Peer-reviewed version available at Bioengineering 2018, 5, 101; doi:10.3390/bioengineering5040101

Article

Accelerated Bioprocess Development of Endopolygalacturonase-Production with *Saccharomyces cerevisiae* using Multivariate Prediction in a 48 Mini-Bioreactor Automated Platform

Annina Sawatzki (https://orcid.org/0000-0002-5125-9093) [1], Sebastian Hans (https://orcid.org/0000-0002-5346-869X) [1], Harini Narayanan [2], Benjamin Haby (https://orcid.org/0000-0001-9461-4414) [1], Niels Krausch (https://orcid.org/0000-0003-2325-6001) [1], Michael Sokolov [2,3], Peter Neubauer (https://orcid.org/0000-0002-1214-9713) [1], Florian Glauche (https://orcid.org/0000-0002-7326-6112) [1] and Mariano Nicolas Cruz Bournazou (https://orcid.org/0000-0001-9461-4414) [1,*]

[1] Department of Bioprocess Engineering, Department of Biotechnology, Technische Universität Berlin, Ackerstr. 71-76, ACK24, D-13355 Berlin, Germany; a.sawatzki@campus.tu-berlin.de (A.S.); sebastian.hans@tu-berlin.de (S.H.); benjamin.haby@tu-berlin.de (B.H.); n.krausch@campus.tu-berlin.de (N.K.); peter.neubauer@tu-berlin.de (P. N.); f.glauche@tu-berlin.de (F.G.); mariano.n.cruzbournazou@tu-berlin.de (M. N. C. B.)

[2] ETH Zürich, Rämistrasse 101, CH-8092 Zurich, Switzerland; nharini@chem.ethz.ch (H.N.); michael.sokolov@chem.ethz.ch (M.S.)

[3] DataHow AG, c/o ETH Zürich, HCl, F137, Vladimir-Prelog-Weg 1, CH-8093 Zurich, Switzerland; m.sokolov@datahow.ch (M.S.)

* Corresponding author

Abstract: Mini-bioreactor systems enabling automatized operation of numerous parallel cultivations have been used to accelerate and optimize bioprocess development. As implementation of fed-batch conditions, multiple options of process control and sample analysis are possible, these systems represent valuable screening tools for large-scale production. However, the dynamic behavior of cultivations has not yet been considered regarding data evaluation and decision making during high-throughput screening in mini-bioreactors. In this study, the characterization of *Saccharomyces cerevisiae* AH22 secreting recombinant endopolygalacturonase is performed in 48 parallel fed-batch cultivations regarding 16 experimental conditions. Automated parallel process control, frequent sampling and analysis were implemented. Data-driven multivariate methods were developed to allow for fast, automated decision making as well as online predictive data analysis regarding endopolygalacturonase production. Using dynamic process information, a cultivation with abnormal behavior could be detected by principal component analysis as well as two clusters of similarly behaving cultivations, later classified according to the feeding rate. By decision tree analysis, cultivation conditions leading to an optimal recombinant product formation could be identified automatically. The developed method is easily adaptable and suitable for automatized process development reducing the experimental times and costs.

Keywords: mini-bioreactors; parallelization; automation; digitalization; multivariate analysis; dynamic processes

1. Introduction

Only few of the molecules developed in early-stage biotechnology research are entering industrial production, for instance due to strain failure in the large scale [1]. Hence, a more consistent bioprocess development from the screening to the production phase to accelerate strain selection while maintaining a high explanatory power of the experiments is needed [1–3]. To reduce the risk of potential failures during scale-up to industrial production, small-scale screening systems should mimic large-scale conditions. But since high throughput (HT) systems in the form of micro well plates focus on increase of throughput, a trade-off must be met sacrificing the sophistication of cultivations monitoring and controls and its relevance for industrial scale.

Mini-bioreactor (MBR) systems, which have been described as scalable to benchtop bioreactors [4], are an effort to fill this gap. In comparison to lab-scale bioreactors, MBRs allow for a higher experimental throughput – e.g. the fast screening of large strain libraries [5] or a great number of experimental conditions [6,7] – while still enabling the implementation of large-scale process conditions such as feeding, closed loop controls and techniques for scale-down simulation [8,9]. Additionally, their integration into liquid handling robots allows for execution of multiple manipulations in parallel based on scripted and pre-programmed protocols – e.g. pH control, feeding as well as automated sampling and at-line analysis – by integration of laboratory devices [10] (Haby et al., 2018). Finally, due to the high number of parallel experiments, multiple experimental set-ups can be tested including replicates, which increases the reliability and transferability of the generated data for scale-up purposes. Still, there is a lack of tools to design and operate efficient experiments in such highly parallelized systems. Although some works address this issue with automation of experimental facilities towards smart platforms [11] and sequential designs, by which experimental data is processed by algorithms to design next experiments [12–14], as well as optimal experimental design methods [15], these methods use either static models (e.g. regression models) or require a thorough understanding of the strains to build mechanistic models (e.g. macro-kinetic growth models).

Cultivations in MBRs have been performed in parallel previously: Two aeration concepts were tested for scalability in 24 MBRs in the ambr 15f system (Sartorius Stedim Biotech, Royston, UK) applying constant feed [16]. Biomass was sampled regularly, however sampling for offline analytics occurred seldomly and the analytics themselves were not automated. Screenings in 48 MBRs in the bioREACTOR 48 fermentation system (2mag AG, Munich, Germany) were performed regarding the performance of recombinant *E. coli* strains with product determination only at the end of the fermentation [5] or regarding recombinant *Bacillus subtilis* strains with samplings every 24 h as well as pH measurement every 30 min [17]. Hortsch et al. [18] performed growth media comparison for E. coli cultivations in 48 MBRs with regular biomass determination but limited offline sampling.

However, to the best of the authors' knowledge, the HT platforms described above have not yet reached a level of automation, parallelization and digitalization, that does not rely on manual steps during the cultivation, at-line and offline sample analysis and data evaluation. Additionally, in HT systems so far, decisions are taken based merely on the connection of the experimental design with the final outcome. Methods for automated data evaluation and decision making based on the dynamic behavior of a culture, have not been combined with MBR cultivations.

In this study, an experimental run with 48 MBR fed-batch cultivations is performed, integrating efficient parallel operation routines, control of temperature, pH, and aeration for all MBRs, as well as multivariate methods for analysis of online and at-line data generated with advanced sensor technologies. Regarding experiments in MBR systems, traditional process control is difficult as the process engineer must operate 48 cultivations simultaneously and eventually optimize the process behavior. Decisions on corresponding operating conditions must be made in parallel for all cultivations, while comparative evaluation of the cultures in real-time is difficult due to the considerable amount of manipulated and controlled process parameters as well as the dynamically evolving process variables. Data-driven statistical methods presented here enable fast, automated and parallel decision making and online predictive data analysis for each cultivation. A digital platform facilitates central information storing, accessible to process engineers and process models.

As a case study, *S. cerevisiae* AH22 secreting a pectinase from *Aspergillus niger* [19] serves as a model organism for development of recombinant productions processes. Still, the developed methods can be applied to a wide range of expression systems.

S. cerevisiae has been a production host for recombinant products such as insulin, Hepatitis B vaccine as well as growth hormones since the 1980s [20] and is still used for a wide range of products, e.g. therapeutical compounds [21], the antioxidant resveratrol [22], and also for production of enzymes in the food industry, for instance pectinases.

Besides rapid growth and a well-characterized genome, *S. cerevisiae* offers greater robustness in industrial processes, GRAS (Generally Regarded As Safe)-status, secretion of very few endogenous proteins and direction of recombinant proteins correctly-folded into the culture supernatant, which simplifies purification [19,21,23,24]. Yeasts are capable of posttranslational modifications, even though the glycosylation pattern is high in mannose [20] and genetic engineering of the secretory pathways in *S. cerevisiae* is difficult [25,26]. Factors, such as promoters, leading sequences and expression cassette copy numbers, leading to high expression of one product in one strain, might not be effective for another strain or product. HT systems can contribute to an accelerated strain characterization and selection. Automated treatment for competence and transformation in high-throughput manner as described in [27,28] will help in constructing heterologous yeast strains producing desired target molecules.

To achieve an optimal production of recombinant products, important factors to consider during the process are the maximization of the biomass and the cultivation time so to improve the product yield. Additionally, the optimal growth rate for protein production, the reduction of ethanol production and similar by-products due to overflow conditions should be considered. *S. cerevisiae* produces ethanol not only during oxygen limitation due to the Pasteur effect, but also aerobically when the substrate concentrations exceeds a critical strain-specific value [29]. This effect known as Crabtree effect or overflow metabolism is caused by a maximum in the respiration capacity [30] and has been described as an escape strategy to avoid accumulation of intracellular sugars to toxic levels [31]. Ethanol is resorbed by the cells when the preferred substrate – namely fermentable sugars such as glucose – is exhausted. However, growth on ethanol is slower than growth on glucose, while increasing the oxygen demand [29,32].

Given that during strain development, parameters such as the maximal growth rate or optimal production conditions are not well known, design of experiments (DoE) is applied to statistically determine which experimental input influences the studied system in a way that leads to the most informative experiments and should guide process optimization through experiments sequentially approaching the optimal target conditions. By these methods experiments not resulting in knowledge gain are avoided, thereby minimizing the time and resources needed for process development. In the case of *E. coli*, model-based approaches for optimal experimental design [34] have been used to identify dynamical cultivation parameters applying an online approach for experimental re-design in mini-bioreactor platforms [15,35,36]. However, these methods require a macro-kinetic model, which is not always available for the organism of interest. Closing that gap, data-driven statistical models are combined with knowledge-driven deterministic models building so-called "hybrid models", which can be flexibly defined given the available process knowledge and data [37].

With regards to the challenging goals of industry 4.0 [2], smart digital solutions should not only be used to support the process development procedure but to operate sophisticated robotic facilities such that the information of each experimental run is maximized. For this purpose, four technical aspects are crucial: i) centralized online data storage and handling for real-time monitoring and visualization, ii) mathematical methods for data analysis to support decisions during operation, iii) a full integration of all operated devices, sensors and stakeholders in one accessible and consistently updated digital platform, and iv) an efficient workflow with proper scheduling assistance and resource availability.

2. Materials and Methods

2.1 Strain and culture storage

Cultivations were performed using *S. cerevisiae* AH22 (leu2-1, leu2-112, his4-519, can1, cir$^+$, mating type a), harboring the plasmid pPG6 constructed for the heterologous expression of endopolygalacturonase (EPG) from *A. niger* [19,38]. The recombinant strain pPG6 M27 showing improved pectin hydrolysis was used previously [39]. Stock cultures were stored in 1 mL aliquots with 20 % glycerol at -80°C.

2.2 Media

For all cultivations modified WMVIII minimal medium containing $NH_4H_2PO_4$ 0.25 g L^{-1}, NH_4Cl 5.48 g L^{-1}, $MgCl_2 \cdot 6H_2O$ 0.25 g L^{-1}, $CaCl_2 \cdot 2H_2O$ 0.1 g L^{-1}, KH_2PO_4 2.0 g L^{-1}, $MgSO_4 \cdot 7H_2O$ 0.55 g L^{-1}, myo-inositol 75 mg L^{-1}, EDTA 11.69 mg L^{-1} , $ZnSO_4 \cdot 7H_2O$ 1.75mg L^{-1}, $FeSO_4 \cdot 7H_2O$ 0.5 mg L^{-1}, $CuSO_4 \cdot 5H_2O$ 0.1 mg L^{-1}, $MnCl_2 \cdot 4H_2O$ 0.1 mg L^{-1}, $Na_2MoO_4 \cdot 2H_2O$ 0.1 mg L^{-1}, nicotinic acid 10 mg L^{-1}, pyridoxin-HCl 25 mg L^{-1}, thiamine-HCl 10 mg L^{-1}, biotin 2.5 mg L^{-1}, calcium pantothenate 50 mg L^{-1}, histidine 100 mg L^{-1} and 0.01% Antifoam 204 (Sigma Aldrich, St. Louis, Missouri, USA) was used [39]. The WMVIII medium was complemented with glucose in different concentrations as a carbon source.

2.3 Precultures

The first preculture was inoculated from one cryo vial into 25 mL modified WMVIII medium, which was supplemented with 40 g L^{-1} glucose, 1.5 g L^{-1} sodium glutamate, and buffered with 5% citrate-phosphate buffer (pH 6) in a 100 mL UltraYield™ flask covered with AirOtop™ Enhanced Seal (both Thomson Instrument Co, Oceanside, California, USA). The preculture was incubated for 24 h at 30 °C and 180 rpm in an orbital shaker (Lab-Therm LT-X, Adolf Kühner AG, Basel, Switzerland; 50 mm amplitude).

A second preculture was inoculated from the first preculture to an optical density at 600 nm (OD_{600}) of 0.3 into 100 mL medium in a 500 mL flask and cultivated under the same conditions as the first preculture for 24 h.

2.4 Main culture

The main culture was inoculated from the second preculture to an OD_{600} of 0.3 into 300 mL of the modified WMVIII medium, which was supplemented with 20 g L^{-1}, respectively 30 g L^{-1} glucose. Under sterile conditions 10 mL of the inoculated medium were transferred into 48 mini-bioreactors of the pre-sterilized bioREACTOR 48 fermentation system (2mag AG, Munich, Germany). The set points of the circulation thermostat and the reflux cooler were 30 °C and 4°C respectively. The cultures were aerated with 5 L min^{-1} of pressurized air and the stirrer speed was kept constant at 2400 rpm during the cultivation. Dissolved oxygen tension (DOT) and pH were measured by fluorescence sensors (PreSens Precision Sensing GmbH, Regensburg, Germany). The pH sensors were calibrated to a range of pH 5-8 and the DO sensors were adjusted with a two-point calibration under oxygen-free conditions by introducing nitrogen and under oxygen-saturation. The pH was controlled at 6.0 by titration of 3.5 M NH_3 (one-sided control).

An initial batch phase of around 12 h was followed by a fed-batch phase with small bolus additions of a concentrated glucose stock solution every 5 min via the Freedom Evo liquid handling station (LHS) (Tecan, Männedorf, Switzerland), as described by (Haby et al., 2018). The concentration of the feed stock was 100 g L^{-1} or 500 g L^{-1} for higher feed rates to lower the volume increase.

After the cultivation, the final volume was measured to calculate the evaporation rate. For determination of the cell dry weight 1.5 mL of culture broth was collected in duplicates in pre-dried Eppendorf tubes and centrifuged. The supernatant was discarded, and the cell pellet was dried at 75°C for more than 48 h.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2018 doi:10.20944/preprints201810.0374.v1

Peer-reviewed version available at Bioengineering 2018, 5, 101; doi:10.3390/bioengineering5040101

2.5 Sampling

From every mini-bioreactor 250 µL of culture volume was taken column-wise with the 8-channel pipette of the Freedom Evo LHS and pipetted into 96 well microtiter plates containing 15 µL dried 2 M anhydrous NaOH per well to inhibit cell activity (Haby et al., 2018).

During the batch phase samples were drawn column-wise in 5-min intervals, the at-line analysis (as described in 2.6) was performed in single determination without replicates. After beginning of the fed-batch phase samples were drawn every 20 min from the eight reactors of one column, consequently each mini-bioreactor is sampled every 2 h. Here, the at-line analysis was performed in double determination.

2.6 At-line and offline analysis

As at-line values OD_{600} and glucose concentration were determined as descripted by Haby et al, 2018. Additionally, the remaining supernatant was immediately sealed and stored at -20°C for offline analysis. The ethanol concentration was determined offline using the Cedex Bio HT Analyzer (Roche Diagnostics International Ltd, Risch, Switzerland). The detection range of ethanol using the Cedex Analyzer Ethanol Bio HT Kit (ETOHB) is 0.5 to 10.1 g L^{-1}. However, as the samples had to be diluted with an equal amount of deionized water prior to analysis due to their small volume, the lower detection limit was 1 g L^{-1}. The volumetric enzymatic activity (EA) of EPG was determined by a colorimetric assay with 2-methyl-2-benzothiazolinonehydrazone (MBTH) in 96-well plate format using the Hamilton LHS as described elsewhere [39].

2.7 Experimental design

The cultivations in the 48 mini-bioreactors were performed under 16 experimental conditions in triplicate based on a fractional factorial design. Two batch lengths – determined by an initial glucose concentration of either 20 g L^{-1} or 30 g L^{-1} –, three feed profiles – exponential, linear and constant –, three feed rates – 0.0875 h^{-1}, 0.125 h^{-1} and 0.35 h^{-1} – and an optional "hunger phase" – a period of 2 h after batch end, where no substrate was supplied – were applied. The full experimental plan is provided in table 1.

Table 1. Cultivation details. The experimental conditions regarding each mini-bioreactor are shown, including the fed-batch profile, feed rate, initial substrate concentration S_0 and occurrence of a hunger phase.

Condition	Profile	Feed rate	S_0	Hunger phase
1	Exponential	0.0875 h^{-1}	30 g L^{-1}	-
2	Exponential	0.175 h^{-1}	30 g L^{-1}	-
3	Exponential	0.35 h^{-1}	30 g L^{-1}	-
4	Constant	0.175 h^{-1}	30 g L^{-1}	-
5	Exponential	0.0875 h^{-1}	20 g L^{-1}	-
6	Exponential	0.175 h^{-1}	20 g L^{-1}	-
7	Exponential	0.35 h^{-1}	20 g L^{-1}	-
8	Constant	0.175 h^{-1}	20 g L^{-1}	-
9	Linear	0.175 h^{-1}	30 g L^{-1}	-
10	Linear	0.35 h^{-1}	30 g L^{-1}	-
11	Linear	0.175 h^{-1}	30 g L^{-1}	2h
12	Linear	0.35 h^{-1}	30 g L^{-1}	2h
13	Linear	0.175 h^{-1}	20 g L^{-1}	-
14	Linear	0.35 h^{-1}	20 g L^{-1}	-
15	Exponential	0.0875 h^{-1}	20 g L^{-1}	2h
16	Exponential	0.35 h^{-1}	20 g L^{-1}	2h

2.8 Calculation of feed rates

The feed rates were calculated based on an initial feed rate F_0:

$$F_0 = \frac{\mu_{set}}{S_i * Y_{X/S}} \cdot X_0 \cdot V_0 \quad (1)$$

where μ_{set} [h^{-1}] represents the set-point of the specific growth rate, S_{in} [g L^{-1}] the glucose concentration in the feed, $Y_{X/S}$ [g g^{-1}] the biomass yield coefficient, V_0 [L] the starting volume and X_0 [g L^{-1}] the biomass concentration present at the start of the cultivation in each mini-bioreactor [g L^{-1}]. $Y_{X/S}$ was estimated to 0.5 [g g^{-1}] according to [38], determined for the same recombinant strain.

In the first feed phase with a duration of 12 h, the feed was increased either exponentially or linearly, or kept constant. The exponential feed F_{exp} [L h^{-1}] was calculated from the initial feed rate F_0, the set growth rate μ_{set} and the time t [h]:

$$F_{exp}(t) = F_0 \cdot e^{\mu_{set}(t)} \quad (2)$$

The linear and constant feed in the first feed phase were based on the total amount of glucose fed in a respective exponential feed (Appendix A). Therefore, all cultivations received the same amount of glucose after the first feed phase (Figure 1).

After 12 h the feed was switched to an equal constant feed for all feed profiles. To ensure the same feeding conditions, the feed rate applied in this feed phase was the same for all feed profiles corresponding to the same feed rate. As the feeding was applied using the semi-continuous method of small bolus additions every 5 min, the feed rates were discretized into pulses of 5-min intervals, adapted from (Anane et al., 2018). Due to the limitation in total volume in the mini-bioreactor, the feed was limited to 30 µL per pulse, shortening the first fed-batch phase in some cultivations.

For calculation of the feed volumes as well as for the data processing and multivariate analysis, explained in the following, MATLAB 2016a, respectively 2017a was used (The MathWorks, Inc., Natick, Massachusetts, USA).

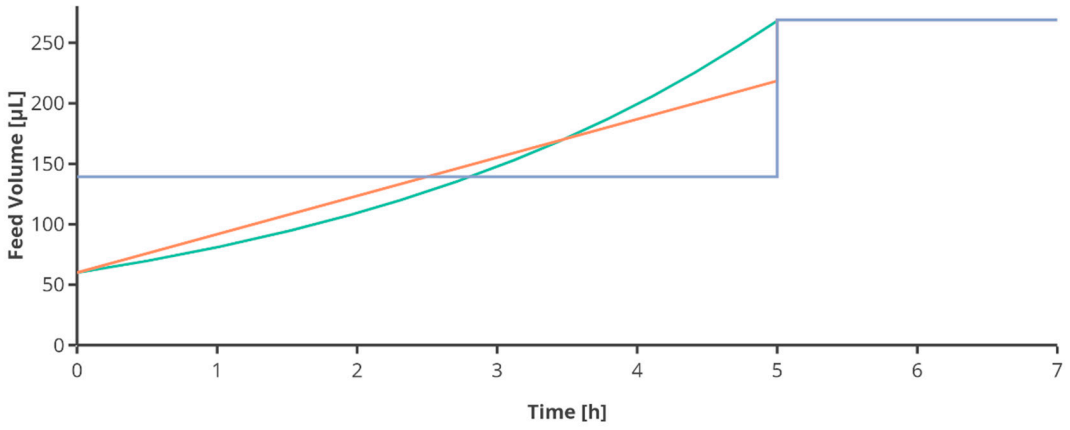

Figure 1. Schematic overview of the feed profiles. In the first feed phase exponential (green), linear (orange) or constant feed (blue) is applied according to the experimental plan, followed by a constant feed (here after 5 h) for all cultivations in the second feed phase. As the calculation of the linear and constant feed is based on the respective exponential feed, all cultivations receive the same amount of glucose during the cultivation, relative to the biomass concentration at the feed start.

2.9 Data processing

To efficiently access the large amount of data, which are produced during this experiment, methods to read the data from the central database (Haby et al., 2018) were developed. Processing and visualization of the cultivation data was implemented using MATLAB. A flexible framework

allows easy adaptation of the number of mini-bioreactors, experimental information and calculations for all mini-bioreactors.

The different feed and evaporation rates led to differing volumes in the mini-bioreactors. To simplify comparison, the OD_{600} values were normalized to the start volume of 10 mL considering the evaporation and the dilution by feed, base and medium addition. All other measurements were not normalized, so the concentrations might be affected by dilution. Regarding the specific EA, the volume difference had no influence and the values therefor did not have to be adapted.

The specific growth rate μ [h^{-1}], the substrate consumption rate q_S [$g_{substrate}$ ($g_{biomass}$ $h)^{-1}$] and the specific product formation rate q_P [U ($g_{biomass}$ $h)^{-1}$] were determined by the following equations:

$$\mu = \frac{\ln(X_{N,2}) - \ln(X_{N,1})}{t_2 - t_1} \quad (3)$$

$$q_S = \frac{S_2 - S_1}{(t_2 - t_1) \cdot X} \quad (4)$$

$$q_P = \frac{EAv_2 - EAv_1}{(t_2 - t_1)} \cdot \frac{2}{X_2 - X_1} \quad (5)$$

where X [g L^{-1}] refers to the biomass concentration, X_N [g L^{-1}] refers to the normed biomass concentration, S [g L^{-1}] to the substrate concentration, t [h] to the cultivation time and EAv [U mL^{-1}] to the volumetric enzyme activity.

2.10 Multivariate statistical analysis

The data generated from the MBR was analyzed using statistical tools to aid process optimization based on the parallel experiments. The resulting three dimensional dataset (runs × variables × times) was unfolded in batch-wise manner [40] to obtain a table with the rows spanning the different experiments and the columns distinguishing different variables at different time points.

Firstly, for a dynamic comparison of the multivariate process behavior of the runs, batch-wise unfolded (BWU) principal component analysis (PCA) [41] using the biomass, glucose and ethanol concentration, pH, DOT, base additions and volumes as input was performed. Score plots were used to abnormally behaving reactors and reactors that performed similarly. Statistically significant clusters were then identified in the score space automatically using k-means clustering algorithm [42]. Correspondingly, clusters were characterized based on the design of experiments using a classification tree analysis [43].

Additionally, for predicting the product characteristic based on the process behavior, a prediction model to estimate the EPG activity was developed using BWU historical partial least square regression (PLSR) model [44]. Thus, for instance, to build a prediction model for EPG activity at 22 h in addition to the design variables, the measured variables, namely biomass, glucose and ethanol concentration, pH, DOT and volumes maximally until this time were used as input for the model. A variable selection routine to identify the most important variables and crucial measurement time point was implemented. Variables were added one at a time and the mean of the 10-fold cross validation [42] error was monitored to identify the variable combination providing the minimum mean root mean square error of cross validation (RMSECV). The RMSECV was computed regarding the total number of cross validations N_{runs}, the predicted values Y^{pred} and the observed values Y^{act} using the following formula:

$$RMSECV = \sqrt{\frac{1}{N_{runs}} \sum_{i=1}^{N_{runs}} \left(\frac{Y_i^{pred} - Y_i^{act}}{Y_i^{act}} \right)^2} \quad (6)$$

Finally, the experimental conditions that could simultaneously minimize ethanol production and maximize EPG activity were identified using regression tree analysis [43]. Four independent regression trees were developed to predict the four targets, ethanol and EPG activity at 22 h and 35 h, using feeding profile, feed rate, initial substrate concentration and hunger phase data. The decision paths, i.e. the applied experimental conditions, leading to high EPG activity and low ethanol

concentrations were identified in all the decision trees. The thus identified optimal experiments for production were evaluated for overlapping cultivation conditions, for example the same feeding rate or profile, to ascertain a suitable design.

3. Results

Screening and process development under fed-batch conditions in stirred MBRs enable strain phenotyping closer to industrial conditions while still reducing the experimental time and effort.

To achieve this, a procedure for process control, sampling and analysis for 48 parallel mini-bioreactor cultivations in a HT platform (Figure 2) was developed. *S. cerevisiae* AH22 producing recombinant EPG was characterized in HT in the milliliter scale regarding the best growth conditions for optimization of the final and specific product concentration.

Sixteen combinations of cultivation conditions – including different batch lengths, feed profiles and rates as well as a hunger phase – were performed in triplicates, and the results were compared regarding growth behavior, glucose consumption, ethanol and EPG production.

Figure 2. The robotic HT platform for cultivation in 48 MBRs and automated sampling and analysis. Shown are the Tecan LHS including the MBR system, the Hamilton LHS connected by a linear transfer unit.

3.1. Growth and carbon metabolism

The fed-batch phase started after 15 h for cultivations with a lower initial substrate concentration of 20 g L^{-1} and after 16.55 h for cultivations with a higher initial substrate concentration of 30 g L^{-1} as well as for cultivations, where a hunger phase was applied. The initial biomass concentration at $t_{feedstart}$ was 1.70 ± 0.28 g L^{-1}.

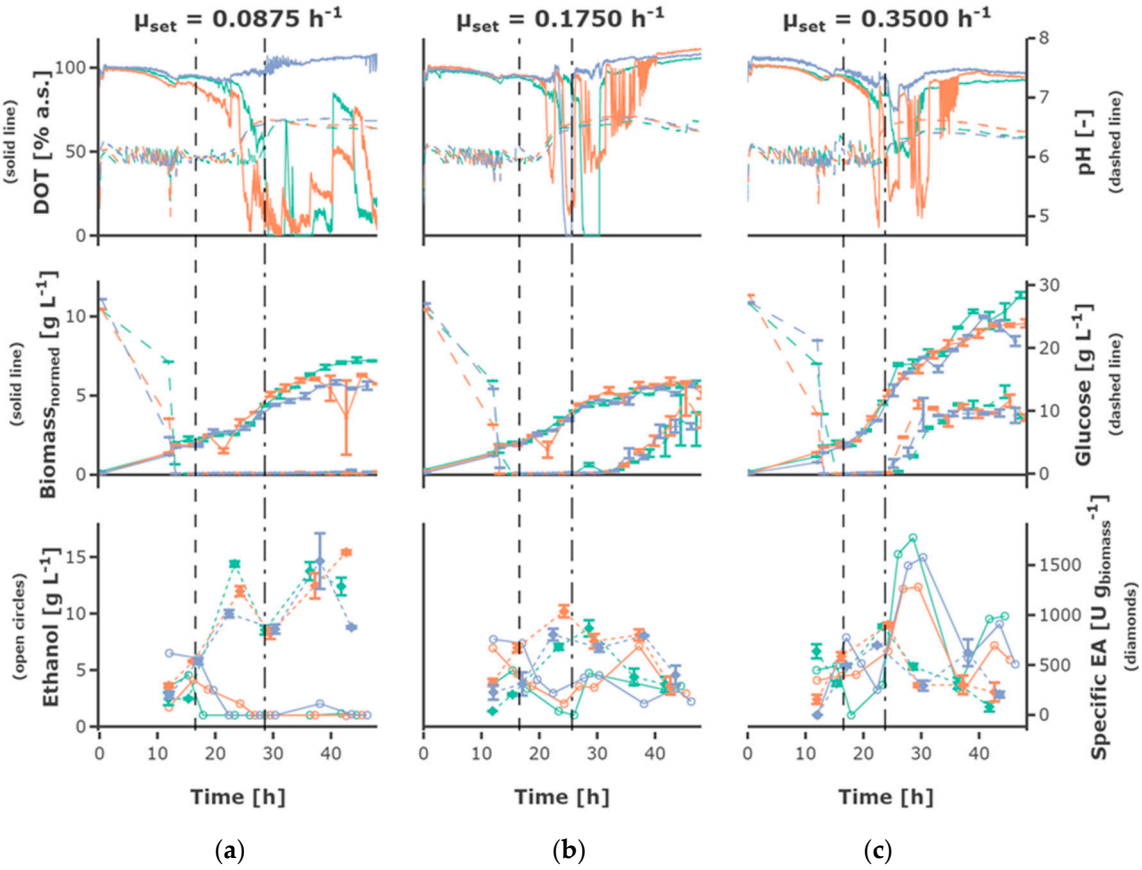

Figure 3. Measurement data is shown regarding the designs 1, 2 and 3 – exponential feed with feed rates of (**a**) 0.0875 h^{-1}, (**b**) 0.175 h^{-1} and (**c**) 0.35 h^{-1}, initial substrate concentration S_0 = 30 g L^{-1} and no hunger phase. The DOT and pH, biomass and glucose concentration as well as ethanol concentration and specific enzyme activity are shown. The mean and standard deviation are shown regarding duplicate measurements for biomass and glucose concentration and regarding triplicate measurements for specific enzyme activity. The three MBR cultivations performed under the same conditions are shown in different colors, the start of the feed and the constant fed-batch phase are shown by the vertical dashed, respectively dashed-dotted line. Interactive versions of the plots can be found online: https://www.tu-berlin.de/?200026

The highest feed rate of 0.35 h^{-1} led to the highest biomass concentrations (Figure 3), the lower feed rates of 0.0875 h^{-1} and 0.175 h^{-1} resulted in comparable growth behavior and final cell density. Regarding exponential growth at different feed rates, similar biomass and glucose concentration were observed for the two different initial substrate concentrations, though a slightly higher biomass concentration was reached for 30 g L^{-1}. The feed profile (Figure 4) and the hunger phase did not have an observable influence. The cultures grew at a growth rate between of 0.18 ± 0.05 h^{-1} in the batch phase (Figure 5a). During the first fed-batch phase, the growth rate was alternating between μ = 0.04-0.1 h^{-1} for cultivations fed at a feed rate of 0.0875 h^{-1}, μ = 0.06 to 0.1 h^{-1} for a feed rate of 0.175 h^{-1} and μ = 0.15 to 0.2 h^{-1} for a feed rate of 0.35 h^{-1}, respectively (Figure S2). The growth rate was lower in the beginning of the first feed phase – especially for cultivations with hunger phase – indicating a lag phase. After the begin of the constant fed-batch phase, the growth rate declined. Especially regarding the biomass concentrations, a high reproducibility between the replicates could be reached (Figure S1).

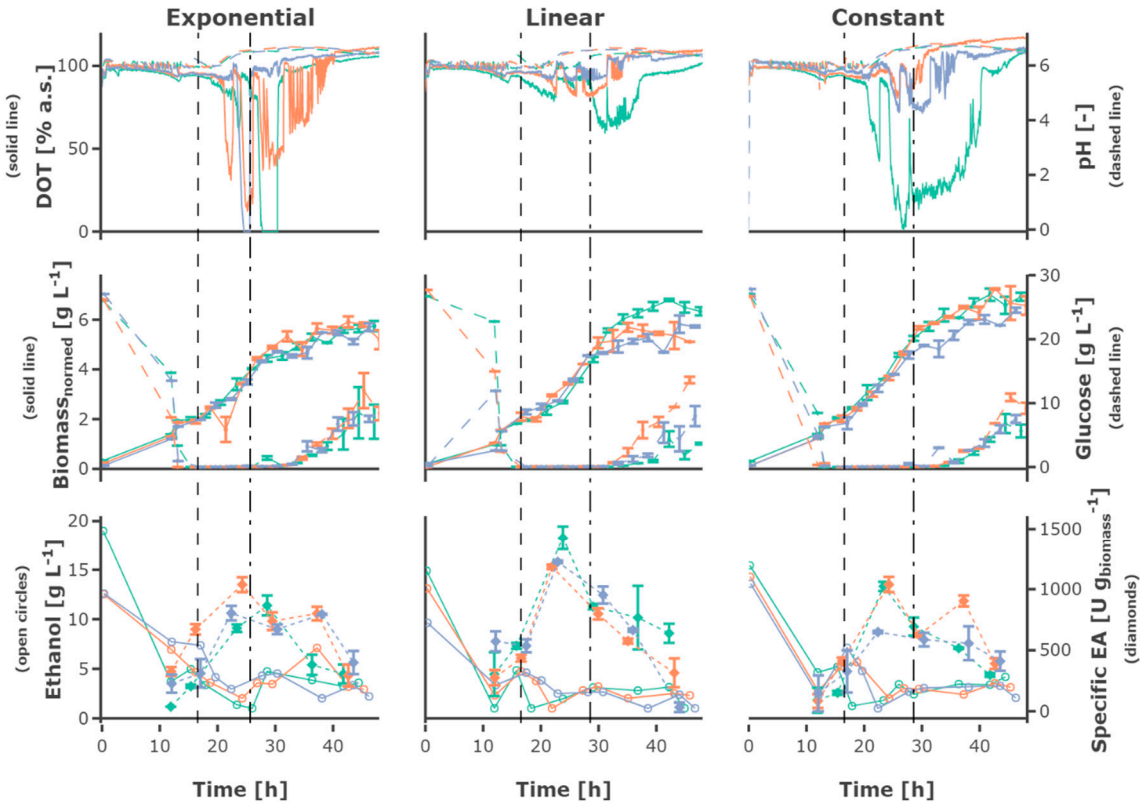

Figure 4. Cultivation data regarding growth at different feed profiles applying a feed rate of 0.175 h^{-1}. Measurement data is shown regarding the designs 2, 4 and 9 – (**a**) exponential, (**b**) linear and (**c**) constant feed at a feed rate of 0.175 h^{-1} with an initial substrate concentration S_0 = 30 g L^{-1} and no hunger phase. The DOT and pH, biomass and glucose concentration as well as ethanol concentration and specific enzyme activity are shown. The mean and standard deviation are shown regarding duplicate measurements for biomass and glucose concentration and regarding triplicate measurements for specific enzyme activity. The three MBR cultivations performed under the same conditions are shown in different colors, the start of the feed and the constant fed-batch phase are shown by the vertical dashed, respectively dashed-dotted line. Interactive versions of the plots can be found online: https://www.tu-berlin.de/?200026

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2018 doi:10.20944/preprints201810.0374.v1

Peer-reviewed version available at Bioengineering 2018, 5, 101; doi:10.3390/bioengineering5040101

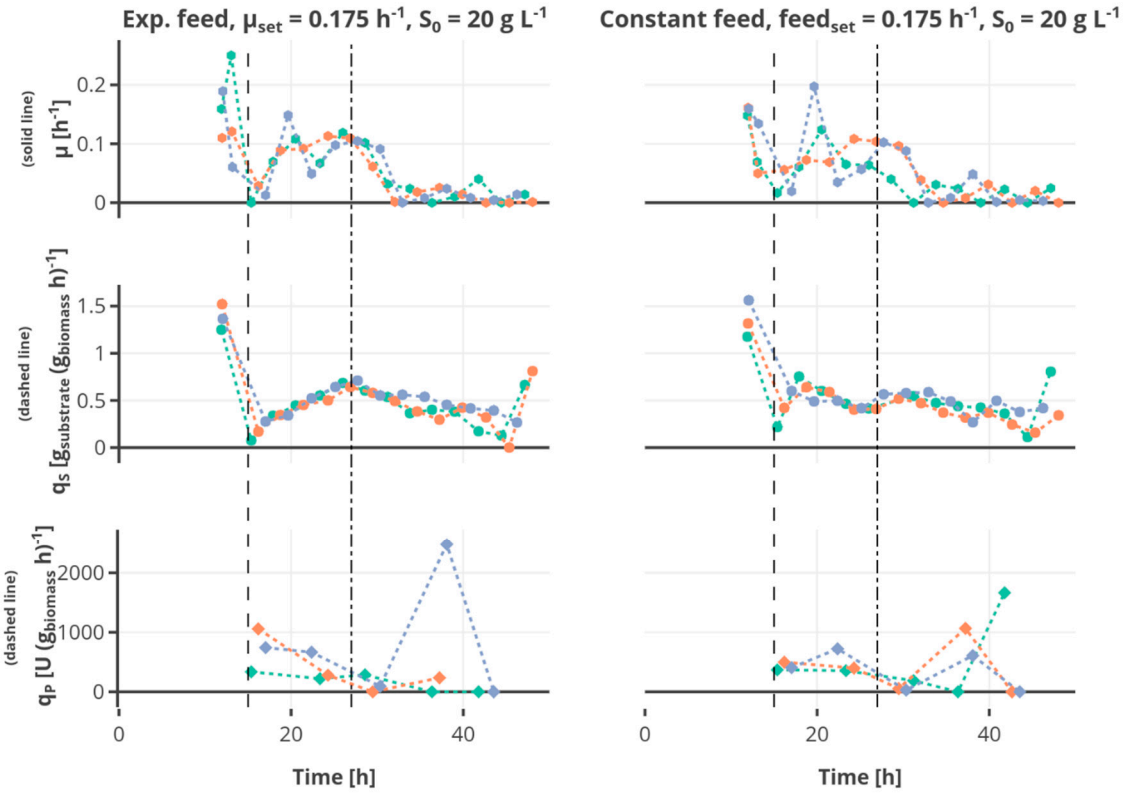

Figure 5. Specific rates for growth rate μ, substrate consumption q_S and product formation q_P are shown regarding (**a**) exponential feed at a feed rate of 0.175 h^{-1}, an initial substrate concentration S_0 of 20 g L^{-1} and no hunger phase, and regarding (**b**) constant feed at a feed rate of 0.175 h^{-1}, an initial substrate concentration S_0 of 20 g L^{-1} and no hunger phase. The rates are calculated using the mean of the respective measurements. The three MBR cultivations performed under the same conditions are shown in different colors, the start of the feed and the constant fed-batch phase are shown by the vertical dashed, respectively dashed-dotted line. Interactive versions of the plots can be found online: https://www.tu-berlin.de/?200026

The substrate uptake rate during the batch was 1.34 ± 0.34 $g_{substrate}$ $(g_{biomass}\ h)^{-1}$. The substrate consumption rate increased in the first feed phase for the cultivations fed exponentially or linearly (see for example Figure 5a). Regarding the cultivations receiving constant feed (see for example Figure 5b), the substrate consumption rate increased in the first part of this feed phase to decrease afterwards due to the decreasing availability of substrate regarding the biomass. After the shift to the constant feed, the substrate consumption rate decreased.

Glucose accumulated depending on the feed rate, for some cultivations fed at the highest feed rate of 0.35 h^{-1} as early as around 20 h cultivation time. Though, glucose accumulation could be observed for all cultivations, even the cultivations fed at the lowest feed rate, after around 40 h of cultivation. After 30-40 h, the cells in all mini-bioreactor cultivations (Figure S1) entered a phase of growth stagnation, although glucose was present in the cultivation medium, partly at high concentrations.

An increase in pH was observed for all cultivations, starting slightly later in cultivations fed at lower feed rates. The DOT signal was very irregular and oxygen limitation was detected for some cultivations for a short time, possibly due to repeated clogging of the aeration ports for the individual MBRs. However, no effects could be seen on growth and ethanol production regarding cultures with and without short-time oxygen limitation – e.g. regarding the triplicates fed at a growth rate of 0.0875 h^{-1} (Figure 3).

3.2. Recombinant protein production

The plasmid for EPG production is equipped with the constitutive *ADHI* promoter, so the enzyme is produced from the beginning of the cultivation.

The volumetric and specific enzyme activity increased during the batch and first h of fed-batch phase, though mainly remained constant or decreased towards the end of the fermentation, only increasing slightly for the cultivations fed at a feed rate of 0.0875 h^{-1}.

Rather constant product formation rates between 150 and 400 U ($g_{biomass}$ h)$^{-1}$ could be observed until 23 h of cultivation. However, the productivity decreased until the end of the cultivation (Figure 4c and 5c).

The highest final yield and final specific yield for EPG were obtained for the cultivations with an initial substrate concentration of 30 g L^{-1}, which were fed exponentially at a μ_{set} = 0.0875 h^{-1} without hunger phase. Specific enzyme activities of up to 1511.9 ± 27.2 U $g_{biomass}^{-1}$ were obtained after around 23 h cultivation time (other replicates: 1239.8 ± 49.8 U $g_{biomass}^{-1}$; 1013.5 ± 37.1 U $g_{biomass}^{-1}$) and 1540.0 ± 278.8 U $g_{biomass}^{-1}$ after around 37 h (other replicates: 1439.7 ± 91.3 U $g_{biomass}^{-1}$; 1293.8 ± 127.1 U $g_{biomass}^{-1}$) (Table S2). The volumetric enzyme activity after 37 h was 9.09 ± 0.58 U mL^{-1} (other replicates: 8.59 ± 1.56 U mL^{-1}; 7.82 ± 0.77 U mL^{-1}).

The cultivations with an initial substrate concentration of 20 g L^{-1} showed a similar EPG expression profile compared to the 30 g L^{-1}.

3.3. Ethanol formation

During the batch phase around 5-10 g L^{-1} ethanol are produced. In cultivations fed at a rate of 0.0875 h^{-1} the remaining ethanol was taken up by co-metabolism of glucose and ethanol, and decreased below the detection rate. Ethanol was present at rather constant concentrations of around 3 g L^{-1} in cultivations fed at a rate of 0.175 h^{-1}, while ethanol accumulating up to 15 g L^{-1} occurs in cultivations fed at a rate of 0.35 h^{-1}. In most cultivations fed at the rate of 0.35 h^{-1} the ethanol concentration declined after a cultivation time of 30 h to increase again until the end of the cultivation.

3.4. Multivariate analysis for information extraction

First, BWU–PCA was performed for each measurement time point by using the historic process information until the considered time point for analysis. Thus, at every time point, the routine was able to identify batches showing abnormal behavior. Figure 6 shows score plots of the BWU-PCA scores for all 48 runs incorporating their process history until 22 h. It can be identified that one run falls distinctly apart from all the other batches at both time points, i.e. that such abnormality can be detected early in the process. This was identified to be run 46 (marked in black in Figure 6a), which experienced a failure during the experiments. The pH and DOT sensors were not working properly resulting in incorrect culture handling. For future analysis, this outlier was removed.

In addition, runs that were similar were identified using the k-means clustering algorithm resulting in three clusters as shown in Figure 6a. A decision tree analysis highlighted that the two characteristic clusters (marked orange and green) were determined by the feeding rate and were segregated into the lower feed rates of 0.0875 h^{-1} and 0.175 h^{-1} and the high feed rate of 0.35 h^{-1}. However, the other manipulated variables in the experimental design, i.e. feeding profile, hunger phase and initial substrate concentration did not show significance in determining the similarity of runs. Nonetheless, these might be indeed important to further understand peculiarities of the process behavior in each of the major clusters as well as to explain the product characteristics. It is important to highlight that the abnormal behavior of run 46 can be detected within the first hour of the process duration, while the distinctly different evolution of the process can be segregated from 22 h. Although in this simple case the outlier can be even detected visually based on the process information, such a tool is generally useful to identify pro-actively abnormal and different process behavior so to suggest on improved operating conditions or abort the process.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 October 2018 doi:10.20944/preprints201810.0374.v1

Peer-reviewed version available at Bioengineering 2018, 5, 101; doi:10.3390/bioengineering5040101

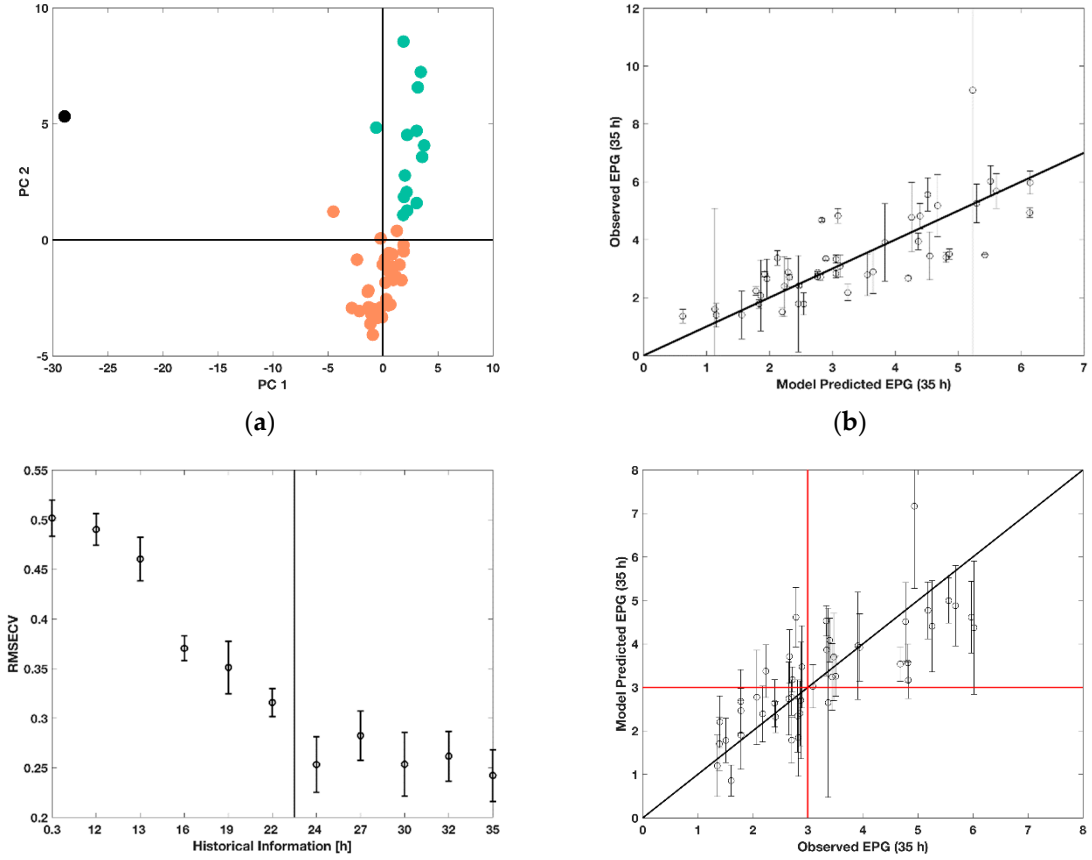

Figure 6. Real-time outlier identification and batch characterization using PCA. (**a**) Clusters 1 (red circles) with 17 observations, cluster 2 (green circles) with 30 observations and the outlier (black circle) detected after 24 h of cultivation; (**b**) Prediction of PLSR model using all history and variable selection against experimental value of volumetric EPG activity at 35 h. (c) RMSECV (see 2.10) of the PLSR models built based on different amounts of history (with variable selection). (d) Experimental values of volumetric EPG activity at 35 h of batch against the model prediction built using the history until 24 h (with variable selection).

With regards to the second goal, historic PLSR models were built to predict the analytically costly EPG activity based on the simple-to-access process measurements. For the EPG prediction at two measurement times, 22 h and 35 h, an average RMSECV of 21.45 ± 12% and 23.30 ± 12% was obtained, respectively. The prediction results for the latter case are additionally visualized in Figure 5b where one can observe that most of the 47 runs are decently predicted while few runs are either under- or over-estimated. The limitation of such predictions can be noise in the measurements of process variables and EPG (highlighted by error bars). Especially, the analytics for the latter should be improved to decrease variability and provide a more consistent basis for prediction. On the other hand, the linear structure of the model is likely not to capture all the peculiarities of the biological system, so that mechanistically valid non-linear terms are likely to be of advantage.

A further goal of this predictive analysis was the evaluation of the possibility to forecast the final volumetric EPG activity based on a shorter duration of process history, i.e. not only to build a soft sensor for the protein activity based on easier-to-access process measurements but also to anticipate the activity in advance so to provide an early basis for decision taking. Figure 6c shows the RMSECV distribution obtained for different amounts of process history used for prediction. One can observe that the process outcome can already be accurately predicted after 24 hours. The two significant drops of the RMSECV distributions at 15 and 21 hours signify that important information on the process characteristics are added here. The latter observation is also in line with the one from Figure 6a where only after 22 h a clear separation of the two process regimes was evident. The first observation is very

likely to be related to the start of feeding, while the second observation could be interpreted as the time point when the response of the fermentation process to the culture conditions including the feeding profile is clearly established. The importance of such analysis is highlighted in Figure 6d, which shows the PLSR model predictions based on process history until 24 hours to forecast the final volumetric EPG activity. Error bars signify the standard deviation of the predictions, while the red cross helps to distinguish low productivity runs (EPG < 3 U mL^{-1}) from high productivity runs (EPG > 3 U mL^{-1}). Although, the error is rather high regarding the model prediction, probably due to deviations between the triplicates of cultivations performed under the same conditions and measurement errors, the model can detect a trend in the culture behavior. With few exceptions, after two thirds of the process duration the model can therefore clearly forecast whether under the given operating conditions, the volumetric EPG activity 11 hours later will be low or high. This enables in the future to optimize the process conditions in real-time based on such predictions. However, further improvements of the model could be achieved by increasing the accuracy of the measurements and improving the outlier detection.

Besides the consideration until which time point the process must be quantified, Table 2 targets the analysis based on the variable selection routine, which time points in particular provide important and unique process information to predict volumetric EPG activity. In fact, the sampling scheme could be drastically reduced, measuring instead of 28 samples (11 for biomass, 11 for glucose and 6 for ethanol) only 17 samples, while retaining the predictive power of the corresponding two models.

Table 2. Important measurement times for biomass, glucose and ethanol concentration that provide crucial information for the volumetric EPG activity prediction model.

	Biomass [h]	**Glucose [h]**	**Ethanol [h]**
EPG (22h)	12, 13, 16, 19, 22	0.3, 16, 19	16, 22
EPG (35h)	16, 22, 27, 32	12, 13, 27, 30	16, 22, 35

Finally, the multi-target characteristics of yeast fed-batch development shall be considered, namely the adaptation of the process operation mode so to have a high product formation while minimizing ethanol production [39]. This was addressed through regression tree analysis of the manipulated process conditions to the four characteristics of interest, i.e. the volumetric EPG activity and ethanol concentration, both measured at 22 h and 35 h. Table 3 shows optimal paths for each of the four characteristics resulting in similar distributions of the considered variable (represented by mean and standard deviation). For instance, for ethanol at 22 h three possible process operation selections were identified, while for the volumetric EPG activity at 35 h one sequence of critical decisions for process variables was obtained. Regarding all four characteristics, recurring conditions were exponential feeding at a feeding rate of 0.0875 h^{-1} with an initial substrate concentration of 30 g L^{-1} and no hunger phase, which were thus identified as the most appropriate conditions. The regression tree analysis confirms the observations made for ethanol accumulation and volumetric EPG activity in the previous sections but additionally identifies other equivalent possibilities. With further targets to be considered in the future as well as additional process parameters tested, this approach offers a stream-lined procedure for model-based decision taking for process optimization.

Peer-reviewed version available at Bioengineering 2018, 5, 101; doi:10.3390/bioengineering5040101

Table 3. Optimal conditions regarding the feed rate and profile, the initial substrate concentration S_0 and the presence of a hunger phase identified by decision tree analysis for high volumetric EPG activity and low ethanol production. Each row in the table corresponds to an optimal combination of process conditions for the considered target variable. The total number of runs performed under these conditions is given as well as he mean and standard deviation of the ethanol concentration and volumetric EPG activity is given regarding those runs. The symbol '-' indicates that any value of this variable is acceptable. Conditions which lead to an optimal result regarding both ethanol concentration and volumetric EPG activity at both time points are underlined.

	Feed rate [h^{-1}]	Feed profile	S_0 [g L^{-1}]	Hunger phase [h]	Number of runs [#]	Mean	Std
Ethanol (22h) (g L^{-1})	0.0875	-	-	-	9	1299	418
	0.175	Linear	-	-	9	1521	577
	0.175	Constant/Exponential	30	-	6	1498	518
Ethanol (35h) (g L^{-1})	0.0875	Linear/Exponential	-	0	6	951	350
	0.175	Linear	-	-	6	1381	568
EPG (22h) (U mL^{-1})	-	Constant/Exponential	30	-	12	2.6	0.7
	0.175	Linear	30	0	6	2.4	0.8
EPG (35h) (U mL^{-1})	0.0875	Constant/Exponential	-	-	6	5	2

4. Discussion

In this study, 16 experimental conditions were carried out in triplicates in 48 MBRs to evaluate the influence of substrate availability and feeding strategy on recombinant protein production in *S. cerevisiae*. Cultivation, sampling and at-line analysis were performed automatically on the high throughput platform. Data handling – from raw data processing to the visualization and predictive modeling – was performed standardized and automated with minimal human input to rapidly gain process information and decision support from the enormous amount of data, which was collected during the experiment.

MBR platforms allow to combine the advantages of both microtiter plates and benchtop-scale bioreactors resulting in high experimental throughput and high information gain [4]. They thus are an important step towards consistent bioprocess development. Multiple replicates of the cultivations with the same experimental conditions were included and high comparability between experiments was achieved as the batch variability was reduced compared to the sequential approach. The latter is also important, as it has been shown that the history of the cells has a strong influence on the results obtained in the following cultivation [39,45].

A maximum specific product formation rate of 400 U $(g_{biomass}\ h)^{-1}$ was observed at the lowest feed rate of 0.0875 h^{-1} and 250 U $(g_{biomass}\ h)^{-1}$ at a feed rate of 0.175 h^{-1}. This is comparable to the rates achieved at similar dilution rates of 0.08 to 0.11 h^{-1} in change-stat cultivations of *S. cerevisiae* AH22 expressing EPG as reported earlier [39]. The actual growth rate was around 30% of the set growth rate and was thus less than expected. Reasons for the growth inhibition might be the metabolic burden of recombinant protein production [46], the negative effect of ethanol on sugar and amino acid transport [47] or – in case of the higher feed rates – of the overflow metabolism [48]. Also, oscillations in substrate availability – which are introduced here by the semi-continuous feed – have been shown to lead to a reduction in growth [49,50]. The growth arrest, occurring here after 30-40 h, could be caused by the metabolic stresses which are applied during recombinant protein production [51], the depletion of some medium components or the accumulation of self-produced toxic by-products, including but not exclusively ethanol [31]. It was shown that accumulation of lactic and acetic acid lead to reduced growth and substrate consumption and increase ethanol production [52]. However, during the cultivations described here, neither lactic nor acetic acid were accumulated to concentrations reported to have a negative influence (Table S3). Ethanol production, reducing the yield, started in the cultivations fed at feed rate of 0.175 h^{-1}, which is lower than the critical dilution rate for ethanol production μ_{crit} = 0.2 h^{-1} in change-stat cultivations [39]. Again, oscillating sugar concentration might be the reason for increased ethanol production [49]. While the intermittent feeding in the MBR system is able to resemble substrate-gradients in large-scale production processes, and thus is a suitable tool for scale-down simulation (Anane et al. 2018), it results in cell stress. An improvement of the cultivation conditions can be achieved by enzymatic glucose release mimicking continuous fed-batch conditions [53].

Principal component analysis supported as a visualization tool the identification of outliers and varying process behavior. In the future, such tools can be pro-actively integrated into the experiments so to stop strongly deviating runs or identify potentially abnormal features to adapt the corresponding process conditions and control them towards the targets. Predictive models based on PLSR showed that soft sensors can be built based on simple-to-access process data to reliably quantify EPG activity. Such predictions can be obtained only after two thirds of the process duration enabling to pro-actively forecast the productivity of given operation conditions and support decision taking to improve or abort the low producing runs. Moreover, the embedded variable selection tool enabled to quantify the minimal number and characteristic time points of measurement yielding sufficient information content, i.e. an effective process analytical scheme. Given the two targets, namely high volumetric yield and low ethanol production, super-imposed regression trees enabled to identify important process parameters and their desired levels to fulfill these targets. A combination of the lowest feed rate of 0.0875 h^{-1}, an exponential feed with an initial substrate concentration of 30 g L^{-1} and no hunger phase resulted in the highest volumetric yield of ~8.5 U mL^{-1} (mean of triplicates) in comparison to a volumetric yield of ~2.1 U mL^{-1} regarding the design with the lowest yield. For future

557 analysis, these tools must be directly implemented into the experimental platform so to enable real-
558 time decision taking and process optimization. Like this, an even better evaluation of the potential of
559 such automated HT technology for efficient process development can be achieved. The real-time
560 adaptations based on the model predictions will pave the path to better process understanding and
561 the creation of a digital twin of the process. Several sequential iterations of such an experiment will
562 enable to not only design an optimal process in the Quality by Design (QbD) perspective, i.e. to
563 identify the settings of the Critical Process Parameters (CPP) resulting in optimal Critical Quality
564 Attributes (CQA), but will also provide a technologically and economically optimal operation
565 procedure with regards to the dynamic control structure, sampling scheme and reporting base for
566 the involved decision takers.
567 Bottlenecks of HT methods still exist but have shifted from experimental throughput to offline
568 analysis, data handling and evaluation [4]. Frequent sampling of up to 48 cultivations in parallel over
569 several days results in many samples, i.e. during the presented cultivation 744 samples were taken.
570 Although methods and devices for HT sample preparation and analysis already exist – e.g. a HT
571 method for cell disruption [11] – exploiting the full potential of HT screening remains a challenge,
572 particularly with regards to online product quality characteristics quantification [54], and requires
573 fast analytical methods suitable for parallelization. Thus, quantitative tools for prediction and
574 decision support will remain a key enabler of the successful realization of such automated
575 technology. In this work, a PLSR based model was developed to predict the EPG activity based on
576 simple-to-access process quantities. More dynamic measurements of this quantity would in the
577 future enable to build a real-time sensor for this product characteristic. Moreover, the developed
578 models can be further intensified through integration of existing process knowledge so to generate
579 hybrid process models [37].
580 The developed method provides a remarkable advancement towards the goals of industry 4.0
581 based on an efficient, parallelized and automated system for cultivation and analytics as well as a
582 predictive digital framework for data management and analysis. A further intensification of the
583 technology towards additional analytical capabilities, complete integration of hardware and software
584 technologies, enabling adaptive process control and integration of all involved stakeholders and
585 process know-how into such a self-learning digital platform, will revolutionize the current
586 procedures of process development through a broadly applicable, automated robotic platform.
587
588

Peer-reviewed version available at Bioengineering 2018, 5, 101; doi:10.3390/bioengineering5040101

Supplementary Materials: The following are available online, Figure S1: Cultivation data regarding growth at different applied cultivation conditions, Figure S2: Specific rates for growth μ, substrate consumption q_S and product formation q_P regarding different applied cultivation conditions, Table S1: Layout of experimental conditions, Table S2: Volumetric and specific enzyme activity (EA) after 23 h and 37 h cultivation time, Table S3: Acetate, lactate, ammonia and phosphate concentration at 32 h cultivation time.

Author Contributions: Conceptualization, A.S., S.H., M.S., F.G., P.N. and M.N.C.B.; methodology, A.S., S.H., B.H., H.N., M.S., F.G., and M.N.C.B.; software, A.S., S.H., B.H., H.N. and M.S.; validation, A.S., S.H., M.N.C.B.; formal analysis, A.S. and N.H.; investigation, A.S., S.H. and B.H.; data curation, A.S., S.H. and M.N.C.B.; writing—original draft preparation, A.S.; writing—review and editing, A.S., S.H., B.H., N.K., N.H., M.S., P.N., F.G. and M.N.C.B.; visualization, N.K.; supervision, S.H., P.N. and M.N.C.B

Funding: This research received no external funding.

Acknowledgments: We thank Terrance Wilms for inputs into the data evaluation scripts and Emmanuel Anane for valuable advice regarding the cultivations. Additionally, we thank Roche Diagnostics International Ltd for supply of the Cedex Bio HT analyzer.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The linear feed F_{lin} [L h^{-1}] was calculated by:

$$F_{lin}(t) = F_0 + a \cdot t \tag{A1}$$

where *a* represents the linear increase in the feed rate. The slope *a* is determined by the integral of the exponential feed *A* as follows:

$$A = \int_{t_0}^{t_{end}} (F_0 + a \cdot t)dt = [F_0 \cdot t + \frac{1}{2} a \cdot t^2]_{t_0}^{t_{end}} = F_0 \cdot (t_{end} - t_0) + \frac{1}{2} a \cdot (t_{end}^2 - t_0^2)$$

$$\leftrightarrow a = \frac{2(A - F_0 \cdot (t_{end} - t_0))}{t_{end}^2 - t_0^2} \tag{A2}$$

The constant feed F_{const} [L h^{-1}] is calculated as follows:

$$F_{const}(t) = b = \frac{A}{t_{end} - t_0} \tag{A3}$$

References

[1] Neubauer, P., Cruz, N., Glauche, F., Junne, S., Knepper, A., Raven, M. Consistent development of bioprocesses from microliter cultures to the industrial scale. *Eng. Life Sci.* **2013**, *13*, 224–238, doi:10.1002/elsc.201200021.

[2] Neubauer, P., Glauche, F., Cruz-Bournazou, M.N. Editorial: Bioprocess Development in the era of digitalization. *Eng. Life Sci.* **2017**, *17*, 1140–1141, doi:10.1002/elsc.201770113.

[3] Haby, B., Glauche, F., Hans, S., Bournazou, M.N.C.-, Neubauer, P., Bioverfahrenstechnik, F., Biotechnologie, I.F.Ü.R., Berlin, T.U. Stammcharakterisierung mittels on-line -Redesign von Experimenten. **2018**, *200*,, doi:10.1007/s12268-018-0889-6.

[4] Hemmerich, J., Noack, S., Wiechert, W., Oldiges, M. Microbioreactor Systems for Accelerated Bioprocess Development. *Biotechnol. J.* **2018**, *13*, 1–9, doi:10.1002/biot.201700141.

[5] Hoefel, T., Faust, G., Reinecke, L., Rudinger, N., Weuster-Botz, D. Comparative reaction engineering studies for succinic acid production from sucrose by metabolically engineered Escherichia coli in fed-batch-operated stirred tank bioreactors. *Biotechnol. J.* **2012**, *7*, 1277–1287, doi:10.1002/biot.201200046.

[6] Faust, G., Janzen, N.H., Bendig, C., Römer, L., Kaufmann, K., Weuster-Botz, D. Feeding strategies enhance high cell density cultivation and protein expression in milliliter scale bioreactors. *Biotechnol. J.* **2014**, *9*, 1293–1303, doi:10.1002/biot.201400346.

[7] Schmideder, A., Cremer, J.H., Weuster-Botz, D. Parallel steady state studies on a milliliter scale accelerate fed-batch bioprocess design for recombinant protein production with Escherichia coli. *Biotechnol. Prog.* **2016**, *32*, 1426–1435, doi:10.1002/btpr.2360.

[8] Neubauer, P., Junne, S. Scale-down simulators for metabolic analysis of large-scale bioprocesses. *Curr. Opin. Biotechnol.* **2010**, *21*, 114–121, doi:10.1016/j.copbio.2010.02.001.

[9] Anane, E., Sawatzki, A., Neubauer, P., Cruz-Bournazou, M.N. Modelling concentration gradients in fed-batch cultivations of *E. coli* -towards the flexible design of scale-down experiments. *J. Chem. Technol. Biotechnol.* **2018**,, doi:10.1002/jctb.5798.

[10] Nickel, D.B., Cruz-Bournazou, M.N., Wilms, T., Neubauer, P., Knepper, A. Online bioprocess data generation, analysis, and optimization for parallel fed-batch fermentations in milliliter scale. *Eng. Life Sci.* **2016**, 1–7, doi:10.1002/elsc.201600035.

[11] Glauche, F., Pilarek, M., Bournazou, M.N.C., Grunzel, P., Neubauer, P. Design of experiments-based high-throughput strategy for development and optimization of efficient cell disruption protocols. *Eng. Life Sci.* **2017**, *17*, 1166–1172, doi:10.1002/elsc.201600030.

[12] Wu, T., Zhou, Y. An Intelligent Automation Platform for Rapid Bioprocess Design. *J. Lab. Autom.* **2014**, *19*, 381–393, doi:10.1177/2211068213499756.

[13] Freier, L., Hemmerich, J., Schöler, K., Wiechert, W., Oldiges, M., von Lieres, E. Framework for Kriging-based iterative experimental analysis and design: Optimization of secretory protein production in Corynebacterium glutamicum. *Eng. Life Sci.* **2016**, *16*, 538–549, doi:10.1002/elsc.201500171.

[14] Glauche, F., Glazyrina, J., Cruz Bournazou, M.N., Kiesewetter, G., Cuda, F., Goelling, D., Raab, A., Lang, C., Neubauer, P. Detection of growth rate-dependent product formation in miniaturized parallel fed-batch cultivations. *Eng. Life Sci.* **2017**,, doi:10.1002/elsc.201600029.

[15] Cruz Bournazou, M.N., Barz, T., Nickel, D.B., Lopez Cárdenas, D.C., Glauche, F., Knepper, A., Neubauer, P. Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities. *Biotechnol. Bioeng.* **2017**, *114*, 610–619, doi:10.1002/bit.26192.

[16] Velez-Suberbie, M.L., Betts, J.P.J., Walker, K.L., Robinson, C., Zoro, B., Keshavarz-Moore, E. High throughput automated microbial bioreactor system used for clone selection and rapid scale-down process optimization. *Biotechnol. Prog.* **2017**, *15*, 1–11, doi:10.1002/btpr.2534.

[17] Vester, A., Hans, M., Hohmann, H.P., Weuster-Botz, D. Discrimination of riboflavin producing Bacillus subtilis strains based on their fed-batch process performances on a millilitre scale. *Appl. Microbiol. Biotechnol.* **2009**, *84*, 71–76, doi:10.1007/s00253-009-1966-z.

[18] Hortsch, R., Weuster-botz, D. Growth and recombinant protein expression with Escherichia coli in different batch cultivation media. **2011**, 69–76, doi:10.1007/s00253-010-3036-y.

[19] Lang, C., Looman, a. C. Efficient expression and secretion of *Aspergillus niger* RH5344 polygalacturonase in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **1995**, *44*, 147–156, doi:10.1007/s002530050533.

[20] Meehl, M.A., Stadheim, T.A. Biopharmaceutical discovery and production in yeast. *Curr. Opin. Biotechnol.* **2014**, *30*, 120–127, doi:10.1016/j.copbio.2014.06.007.

[21] Krivoruchko, A., Nielsen, J. Production of natural products through metabolic engineering of Saccharomyces cerevisiae. *Curr. Opin. Biotechnol.* **2015**, *35*, 7–15, doi:10.1016/j.copbio.2014.12.004.

[22] Li, M., Kildegaard, K.R., Chen, Y., Rodriguez, A., Borodina, I., Nielsen, J. De novo production of resveratrol from glucose or ethanol by engineered Saccharomyces cerevisiae. *Metab. Eng.* **2015**, *32*, 1–11, doi:10.1016/j.ymben.2015.08.007.

[23] Hitzeman, R., Leung, D., Perry, L., Kohr, W., Levine, H., Goeddel, D. Secretion of human interferons by yeast. *Science (80-.).* **1983**, *219*, 620–625, doi:10.1126/science.6186023.

[24] Romanos, M.A., Scorer, C.A., Clare, J.J. Foreign gene expression in yeast: a review. *Yeast* **1992**, *8*, 423–488, doi:10.1002/yea.320080602.

[25] Kim, H., Yoo, S.J., Kang, H.A. Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS Yeast Res.* **2015**, *15*, 1–16, doi:10.1111/1567-1364.12195.

[26] Nielsen, J. Production of biopharmaceutical proteins by yeast. *Bioengineered* **2013**, *4*, 207–211, doi:10.4161/bioe.22856.

[27] Liu, G., Lanham, C., Buchan, J.R., Kaplan, M.E. High-throughput transformation of Saccharomyces cerevisiae using liquid handling robots. *PLoS One* **2017**, *12*, 1–15, doi:10.1371/journal.pone.0174128.

[28] Hans, S., Gimpel, M., Glauche, F., Cruz-Bournazou, M.N. Automated Cell Treatment for Competence and Transformation of Escherichia coli in a High-Throughput Quasi-Turbidostat Using Microtiter Plates. *Microorganisms* **2018**,, doi:10.3390/microorganisms6030060.

[29] Sonnleitnert, B. Growth of Saccharomyces cerevisiae Is Controlled by Its Limited Respiratory Capacity : Formulation and Verification of a Hypothesis *. *Biotechnol. Bioeng.* **1985**,.

[30] Thi, H., Pham, B., Larsson, G., Enfors, S. Growth and Energy Metabolism in Aerobic Fed-Batch Cultures of Saccharomyces cerevisiae : Simulation. **1998**,.

[31] Mazzoleni, S., Landi, C., Cartenì, F., de Alteriis, E., Giannino, F., Paciello, L., Parascandola, P. A novel process-based model of microbial growth: Self-inhibition in Saccharomyces cerevisiae aerobic fed-batch cultures. *Microb. Cell Fact.* **2015**, *14*, 1–14, doi:10.1186/s12934-015-0295-4.

[32] Paalme, T., Elken, R., Vilu, R., Korhola, M. Growth efficiency of Saccharomyces cerevisiae on glucose/ethanol media with a smooth change in the dilution rate (A-stat). *Enzyme Microb. Technol.* **1997**, *20*, 174–181, doi:10.1016/S0141-0229(96)00114-7.

[33] Enfors, S.-O. Fermentation Process Engineering, **2011**.

[34] Van Daele, T., Gernaey, K. V., Ringborg, R.H., Börner, T., Heintz, S., Van Hauwermeiren, D., Grey, C., Krühne, U., Adlercreutz, P., Nopens, I. Application of iterative robust model-based optimal experimental design for the calibration of biocatalytic models. *Biotechnol. Prog.* **2017**, *33*, 1278–1293, doi:10.1002/btpr.2515.

[35] Barz, T., López Cárdenas, D.C., Arellano-Garcia, H., Wozny, G. Experimental evaluation of an approach to online redesign of experiments for parameter determination. *AIChE J.* **2013**, *59*, 1981–1995, doi:10.1002/aic.13957.

[36] Anane, E., López C, D.C., Neubauer, P., Cruz Bournazou, M.N. Modelling overflow metabolism in Escherichia coli by acetate cycling. *Biochem. Eng. J.* **2017**, *125*, 23–30, doi:10.1016/j.bej.2017.05.013.

[37] Stosch, M. Von, Hamelink, J., Oliveira, R. Hybrid modeling as a QbD / PAT tool in process development : an industrial E . coli case study. *Bioprocess Biosyst. Eng.* **2016**, *39*, 773–784, doi:10.1007/s00449-016-1557-1.

[38] Lang, C., Göllnitz, C., Popovic, M., Stahl, U. Optimization of fungal polygalacturonase synthesis by *Saccharomyces cerevisiae* in fed-batch culture. *Chem. Eng. J.* **1997**, *65*, 219–226, doi:10.1016/S1385-8947(97)00029-6.

[39] Glauche, F., Glazyrina, J., Bournazou, M.N.C., Kiesewetter, G., Cuda, F., Goelling, D., Raab, A., Lang, C., Neubauer, P. Detection of growth rate-dependent product formation in miniaturized parallel fed-batch cultivations. *Eng. Life Sci.* **2017**, 1–6, doi:10.1002/elsc.201600029.

[40] Nomikos, P., MacGregor, J.F. Multi-way partial least squares in monitoring batch processes. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 97–108, doi:10.1016/0169-7439(95)00043-7.

[41] Abdi, H., Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459, doi:10.1002/wics.101.

[42] Hastie, T., Tibsharani, R., Friedman, J. Springer Series in Statistics The Elements of. *Math. Intell.* **2009**, *27*, 83–85, doi:10.1007/b94608.

[43] Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. Classification and Regression Trees. *Cole Publ. Monterey* **1984**, *535*, 358, doi:10.1002/cyto.990080516.

[44] Sokolov, M., Soos, M., Neunstoecklin, B., Morbidelli, M., Butt??, A., Leardi, R., Solacroup, T., Stettler, M., Broly, H. Fingerprint detection and process prediction by multivariate analysis of fed-batch monoclonal antibody cell culture data. *Biotechnol. Prog.* **2015**, *31*, 1633–1644, doi:10.1002/btpr.2174.

[45] Brand, E., Junne, S., Anane, E., Nicolas, M., Bournazou, C., Neubauer, P. Importance of the cultivation history for the response of Escherichia coli to oscillations in scale-down experiments. *Bioprocess Biosyst. Eng.* **2018**, *0*, 0, doi:10.1007/s00449-018-1958-4.

[46] Neubauer, P., Lin, H.Y., Mathiszik, B. Metabolic load of recombinant protein production: Inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in Escherichia coli. *Biotechnol. Bioeng.* **2003**, *83*, 53–64, doi:10.1002/bit.10645.

[47] Thomas, D.S., Rose, A.H. Inhibitory Effect of Ethanol on Growth and Solute Accumulation by Saccharomyces cerevisiae as Affected by Plasma-Membrane Lipid Composition. *Arch. Microbiol.* **1979**, *55*, 49–55, doi:10.1007/BF00408045.

[48] Habegger, L., Rodrigues Crespo, K., Dabros, M. Preventing Overflow Metabolism in Crabtree-Positive Microorganisms through On-Line Monitoring and Control of Fed-Batch Fermentations. *Fermentation* **2018**, *4*, 79, doi:10.3390/fermentation4030079.

[49] George, S., Larsson, G., Enfors, S.O. A scale-down two-compartment reactor with controlled substrate oscillations: Metabolic response of Saccharomyces cerevisiae. *Bioprocess Eng.* **1993**, *9*, 249–257, doi:10.1007/BF01061530.

[50] Marbà-Ardébol, A.M., Bockisch, A., Neubauer, P., Junne, S. Sterol synthesis and cell size distribution under oscillatory growth conditions in Saccharomyces cerevisiae scale-down cultivations. *Yeast* **2018**, *35*, 213–223, doi:10.1002/yea.3281.

[51] Mattanovich, D., Gasser, B., Hohenblum, H., Sauer, M. Stress in recombinant protein

producing yeasts. *J. Biotechnol.* **2004**, *113*, 121–135, doi:10.1016/j.jbiotec.2004.04.035.

[52] Narendranath, N. V, Thomas, K.C., Ingledew, W.M. Effects of acetic acid and lactic acid on the growth of Saccharomyces cerevisiae in a minimal medium. *J. Ind. Microbiol. Biotechnol.* **2001**, *26*, 171–177, doi:10.1038/sj/jim/7000090.

[53] Krause, M., Neubauer, A., Neubauer, P. The fed-batch principle for the molecular biology lab: Controlled nutrient diets in ready-made media improve production of recombinant proteins in Escherichia coli. *Microb. Cell Fact.* **2016**, *15*, 1–13, doi:10.1186/s12934-016-0513-8.

[54] Pais, D.A.M., Carrondo, M.J.T., Alves, P.M., Teixeira, A.P. Towards real-time monitoring of therapeutic protein quality in mammalian cell processes. *Curr. Opin. Biotechnol.* **2014**, *30*, 161–167, doi:10.1016/j.copbio.2014.06.019.

Submitted manuscripts:

Anane et al., 2018
Haby et al., 2018