution," 2004.
- [15] D. J. Karst *et al.*, "Intracellular CHO Cell Metabolite Profiling Reveals Steady-State Dependent Metabolic Fingerprints in Perfusion Culture," *Biotechnol. Prog.*, vol. 33, no. 4, pp. 879–890, 2017.
- [16] X. Pan, C. Dalm, R. H. Wijffels, and D. E. Martens, "Metabolic characterization of a CHO cell size increase phase in fed-batch cultures," *Appl. Microbiol. Biotechnol.*, vol. 101, no. 22, pp. 8101–8113, 2017.
- [17] P. M. J. Jedrzejewski, I. J. del Val, K. M. Polizzi, and C. Kontoravdi, "Applying quality by design to glycoprotein therapeutics: experimental and computational efforts of process control," *Pharm. Bioprocess.*, vol. 1, no. 1, pp. 51–69, Apr. 2013.
- [18] J. K. Ryu, H. S. Kim, and D. H. Nam, "Current status and perspectives of biopharmaceutical drugs," *Biotechnol. Bioprocess Eng.*, vol. 17, no. 5, pp. 900–911, Oct. 2012.
- [19] H. Schellekens, "Biosimilar therapeutics--what do we need to consider?," *NDT Plus*, vol. 2, no. Supplement 1, pp. i27–i36, Jan. 2009.
- [20] R. Jefferis, "Recombinant antibody therapeutics: the impact of glycosylation on mechanisms of action," *Trends Pharmacol. Sci.*, vol. 30, no. 7, pp. 356–362, Jul. 2009.
- [21] N. Alt *et al.*, "Determination of critical quality attributes for monoclonal antibodies using quality by design principles," *Biologicals*, vol. 44, no. 5, pp.

291–305, Sep. 2016.

- [22] P. Hossler, S. F. Khattak, and Z. J. Li, “Optimal and consistent protein glycosylation in mammalian cell culture,” *Glycobiology*, vol. 19, no. 9, pp. 936–949, Sep. 2009.
- [23] C. T. Goudar, N. Titchener-Hooker, and K. Konstantinov, “Integrated continuous biomanufacturing: A new paradigm for biopharmaceutical production.,” *J. Biotechnol.*, vol. 213, pp. 1–2, Nov. 2015.
- [24] P. Zhang *et al.*, “Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs,” *Drug Discov. Today*, vol. 21, no. 5, pp. 740–765, May 2016.
- [25] M. Gawlitzek, M. Estacio, T. Fürch, and R. Kiss, “Identification of cell culture conditions to control N -glycosylation site-occupancy of recombinant glycoproteins expressed in CHO cells,” *Biotechnol. Bioeng.*, vol. 103, no. 6, pp. 1164–1175, Aug. 2009.

Appendix

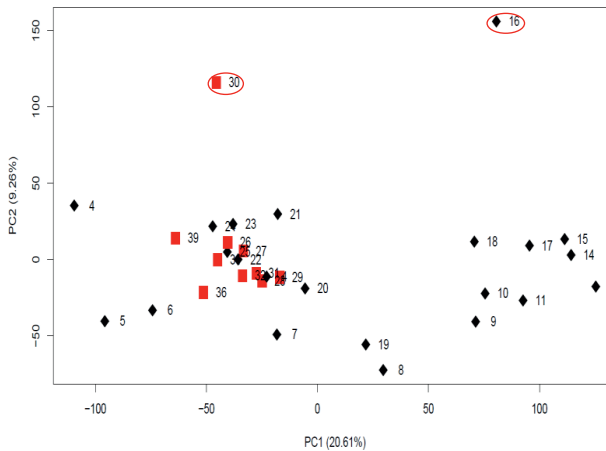


Figure A.1: Principle Component Score plot showing PC1 vs PC2 including the two outliers on day 30 and day 16, day 4-25 represent the pre steady state (black), and day 26-39 represent the steady state phase (red).

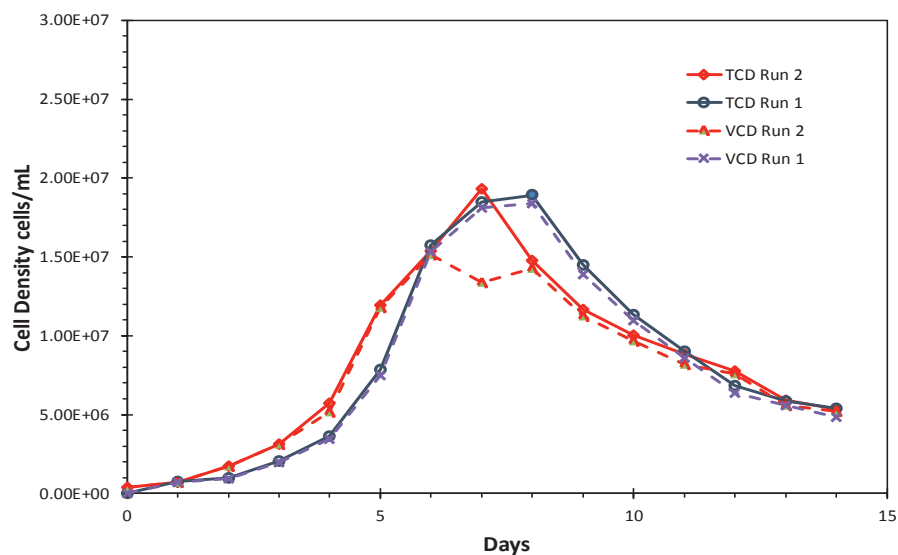


Figure A.2: The total cell density (TCD) and viable cell density (VCD) for the validation run (Run 2) and the experimental run (Run 1) until day 14. Run 2 was terminated on day 14 due to a contamination,. Run 1 continued until day 40.

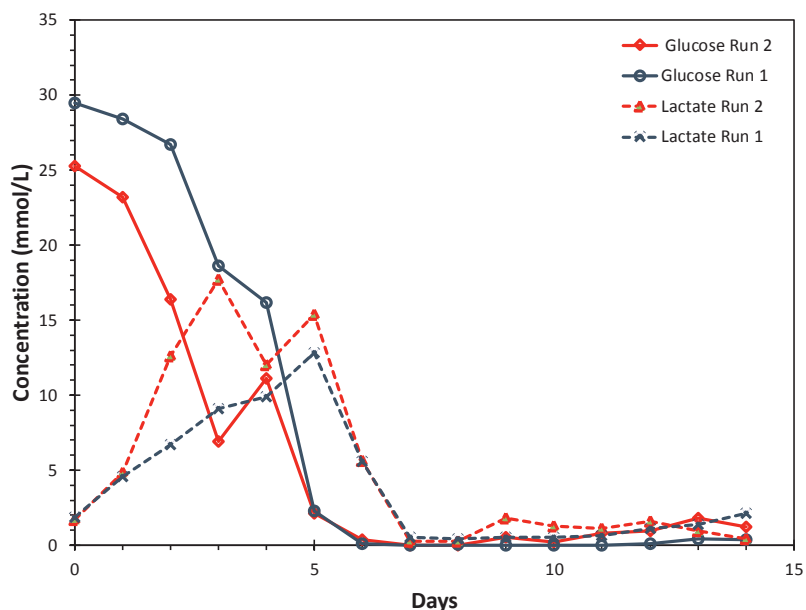


Figure A3: Glucose and Lactate concentration for the validation run (Run 2) and the experimental run (Run 1) until day 14. Run 2 was terminated on day 14 days due to a contamination,. Run 1 continued until day 40.

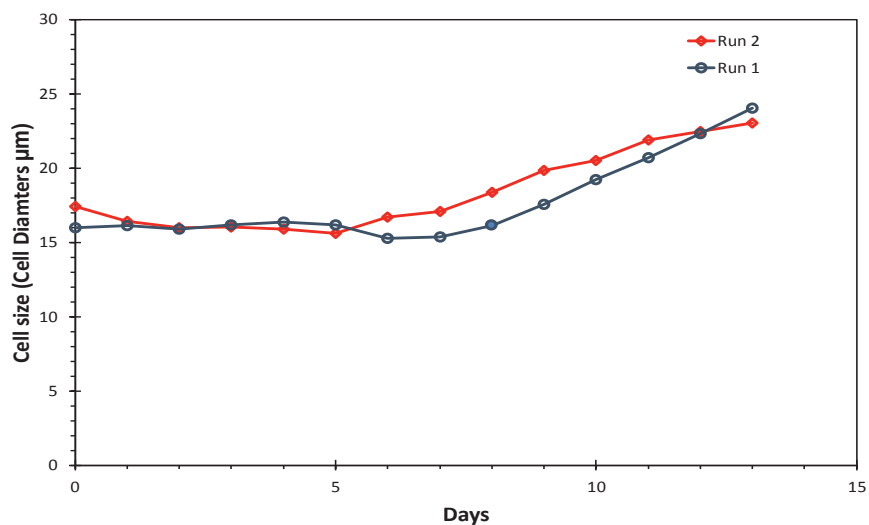


Figure A4: Cell size for the validation run (Run 2) and the experimental run (Run 1) until day 14. Run 2 was terminated on day 14 due to a contamination,. Run 1 continued until day 40.

Chapter 6

General discussion

6.1 Background

In recent years, a better understanding of disease progression and on the design of proteins have led to a major increase in the number of antibody therapeutics [1]. In combination with improvements in manufacturing this has led to new treatments for complex diseases such as cancer and autoimmune diseases. However, the costs per patient of antibody therapeutics are still very high, which at least in part is due to the high development costs of a new product including clinical trials and the relatively high failure rate of new potential medicines. Part of the development costs as well as causes for the high failure rate are related to the development of the manufacturing process. Development of the manufacturing process must be fast and robust to quickly have material for clinical testing and be on the market as quick as possible. Currently, many of the first developed glycoprotein therapeutics products will be coming off patent or are already off patent, which allows for the development of biosimilars. A biosimilar is a biological product that is highly similar to the original product (reference product) in terms of safety, efficacy and quality and is usually made by a different company and process [2]. In general the biosimilar regulatory path requires less clinical testing leading to lower costs and faster development. Thus, the expectation is that the price of biosimilars can be on average about 30% lower than the reference products [3]. Currently, more than 41 biosimilar products have received marketing authorization in the European Union (EU) [4]. Fourteen of these products are biosimilars of which six are biosimilars for reference antibodies, namely infliximab, adalimumab, rituximab, etanercept, trastuzumab, and bevacizumab (see Table 1). Chinese Hamster Ovary (CHO) cells are considered to be the most reliable host for producing proteins with complex structures and for ensuring correct post-translational modification (PMT) [5]. In addition, they are known to have a low susceptibility to certain viral infections and have a low risk of triggering unwanted immune reactions [6]. Consequently, CHO cells are the production platform of 70% of the glycoprotein therapeutics [7], and 19 of the biosimilars currently approved by the EMA use CHO cells as production host (12 Antibodies, 5 Enzymes (Epotein) and 2 Hormones (Foliotropin)). Currently, there is only one approved biosimilar where the host cell differs from that of the reference product [8], which is Flixabi®. This biosimilar is produced in CHO cells, while the reference monoclonal antibody Infliximab (Remicade®) is produced in mouse cells.

The production of biopharmaceuticals in mammalian cells is commonly done in either batch or fed-batch mode. While the perfusion cultivation mode has recently received increasing attention from the biopharmaceutical industry, the fed-batch mode remains the most popular mode for production on a commercial scale [9]. Thus, the antibody biosimilars that are currently approved almost all use the fed-batch mode. Nevertheless, the biopharmaceutical industry shows an increasing interest in continuous manufacturing and perfusion technology, due to the many potential advantages, such as shorter production time, increased flexibility and agility in response to manufacturing needs, more efficient use of equipment, and reducing manufacturing scale and the size of the production facilities. In addition, single-use bioreactors are available that fit effectively with the continuous biomanufacturing practice. Finally, in perfusion systems a steady state can be reached where conditions and cellular physiology are constant, which is beneficial for the product quality and also may facilitate process transfer during the scale-up better than the dynamic fed-batch. All this is expected to lead to a reduction in production costs. However, the introduction of continuous manufacturing technology requires proof of robustness with respect to delivering a high-quality product in a reproducible way. Therefore, implementing quality by design elements in continuous biomanufacturing will provide confidence in this manufacturing technology.

Table 1: List of approved antibody biosimilars products [10]

Medicine Name	Active Substance	Marketing Authorization Holder	Authorization date	Host cells
Humira (reference)	adalimumab	AbbVie Deutschland GmbH & Co. KG	08/09/2003	CHO
Amgevita	adalimumab	Amgen Europe B.V.	22/03/2017	CHO
Cyltezo	adalimumab	Boehringer Ingelheim International GmbH	10/11/2017	CHO
Imraldi	adalimumab	Samsung Bioepis UK Limited (SBUK)	24/08/2017	CHO
Avastin (reference)	bevacizumab	Roche Registration GmbH	12/01/2005	CHO
Mvasi & Solymbic	bevacizumab	Amgen Europe B.V.	12/01/2018	CHO
Enbrel (reference)	etanercept	Pfizer Limited	03/03/2000	CHO
Benepali	etanercept	Samsung Bioepis UK Limited	14/01/2016	CHO
Erelzi	etanercept	Sandoz GmbH	23/06/2017	CHO
Remicade (reference)	infliximab	Janssen Biologics B.V.	13/08/1999	Mouse
Flixabi	infliximab	Samsung Bioepis UK Limited (SBUK)	26/05/2016	CHO
Inflectra	infliximab	Hospira UK Limited	10/09/2013	Mouse
Remsima	infliximab	Celltrion Healthcare Hungary Kft.	27/09/2013	Mouse
MabThera (reference)	rituximab	Roche Registration Ltd	02/06/1998	CHO
Blitzima & Ritemvia	rituximab	Celltrion Healthcare Hungary Kft.	13/07/2017	CHO
Rituzena (previously Tuxella)	rituximab	Celltrion Healthcare Hungary Kft.	13/07/2017	CHO
Rixathon & Riximyo	rituximab	Sandoz GmbH	15/06/2017	CHO
Truxima	rituximab	Celltrion Healthcare Hungary Kft.	17/02/2017	CHO
Herceptin (reference)	trastuzumab	Roche Registration Ltd	28/08/2000	CHO
Ontruzant	trastuzumab	Samsung Bioepis UK Limited (SBUK)	15/11/2017	CHO

6.2 Quality by design (QbD)

Quality by design (QbD) is a holistic manufacturing approach to increase the efficiency of biopharmaceutical manufacturing while maintaining a high standard of quality control. In general, the efficiency of biopharmaceutical production is considered low (2-3 sigma, 25–35% defects) compared to many other well-established industries such as chemical and automobile manufacturing [11]. Therefore, pharmaceutical and biopharmaceutical companies are encouraged to adapt process improvement strategies [12]. In 2011, the EMA and FDA launched a joint pilot program for the parallel assessment of QbD applications [13]. This joint pilot program to implement QbD was a milestone in the development of pharmaceutical and biopharmaceutical products. The QbD approach has also helped to shift the attention of biopharmaceutical manufacturers toward the process, and away from their heavy reliance on post-manufacturing testing and quality control.

Atezolizumab (Tencentriq®), and obinutuzumab (Gazyvaro®/Gazyva®) are currently the only two approved mAbs with both FDA and EMA that used a full QbD approach including definition of the design spaces. Despite the limited number of approved products using a full QbD approach, the development of almost all new biopharmaceutical and biosimilar products implements certain elements of QbD as described in ICH guidelines: Q8, 9, 10, 11 such as risk assessment. The most critical element or principle in the QbD approach is the concept of a well-defined design space. The design space is formed by a value range for all critical process parameters (CPP), including their interaction effects, within which the product quality is good. Any changes within the approved design space would not require regulatory notifications [14]. Defining this design space for the cultivation process remains a challenging task because of the high number of critical process parameters (CPPs), which can affect any one of the well-defined critical quality attributes (CQAs) of the glycoprotein therapeutics. All these parameters must be identified after which the valid range must be established, and proper monitoring and control must be implemented. Finally, continuous improvement throughout the life cycle of the product must be implemented. Usually the design space is identified using a design of experiments (DOE) approach resulting in detailed statistical relations between CPPs and CQAs. These

relationships do, however, not provide cellular mechanistic knowledge although they are a good basis to obtain this knowledge.

The identification of a design space usually also includes traditional routine measurements or key performance indicators (KPIs), like cell counts, viability, nutrient consumption, by-product concentrations, and product concentration profiles, that are important for determining the status of the culture and the cells. They can give more mechanistic insight in the relation between CPPs and CQAs and also form the basis of many critical decisions in cell culture practice, such as determining the moment of feeding, and harvesting in fed-batch cultures. Regulatory bodies require biomanufacturers to use state-of-the-art techniques to ensure consistency and reproducibility. Transcriptomic profiling is such a state of the art technique and a powerful tool that can be used to characterize the physiological state of the cells and to understand the biology behind the relation between CPPs, KPIs and CQAs in the cultivation process. Dynamic transcriptomic profiling has the potential to be a major KPI for cellular performance and can be implemented in process development, technology transfer, i.e. the transfer of the developed process to the manufacturing site or from one to another manufacturing site, and during the actual manufacturing itself. Such knowledge can improve the validity of the design space and reduce the number of experiments that are needed to identify this design space.

6.3 Dynamic transcriptomic profiling

The last decade techniques to analyze the whole transcriptome, like RNAseq and microarrays, have rapidly improved and become cheaper and are now well-established techniques to measure the expression of thousands of genes [15]. Such transcriptomic expression profiling techniques have had a significant impact on the fundamental understanding of physiological responses to cultivation conditions, and have the potential to improve cell culture, in terms of cell line development, and process design [16]. In 2011, the genome sequence of the CHO-K1 cell line was published, opening up promising opportunities for studying cellular physiology [17]. In addition, CHO specific microarrays became available, making it easier to do transcriptome analysis for these cells. Although,

transcriptomic profiling is used intensively in several biopharmaceutical production platforms such as *E.coli* and yeast, the use in CHO cell cultivation is still limited.

6.4 Aim

The aim of this thesis is to evaluate the value of transcriptome analysis for process development for the upstream cultivation process and give an outlook on its possible use in Quality by Design. In the experiments presented in this thesis, a commercial Affymetrix® Chinese Hamster Ovary (CHO) Gene 2.1 ST Array Plate was used for transcriptomic profiling of an antibody producing CHO cell line cultured in batch, fed-batch and perfusion mode. In principle, data obtained on transcriptomic profiling can be used in two ways: (1) to carry out process fingerprinting using an unsupervised methodology such as principal component analysis (PCA) and thereby identifying specific expression patterns and/or combinations of specific marker genes; and (2) to gain cellular mechanistic understanding by linking CPPs to CQAs and KPIs through gene expression. In the following section, five typical cases of these approaches will be presented and critically discussed.

6.5 Process fingerprinting using multivariate data analysis

Case 1. Detection of process deviations

Gene expression data can be coupled with multivariate data analysis (MDVA) tools such as principal component analysis (PCA) to fingerprint the cultivation process on a cellular level. In **case 1**, six fed-batch bioreactors were run in parallel under the same process conditions (all the bioreactors were operated according to material and methods described in Chapter 3), with samples taken at three time points, namely days 3, 5, and 7. Traditional KPI measurements showed these runs to be similar in terms of cellular performance, nutrient and product concentrations, and glycosylation patterns (see supplementary case 1). However, at the gene expression level, the measurements showed that two of the bioreactors deviated from the others, as illustrated in the PCA plot in Figure 1. Bioreactor 1 started to deviate on day 5 and continued to deviate, while bioreactor 3 deviated on day 7 only. Upon closer inspection it was found that these deviations were caused by an

unintentional increase in gas flow due to a system fault in the DO control cascade. This demonstrates that deviations in the process can be easily detected using such a transcriptomic fingerprinting approach. If it is possible to perform the analysis within a few hours it could be even used for correcting the deviation or abort the cultivation in time. The next step would be to link this to the critical quality attributes of the product like the glycosylation pattern.

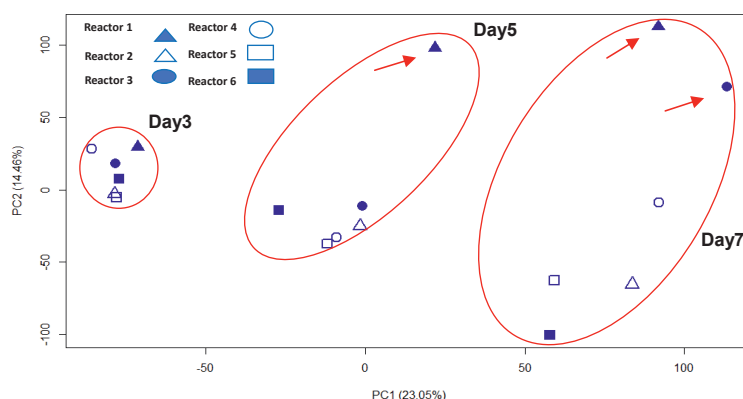


Figure 1. PCA score plot of gene expression for 6 fed-batch bioreactor runs, for three time points (days 3, 5, and 7) as indicated by the red circles in the graph. Arrows indicate the deviating cultures.

Case 2. Down-scale validation

In this **case 2**, examples are presented where transcriptomic analysis was used as a fingerprinting tool to compare two different reactor scale-down models. Validating the scale-down model is considered one of the critical concerns of implementing design space, since differences in conditions between the small scale and production scale reactor may result in unexpected changes in the design space upon scale-up of the process and failure of the product. Figure 2. shows the transcriptome data of two different scale-down comparisons; (1) 250mL shake flasks (30mL working volume) compared to

1L bioreactors (Chapter 3). (2) The ambr 15[®] system (15ml working volume) compared to 10 L bioreactors (Chapter 4). Both comparisons were done for a fed-batch process. As shown in Figure 2.A&B (data described in Chapter 3 and Chapter 4), the variation in gene expression is larger between the shake flasks and the 1L bioreactors as compared to the variation between the ambr and the 10 L bioreactors. This is to be expected as the ambr system has control of pH and DO and agitation like in the 1 and 10 liter bioreactors, whereas shake flasks have no control of pH and DO and a different way of agitation. Although the medium system is different for both comparisons, the PCA indicates that the ambr is a better scale down model than the shake flask and that the higher level of control in the ambr is relevant. This case also indicates the high sensitivity of the transcriptome analysis. At different scales also the preculture for that scale will be different as it is prepared in a different bioreactor system. The preculture is considered a critical step in batch-to-batch reproducibility. In the second example of this case 2 we assess the effect of differences in the preculture on the scale comparison. The precultures for two cultivations that were run in the comparison of the ambr15[®] and 10 L Sartorius stainless steel bioreactors (Chapter 4) came from two different vials from the same working cell bank (full data provided in Chapter 4). The cells for the ambr system were precultured in a shake flask, while the cells for the 10 liter system were precultured in a wave bag. Comparative gene expression analysis of the precultures and of cells taken from the ambr15[®] and the 10 L Sartorius bioreactors on day 3 are shown in the PCA plot in Figure 2.B. PC1 represents time variation with 32.8% explained variance. With respect to PC1 the data points of the preculture samples overlapped with those of day 3 in the ambr and 10L system, since the preculture was used for inoculation also at day 3. PC3 (6.8% explained variation) represents variation between both precultures. As can be seen the difference that exists between the precultures is not present anymore on day 3 in both reactors. The Venn diagram in Figure 2.C compares the number of genes that were significantly up- or down-regulated in the precultures and on day 3 of the ambr and 10L bioreactors. Also here it can be seen that the difference between the precultures (about 1100 genes) has disappeared on day 3 in the reactor systems (150 genes). Moreover the genes that are different between both reactors are different genes than those that are different between

the precultures. Thus on day 3, the effects of the precultures on gene expression levels have disappeared and the preculture differences do not affect the scale comparison.

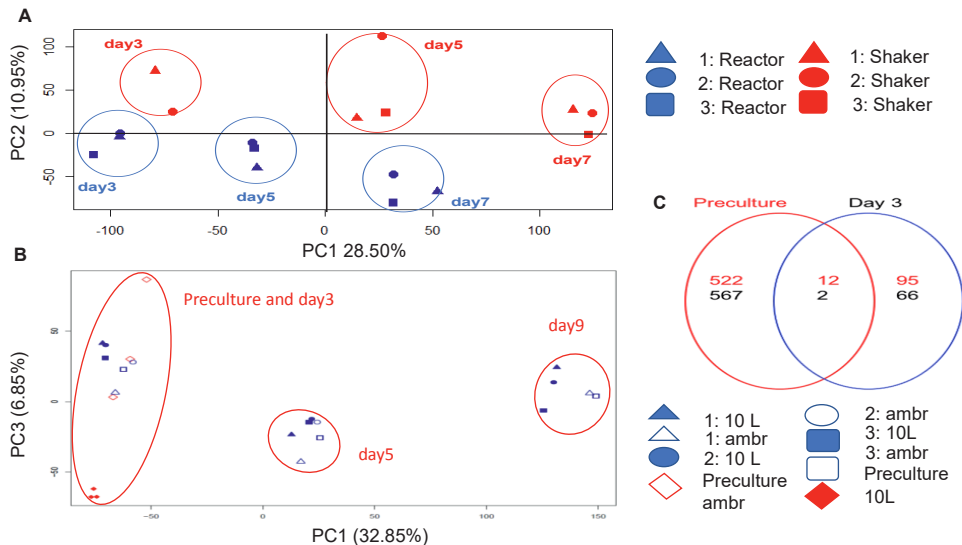


Figure 2. (A) Principle component analysis (PCA) score plot of gene expression data from shake flasks (red), and bioreactors (blue) for three time points (days 3, 5, and 7) , (B) PCA score plot of gene expression data for three time points (days 3, 5, and 9) and for the preculture. Open shapes indicate measurements from the ambr bioreactor; filled shapes indicate those from the 10 L Sartorius bioreactor; red shapes indicate preculture measurements. (C) Venn diagram comparing gene expression data form the 10 L reactors to the ambr (full data are provided in Chapter 4). Number of up-regulated genes indicated in red, down-regulated in black. (Figure 2.A is adapted from figure 3 in chapter 3 and figure 2.B is adapted from figure 6 in chapter 4).

Case 3. Assessing process modifications

In case 3 we use transcriptome fingerprinting to assess the effect of the feeding strategy in a fed-batch process. Commercial fed-batch production processes commonly use bolus feeding. However, industrial biomanufacturing practices are increasingly employing a continuous feeding strategy in order to limit fluctuations in concentrations of nutrients, waste metabolites and pH and thereby maintain a more constant environment for the cell

and thus a more constant product quality. Reliable and state-of-the-art pumps are available, which provide accurate feeding control and can be adjusted to respond to cellular needs. However, experiments during early process development are performed at ml scale at which continuous feeding is often technically not possible and thus replaced by bolus feeding. Therefore, designing continuous feeding at large scale requires an assessment of how and to what extent the applied feeding strategy affects the cell physiology and product quality. Case 3 presents a comparison between the bolus and continuous feeding strategies using a 1 L bioreactor in fed-batch cultivation mode. The cellular performance and the glycosylation patterns showed no differences between the two feeding strategies (further details provided in Appendix figure A.1 and figure A.2), an outcome that suggests the feeding strategy has no impact on quality. The PC1 (27.9% explained variation) of the PCA analysis shown in Figure 3 again represents the variation over time on days 3, 5, and 7. The PC2 of the PCA score plot of this experiment represents the feeding strategy related variation indicating an increasing separation over time in terms of gene expression levels between these two feeding strategies on days 5 and 7 (8.4% explained variation). Detecting such an effect of feeding on gene expression illustrates again the high sensitivity of transcriptomic profiling.

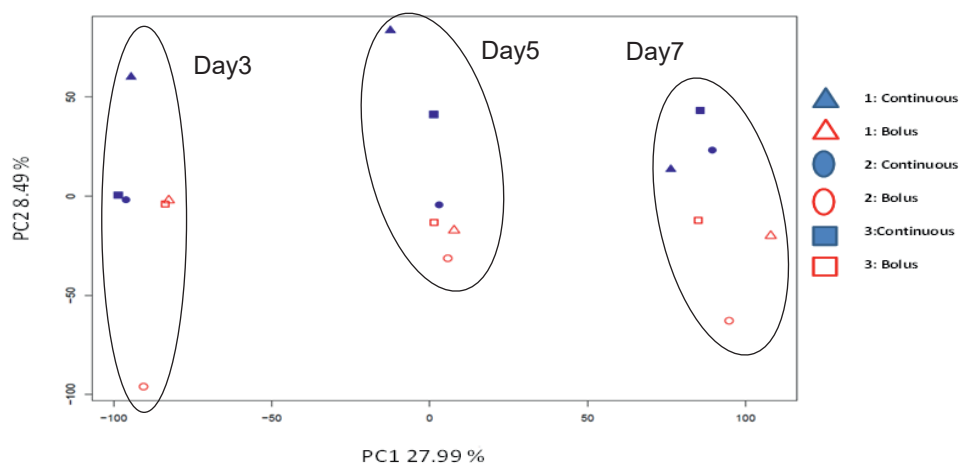


Figure 3. PCA score plot of gene expression data taken for three time points (days 3, 5, and 7) of a fed-batch culture. Solid blue shapes indicate continuous feeding, and open red shapes indicate bolus feeding (data are provided in appendix).

Case 4. Effect of different media

The effect of the choice of medium on cell physiology is another parameter that we can assess using transcriptomic profiling. For this case we used data collected using a single clone cultivated in two different media for different reactors and different process modes. The PCA score plot is shown in Figure 4 (data generated from Chapters 3, 4, and 5, and from non-published experiments). PC1 (21.4 %) again represents time variation and PC2 (13.0%) mainly represents the variation due to the use of the two different media. The difference in medium has, after process time, by far the biggest impact on gene expression levels being more than any of the other process variations named above. Moreover, the differences between both media increased with time.

Case 5: Process characterization

In **case 5**, we use the gene expression data to characterize the different states of a perfusion cultivation (data described in Chapter 5). The PCA in Figure 5 shows a clear separation between different states of the culture. Furthermore PC3 versus PC1 shows a clear spiral like development towards steady state, with the steady state points, as defined from the other measured parameters, grouping together and separate from the other points. Also in the PC2 vs PC1 plot the steady state points group together. Another example of the use of dynamic transcriptomic profiling for process characterization is shown In Figure 6 where we do a direct comparison between two different production modes fed-batch and perfusion. The PCA shows a clear separation between the two processes. Furthermore during the dynamic period of the perfusion system the variation in gene expression is comparable to the fed-batch. However, it is immediately clear that as expected the steady state of the perfusion shows much less variation than the fed-batch.

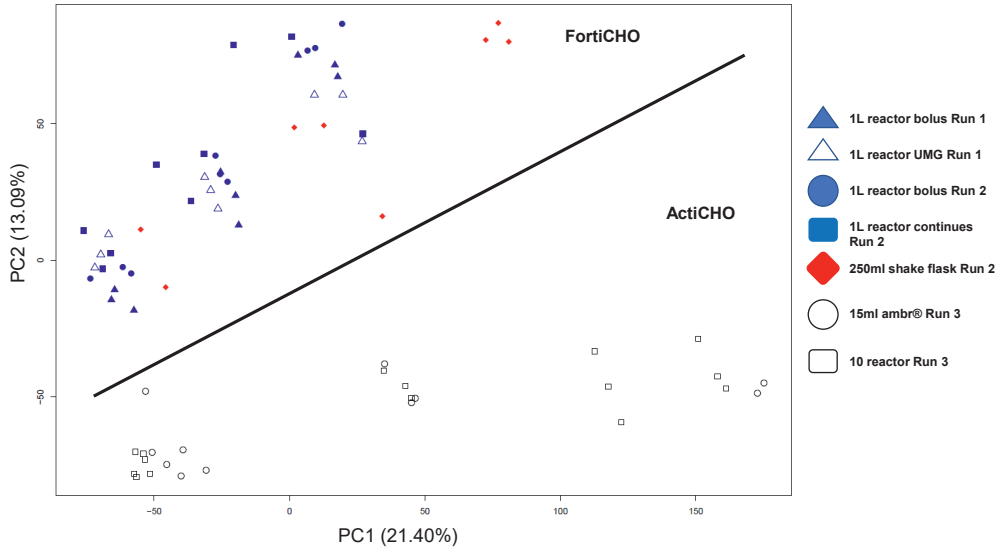


Figure 4. PCA score plot of gene expression data obtained from cultivations with a variety of process conditions in two different commercial well-defined media. Seven different cultivation processes with biological triplicates (as described in Table 2). All the cultivation used the same clone.

Table 2

Groups	Cultivation system	medium	Symbol
Run 1	1 L bioreactors (bolus feeding)	FortiCHO	Filled blue triangle
Run 1	1L bioreactor (additive feed Uridine, Manganese, Galactose (UMG)	FortiCHO	Open blue triangle
Run 2	250 ml shake flask (30ml working volume)	FortiCHO	Filled red diamond
Run 2	1 L bioreactor (bolus feeding)	FortiCHO	Filled blue circle
Run 2	1 L bioreactor (continuous feeding)	FortiCHO	Filled blue square
Run 3	Ambr15® microscale reactor	ActiCHO	Open black circle
Run 3	10 L bioreactor	ActiCHO	Open black square

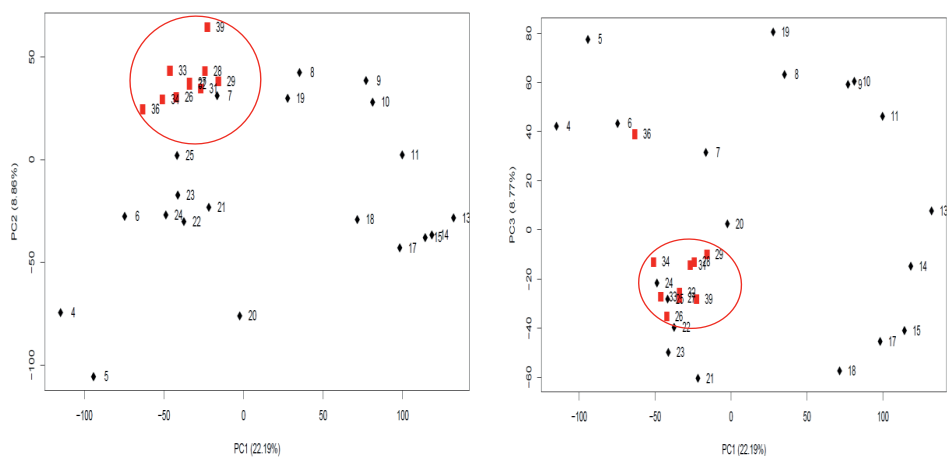


Figure 5. PCA score plot for a perfusion culture showing PC1 (22.19%), PC2 (8.86%), and PC3 (8.77%). Day 4-23 dynamic phase (black), Day 24-39 steady state phase (red). (This figure is adapted from figure 4 in chapter 5).

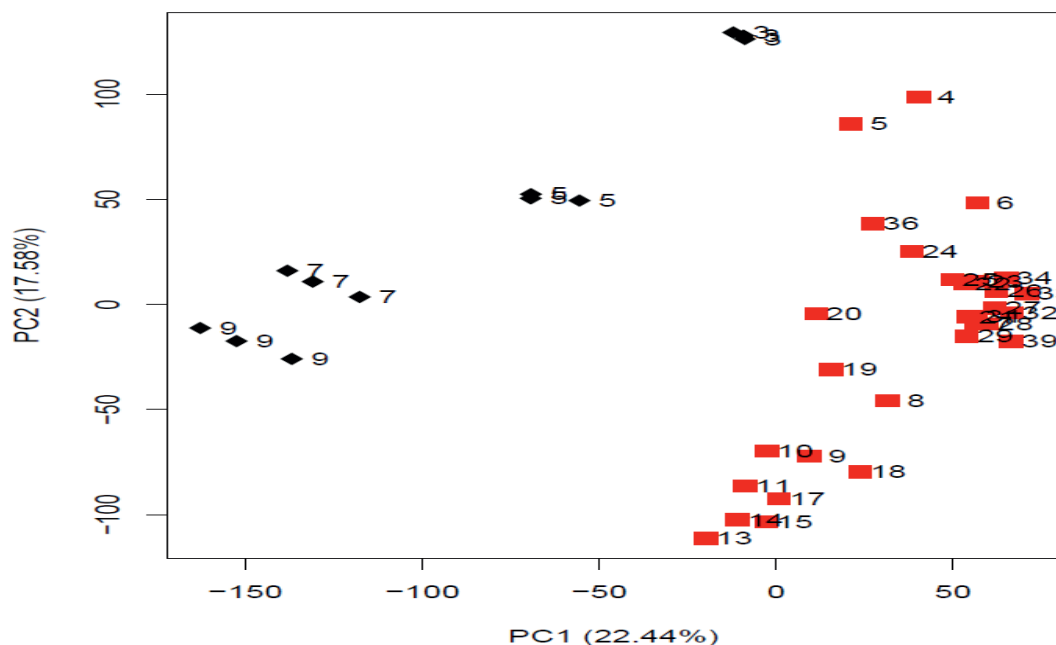


Figure 6 PCA score plot for gene expression data form a fed-batch and perfusion cultivation for PC1 (22.44%), and PC2 (17.56%), for the 1.5L perfusion cultivation (4-39 days except 12,16,30,35) (Red square), and 10L fed-batch cultivation (days:3, 5, 7, and 9)(Black diamond).

6.6 Using gene expression to understand cellular mechanisms

Understanding the interaction between CQAs and CPPs forms an essential basis for the design space and control strategy. The design space is mainly formed by the process parameter ranges within which the process output is certainly good. The CQAs are influenced by many factors and complex interactions. Therefore, implementing QbD elements in cultivation steps started with building a knowledge management (KM) system with more emphasis on understanding the dynamics of the biological processes in the cell during the cultivation process using omics tools. One way of addressing this is to use transcriptomic profiling to identify the genes and pathways that are significantly up- or down-regulated and link the phenotype with process conditions.

As an example of this, in **chapters 2 and 3**, a comparison between shake flasks and bioreactors showed that the number of genes that are up- or down-regulated is higher in shake flasks than in bioreactors, indicating that in shake flasks more pathways are being affected than in bioreactors. The most frequently up- or down-regulated pathways are related to stress. Furthermore also a pathway related to oxygen limitation was regulated (see Chapters 2 and 3). For the batch the development of the shake-flasks and bioreactor cultivations was similar over time and differences were due to system differences. The fed-batch shake flasks had, however, a different development of the cultivation over time being that they are ahead of the fed-batch bioreactor. Therefore, some of the differences in pathways were due to time transitions and not directly to the system itself. It can be that the difference between the systems itself is responsible for the time-delay and thus indirectly for the difference in gene expression related to time. However, since the difference in time-development is not seen for the batch, it must be a combination of fed batch and system differences. Appendix-table A.1, provides an overview of the top up- and down-regulated pathways in the shake flasks and bioreactors (also provided in Chapter 3). The stress related genes are probably related to the fact that the cells in the shake flasks grew faster and were already in the death phase while the cells in the bioreactor were still in the stationary phase. In the top of the list of up-regulated pathways is the oxidative phosphorylation pathway. The oxidative phosphorylation Pathview [18] analysis (see Figure 7) shows that the NADH dehydrogenase gene family and the V type ATPase family are only up-regulated in the shake flasks. This may be an indication of oxygen limitation, which is linked to the process of cellular respiration in the mitochondria and is likely due to levels of dissolved oxygen (DO) in the shake flask dropping to very low levels (unmeasured) towards the end of the culture.

In biomanufacturing, technology transfer, as for example the transfer of the process to manufacturing, and optimization can lead to unacceptably high variations in glycosylation patterns. The glycosylation composition is known to be highly sensitive to changes in the cultivation process [19]. Several studies have been done to understand the interaction between glycosylation and the cultivation process [20]–[22]. Many empirical approaches have been used to control the composition level of glycosylation, such as using additive feed with different concentrations of uracil,

acceptable purely based on a fingerprint. The first problem can be solved by including additional measurement techniques like for example proteomics or metabolomics. The second problem can be solved by building up a knowledge management system (KM) so that in the end based on historical data one can give meaning to the observed variance. Moreover, the expression fingerprint possibly can be linked to product quality attributes and be used as an additional process key performance indicator to define the design space. In addition, to this also specific marker genes or pathways could be correlated to product quality. For example Van der Waterbeemd et al. [24] defined a quality score based on gene expression to determine the harvest point. This can be used in process development and to check whether a production process performed well and did not deviate from what was expected. Possibly, if the analysis of gene expression is fast enough, it could also be used to timely detect deviations in the process and act on this. Whereas for fast growing yeasts and bacteria this may be difficult it may be possible for the slower growing CHO cells. It could in principle already be applied to timely detect deviations from steady state in perfusion cultures. (Chapter 5).

The above approach is based on statistical relations between process parameters gene expression and product quality and not on mechanistic understanding of cell physiology. A mechanistic understanding on how process parameters affect PQAs will result in an increased confidence in the relations and may also be able to predict effects of process parameters or combinations of process parameters that were not studied yet. Analysis of pathways and genes that are differentially expressed can help in getting this mechanistic understanding. Furthermore, such mechanistic knowledge may reduce the number of experiments in a design space exploration simply because the effect of some parameters or interaction of parameters on PQAs can be predicted based on biological mechanisms. This information can also be used to further improve the CHO cell as a host for pharmaceutical protein production. For example, for yeast cells Huang et al. [25] use RNAseq and inverse metabolic engineering to show that by tuning metabolism yeast cells are able to efficiently secrete proteins. A specific problem with using transcriptome analysis for biological understanding is the lack of annotation for a number of genes. In addition, even for annotated genes not all functions are known. This is clearly more difficult for CHO cells than for yeasts and

bacteria due to the higher level of complexity. Finally, like with the fingerprinting approach not everything is regulated on the expression level.

.

6.8 Conclusions

Dynamic transcriptome profiling of CHO cells, the primary host for production glycoprotein biosimilars, is a promising step for better scientific understanding of the production process. This will support the implementation of QbD as a holistic approach to increase the efficiency of manufacturing and will result in faster process development. Apart from getting more insight into CHO cell physiology, the transcriptome can also be used as a fingerprint of the CHO cell physiological state. As such it can be used for various purposes like comparing two processes or reactors, detecting deviations in the process and establishing whether a steady state is reached. In the end it may be possible to relate certain fingerprints directly to the product quality. Not only will this lead to better and cheaper processes, but it will also lead to more trust with regulatory bodies that the manufacturing process is well controlled.

6.9 Reference

- [1] A. A. Shukla, L. S. Wolfe, S. S. Mostafa, and C. Norman, "Evolving trends in mAb production processes," *Bioeng. Transl. Med.*, vol. 2, no. 1, pp. 58–69, Mar. 2017.
- [2] European Medicines Agency, "Guideline on similar biological medicinal products," 2014.
- [3] J. P. F. Erwin A. Blackstone, "The Economics of Biosimilars," *Am. Heal. Drug Benefits*, vol. 6, no. 8, pp. 469–478, Nov. 2013.
- [4] European Medicines Agency, "Biosimilars in the EU," 2017. [Online]. Available: http://www.ema.europa.eu/docs/en_GB/document_library/Leaflet/2017/05/WC500226648.pdf. [Accessed: 20-Aug-2001].
- [5] G. Walsh, "Post-translational modifications of protein biopharmaceuticals," *Drug Discov. Today*, vol. 15, no. 17–18, pp. 773–780, Sep. 2010.
- [6] J. Dumont, D. Euwart, B. Mei, S. Estes, and R. Kshirsagar, "Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives," *Crit. Rev. Biotechnol.*, vol. 36, no. 6, pp. 1110–1122, Nov. 2016.
- [7] J. Y. Kim, Y.-G. Kim, and G. M. Lee, "CHO cells in biotechnology for production of recombinant proteins: current state and further potential," *Appl. Microbiol. Biotechnol.*, vol. 93, no. 3, pp. 917–930, Feb. 2012.
- [8] European Medicines Agency, "CHMP assessment report Flixabi," *EMA/CHMP/272283/2016*, 2016. [Online]. Available: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/004020/WC500208358.pdf.
- [9] F. Lu *et al.*, "Automated dynamic fed-batch process and media optimization for high productivity cell culture process development," *Biotechnol. Bioeng.*, vol. 110, no. 1, pp. 191–205, Jan. 2013.
- [10] European Medicines Agency, "European public assessment reports list of biosimilars," 2018. [Online]. Available: http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/landing/epar_search.jsp&mid=WC0b01ac058001d125.
- [11] R. M. Haleem, M. Y. Salem, F. A. Fatahallah, and L. E. Abdelfattah, "Quality in the pharmaceutical industry – A literature review," *Saudi Pharm. J.*, vol. 23, no. 5, pp. 463–469, Oct. 2015.

- [12] O. P. Sharma, V. Gupta, G. S. Rathore, N. K. Saini, and K. Sachdeva, "Six Sigma in Pharmaceutical industry and Regulatory Affairs : A Review," *J. Nat. Conscientia*, vol. 2, no. 1, pp. 273–293, 2011.
- [13] E. M. Agency *et al.*, "Report from the EMA-FDA QbD pilot program," vol. 44, no. April, pp. 3–4, 2017.
- [14] Ich, "Pharmaceutical Development Q8," *ICH Harmon. Tripart. Guidel.*, vol. 8, no. August, pp. 1–28, 2009.
- [15] M. Castro-Melchor, H. Le, and W.-S. Hu, "Transcriptome Data Analysis for Cell Culture Processes," in *Genomics and Systems Biology of Mammalian Cell Culture*, vol. 123, no. July 2015, Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 27–70.
- [16] P. Gupta and K. H. Lee, "Genomics and proteomics in process development: opportunities and challenges," *Trends Biotechnol.*, vol. 25, no. 7, pp. 324–330, Jul. 2007.
- [17] X. Xu *et al.*, "The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line," *Nat. Biotechnol.*, vol. 29, no. 8, pp. 735–741, Aug. 2011.
- [18] W. Luo and C. Brouwer, "Pathview: An R/Bioconductor package for pathway-based data integration and visualization," *Bioinformatics*, vol. 29, no. 14, pp. 1830–1831, 2013.
- [19] S. Sha, C. Agarabi, K. Brorson, D. Lee, and S. Yoon, "N-Glycosylation Design and Control of Therapeutic Monoclonal Antibodies," *Trends Biotechnol.*, vol. 34, no. 10, pp. 835–846, Oct. 2016.
- [20] R. Jefferis, "Glycosylation of Recombinant Antibody Therapeutics," *Biotechnol. Prog.*, vol. 21, no. 1, pp. 11–16, Sep. 2008.
- [21] G. Walsh and R. Jefferis, "Post-translational modifications in the context of therapeutic proteins," *Nat. Biotechnol.*, vol. 24, no. 10, pp. 1241–1252, Oct. 2006.
- [22] P. Hossler, S. F. Khattak, and Z. J. Li, "Optimal and consistent protein glycosylation in mammalian cell culture," *Glycobiology*, vol. 19, no. 9, pp. 936–949, Sep. 2009.
- [23] M. J. Gramer *et al.*, "Modulation of antibody galactosylation through feeding of uridine, manganese chloride, and galactose," *Biotechnol. Bioeng.*, vol. 108, no. 7, pp. 1591–1602, Jul. 2011.

- [24] Van de Waterbeemd B, Streefland M, Pennings J, van der Pol L, Beuvery C, Tramper J, Martens D. "Gene-expression-based quality scores indicate optimal harvest point in *Bordetella pertussis* cultivation for vaccine production". *Biotechno. Bioeng.* 103:900-908, Aug. 2009.
- [25] Huang, M, Bao, J, Hallström, B M., Petranovic, D, Nielsen, J. "Efficient protein production by yeast requires global tuning of metabolism". *Nature communications*, 8:1131-1142, Oct. 2017

Appendix

Table A.1. Comparison of top up- and down-regulated pathways in fed-batch shake flask and bioreactor. (FDR-BH< 0.5, fold change > ±1.4) (chapter 3)

	Common	Day3_5 (A)						Common	Day6_7 (B)						Common	Day3_7 (C)						Uniquely expressed (D)					
		Reactor			Shaker				Reactor			Shaker				Reactor			Shaker			Common		Reactor		Shaker	
		Up	Down	Up	Down	Up	Down		Up	Down	Up	Down	Up	Down		Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down
		Up regulated in shake flask																									
Amoebiasis	95	0	1	0	2	*6	0	1	0	0	2	4	0	*8	2	2	2	*12	*3	*7	3	1	3	*12	3		
Apoptosis	161	0	1	4	1	*8	6	1	0	1	1	*10	5	*8	4	7	12	*17	*8	5	5	7	*14	*17			
Autophagy - animal	150	0	0	1	2	*9	1	2	0	1	1	*9	0	*10	4	1	2	*15	*9	1	4	*11	4	*11			
Axon guidance	188	1	0	1	2	*12	4	1	1	2	2	*8	2	*12	6	*8	1	*19	*10	*12	6	*7	1	*23	11		
Galactose metabolism	33	0	0	0	0	1	0	1	0	0	0	2	0	2	0	0	0	4	1	2	0	0	0	4	1		
Glutathione metabolism	70	0	0	1	1	3	1	1	0	1	0	*8	4	8	7	0	1	9	1	8	7	0	1	9	1		
Lysosome	133	2	1	3	1	*16	0	*10	1	1	2	*27	0	*34	1	5	1	30	1	*36	1	5	2	*29	1		
Oxidative phosphorylation	131	0	0	0	0	2	0	2	0	0	0	*8	1	6	3	1	2	15	3	7	3	1	2	*14	4		
Peroxisome	88	0	0	1	2	*4	0	1	0	1	0	7	0	*10	2	3	2	11	0	*12	3	2	4	*11	0		
Phagosome	212	2	2	1	1	*10	0	*11	0	1	1	*17	6	*22	2	1	6	24	5	*23	3	2	4	*26	7		
	Common	Day3_5 (A)						Common	Day6_7 (B)						Common	Day3_7 (C)						Uniquely expressed (D)					
		Reactor			Shaker				Reactor			Shaker				Reactor			Shaker			Common		Reactor		Shaker	
		Up	Down	Up	Down	Up	Down		Up	Down	Up	Down	Up	Down		Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down
		Down regulated in shake flask																									
Cell cycle	129	0	*5	0	0	2	*32	0	11	0	1	0	*36	2	39	0	1	5	*19	2	*39	0	2	*6	*21		
DNA replication	38	0	*4	0	0	2	*19	1	10	0	1	0	*16	1	23	0	0	1	*10	1	*23	0	0	1	*10		
Fatty acid elongation	30	0	0	0	0	1	3	1	0	0	0	1	0	0	3	1	3	2	0	1	*3	0	*3	3	0		
Histidine metabolism	27	0	0	0	0	2	0	0	0	0	0	1	2	2	0	0	1	2	0	1	2	0	0	1	2		
Homologous recombination	42	0	*2	0	0	2	*13	1	2	0	0	*14	1	11	1	0	1	*16	0	*15	0	0	1	*16	1		
Mismatch repair	26	0	*2	0	0	3	*9	1	1	1	0	0	*9	2	10	0	0	1	*8	2	*10	0	0	1	*8		
mRNA surveillance pathway	99	0	0	0	0	1	*8	1	0	0	0	2	*9	2	15	0	1	*12	2	*15	0	1	1	*14	2		
p53 signaling pathway	72	0	1	0	0	2	*8	0	3	0	1	1	*13	3	11	2	1	*9	*9	3	*11	2	1	3	*11		
fucose and glucuronate interconversion	32	0	0	0	0	1	1	1	0	0	0	2	0	2	1	0	0	1	3	2	1	0	0	1	*3		
Proteasome	47	0	0	0	2	0	0	0	0	0	0	1	0	0	4	0	8	1	1	4	0	0	8	*1	1		
RNA degradation	79	0	*3	0	1	0	*6	2	1	0	0	1	*6	1	16	0	6	*8	1	*16	0	*10	3	*9	1		
RNA transport	185	0	*4	1	1	1	*13	3	6	0	1	1	*17	4	36	0	2	*32	4	*36	0	2	0	0	*35		
Steroid biosynthesis	21	1	*3	0	1	0	0	1	2	0	0	0	1	1	5	0	4	1	0	1	5	0	*4	1	0		

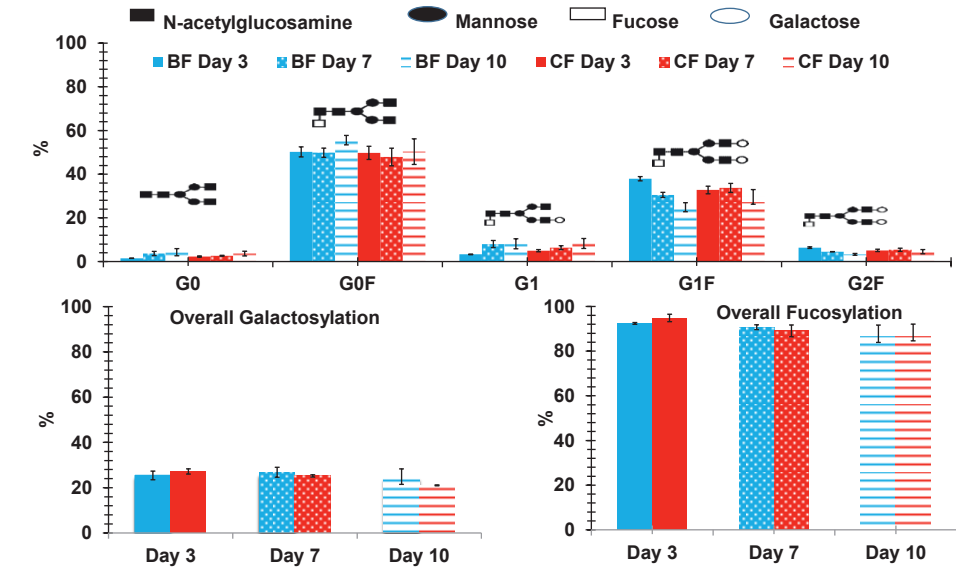


Figure A.1. Average of 3 replicate glycosylation Patterns for the between continuous feeding (CF) and bolus feeding (BF).

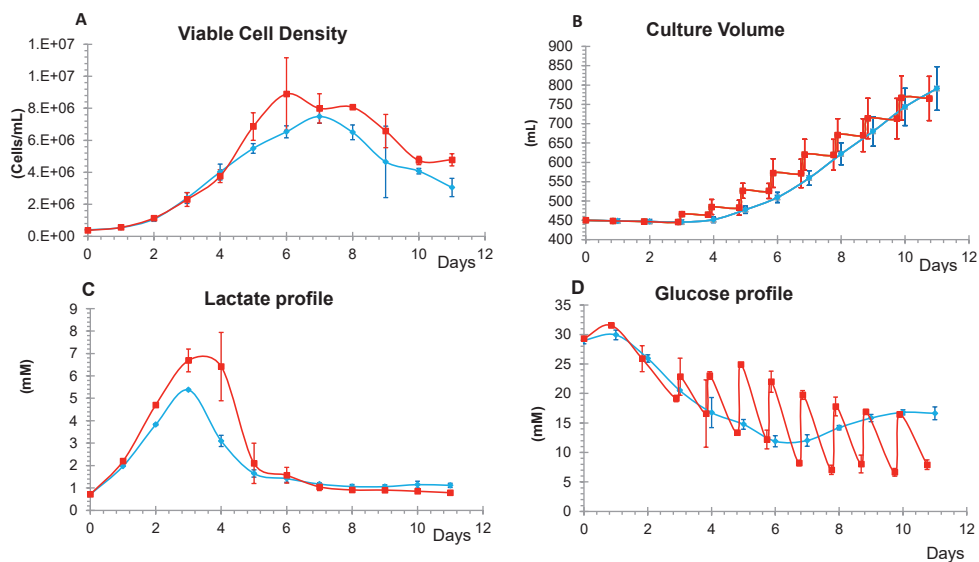


Figure A.2: Comparison of cell performance between continuous feeding (CF) in blue and bolus feeding (BF). A: Viable cell density, B: Culture volume, C: Lactate profile, D: Glucose profile. Data points are the average of 3 replicate shake flask and bioreactor runs. Error bars show the standard deviation.

Summary

Summary

Global healthcare demands for affordable and accessible biological medicines are rapidly increasing as a result of an increasing number of patients. Biosimilars products have a potential to help decrease the price and increase the availability of the biological medicines. Monoclonal antibodies are a special class of biopharmaceutical proteins that are used to treat life-threatening diseases like cancer. Mammalian cells and more specific Chinese Hamster Ovary cells are the main production platform for these monoclonal antibodies. Many of the first products, which were developed in the early 1990s are going off patent opening the possibility to develop biosimilars. Nowadays, for the development of biosimilars and also completely new biopharmaceuticals it is required for pharmaceutical companies to include (aspects of) Quality by Design (QbD). This means that they need to have a good scientific understanding of the relation between critical process parameters (CPPs) and critical quality attributes (CQAs) of the products as well as with key performance indicators (KPIs), like the cell concentration. These relations are usually statistical relations that are not based on detailed knowledge of biological mechanisms, although they are a good basis to acquire such knowledge. Transcriptomics is a state of the art technique to measure the global gene expression of a cell population and can give insight into the biological mechanisms that are behind the relation between CPPs, CQAs and KPIs. Furthermore, the transcriptome is a good representation of the physiological state of the cells and can thus be used for comparison of for example reactors of different scales, different processes, or comparison of a new bioreactor production run to past runs.

The aim of this thesis is to study the value of transcriptome analysis for process development and the quality by design approach for the upstream cultivation process. All experiments in this thesis use the same CHO cell clone producing a monoclonal antibody. All transcriptome measurements are done using commercially available Affymetrix CHO Gene 2.1 ST arrays.

Early process development for biopharmaceuticals is often partly done in small scale systems in a high throughput approach. Transcriptome analysis is used in this thesis

to assess whether the small scale systems are representative for the production scale reactor.

The focus of **Chapter 2** is to have a better understanding of the differences in cell performance between uncontrolled (shake flask) and controlled (bioreactor) systems. In this chapter, we evaluated differences in gene expression profiles between shake flask and bioreactor cultures at three different time points during the exponential and stationary phase of a batch cultivation. Both systems showed a similar large variation in gene expression over time, meaning the differences between two cultivation systems are small. However, the small gene expression difference between the two systems occurred during a short period of cultivation during batch cultivation and could be directly linked to the absence of control of some of the process cultivation parameters such as pH and DO in the shake flask.

For fed-batch processes larger differences are expected to be present between controlled and uncontrolled systems as compared to batch due to the higher cell densities, bolus addition of concentrated feeds and longer process duration. Therefore, we investigate in **Chapter 3**, the differences in gene expression between, shake flasks and 1 L bioreactors for a fed-batch cultivation process. The shake flask cultures grew faster than the bioreactor cultures and went earlier into the stationary and death phase. Using PCA the difference in time development was represented in PC1, while also a time independent difference was observed and represented by PC2. Although transcriptome data can also identify differentially expressed pathways, which may lead to the root cause of the differences, here we were not able to identify this root cause.

In **Chapter 4**, transcriptome analysis was used to evaluate the down scale of a CHO cell fed-batch process from a 10 L bioreactor to an ambr 15® (ambr) system. In this case both systems have control of pH and DO. The results of this comparative transcriptomic study showed that the variation in gene expression was less than 6% based on PCA. Moreover, the gene and pathway analysis did not reveal a direct relation of this difference with scale differences or specific process conditions. To be able to obtain an objective meaning to this 6% difference more transcriptome experiments need to be done. In this Chapter we also show that differences in gene

expression in the preculture quickly disappear in the culture systems, showing that the preculture did not have an effect on the comparison.

In **Chapter 5**, we explore the use of transcriptomic profiling for monitoring long term production campaigns such as occur using a perfusion cultivation mode. The results show that transcriptome data can be linked to specific phases of a perfusion process like a period characterized by a cell size increase and a period with a specific nutrient limitation. More importantly, quantitative transcriptome data visualized using PCA, show a clear separation between the steady state data points and the other data points and can thus be used to identify the steady state.

In **Chapter 6**, transcriptomic profiling as a tool for quality by design in the cultivation process of antibody biosimilars is discussed. Comparative transcriptomic analysis could play a role in validation of the scale down system by demonstrating similar responses of biological processes for both systems. Likewise it can be an useful tool to detect process deviations, evaluate process modifications, and process characterization. Furthermore, in combination with other analyses it can give insight into biological processes that link CCPs to CQAs. Thus, transcriptomic profiling can be part of process performance indicators and combined with other bioanalytical measure give a better understanding and faster development of the design space. When the analysis can be done routinely and sufficient fast it can possibly also be used at line to evaluate and control running production runs

Acknowledgement

There are many people who I would like to thanks for their contributions on this thesis and in my Ph.D life. First, I want expressed my gratitude and appreciation to **René Wijffels** and **Dirk Martens** for the opportunity to be part of Bioprocess engineering group. Thank you for all scientific discussions and mentorship throughout my PhD project. Again **Dirk**, I was lucky to have you as supervisor, words will never express my feeling, Thank you !!

Jos, thank you for your time and supervising me on statistics and data analysis. **Xiao**, we shared many things during our PhD years (lab, bioreactors, UPLC, cell line, etc.) Thank you Xiao, I had a lot of fun and good memories working with you.

Mathieu, thank you for introducing me to BPE and for everything during the first 1.5 year of my PhD project (I learn a lot from you in critical thinking and quality by design and transcriptomic). **Guido**, thank you for your help in R-Bioconductor and writing R-codes, you always find time for me in your busy schedule.

Wendy and **Fred** and **Sebastiaan**, **Sbezana**, and **Miranda**, thank you for your help and support.

I would like to thank all my MSc thesis students that I supervised during Ph.D. thesis, **Jeroen Brouwer**, **Jarmo Hoogendoorn**, **Tjasa Marolt**, **Stefanie Laake**, **Ivan Mijangos**, **Timo Vos**, **Jort Altenburg**, **Jianan Wang**, **Chegnan Huang**, **Deniel Komuczki**, I appreciate your direct or indirect contributions in my thesis.

To all BPE friends and colleagues, Thank you for all enjoyments during lunch break, coffee break, all other activities, and special thank you, **Fabian**, **Youri**, **Enrico**, **Goao**, **Pauline**, **Xiao**, **Ward**, **Agi**, **Carl**, **Edgar**, **Luci**, **Camilo**, **Kylie**, **Rafa**, **Wendy**, **Stephaine**, **Cata**, **Iago**, **Jeroen**, **Elisa**, **Jorijn**, **ILse**, **Jort**, **Pieter**, **Anna**.

Finally, I would like to dedicate this thesis to my **Father Abdullah**, and My **Mother Tarafah**, and to my **wife Laila**, to my daughters (**Alya** and **Bassmah**) and to my Sons (**Abdullah** and **Omar**). Thank you for the support and being part of my life.

Overview of completed training activities



Disciplines specific activities

NBC13 (Ede, the Netherlands, 2014)
NBV national conference (2013, 2015)
Advance course bioprocess design (2014)
Advanced visualization, integration & biological interpretation of -omics data (2015)
Multivariate Data Analysis (Oslo, Norway, 2014)
Bio Process International EU (Amsterdam and Vienna, 2016, 2017)

General courses

Project and time management (2013)
Scientific writing (2014)
The essentials of writing and presenting a scientific paper (2014)
Writing Grant Proposals W15-513 (2015)
Workshop programme Entrepreneurship in and outside science (2016)

Optional activates

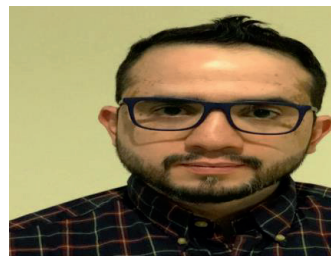
Preparation project proposal (2013)
PhD student day (2013, 2014, 2015, and 2016)
Animal Cell Technology (2015)
QbD for biopharmaceuticals (2014)

This thesis was financially supported by Saudi Food and Drug Authority (SFDA). The research in this thesis was performed at Bioprocess Engineering, Wageningen University, Wageningen the Netherlands.

Printed by
ProefschriftMaken || DigiForce

Curriculum Vitae

Abdulaziz A Alsayyari was born on 2nd of May 1980 in San Diego, California, United States of America. He did his bachelor degree in Clinical Laboratory Science CLS (Medical Technology) at Winston Salem State University, Winston Salem, North Carolina, United States of America. In 2009, he moved to Nottingham,



United Kingdom, to pursue his master degree in Applied Biomolecular Technology (ABT) for biopharmaceutical and biotechnology industrial in the University of Nottingham. In 2010, Abdulaziz worked as biological medicine evaluator within Saudi Food and Drug Authority, Riyadh, Saudi Arabia. In March 2013, started in his position as PhD researcher within Bioprocess Engineering, Wageningen University and Research. His PhD project was involved in Transcriptomic profiling as a tool to characterize CHO cell processes. The results of his research are presented in this thesis.

In July 2017- August 2018, he become project manager for developing new drug evaluation model after promising outcomes. Abdulaziz currently work as Director for biological medicine evaluation within Saudi Food and Drug Authority (SFDA), Drug Sector, Riyadh, Saudi Arabia.

