Die approbierte Originalversion dieser Dissertation ist in der Hauptbibliothek der Technischen Universität Wien aufgestellt und zugänglich.
http://www.ub.tuwien.ac.at

The approved original version of this thesis is available at the main library of the Vienna University of Technology.
http://www.ub.tuwien.ac.at/eng

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DISSERTATION

Increased information to effort ratio through physiological bioprocess development

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Prof. Dr. Christoph Herwig

E166

Institut für Verfahrenstechnik, Umwelttechnik und Technische Biowissenschaften

eingereicht an der Technischen Universität Wien,

Fakultät für Technische Chemie

von

Wieland Reichelt

0512551

Thaliastraße 7/20, 1160 Wien

Wien, am   eigenhändige Unterschrift

Deutsche Kurzfassung

Escherichia coli ist der am häufigsten genutzte Organismus für die rekombinante Proteinproduktion mit Bioprozessen [1, 2]. Während der Prozessentwicklung im Labor bis hin zum Produktionsmaßstab, werden kritische Prozessparameter identifiziert und anschließend untersucht. Das Ziel ist die Gernerierung von möglichst produktunabhängigem und damit transferierbarem Prozessverständnis um den Aufwand für die neuerliche Bioprozessentwicklung zu minimieren. Derzeitig werden Bioprozesse häufig anhand technischer Parameter entwickelt z.B. volumetrischer Fütterungsraten, was die Transferierbarkeit des Prozessverständnis limitiert. Nach Jahrzehnten der technologieorientierten Prozessentwicklung findet nun eine Umorientierung auf physiologische Parameter statt [3-6]. Unterteilt in einen analytischen und einen methodischen Teil, wird in dieser Dissertation der Mehrwert physiologischer Bioprozessentwicklung im Vergleich zu herkömmlichen Ansätzen analysiert und bewertet.

1) Methodenentwicklung und Bewertung zur Quantifizierung physiologischer Prozesse und Phänomene bzgl. Robustheit und Sensitivität:
   - Als Begeleiterscheinung von Zelllyse werden zytosolische Proteine freigesetzt [7]. Der analytische Fehler der Proteinquantifizierung wurde auf Probenmatrixeffekte zurückgeführt und durch methodische Adaption von >200% auf <50% reduziert.
   - Protein Expression führt häufig zum physiologischen Phänomen der Inclusion body (IB) Bildung. Um das Wachstum der IBs als Auswirkung der Expressionsrate quantifizierbar zu machen, wurde Nano Particle Tracking Analysis als neue Methode für die Größenbestimmung von IBs etabliert und verifiziert.
   - Physiologische Bioprozesskontrolle bedarf einer akkuraten Bestimmung der Biomasse bereits in der frühen Phase der Bioprozessentwicklung. Für Biomasseschätzung in Echtzeit, hat sich die per gewichtetem Mittelwert kombinierte, massenbilanzbasierede Softsensorenschätzung als praktikabelste Methode (Information/Aufwand) erwiesen.

2) Analyse der Vor und Nachteile physiologischer Bioprozessentwicklung anhand industriell relevanter Produktionsprozesse:
   - Physiologische Bioprozesentwicklung fußt auf numerischen, physiologischen Deskriptoren, spezifisch für definierte Prozessphasen. Eine neue substratverbrauchsbasierte Grundlage zur Phasendefinition wurde vorgestellt und ein Schema zur Integration von vorhandenem Prozesswissen in die Versuchsplanung anhand eines Beispiels illustriert.
   - Eine physiologische Fütterungsstrategie bedarf der präzisen Definition physiologischer Limits z.B. des kritischen q_S (q_{Scrit}). Mit Hilfer kontrollierter Oszillationen von qs, konnte eine starke Abhängigkeit von der Zeit sowie von der metabolischen Aktivität gezeigt werden. Folglich bedarf es einer prozesstechnologischen Strategie um das dynamische Verhalten von q_{Scrit} auch in Echtzeit zu erkennen und abzufangen.
   - Basierend auf einer Kombination aus Softsensoren wurde eine Closed-Loop-Echtzeit Kontrollstrategie für q_S etabliert. Mit dieser Kontrollstrategie war es möglich die sonst verbreitete Anhäufung von Substrat in der späten Induktionsphase zu vermeiden.

In dieser Dissertation werden die Vor- und Nachteile physiologischer Bioprozessentwicklung umfassend analysiert und diskutiert. Physiologische Bioprozessentwicklung ermöglicht tiefe Einblicke in physiologische Prozesse und fördert das generelle Verständis für das Verhalten des Produktionsstammes unabhängig von einem Produktivitätsgewinn. Physiologische Bioprozesskontrolle stellt das Rückgrat für physiologische Prozessentwicklung dar und kann Experimente zur Stammcharakterisierung offenbar sogar ersetzen. Sie eröffnet neue Möglichkeiten in der Prozessentwicklung, möglicherweise auch eine Steigerung der Raum/Zeitausbeute. Physiologische Bioprozessentwicklung verlangt anfänglich nach einem erhöhten Aufwand für die Implementierung der Analytik sowie der Kontrollalgorithmen, doch nachhaltig lohnt sie sich und erhöht das Informations zu Aufwandsverhältnis substantiell.

Acknowledgements

First of all, I would like to express the deepest appreciation to Prof. Christoph Herwig for giving me the opportunity to carry out my thesis as a member of the bioprocess engineering group at Vienna University of Technology. Without his guidance, support and inspiration this thesis would not have been possible.

Special thanks to the core members of our project team Paul Kroll, Jens Fricke, Patrick Wechselberger, Daniela Ehgartner. Thank you for the excellent working atmosphere and the rewarding scientific discussions.

Thanks also to my colleagues Lukas Neutsch, Patrick Sagmeister, Aydin Golabgir, Tobias Klein and Lukas Marschal for the great discussions and scientific collaborations in various research projects.

Furthermore, I would like to thank Florian Haas, Julian Kager, Claudia Hattinger, Peter Thurrold, Markus Brillmann, Anna Bankuti, Andreas Kaineder and Julius Albrecht. Thank you for joining our team and sharing the enthusiasm to pull through all the necessary experiments and publications. Thanks to the Christian Doppler Society for the financial support within the Laboratory for Mechanistic and Physiological Methods for Improved Bioprocesses and the Sandoz GmbH. Thanks also to the Lonza AG for financial support.

Thanks to my family, especially to my parents Kristin and Norbert, my grandparents Rosemarie, Hans, Christel and Klaus. Thanks for raising me in a vivid atmosphere, embracing curiosity and the patience for discussing all the mysteries of everyday life. Thank you for your generous, lifelong support which also reflects in this thesis.

At last, I would like to express my gratitude and love to my fiancée Doris. Thank you for your patience during the times I spend nights out in the laboratory or working in the office. Your support and encouragement were the footing facilitating this thesis.

Keywords: physiological bioprocess development, physiological bioprocess control, fermentation technology, inclusion body characterization, protein quantification, inclusion body, physiological feedback control, specific substrate uptake rate, substrate accumulation, real time biomass estimation, first principle softsensor.

Index

Abstract 5
Introduction: 6
Background 6
Goals 9
Roadmap 9
Results and discussion 11
Analytics for physiological process development 11
Discussion 69
Effectivity of physiological control 70
Discussion 154
Overall scientific Novelty and Author contributions 155
Discussion and Conclusions 157
References which formed this cumulative thesis 162
Contributions as first Author 162
Contributions as Co-Author 163
References 161
Curriculum Vitae 164

Abstract

Escherichia coli is one of the most exploited organisms for industrial production of recombinant proteins using bioprocesses [1, 2]. Within process development, critical process parameters are identified and consequently investigated from lab to production scale. In order to minimize process development effort the generation of product independent, transferable prior process knowledge is of utmost interest. Currently, bioprocesses are commonly developed based on technical process parameters as e.g. volumetric feeding rates, generating hardly transferable, technology oriented process knowledge. After decades of focusing on technical parameters, a more physiological approach has emerged [3-6]. Structured in two parts, this thesis aims to establish and assess physiological bioprocess development as well as to analyze whether it bears significant advantages compared to conventional approaches.

1) Establishment and investigation of analytical methods to quantify and detect physiological processes and phenomena with respect to accuracy and robustness:

- Cell lysis, as physiological event, features cytosolic protein release [7]. For protein quantification in complex sample matrixes the error of the method was reduced from >200% to <50%.
- High titer expression of protein frequently features the physiological phenomena of inclusion body (IB) formation. To analyze the IB growth as an effect of expression rates, the novel method of nano particle tracking analysis for IB sizing was established and successfully verified.
- Physiological bioprocess control requires accurate biomass estimation. In the context of early bioprocess development, a weighted average combination of first principle soft sensors was proven the most suitable approach for real time biomass estimation.

2) Analysis of the advantages and challenges of physiological bioprocess development at hand of industrial relevant production processes.

- Physiological bioprocess development requires single numerical descriptors of physiology representing distinct process phases. Therefore, a novel variable for physiological phase definition and a workflow to increase process knowledge integration was illustrated.
- A physiological feeding strategy based on the specific substrate uptake rate (q_S), in comparison to technological feeding profiles, was shown to be highly beneficial in terms of product titer.
- Physiological process control requires the accurate definition of physiological limits e.g. the critical q_S (q_{Scrit}). Using controlled oscillations of q_S, q_{Scrit} was shown to be highly dependent on time after induction and on the average metabolic activity q_{Smean}. The latter finding calls for technological strategies to cope with the dynamic nature of q_{Scrit}.
- Using a combination of first principle softsensors a closed loop real time control approach of q_S was established. Hereby, substrate accumulation in late process phases was effectively avoided.

This thesis comprehensively discusses advantages and challenges of physiological bioprocess development. Physiologic process development grants deeper insights into relevant physiological processes and fosters the general understanding of the behavior of the production strain regardless of an associated titer increase. Physiological bioprocess control approaches, as the backbone of physiological bioprocess development are even able to substitute strain characterization experiments. It grants additional degrees of freedom and thereby potentially allows for higher time space yields. It can be concluded, that physiological bioprocess development asks for a one time effort investment for the establishment of sensitive analytics and physiological control approaches but on the long term it rewards with a substantial increase in information to effort ratio.

Introduction

Biopharmaceuticals are defined as "A protein or nucleic acid based pharmaceutical substance used for therapeutic or in vivo diagnostic purposes, which is produced by means other than direct extraction from a native (non-engineered) biological source" [8]. Besides conventional, high volume drugs, as insulin, biopharmaceuticals are often key drugs for frequent and deleterious diseases as neurodegenerative diseases e.g. Alzheimer´s [9] or even cancer [10]. Biopharmaceutical drug production within bioprocesses is a growing, high-volume but also highly competitive market [2].

Progressively expiring patent protection of conventional drugs brings forward generic drug industry and heats up competition for the most productive and robust bioprocesses [11]. While in the last decades products targeted high volume markets, more individualized products have been emerging. Despite a smaller market size per product, the demands regarding process knowledge for a robust and productive bioprocess remain constant or are even increasing.

In contrast to the struggle for cost efficiency and productivity of the pharmaceutical industry, regulatory authorities are mainly motivated by safety concerns. High and constant product quality as well as process reproducibility [12] are the superimposed demands which pharmaceutical industry has to meet. In return, regulatory authorities reward sound science-based bioprocess development and bioprocess understanding [13, 14] with greater manufacturing flexibility. Process understanding is commonly demonstrated at hand of mechanistic process knowledge i.e. information regarding the physiological basis of observed interrelations of process parameters and productivity/product quality. Summarizing, in an environment of decreasing market size for single products and intensifying competition, growing demands regarding process knowledge have to be met by manufacturers. This situation explains the crucial role of bioprocess development for pharmaceutical industry.

Setup, host and product are the higher level variables common to every bioprocess and are usually defined prior to entering the bioprocess development phase. For the sake of cost efficiency, production units are commonly designed as host specific, multipurpose (different products) units rather than product dedicated. *Escherichia coli* is one of the most industrially exploited procaryotic production hosts for heterologous protein production [1, 15]. This is owned to *E.coli* inherent attributes of fast growth, simplistic genetic engineering, inexpensive media and the possibility of high cell density cultivations.

Background

For a novel drug, a product specific strain is generated and selected during strain engineering. Subsequently, the bioprocess development phase is entered. The main characteristic of early bioprocess development is commonly the lack of strain specific prior knowledge as e.g. biomass yield and physiological capacities. Within bioprocess development, process parameters are investigated as factors using a design of experiment approach (DoE). Prerequisite for the investigation is hereby the ability to control or to adjust the respective factor to a discrete level e.g. medium recipe, pH, temperature, substrate

supply [16]. Hereby, the respective factors are investigated regarding the response of productivity and product quality. Factor selection is commonly based on a risk assessment, which is merely of qualitative nature and commonly fully based on theoretical expert knowledge [17]. Using lab scale experiments within screening DoEs, the investigated factors are subsequently qualitatively categorized into critical and non-critical process parameters regarding their impact on productivity and product quality [18]. To optimize productivity, the quantitative relationship of process parameters and target responses is established by conducting additional experiments. Post lab scale process development, the process is up-scaled to industrial relevant scales.
In bioprocess development, laborious experiments are the only source of information and consequently knowledge, necessary to satisfy industrial as well as regulatory demands regarding a specific product. These experiments are regarded as the main cost drivers in bioprocess development, since they evoke substantial investments in terms of time, equipment and human resources. Consequently, the reduction of experiments necessary for process development is of great interest, for the sake of cost efficiency and a competitive edge.

While the regulatory demands are high product quality and reproducibility, the inherent demand of pharmaceutical industry is cost efficiency. Aligning these demands appears to constitute a contradiction. But the demand for increased process understanding might potentially turn out as synergistic in respect of the struggle for cost efficiency even despite the admittedly scarce reports on the benefits of mechanistic knowledge [19, 20]. The elucidation of the root cause of an observed phenomenon (mechanistic knowledge [19, 20]) generates a more general validity/scope than the mere correlation of process variables (technological knowledge [21, 22]), regardless of the field of science.

Theoretically, the amount of information necessary to satisfy industrial and regulatory demands is independent of the higher level variables (host, setup, product). Consequently, the number of necessary experiments can be reduced by increasing the amount of available prior knowledge. To do so, the respective knowledge and conclusions derived from bioprocess development need to be transferable. The independency of one of the higher level variables of a bioprocess is regarded as the main characteristic of transferability of process knowledge. Transferability can consequently be achieved at different levels. Transferability of process knowledge in-between products is the most sought for type of transferability. Only product independent knowledge adds to the amount of prior knowledge which can help to reduce the amount of effort for process development of a consecutive product. Nevertheless, even transferable, product independent process knowledge remains dependent on the genetic background of the host. Consequently, the host is usually standardized for certain product categories, as in this case *E.coli*. As further measure, industry has adopted the use of highly engineered strains as production platforms for various products [23]. Hereby, the use of a standardized genetic background facilitates (e.g. BL21, K12) an increasingly predictable physiological behavior.
Transferability of process knowledge in-between setups (up-scaling) can be regarded a pre-requisite. This transferability of knowledge in between scales is commonly congruent with the demand of scalability of processes and correlated control approaches. While setups change repeatedly during up-scaling throughout process development, the production strain remains constant. Rather than the reactor, the actual producers of the product in

bioprocesses are the cells, since product formation is generally regarded as physiological process. Consequently, in favor of transferability it appears logical to focus on the physiology of the cells during bioprocess development, rather than on setup specific technological variables. For the sake of information to effort ratio a shift from control of technology to control of physiology would be indicated, yet an alternative route for process development is needed. Physiologic process development targets the elucidation of physiologic interrelations using physiologic process control. Hereby, physiologic process control shall be defined as the control of biomass specific variables such as the specific substrate uptake rate, for the sake of robustness, preferably in real time.

Regarding the aim of decreasing the amount of necessary effort for bioprocess development this thesis investigates the hypothesis that physiological bioprocess development grants a greater information to effort ratio than conventional approaches.

Fundamental, physiological interrelations appear to feature a greater general scope and consequently transferability in-between products than e.g. the finding of a positive correlation of productivity and volumetric substrate supply. In this context, from a theoretical point of view the adoption of physiologic process development appears self-evident. Nevertheless, conventional technological process development still mainly focuses on the control of technological variables as the substrate feed rate [22] into the reactor system instead of focusing on the substrate uptake into the cells. Although a substantial amount of literature on physiologic bioprocess development approaches is available, reports of industrial adoption of physiological bioprocess development remain scarce.

Instead, in an industrial environment, development approaches remain focused on technical variables [24]. The hesitant industrial adoption of physiological process development may be attributed to perceived gaps in (1) analytics for physiological process development and (2) evidence of feasibility of physiological control.

Physiologic bioprocess development relies on the physiological variables which commonly comprise more data than technological variables (e.g. biomass concentration). In this context error propagation of individual input data amplifies the impact of measurement noise. Consequently, the demands regarding the sensitivity and robustness of analytical methods for physiological bioprocess development are substantial. In this context the challenge for bioprocess development is the establishment of sensitive analytical methods. This applies for process and product related responses (1a) as well as for real time estimation of e.g. biomass concentration (1b). While a lot of effort is commonly invested into timely resolved measurement e.g. by automation efforts, hardly any effort is invested into (re)assessment and de-novo establishment of analytical methods.

Based on sensitive analytical methods, physiological process control becomes feasible, Subsequently generated large data sets renders physiological process evaluation is challenging (2a) and requires a clear roadmap to standardize data evaluation. Hereupon, following the same process development routine for different products can facilitate an assessment of the transferability of physiologic findings (2b) between different products. Additionally, being able to evaluate processes irrespective of their control strategies, the standardized data evaluation facilitates the direct comparison of technological and physiological control strategies (2c). Any type of DoE requires a definition of boundaries, which can be of technological as well as physiological nature. In this context, especially the quantification of the physiological feasible space, which is limited by physiological

capacities, is challenging and laborious (2d). Commonly physiological process development is perceived to be correlated to substantial additional effort for process control. These methods are commonly perceived as hardly robust but highly complex. The consequently unclear information to effort ratio (2e) appears to one of the main obstacles for the adoption of physiological bioprocess development approaches in industry.

To overcome these hurdles, the goal of this thesis is the comprehensive discussion of the benefit and challenges of physiological bioprocess development using exemplary industrial processes.

Goals

(1) Analytics for physiological process development

a) Sensitive at-line analytics – product/process related

b) Real time quantification of biomass

(2) Feasibility of physiological control

a) Workflow for physiological process evaluation

b) Transferability of physiologic findings

c) Benefit of physiological process control

d) Definition of physiologic capacities/boundaries for DoE

e) Information to effort ratio

Roadmap

(1) Analytics for physiological process development

Physiological bioprocess development requires sensitive and accurate analytical methods as foundation for subsequent physiological conclusions. In addition to analytical methods for response evaluation, real time analysis is necessary to enable physiological control strategies. But physiological control strategies are often correlated to a substantial increase in effort compared to controlling a technical parameter as e.g. the volumetric feed rate. Consequently, analytical methods are re-assessed, verified or established to quantify and detect physiological phenomena as protein release, protein aggregation and biomass growth.

(2) Physiological process control

To increase productivity in microbial bioprocesses one of the most promising factors, controlled and frequently investigated, is the substrate feed rate [25-27, 29-31]. In the context of physiological bioprocess control a lot of effort has been invested in order to develop generic control approaches to control the specific growth rate [3, 12, 32-34] as well as the specific substrate uptake [35-37] via the feeding rate. Nevertheless, comprehensive studies comparing technological to real time controlled physiological approaches with respect to productivity to this date missing in literature. To investigate the benefits and limitations of physiological bioprocess development, the physiological approach is compared to the state of the art process development routine in terms of process phase definition, workflow, information to effort ratio, physiological constraints and control approaches.

To establish and to investigate the benefits/challenges of physiological process development this thesis is structured into the following subsections according to Figure 1.

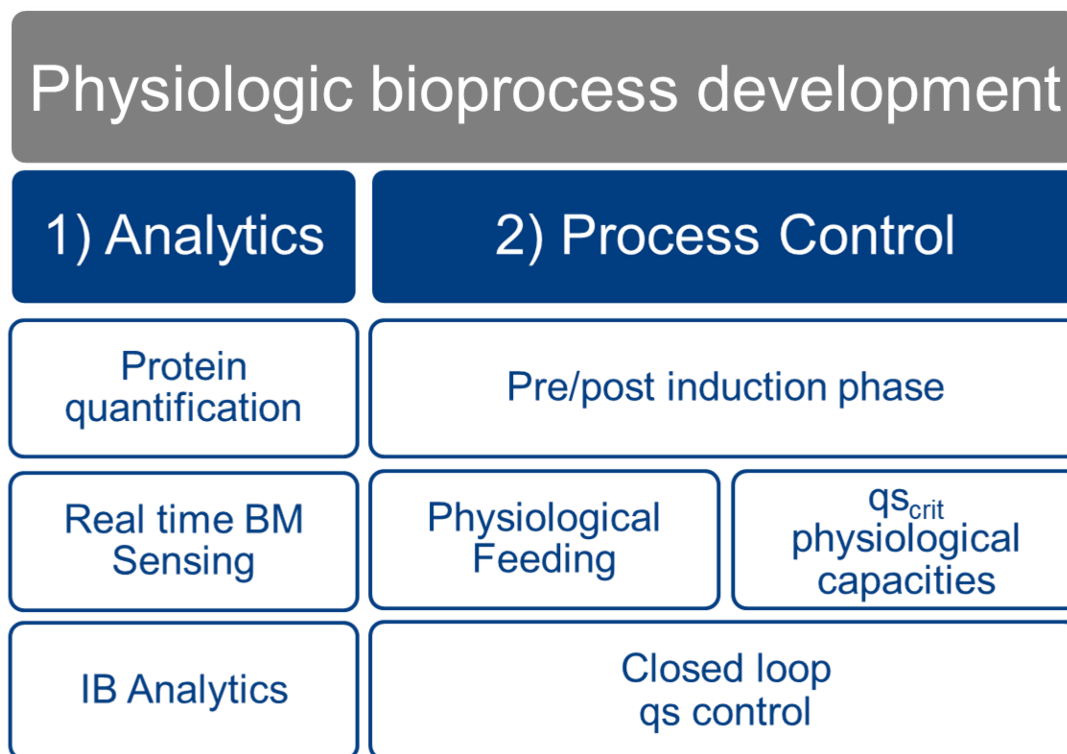


Figure 1: General structure of the thesis to assess physiological bioprocess development: 1) Analytics of product and process related variables, 2) Process control in terms of data evaluation, the impact of physiological feeding on productivity, the nature of physiological capacities and a control strategy independent of fixed strain specific variables besides the biomass composition; Discussion these aspects the challenges and the benefits of physiological bioprocess development shall be illustrated.

Results and discussion

1) Analytics for physiological process development

With limited exceptions [38], available literature in bioprocess development hardly ever investigates the physiological reason of the observed phenomena e.g. the strong impact of substrate supply on productivity [33]. This might be owned to a lack of sensitivity of analytical methods, utilized to detect and to quantify physiology and correlated phenomena. As outlined in the introduction, physiological bioprocess development relies on a greater amount of input data. In turn this makes the calculation of physiological variables increasingly sensitive to measurement noise since the noise is being amplified by error propagation. Within this section we address the respective challenges (Figure 1). The necessity for sensitive analytics applies for sensing major physiological events as protein secretion as well as for the formation of protein aggregates (inclusion bodies) as response to protein overexpression or simple biomass growth. Interestingly, although various methods are available, hardly any contribution assesses the sensitivity of published methods within the specific area of application. In this context comprehensive revision of analytical methods might potentially comprise the key to a sensitivity increase enabling to novel physiological or process relevant findings.

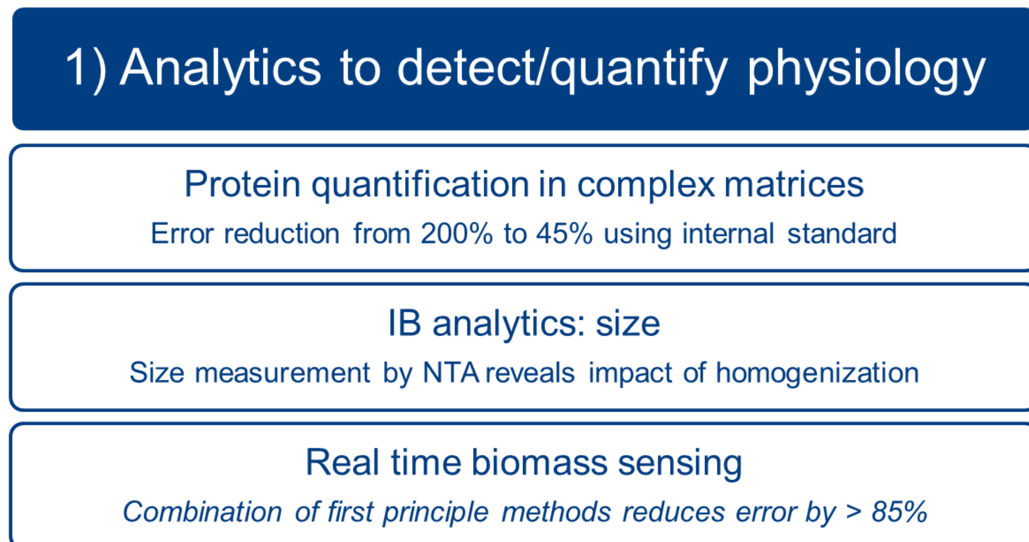


Figure 1: Graphical representation of the structure of the analytical section: Analytical challenges as in protein quantification, real time biomass estimation and inclusion body analytics are investigated and discussed.

i. Protein quantification in complex sample matrices

In biotechnology and numerous other scientific areas, a precise measurement of the protein concentration is of great relevance (Walker, 1994).Especially in the recombinant production of biopharmaceuticals and other high-value added compounds, the total protein concentration serves as a key variable for process development and quality control purposes (Han et al., 2003; Jazini and Herwig, 2013, 2011). The total protein release into the culture supernatant can not only give a direct estimate of productivity in case of secreted proteins but also provides valuable information on the overall cell physiology during a bioprocess. This is owned to the putative source of soluble protein in the supernatant: cellular lysis or protein secretion. In both cases the onset of these major physiological events constitutes valuable information targeting sound science based understanding of the impact of feeding strategies on physiology. Consequently, we challenged the state of the art approach of wet chemical protein quantification in complex sample matrices and were able to reduce the analytical error significantly.

Despite substantial efforts to identify and remove interfering substances (Schoel et al., 1995; Morton and Evans, 1992; Brown et al., 1989), complex sample matrices still greatly impair to the commonly used assays for total protein quantification. In contrast to existing literature the use of system relevant sample matrix instead of artificial sample matrix facilitated the identification of significant measurement bias. This finding highlights the sensitivity of underlying hypothesis if analytical assays are tested with artificial matrices – all occurring substances in the sample matrix have to be identified *a priori*. Using the sample matrix directly derived from the analytical area circumvents this threat. Unfortunately, the biasing influence of media components is often neglected in bioprocess monitoring and appropriate controls are not included.

J Ind Microbiol Biotechnol
DOI 10.1007/s10295-016-1796-9

CrossMark

FERMENTATION, CELL CULTURE AND BIOENGINEERING

SIMB
Society for Industrial Microbiology and Biotechnology

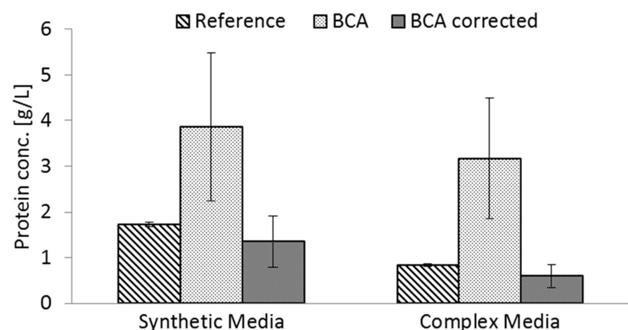
Bioprocess monitoring: minimizing sample matrix effects for total protein quantification with bicinchoninic acid assay

Wieland N. Reichelt[1] · Daniel Waldschitz[1] · Christoph Herwig[1,2] · Lukas Neutsch[2]

Received: 1 March 2016 / Accepted: 31 May 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Determining total protein content is a routine operation in many laboratories. Despite substantial work on assay optimization interferences, the widely used bicinchoninic acid (BCA) assay remains widely recognized for its robustness. Especially in the field of bioprocess engineering the inaccuracy caused by interfering substances remains hardly predictable and not well understood. Since the introduction of the assay, sample pre-treatment by trichloroacetic acid (TCA) precipitation has been indicated as necessary and sufficient to minimize interferences. However, the sample matrix in cultivation media is not only highly complex but also dynamically changing over process time in terms of qualitative and quantitative composition. A significant misestimation of the total protein concentration of bioprocess samples is often observed when following standard work-up schemes such as TCA precipitation, indicating that this step alone is not an adequate means to avoid measurement bias. Here, we propose a modification of the BCA assay, which is less influenced by sample complexity. The dynamically changing sample matrix composition of bioprocessing samples impairs the conventional approach of compensating for interfering substances via a static offset. Hence, we evaluated the use of a correction factor based on an internal spike measurement for the respective samples. Using protein spikes, the accuracy of the BCA protein quantification could be improved fivefold, taking the BCA protein quantification to a level of accuracy comparable to other, more expensive methods. This will allow reducing expensive iterations in bioprocess development to due inaccurate total protein analytics.

Graphical abstract



Keywords Total protein measurement · Bioprocess analytics · BCA measurement in complex sample matrix · BCA assay interference · TCA protein precipitation

Electronic supplementary material The online version of this article (doi:10.1007/s10295-016-1796-9) contains supplementary material, which is available to authorized users.

✉ Christoph Herwig
christoph.herwig@tuwien.ac.at

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

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

Introduction

In biotechnology and numerous other scientific areas, a precise measurement of the protein concentration is of great relevance [44]. Especially in the field of recombinant production of biopharmaceuticals and other high-value added compounds, the total protein concentration serves as a key variable for process development and quality control

Published online: 17 June 2016

purposes [16–18]. The total protein release into the culture supernatant can not only give a direct estimate of productivity in case of secreted proteins but also provides valuable information on the overall cell physiology during a bioprocess. As is the case for other critical variables, the strive for deeper bioprocess understanding calls for analytical methods that are accurate, sensitive, robust and cost efficient. Despite substantial efforts to identify and remove interfering substances, complex sample matrices still put limitations on the commonly used assays for total protein quantification. Unfortunately, the biasing influence of medium components is often neglected in bioprocess monitoring and appropriate controls are not included.

Commonly, two approaches for total protein quantification are employed: non-colorimetric or colorimetric assays. Non-colorimetric measurements of the protein concentration, e.g. amino acid analysis [38], size exclusion chromatography or mass spectrometry are usually linked to a high instrumental expense and effort in preparatory work [14]. The high protein specificity of the latter methods is advantageous for target protein quantification. However, the same, high sensitivity towards different types of proteins and the associated need to use appropriate standards are significant drawbacks in the context of total protein quantification. In case of methods that are less specific, e.g. UV/Vis-based platform SoloVPE [30] instrumental advances have allowed for an increase in sensitivity and decrease of the sample volume. Nevertheless, protein quantification via UV/VIS absorption is often hindered by sample matrices containing unsaturated fatty acids [47]. In colorimetric assays, lab-on-a-chip systems [2, 14, 35] have led to substantial progress in terms of sensitivity and reproducibility. These techniques combine a chromatographic separation phase to the colorimetric detection step, leading to good resolution and sensitive quantification. However, the chromatographic separation step has to be specifically adapted to the sample matrix. In case of bioprocess samples this sample matrix can be subjected to dynamical changes over process time. This requires case-by-case adaptations of the chromatographic separation step and makes total protein quantification via such systems tedious. Additionally, owing to the need for advanced microfluidics in the chip technology, these assays are linked to substantial investments and higher consumable expenses as compared to conventional colorimetric assays.

Wet-chemical assays are more cost efficient and, although involving several handling and preparation steps, usually allow for high-throughput analysis. The underlying principle of a more or less uniform protein staining, based merely on amino acid residues, is an advantage in the context of total protein quantification. In combination with their simplicity the latter characteristics are the reason for the wide usage of these wet-chemical assays for total protein quantification [6, 13, 29, 32, 47]. Bradford, Lowry and the Bicinchoninic acid (BCA) assay are the most commonly used colorimetric assays. Especially in microbial bioprocesses the composition of the supernatant sample usually becomes increasingly complex throughout the fermentation time course, mainly due to a gradual accumulation of sugars, phospholipids, DNA and salts. Considerable research effort has been devoted to the direct comparison of the available colorimetric assays, leading to some general recommendations regarding assay usage [11, 22, 29, 37].

The Bradford or Coomassie Blue assay is based on a residue-specific stain, first described by Bradford [4]. Via hydrophobic interactions, Coomassie Brilliant Blue G-250 [11] binds to arginine, histidine, phenylalanine, tryptophan and tyrosine residues [8] at acidic pH. Disadvantages of this assay include sensitivity to different reagent formulations [33] as well as the high sensitivity to varying amino acid composition [8]. This sensitivity to the amino acid distribution renders the method less applicable for the generic quantification of the total protein content in biotechnology.

The Lowry assay is based on a two-step chemical reaction: first, a reduction of cupric ions to cuprous ions under alkaline conditions, and second, a reduction of protein residues [24]. This reduction is followed by a reaction with the Folin–Ciocalteu reagent, resulting in a blue complex absorbing at 750 nm [31]. Since the color formation is not only induced by cuprous ions, but also by chromophoric amino acids such as tyrosine, tryptophan, phenylalanine [48] as well as cysteine residues [10], differences in the content of the various amino acids can cause high protein-to-protein variation. Nonionic detergents have been reported to form a precipitate with the Folin–Ciocalteu reagent and the use of anionic detergents such as sodium dodecylsulphate (SDS) or sodium deoxycholate (DOC) has been proposed to counteract this problem [9]. Adopting the use of DOC without further investigation, a precipitation step has been found beneficial for the removal of interfering substances from artificial samples [3]. More advanced modifications of the Lowry assay have been developed to improve robustness against interfering substances, as well as linear range. Nevertheless, the assay is still being outmatched by the BCA assay in terms of linear range and sensitivity [6].

In the BCA assay, the Folin–Ciocalteu reagent is replaced with bicinchoninic acid as described by Smith [40]. Unlike the Bradford and the Lowry, the BCA assay features a relatively small protein-to-protein variation [11, 28], making it the most suitable assay for total protein quantification. As described in literature [12, 22, 44], the BCA assay is the best choice for samples with undefined protein content in the presence of detergents. Even in combination with a DOC-TCA precipitation step, the Lowry assay is outmatched by the standard BCA assay in terms

of robustness towards interfering detergents [34]. This is of particular relevance when analyzing the supernatant of culture medium, which often contains significant amounts of biological (e.g. DNA and phospholipids of cellular origin) and synthetic surface-active compounds, e.g. nonionic detergents such as antifoam additives.

Several other substances are known to cause interference with the BCA measurement in bioprocess samples, including medium components like ethylenediaminetetraacetic acid (EDTA) [45], reducing sugars [5, 43] like fructose or lactose [34] and metabolites as phospholipids [19] or biogenic amines. Already the work of Smith [40] highlighted the need to implement proper controls and, if necessary, pre-treatment steps to avoid interference. Efforts have been undertaken to remove interfering substances, e.g. by precipitation with TCA [5, 26, 39]. Hereby, DOC has occasionally been used in combination with TCA [5, 39] referring to work based on the Folin–Ciocalteu reagent [3]. However, to our knowledge, up to now no statistical significant data has been provided in literature indicating the benefit of the additional use of DOC compared to the mere TCA precipitation in the context of the BCA assay. To account for interference of the sample matrix, countermeasures have been described which aim at the identification of the interfering substances [27]. Once identified, the components can be accounted for during calibration. Unfortunately, bioprocess samples often are subjected to unpredictable changes in the amount and nature of the interfering components, leading to substantial bias. Despite multiple accounts in literature pointing out the risk and impact of sample matrix interference for the BCA assay, many researchers are too confident regarding the universal applicability of this long-established standard procedure [18, 20, 25, 36].

This work presents an application-oriented re-assessment of the BCA assay as the current state-of-the-art method in bioprocess protein quantification. We demonstrate the substantial bias caused in total protein quantification when following standard protocols over the course of typical fed batch cultivations and demonstrate how simple adjustments to the method can lead to remarkable improvements in measurement accuracy,

Materials and methods

Media

One industrially relevant complex and one synthetic culture medium were tested in a typical fed batch bioprocess [23]. *Escherichia coli* was cultivated at controlled pH (7), DO_2 (>30 %) and temperature (30 °C) to high cell density (biomass concentration >40 g/L). In the complex medium the *E. coli* strain K12 was grown with glycerol as C-source, producing a Fab antibody as soluble intracellular protein (~24 kDa) throughout an induction phase of 48 h. The complex medium was based on the formulation given in Wilms et al. [46], supplemented with complex medium components. In the synthetic medium, based on the formulation of Korz et al. [21] the *E. coli* strain BL21 DE3 was grown on glucose as a C-source. During induction phase an intracellular protein (~30 kDa) was expressed which led to the formation of inclusion bodies.

Samples

Time-resolved fermentation samples were taken throughout induction phase of the experiments and labeled from A-I. The samples were cleared from cells and other debris (10,000 rpm; 10 min, 4 °C). The clear supernatant served as sample for further investigation and was stored at −20 °C.

Trichloroacetic acid (TCA) precipitation

Prior to protein quantification by BCA assay, the protein was isolated via TCA precipitation [42]. 500 µL of 10 % TCA solution (Carl Roth, Austria, 8789) in MilliQ was added to 500 µl of sample. After 10 min incubation on ice the samples were centrifuged (10,000 rpm; 10 min, 4 °C). Subsequently, the supernatant was discarded and the pellet re-dissolved in 1 mL of the reference sample buffer 0.1 M NaOH/1 % SDS (NaOH/SDS).

BCA assay

Using a commercial BCA assay kit (Sigma, Austria, B9643) assay according to [40] the samples were incubated at 60 °C for 15 min to ensure the lowest protein-to-protein variations. After incubation, the samples were equilibrated for 10 min at room temperature prior to absorbance measurement within the linear range from 0.1 to 0.7 relative absorption units (rAU). The correlation between signal and protein concentration was established based on a separate calibration from 0.05 to 1 g/L BSA in NaOH/SDS. The limit of detection (LOD) was determined at 0.2 g/L.

Protein spiking

In contrast to “uncorrected” native samples the “spiked” samples were spiked with bovine serum albumin (BSA) (Carl Roth, Austria, 3737, >98 % purity, IgG and protease free) in a concentration range of 0–10 g/L. Two different spike levels were used to correct for matrix effects in the BCA assay (detailed below). In brief, each sample was diluted 1:1 with a BSA stock solution (1000 or 500 µg/mL) after the TCA precipitation step. Four different sample dilutions (in NaOH/SDS 1:4, 1:8, 1:16 and 1:32) were analyzed

in each run. The protein concentrations were calculated from the mean values of the repetitive measurements. The number of replicates is indicated in each section. The quotient of measured and theoretic spike concentration was calculated to serve as a correction factor. For quantitative evaluation of the method accuracy, the relative deviation of corrected/uncorrected protein concentrations from the reference protein concentration was determined. Reference protein quantification is described below ("Quantification of total nitrogen").

Efficiency assessment of the TCA precipitation step

The BSA standard used for the spikes was supplemented (1:100) with FITC-labeled fBSA (Sigma, Austria, A9771). Fluorescence was measured with an Infinite M200 plate reader (Tecan Group Ltd) in a dilution of the sample 1:10 with NaOH/SDS in 96 multiwell plates (M&B Stricker, Germany, GRE-655101) with an excitation wavelength of 485 nm and an emission wavelength of 525 nm. The fluorescence signal of the samples before precipitation was compared to the fluorescence signal of the precipitated and re-suspended sample.

BCA assay corrected with one spike level

All samples were spiked with 500 µg/mL BSA. The average was calculated from triplicate measurements within the linear range. To account for the effect of matrix components, the measured protein concentration of the unspiked samples was subtracted from the measured protein concentration of the spiked samples to determine the contribution of the added spike (Eq. 1). The quotient of theoretic and measured spike concentration served as correction factor (Eq. 2) of the measured protein concentration of each sample (Eq. 3).

BCA assay corrected with two spike levels

All samples were spiked (see "Protein spiking") separately with 250 and 500 µg/mL BSA. All measurements (incl. dilutions) were measured in triplicates. The correction factor k corresponds to the slope of the correlation of measured and theoretic concentrations of 0/250/500 µg/mL spikes. k was calculated separately for each sample and for each dilution (Eq. 2). Finally, the mean of the corrected protein concentration calculated over all dilutions within the linear range, yielded the final protein concentration.

$$C_{sm} = C_{ps} - C_p \tag{1}$$

The measured spike concentration (c_{sm}) is derived from the measured protein concentration of the spiked sample (c_{ps}) and the measured protein concentration of the unspiked sample (c_p).

$$\frac{C_{st}}{C_{sm}} = k \tag{2}$$

Accounting for matrix effects with two spike levels. The theoretic spike concentration (c_{st}) correlates to the measured spike concentration (c_{sm}) by the factor (k). In case of one spike (k) is simple a proportionality factor. In case of two spikes (k) corresponds to the slope of the correlation (0/250/500 µg/mL BSA) of c_{st} and c_{sm} for the utilized spike concentrations.

$$C_{pc} = C_p \times k \tag{3}$$

The corrected protein concentration (c_{pc}) is calculated from the measured protein concentration of the unspiked sample (c_p) and the correction factor (k).

Quantification of total nitrogen (TN)

For verification purposes, measurements of the total nitrogen bound (TN) were conducted. The total nitrogen content was quantified by an adapted method based on peroxodisulfate oxidation of nitrogen compounds in water to nitrate, with consequent detection with copperized cadmium according to DIN EN ISO 11905-1 (Technical Committee ISO/TC 147 [41]. Samples were pre-diluted to an approximate concentration of 5–50.00 mg/mL total nitrogen. The LOD of the method was determined at 5.27 mg/L total nitrogen. Data below the LOD were set to 0 mg/L. According to a calibration (Supplemental 1) with BSA as standard protein the total protein content of the sample was calculated based on the TN content of each sample.

Statistical data analysis

Data were subjected to statistical analysis (2 sample F test, 2 sample t test, Welch test) Datalab Version 3.5 (distributed by Epina http://datalab.epina.at/). Based on an $\alpha = 0.05$ the significance of the correlation was evaluated at hand of the p value.

Results

BCA-based protein quantification is significantly impacted by sample matrix composition

To demonstrate the impact of sample matrix interference, a dilution row of bovine serum albumin (BSA) was prepared in reference buffer NaOH/SDS (Fig. 1a) and in fermentation supernatant (Fig. 1b). Measuring the concentrations of BSA in the background of NaOH/SDS via the BCA assay yielded highly accurate results. This confirms the general capability of the BCA method to quantify total protein with

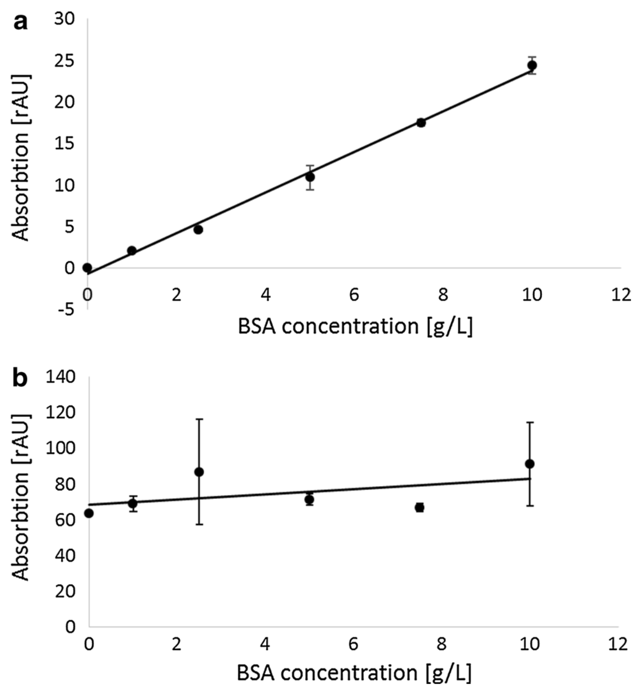


Fig. 1 Protein quantification by BCA assay is highly sensitive to sample matrix composition; protein quantification of BSA dilution rows in reference buffer versus spent synthetic culture medium as sample matrix. The standard deviations for each sample ($n = 5$) are indicated as *whiskers*; **a** dilution row of BSA measured in the background NaOH/SDS yielding a R^2 of 0.995 and a mean relative standard deviation of 5.43 %. **b** Dilution row of BSA standards measured in synthetic culture medium yielding a R^2 of 0.255 and a mean relative standard deviation of 12.6 %. In fermentation medium the 10 g/L spike signal is not significantly larger than the 1 g/L spike level (Welch test: $p(t) = 0.0796$)

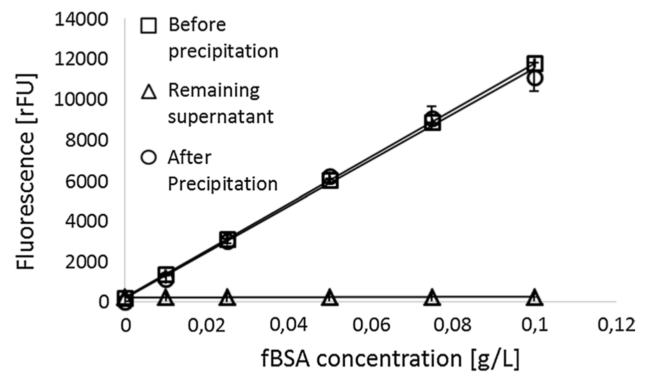


Fig. 2 Protein precipitation by TCA is quantitative; fluorescence measurements of BSA dilution rows of TCA-precipitated, synthetic process media supplemented with fBSA are displayed. All samples were measured in quadruplicates ($n = 4$); the standard deviations are indicated as *whiskers*. 99.7–97.5 % of the added fBSA was recovered in the reference buffer (after precipitation), and only 0.3–2.5 % of the initial fluorescence was found in the supernatant (remaining supernatant). The fluorescence intensity before precipitation (before precipitation) and after precipitation (after precipitation) correlated with the nominal concentration of the stock solution ($R^2 > 0.99$). The fluorescence intensity found in the supernatant is almost negligible and not correlated with the spike concentration ($R^2 = 0.24$)

high reproducibility under ideal conditions. However, if synthetic *E. coli* culture medium from actual process samples was used as matrix, it was not possible to resolve differences in protein concentration up to 10 g/L.

Quantitative protein precipitation by TCA

The lack in sensitivity of the BCA protein quantification method in complex sample matrixes (Fig. 1) may be improved by removal of the interfering substances and error compensation. The basic aim of introducing a precipitation step is to remove interfering substances from the sample matrix. While interfering substances should be retained in the supernatant, protein shall be quantitatively precipitated in the pellet. Commonly, such matrix replacement is performed by TCA precipitation, followed by re-suspension in a defined reference buffer such as NaOH/SDS. However, to reliably exclude measurement bias caused by the precipitation step itself, the efficiency of the TCA precipitation procedure first has to be evaluated. To this end, a dilution row of BSA in fermentation supernatant was additionally supplemented with a defined amount of fBSA, which allows for identifying potential protein loss during the workup procedure via direct fluorescence readout. Samples were precipitated, the pellets dissolved in fresh buffer and all resulting fractions were analyzed for fluorescence intensity (Supplemental 2). 99.7–97.5 % of the added fBSA was recovered after precipitation in the reference buffer (Fig. 2) from culture medium supernatant 48 h post-induction. Native medium and medium 24 h post-induction gave the same results (data not shown). Based on the results it can be concluded that the precipitation of protein by TCA is highly efficient and unlikely to cause significant measurement bias due to uncontrolled loss of protein.

Interfering substances accumulate in the culture medium

After substantiating the quantitative precipitation of protein by TCA (Fig. 2) the potential origin of the observed interference on BCA assay readout was investigated in further detail. In principle interfering substances may originate from the biomass or be contained in the medium formulation. To elucidate the basic root cause of the interference, the supernatant of precipitated culture medium from 0/24/48 h after induction was subjected to total protein determination. Equal dilution rows of BSA were prepared in the supernatant of TCA-precipitated culture medium. As shown in Fig. 3, no interfering substances seemed to accumulate in the supernatant over the first 24 h of the process,

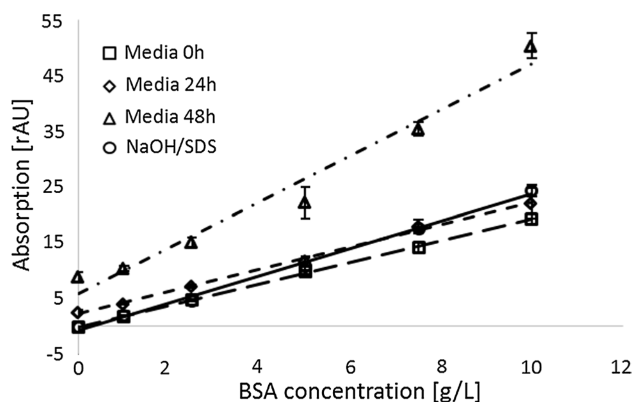


Fig. 3 The impact of interferences increases over process time and traces back to cell-related processes: BSA spikes (concentrations 0–10 g/L) were added to precipitated fermentation samples obtained from different time points of a cultivation performed in synthetic medium. Comparison to BSA concentrations measured in reference buffer NaOH/SDS. For late time points (48 h) of the fermentation, the correlation of signal-to-protein is altered substantially. All samples were measured in quadruplicates ($n = 4$); the standard deviations are indicated as *whiskers*

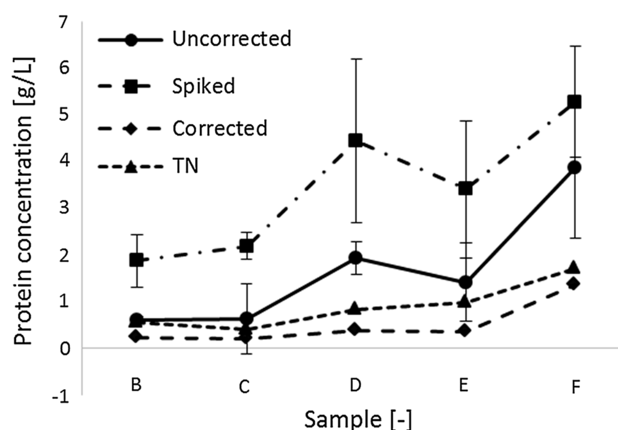


Fig. 4 Correction of protein determination based on spike addition leads to an increase in accuracy: samples from consecutive time points during the fermentation in synthetic medium between 0 and 24 h after induction (*B*–*F*). All measurements were performed after TCA precipitation. *uncorrected* measured protein concentration of native samples; *spiked* measured protein concentration of samples with spike (500 µg/mL); *TN* measured reference protein concentration derived from TN based protein quantification; *corrected* calculated protein concentrations calculated according to Eq. 3. *Lines* between measurement points have been included to ease orientation. The relative differences of the corrected protein concentration from the TN derived protein concentrations are significantly smaller than the respective relative differences of the uncorrected concentrations [$p(t) = 0.008$]. The relative standard deviation of the respective differences is for the corrected values (16 %) significantly [$p(F) = 0.004$] smaller than of the relative uncorrected protein concentration (85 %). BCA protein quantification was performed in triplicates ($n = 3$); the mean values were used for calculation. The standard deviation is indicated as *whiskers*

since the signal-to-protein correlation was not significantly altered in comparison to the reference buffer. However, a clear change in signal correlation became visible after a process time of 48 h. Regardless of the identity of the interfering substances present in spent culture medium, their persistent biasing effect has to be accounted for.

TCA precipitation alone is not sufficient to avoid interference

A constant impact of the interfering substances, without changes over process time, would permit straightforward correction of the BCA measurements via a given, predefined factor. It was thus important to check in how far the distortion of the signal-to-protein ratio changes over process time (*B*–*F*). The BCA analysis was compared to total nitrogen measurements (TN) as an orthogonal method for protein quantification (Fig. 4). However, the correlation between the uncorrected protein concentration obtained via the BCA measurement and the protein concentration determined via TN changed over process time, substantiating the need for a sample-specific compensation strategy. The direct comparison of the uncorrected protein concentrations derived from TCA-precipitated samples to the reference protein concentration (TN) yielded an enormous average relative deviation of 212 % (Figs. 4, 5, 6).

It can thus be concluded that TCA precipitation only is an insufficient strategy to avoid sample matrix interferences in bioprocess analysis and additional corrective actions are required to avoid misestimation of total protein content. By individual spiking of each sample the process time-dependent impact of matrix components on TCA-precipitated samples can be corrected (Fig. 4). Despite overcompensation, the correction led to a substantial increase in convergence of the BCA assay derived protein concentrations and the actual protein concentration (TN).

Having established the qualitative benefit of corrections via spike addition (Fig. 4), a quantitative evaluation was the next step to conclude on the practical usability of the modified protocol. In order to prove the generic applicability, we tested the approach for two different medium formulations. Interestingly, in complex medium the apparent total protein concentration in [g/L] was found to be in average two- to threefolds higher as compared to synthetic medium (data not shown). Figure 5 displays the deviation of the uncorrected and corrected protein concentrations from the protein concentrations derived from TN measurement. By correcting the values of the unknown samples according to Eq. 3, the deviance was substantially reduced from 212 to 41 % for synthetic medium as well as for complex medium. Moreover, the method error became significantly more systematic, with the variance in deviation decreasing from 127 % to 14 % for both options.

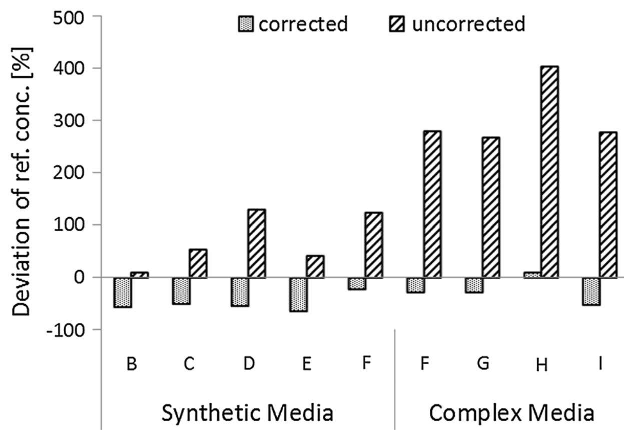


Fig. 5 Relative error of measurement is reduced from 212 to 41 % in average by the use of one spike: samples from consecutive time points during the fermentation in a complex and a synthetic culture medium. The *letters B–I* refer to different time points during the fermentation. Differences of protein concentrations derived from BCA measurements (*corrected/uncorrected*) compared to protein concentrations according to TN method are plotted on the *y* axis [deviation from ref. conc. (%)]. The relative differences of the corrected protein concentration (41 %) from the TN derived protein concentrations are significantly smaller [$p(t) = 0.0001$] than the respective relative differences of the uncorrected concentrations. The standard deviation of these respective differences is for the corrected values (14 %) significantly smaller [$p(F) = 0.0000$] than the standard deviation for the uncorrected protein concentration (127 %). All values under the LOD of the TN measurement of 5.27 mg/L were set to zero and are not displayed. All samples were measured in triplicates ($n = 3$), the mean values were used for calculation

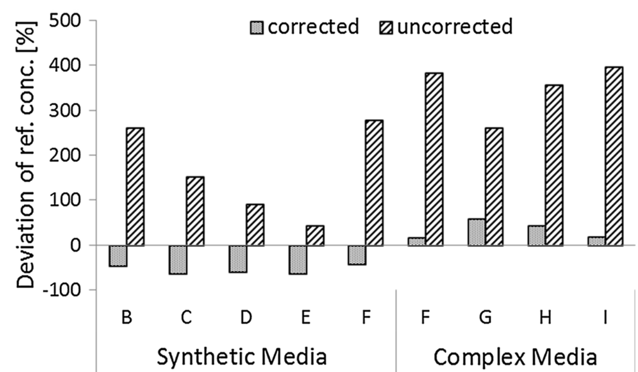


Fig. 6 Relative error of measurement is reduced from 212 to 46 % in average by the use of two spikes: Samples from consecutive time points during the fermentation in a complex and a synthetic culture medium. The *letters B–I* refer to different time points during the fermentation. Differences of protein concentrations derived from BCA measurements (*corrected/uncorrected*) compared to protein concentrations according to TN method are plotted on the *y* axis [deviation from ref. conc. (%)]. The relative differences of the corrected protein concentration (46 %) from the TN derived protein concentrations are significantly smaller than the respective relative differences of the uncorrected concentrations [$p(t) = 0.0001$]. The standard deviation of these respective differences is for the corrected values (17 %) significantly smaller [$p(F) = 0.0001$] than the standard deviation for the uncorrected protein concentration (112 %). All values under the limit of detection (LOD) of the TN measurement of 5.27 mg/L were set to zero and are not displayed. All samples were measured in triplicates ($n = 3$) and the mean values were used for calculation

The results shown in Fig. 5 led to the question whether assay accuracy could be further improved by the use of an additional spike level. The benefit of measuring two internal spike levels per sample is exemplified in Fig. 6. Using two spikes, the deviation was reduced to 45 % in respect to the uncorrected values. However, in case of the synthetic medium the relative deviations of the uncorrected protein quantification declined over time in contrast to the trajectory of the deviations for one spike level. This may be attributed to the generally low protein concentrations for the strain grown in synthetic medium. Especially samples B and C displayed protein concentrations close to the limit of detection of the BCA assay.

The overestimation of protein content for the corrected values can presumably be attributed to dilution effects, which may in this case be more severe owing to the genuinely higher protein concentrations in complex medium. In comparison to one spike level, the dynamic range of the assay did not allow the measurement of the native sample and the two different spike levels within one dilution. Two spikes yielded a variance of deviation not significantly smaller [$p(F) = 0.207$] than for one spike. Concluding, the use of two spike levels does not lead to any significant improvements in terms of measurement accuracy.

Discussion

Far too often, a widely used standard procedure like the BCA or Bradford assay is adopted in the erroneous assumption that straightforward method transfer between different applications is possible. Bioprocess samples are especially challenging in this regard. Routine biotechnological monitoring strategies typically cover a time series analysis of multiple consecutive samples over the duration of the process. A plethora of uptake and secretion and release processes related to metabolic turnover, as well as time-dependent cellular lysis can lead to substantial, yet gradually evolving changes in the chemical composition of the culture supernatant. If one or several of the changing factors happen(s) to have an impact on measurement accuracy, this biasing effect(s) too will evolve gradually without brisk steps being visible in the signal-over-time-curve. Many researchers tend to focus on the smoothness of measurement values over time as a primary indicator for data quality and hence will fail to detect such errors. The required, thorough method qualification is frequently omitted for sake of time, regardless of the important role of total protein determination in bioprocess engineering. Total protein often serves as the key variable to conclude on culture physiology, and critical decisions in process development

as well as strain screening are based on the protein data. In at-line process monitoring, the determination of the ideal point of harvest or detection of unintended cell lysis events rely on a robust method for protein quantification.

Unfortunately, even if respective controls at early and late process times are being included, a negative check for interfering substances has to be repeated as soon as any major change is brought to the process setup that may lead to a change in matrix composition. This would lead to significant complications in the usual, iterative workflow of process optimization. There is thus a substantial need for refined analytical protocols that allow for taking such factors into account, yet without increasing operator workload beyond a reasonable extent. Against this backdrop, the aim of this study was to evaluate and to illustrate the impact of matrix components on protein quantification by the means of the BCA assay, as well as the elaboration of a rapid and generally applicable method to compensate for the biasing effects.

In our studies, the original BCA assay was found incapable of resolving an addition of up to 10 g/L BSA in complex sample matrices. This substantial loss in sensitivity underlines the necessity to remove interfering substances, as it has been advised in the past [5, 8]. This shortcoming of the standard BCA assay has also been reported recently in other context [47], albeit in this case assay performance could be remarkably improved via TCA precipitation. This was not the case for the systems investigated here, as well as for several other bioprocess setups that were evaluated in our and other laboratories (personal communication to the authors). We hence speculate that a considerable number of biotechnology R&D projects will experience similar problems, often without being aware of it.

One potential cause for the failure of the TCA protocol to improve measurement consistency in case of the investigated bioprocess media could lie in a changing efficiency of the protein denaturation, precipitation or re-solubilisation step. We were, however, able to show that the loss of protein is far too low to account for the observed bias. Also a standardization of the pH value after TCA precipitation, which is a known cause for variations in the dye-protein reaction [40], was found to remain without consequences for signal quality in the present case. Several modifications and fine-adjustments of the TCA protocol, including wash or solubilization steps with pH-stabilizing reagents such as NaOH or HCl, that were successfully employed in other settings [5], evidently could not remove the source of bias in complex culture medium. It should be noted in this regard that an interplay of multiple biasing substances, may account for the observed interference, which is why wash protocols from more defined applications may fail. Also others have reported such continuing interference after TCA precipitation, but in this case acid wash let to a substantial reduction of interferences [39].

The dynamically changing impact of sample matrix components, illustrated in this work, is indicative of a highly complex matrix composition, presumably not only in terms of concentration but also regarding the chemical nature of the individual agents. Given the complete lack of knowledge on type and amount of the interfering agents, it would be risky and probably counterproductive to include time-intensive purification protocols (e.g. by dialysis) in the workup chain, as was proposed for BCA assays when applied to bioprocess monitoring [27]. In direct comparison to TN quantification as a reference method, the spike-corrected BCA measurement protocol led to a systematic underestimation of the protein concentration. Although the correction leads to an underestimation of the protein concentration, it yields more accurate and reproducible values for all tested strains in all tested media. The systematic underestimation may partially trace back to differences in both methods, regarding the sensitivity for the BSA standard, since, depending on the molecular weight, the average nitrogen content in proteins found in the culture supernatant may differ from the nitrogen content of BSA [7]. Principally, such deviations could be corrected for via an error offset that could be determined for each fermentation run by parallel analysis of chosen samples via the TN method. However, a prerequisite for this correction approach would be a constant nitrogen level in the supernatant, which may be critical especially in processes where NH4 is used for pH correction after acetate production.

Other methodological alternatives proposed for total protein quantification in fresh and complex cell culture media include fluorescence anisotropy as proposed by Groza et al. [15]. However, sensitivity of the method in the dynamic environment of microbial bioprocesses was not investigated up to now and remains to be demonstrated. In comparison to the relative standard deviation of 15–46 % in BSA protein quantification achieved with a bioanalyzer© system [1], the relative standard deviation of 41 % for process samples obtained via the herein proposed improved measurement protocol is within an attractive range. Especially with regard to the limited effort for data processing and instrumental costs, the BCA assay appears highly suited for a broad range of applications. In conclusion, the proposed method of compensation renders the BCA assay a highly cost- and time-efficient method for total protein quantification in complex sample matrices. In the context of bioprocess monitoring and development, the refined approach can be expected to help to improve existing control strategies and reduce the effort in development iterations.

Acknowledgments We gratefully acknowledge financial support by the Christian Doppler Society Austria.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Agilent Technologies Technical Note (2008) Agilent Technologies Technical Note http://www.chem-agilent.com/cimg/5989-8940EN.pdf: performance characteristics of the high sensitivity protein 250 assay for the Agilent 2100 Bioanalyzer: http://www.chem-agilent.com/cimg/5989-8940EN.pdf. Accessed 20 Sept 2015
2. Anema SG (2009) The use of "lab-on-a-chip" microfluidic SDS electrophoresis technology for the separation and quantification of milk proteins. Int Dairy J 19:198–204. doi:10.1016/j.idairyj.2008.11.001
3. Bensadoun A, Weinstein D (1976) Assay of proteins in the presence of interfering materials. Anal Biochem 70:241–250. doi:10.1016/S0003-2697(76)80064-4
4. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254
5. Brown RE, Jarvis KL, Hyland KJ (1989) Protein measurement using bicinchoninic acid: elimination of interfering substances. Anal Biochem 180:136–139. doi:10.1016/0003-2697(89)90101-2
6. Burgess RR, Deutscher MP (2009) Chapter 8: Quantitation of protein. In: Burgess RR, Deutscher MP (eds) Methods in enzymology guide to protein purification, 2nd edn. Academic Press, London, pp 73–95
7. Carugo O (2008) Amino acid composition and protein dimension. Protein Sci A Publ Protein Soc 17:2187–2191. doi:10.1110/ps.037762.108
8. de Moreno MR, Smith JF, Smith RV (1986) Mechanism studies of coomassie blue and silver staining of proteins. J Pharm Sci 75:907–911
9. Dulley JR, Grieve PA (1975) A simple technique for eliminating interference by detergents in the Lowry method of protein determination. Anal Biochem 64:136–141. doi:10.1016/0003-2697(75)90415-7
10. Everette JD, Bryant QM, Green AM, Abbey YA, Wangila GW, Walker RB (2010) Thorough study of reactivity of various compound classes toward the Folin–Ciocalteu reagent. J Agric Food Chem 58:8139–8144. doi:10.1021/jf1005935
11. Fountoulakis M, Juranville JF, Manneberg M (1992) Comparison of the Coomassie brilliant blue, bicinchoninic acid and Lowry quantitation assays, using non-glycosylated and glycosylated proteins. J Biochem Biophys Methods 24:265–274
12. Gates RE (1991) Elimination of interfering substances in the presence of detergent in the bicinchoninic acid protein assay. Anal Biochem 196:290–295. doi:10.1016/0003-2697(91)90468-9
13. Glyk A, Heinisch SL, Scheper T, Beutel S (2015) Comparison of colorimetric methods for the quantification of model proteins in aqueous two-phase systems. Anal Biochem 477:35–37. doi:10.1016/j.ab.2015.02.007
14. Goetz H, Kuschel M, Wulff T, Sauber C, Miller C, Fisher S, Woodward C (2004) Comparison of selected analytical techniques for protein sizing, quantitation and molecular weight determination. J Biochem Biophys Methods 60:281–293. doi:10.1016/j.jbbm.2004.01.007
15. Groza RC, Calvet A, Ryder AG (2014) A fluorescence anisotropy method for measuring protein concentration in complex cell culture media. Anal Chim Acta 821:54–61. doi:10.1016/j.aca.2014.03.007
16. Han M-J, Jeong KJ, Yoo J-S, Lee SY (2003) Engineering *Escherichia coli* for increased productivity of serine-rich proteins based on proteome profiling. Appl Environ Microbiol 69:5772–5781. doi:10.1128/AEM.69.10.5772-5781.2003
17. Jazini M, Herwig C (2011) Effect of post-induction substrate oscillation on recombinant alkaline phosphatase production expressed in *Escherichia coli*. J Biosci Bioeng 112:606–610. doi:10.1016/j.jbiosc.2011.08.013
18. Jazini M, Herwig C (2013) Effects of temperature shifts and oscillations on recombinant protein production expressed in *Escherichia coli*. Bioprocess Biosyst Eng 36:1571–1577. doi:10.1007/s00449-013-0927-1
19. Kessler RJ, Fanestil DD (1986) Interference by lipids in the determination of protein using bicinchoninic acid. Anal Biochem 159:138–142
20. Konstantinidis S, Heldin E, Chhatre S, Velayudhan A, Titchener-Hooker N (2012) Strategic assay deployment as a method for countering analytical bottlenecks in high throughput process development: case studies in ion exchange chromatography. Biotechnol Prog 28:1292–1302. doi:10.1002/btpr.1591
21. Korz DJ, Rinas U, Hellmuth K, Sanders EA, Deckwer WD (1995) Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. J Biotechnol 39:59–65
22. Krohn RI (2002) The colorimetric detection and quantitation of total protein. Curr Protoc Cell Biol Appendix 3:Appendix 3H. doi:10.1002/0471143030.cba03hs15
23. Lee SY (1996) High cell-density culture of *Escherichia coli*. Trends Biotechnol 14:98–105. doi:10.1016/0167-7799(96)80930-9
24. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the Folin phenol reagent. J Biol Chem 193:265–275
25. McMillan JD, Jennings EW, Mohagheghi A, Zuccarello M (2011) Comparative performance of precommercial cellulases hydrolyzing pretreated corn stover. Biotechnol Biofuels 4:29. doi:10.1186/1754-6834-4-29
26. Morton RE, Evans TA (1992) Modification of the bicinchoninic acid protein assay to eliminate lipid interference in determining lipoprotein protein content. Anal Biochem 204:332–334
27. Niazi SK, Brown JL (2015) Quality Consideration. In: Fundamentals of modern bioprocessing, CRC Press, London. https://books.google.at/books?id=wiXSCgAAQBAJNiazi. (ISBN: 9781466585737)
28. Noble JE, Knight AE, Reason AJ, Di Matola A, Bailey MJA (2007) A comparison of protein quantitation assays for biopharmaceutical applications. Mol Biotechnol 37:99–111
29. Olson BJ, Markwell J (2007) Assays for determination of protein concentration. Curr Protoc Protein Sci Chapter 3:Unit 3.4. doi:10.1002/0471140864.ps0304s48
30. Pena-Pereira F, Costas-Mora I, Romero V, Lavilla I, Bendicho C (2011) Advances in miniaturized UV-Vis spectrometric systems. TrAC Trends Anal Chem 30:1637–1648. doi:10.1016/j.trac.2011.04.018
31. Peterson GL (1979) Review of the Folin phenol protein quantitation method of lowry, rosebrough, farr and randall. Anal Biochem 100:201–220. doi:10.1016/0003-2697(79)90222-7
32. Pinto IF, Rosa SASL, Aires-Barros MR, Azevedo AM (2015) Exploring the use of heparin as a first capture step in the

purification of monoclonal antibodies from cell culture supernatants. Biochem Eng J. doi:10.1016/j.bej.2015.05.011
33. Read SM, Northcote DH (1981) Minimization of variation in the response to different proteins of the Coomassie blue G dye-binding assay for protein. Anal Biochem 116:53–64
34. Redinbaugh MG, Turley RB (1986) Adaptation of the bicinchoninic acid protein assay for use with microtiter plates and sucrose gradient fractions. Anal Biochem 153:267–271. doi:10.1016/0003-2697(86)90091-6
35. Reichmuth DS, Shepodd TJ (2005) Microchip HPLC of peptides and proteins. J Anal Chem 77:2997–3000. doi:10.1021/ac048358r
36. Routledge SJ, Hewitt CJ, Bora N, Bill RM (2011) Antifoam addition to shake flask cultures of recombinant *Pichia pastoris* increases yield. Microb Cell Factories 10:17. doi:10.1186/1475-2859-10-17
37. Sapan CV, Lundblad RL, Price NC (1999) Colorimetric protein assay techniques. Biotechnol Appl Biochem 29(Pt 2):99–108
38. Schegg KM, Denslow ND, Andersen TT, Bao Y, Cohe SA, Mahrenholz AM, Mann K (1997) Quantitation and identification of proteins by amino acid analysis: ABRF-96AAA collaborative trial. Tech Protein Chem VIII:207–216
39. Schoel B, Welzel M, Kaufmann S (1995) Quantification of protein in dilute and complex samples: modification of the bicinchoninic acid assay. J Biochem Biophys Methods 30:199–206
40. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC (1985) Measurement of protein using bicinchoninic acid. Anal Biochem 150:76–85
41. Technical Committee ISO/TC 14713.060.5013.060.50 (1997) Technical Committee ISO/TC 147: Water quality—determination of nitrogen—part 1: method using oxidative digestion with peroxodisulfate, 90.93rd ed, 13.060.50. Accessed 5 May 2015
42. Thermo Scientific (2010) TECH TIP #8 Eliminate interfering substances from samples for BCA protein assays, TR0008.3. https://tools.thermofisher.com/content/sfs/brochures/TR0008-TCA-acetone-precip.pdf. Accessed 8 Dec 2014
43. Vlasenko EY, Ryan AI, Shoemaker CF, Shoemaker SP (1998) The use of capillary viscometry, reducing end-group analysis, and size exclusion chromatography combined with multi-angle laser light scattering to characterize endo-1,4-β-D-glucanases on carboxymethylcellulose: a comparative evaluation of the three methods. Enzyme Microbial Technol 23:350–359. doi:10.1016/S0141-0229(98)00052-0
44. Walker JM (1994) Basic protein and peptide protocols, methods in molecular biology, vol 32. Humana Press, Totowa. doi:10.1385/089603268X
45. Wiechelman KJ, Braun RD, Fitzpatrick JD (1988) Investigation of the bicinchoninic acid protein assay: identification of the groups responsible for color formation. Anal Biochem 175:231–237
46. Wilms B, Hauck A, Reuss M, Syldatk C, Mattes R, Siemann M, Altenbuchner J (2001) High-cell-density fermentation for production of L-*N*-carbamoylase using an expression system based on the Escherichia coli rhaBAD promoter. Biotechnol Bioeng 73:95–103
47. Wiśniewski JR, Gaugaz FZ (2015) Fast and sensitive total protein and Peptide assays for proteomic analysis. Anal Chem 87:4110–4116. doi:10.1021/ac504689z
48. Wu AM, Wu JC, Herp A (1978) Polypeptide linkages and resulting structural features as powerful chromogenic factors in the Lowry phenol reaction. Studies on a glycoprotein containing no Lowry phenol-reactive amino acids and on its desialylated and deglycosylated products. Biochem J 175:47–51

i. Inclusion body analytics

Heterologous protein expression in the cytosol of E.coli facilitates protein titers of up to 10 g/l[24]. The high expression rate facilitates high titers but results in high amounts of unfolded protein. Thereby, high titer expression often triggers aggregation of unfolded protein as biologic nano particles also known as inclusion bodies (IB). Owned to the generally presumed close correlation of IB size to expression rates and consequently physiology, the quantification of IB size has been of interest in literature for some time [39-42]. Consequently, a lot of effort has been invested into the characterization of IBs in general and especially in the quantification of IB size [43-45]. Despite the substantial effort the analysis of IB size, described methods mostly lack verification by an orthogonal method sizing a significant amount of IBs. A final conclusion concerning the information content of IB size analysis is still not available. We have addressed this topic introducing a novel method for IB sizing, using an orthogonal verification method based on transmission electron microscopy.

Using transmission electron microscopy in combination with a highly sensitive fixation method the shape of IBs was found to be of sponge like nature rather than the dense structure of IBs propagated elsewhere in literature. Hereby, we have been able to establish nano particle tracking analysis as valid method for the analysis of IB size. Interestingly, the impact of high pressure homogenization has been identified to substantially interfere with the size of IBs. Consequently, the basic hypothesis that measuring the IB size post homogenization contains DSP relevant information had to be questioned. But the ratio of fixated and non - fixated particles might offer a measure of relative stickiness and therefore an industrial relevant analytical parameter.

Research Article

High throughput inclusion body sizing: Nano particle tracking analysis

Wieland N. Reichelt[1], Andreas Kaineder[1], Markus Brillmann[1], Lukas Neutsch[2], Haider Samir[4], Hans Lohninger[3], Christoph Herwig[1,2]

[1] Christian Doppler Laboratory for Mechanistic and Physiological Methods for Improved Bioprocesses, Institute of Chemical Engineering, Vienna University of Technology, Vienna, Austria

[2] Research Division Biochemical Engineering, Institute of Chemical Engineering, Vienna University of Technology, Vienna, Austria

[3] Research group electronic media; Institute of Chemical Technologies and Analytics; Vienna University of Technology, Vienna, Austria

[4] Division of Clinical Pharmacy and Diagnostics, University of Vienna, Vienna, Austria

Correspondence: Prof. Christoph Herwig, Institute of Chemical Engineering, Vienna University of Technology, Getreidemarkt 9/166, A-1060 Vienna, Austria

E-Mail: christoph.herwig@tuwien.ac.at

Keywords: inclusion body; inclusion body analysis; nano particle tracking analysis; electron microscopy; grey scale image segmentation

List of Abbreviations

TEM	Transmission electron microscopy
DSP	Down stream processing
USP	Up stream processing
IB	Inclusion body
NTA	Nano particle tracking analysis
DLS	Dynamic light scattering
PBS	phosphate buffered saline
BSA	bovine serum albumin
EDTA	Ethylenediaminetetraacetic acid
HPF	High pressure freezing
IPTG	Isopropyl-β-D-thiogalactopyranosid
FL	Fluorescence

Abstract

The expression of pharmaceutical relevant proteins in *E.coli* frequently triggers inclusion body (IB) formation caused by protein aggregation. In literature substantial effort has been invested into the quantification of IB size. But the lack of particle based methods to size representative numbers of IBs in combination with the lack of an orthogonal verification method impaired profound method assessment and establishment.

Using high pressure freezing and automated freeze substitution the native, cytosolic inclusion body structure was preserved for transmission electron microscopy (TEM). TEM imaging in combination with grey scale image segmentation allowed the quantification of relative areas covered by the inclusion body within the cytosol. As high throughput method nano particle tracking analysis (NTA) derives the hydrodynamic diameter of particles, based on a measurement of the Brownian motion. Using the NTA with TEM as orthogonal method we were able to illustrate that chemical fixation leads to a condensation of the native poriferous IB shape. Comparing the NTA results of fixated and native IBs it can be concluded that high pressure homogenization annihilates the native physiological shape of IBs. Nevertheless, the ratio of particle counts of native and fixated samples could potentially serve as factor for particle stickiness.

Concluding, with the image segmentation of TEM pictures we have established an orthogonal method to size biologic particles in the cytosol of cells. Moreover, NTA has been established as high throughput method for analysis of 1000-3000 particles within 20 min, facilitating a much more representative analysis than currently available methods

1. Introduction

The production of bio-similars is one of the main growth markets in pharmaceutical industry [cite]. Especially *Escherichia Coli,* as well characterized expression host, has been established as easy accessible host for fast and efficient, high titer protein production. Hereby, high titer expression of heterologous protein frequently leads to inclusion body (IB) formation. This protein aggregation either coincides with high cytosolic concentrations of unfolded protein or is induced using a protein tag in order to produce an otherwise toxic protein. But while USP is hardly affected by IB formation, DSP constitutes the bottleneck in IB related production processes [1] and causes the bigger share of the total production costs.

Integrated bioprocess development [2] targets an efficiency increase is by addressing the impact and interrelation of USP on DSP [3, 4]. This calls for sensitive response parameters describing the characteristics of IBs as intermediate product of USP and DSP. A lot of effort has been invested especially into the quantification of IB size, but until now i) the methods have hardly been challenged by an orthogonal verification method, ii) the number of analyzed IBs was comparably low iii) presented methods lacked sensitivity. Within this contribution we aim to establish an orthogonal verification method to analyze the size of an inclusion body, and we assess sensitivity and information content of the nano particle tracking analysis (NTA) as automated method to analyze large numbers of IBs.

As end product of USP, IBs are isolated from the cell during cell disruption in DSP. Industrial cell disruption is commonly conducted using high pressure homogenization [5] and a continuous centrifuge. Hereby, the continuous addition of washing buffer allows the combination of cell disruption and the removal of cellular debris. Post isolation IBs are commonly solubilized in a chaotropic solubilisation buffer prior to refolding of the protein into the native and therefore active protein conformation. In comparison to the high yield protein production during USP, DSP is challenged by comparably low and moreover variable product yields. In this context especially refolding and its efficiency has been investigated comprehensively. But besides refolding the laborious isolation process prior solubilisation might potentially play a critical role as well. Given the duration of several hours of IB release and isolation, high solubility of IBs in the washing buffer as well as a tendency to adhere to surfaces (stickiness) could cause significant product loss. Consequently, high solubility and a tendency to stickiness could directly impact process efficiency, which substantiates the necessity of a sensitive characterization of IB particle properties. To this date, a lot of effort has been invested into the challenging topic of characterizing IBs as biologic nanoparticles to provide sensitive response parameters for integrated bioprocess development.

1.1 State of the Art: Inclusion body analytics

Various methods have been investigated in order to characterize IBs, addressing different physical or chemical properties of IBs. IB purity directly affects the necessary effort for further purification post refolding and can be easily analyzed by SDS-PAGE [6]. Furthermore, IB solubility is critical for DSP performance, since highly soluble IBs would dissolve during washing steps. In contrast, hardly soluble IBs require high amounts of chaotrope reagents during solubilisation, which increases buffer volume for refolding [7]. The increased volume in turn calls for bigger column diameters of economically expensive DSP purification columns [8]. Recent developments have enabled the concise measurement of solubility in respect to time [3, 9] as well as in respect to the concentration of chaotrope reagents [10]. Size and stickiness of IBs presumably determines yields especially during IB release. But despite substantial effort, analytical methods to quantify physical properties like size and shape of IBs are up to date less developed.

1.2 Inclusion body analytics: Size

Besides sensitivity, a suitable method for the characterization of the physical properties of IBs needs to be robust and reproducible. But only if the method is sufficiently simplistic, method adoption in academic and industrial labs becomes probable. Highly sophisticated methods often lack transferability and comparability owned to a certain degree of operator dependency.

Early studies used centrifugation techniques as centrifugal disc photo sedimentation [11] or cumulative sedimentation analysis [12] but require a particle density for the calculation of a size distribution of IBs. The more recently discussed approach of using an analytical centrifuge for IB Sizing also relies on the density [13] calling for additional analytical method and sample processing which makes the method a less direct method to analyze the size of IBs.

Dynamic light scattering (DLS) has extensively been utilized to size biological nano particles [7, 10, 14-17]. Nevertheless, since this method only measures one variable its sensitivity is limited by multimodal distributions as well as by background particles [12]. Addressing this shortcoming, the samples have been purified by serial washing steps [7, 10] or combined with full-grown purification techniques as ultracentrifugation [14].

Field flow fractionation (FFF) as separation or purification technique, as described elsewhere [18], has a wide dynamic range from 0.3-100 µm of particle separation capacity. The separation mechanism is a combination of Brownian motion, sedimentation and hydrodynamic lift forces [18] and facilitates bulk separation of nano particles according to the respective size and mass. Luo et al used asymmetrical FFF in combination with multi angel light scattering in order to

analyze the size distribution of GFP inclusion bodies in response to induction time and temperature [15]. Using a UV-Vis detector Margreiter et al used a sedimentation FFF to investigate the impact of inducer concentration and induction time on IB size [19]. Hereby, an increase in the median spherical diameter of up to 140 nm over induction time was observed.

1.3 Imaging/TEM

The majority of the previously described methods and contributions feature transmission electron microscopy in an attempt to verify made conclusions based on the respective method. TEM facilitates conclusions based on single particle analysis by making single IBs visible. Given the overall goal of IB analytics of characterizing IBs as product of USP, the IBs should be analyzed in the most native conformation possible. Hereby, imaging in of IBs in the cytosol would exclude most of the otherwise necessary sample preparation and therefore a potential analytical bias. Sizing IBs using TEM of the respective IB sample is commonly based on a manual image analysis based on the TEM image [19, 20]. Nevertheless the effort for sample preparation the analytical technique, and image evaluation is substantial and basically denies the usage of the state of the art TEM method as routine analytical technique.

Within this contribution we introduce nano particle tracking analysis (NTA) as method to analyze and size a large number of biologic nano particles individually. For biologic particles the dynamic range of NTA spans 100-1500 nm; which fits the reported size range of IBs from 170 to 1300 nm [10, 11, 19, 20]. NTA uses a laser as light source, which passes through the sample particle suspension and illuminates the particles. In scatter mode the scattered light, in fluorescence mode the emitted light is recorded by a high speed camera through a microscope. Owned to the angle between light beam and the camera axis individual particles can be tracked and analyzed. At a constant temperature and a constant viscosity of the liquid, the size of each particle correlates to the particle movement. Using the Stokes-Einstein equation the individual particle size can consequently be calculated resulting in a histogram of the particle size distribution of the particles in suspension.

1.4 Goals

The overall goal of this contribution is to provide a cost and time efficient method to quantify IB size. Firstly, for method verification, we aim to establish grey scale image segmentation of transmission electron microscopy pictures as orthogonal method to assess IB size. Secondly, as easily transferable, cost and time efficient method NTA is assessed as method to quantify

IBs and their size in the background of cell debris. Finally, as an exemplary application the growth of IBs is illustrated over process time.

2. Materials and methods

2.1 Bioreactor system

The fermentations were conducted in a DASGIP multi-bioreactor system (4Force; Eppendorf; Germany) with a working volume of 2 L each. The DASGIP control software v4.5 revision 230 was used for data logging and control: pH (Hamilton, Reno, USA), pO2 (Mettler Toledo; Greifensee, Switzerland; module DASGIP PH4PO4), temperature and stirrer speed (module DASGIP TC4SC4), aeration (module DASGIP MX4/4) and pH (module DASGIP MP8). CO2, O2 concentrations in the off-gas were quantified by a gas analyzer (module DASGIP GA4) using the non-dispersive infrared and zircon dioxide detection principle, respectively.

2.2 Cultivations

A recombinant BL21 DE3 *E.coli* strain was cultured, producing an intracellular protein (~30 kDa) in form of inclusion bodies, after a one-time induction with IPTG (1 mM). The synthetic media was based on the recipe of Korz, Rinas et al. [21], where the limiting C-source was glucose.

Pre cultures were grown to a OD600 of approx. 1.5 in 150 mL batch media. 2.5% of the batch volume was added as pre-culture for inoculation. The stain was cultivated at controlled pH, dissolved oxygen DO2 (>30%) and temperature. The DO2 was kept over 30% by supplementing oxygen to the air. After depletion of the C-source in an initial batch phase, the pre-induction fed-batch was started. The pre-induction feeding strategy was based on an exponential feed forward profile to maintain a predefined growth rate [22]. Upon induction stirrer speed was set to 1400 rpm and aeration to 1.4 v/v/m for the whole process. The pH was maintained by adding 12.5% NH4OH, which also served as nitrogen source.

2.3 Imaging

For High pressure freezing (HPF) *E.coli* samples were pelleted and re-suspended in 5% BSA. After a second centrifugation step the pellet was immediately frozen in a high-pressure freezer (HPF Compact 01; Wohlwend; Switzerland). The samples were then transferred into a freeze substitution unit (EM AFS2; Leica Microsystems; Germany) for water substitution with 2% uranyl acetate in anhydrous acetone over 5 days.

For chemical fixation the supernatant of the pelleted *E.coli* samples was carefully aspirated and cells were fixed using 2.5% glutaraldehyde in 100mM cacodylate buffer at pH 7.4 for 1 h at room temperature. After washing in the same buffer samples were post-fixed in 2% osmium tetroxide in cacodylate buffer, washed and dehydrated in a graded series of ethanols.

The dehydrated specimens were embedded in agar 100 resin (AGR10131; Agar Scientific Ltd) and after hardening ultrathin sections (70 nm) were prepared. Sections were post-stained with 2% aqueous uranyl acetate, followed by incubation with Reynold's lead citrate. Images were collected using a transmission electron microscope (Morgagni 268D; FEI; The Netherlands) operated at 80 kV and equipped with an 11 megapixel camera (Morada CCD; Olympus-SIS; Germany).

2.4 Image segmentation

To quantify the inclusion body size, the relative area of IB per cell was quantified based on grey scale image segmentation. The thresholds for background and IBs were selected manually by the operator, specific for each picture. The difference in area of the background and total image area corresponds to the area covered by cells. A pre-test with a larger number of operators substantiated that the image segmentation is insignificantly impacted by the operator and can be regarded as transferable in-between operators. Image segmentation of the 17 different samples and one negative sample with 3-6 images for each sample was conducted in Image Lab. 160-380 individual cells were repetitively analyzed per sample.

2.5 Cell disruption

Two mL of the fresh culture broth were centrifuged (4500 x g; 10 min; 4°C). The cell pellets were re-suspended in 20 mL 0.1 M Tris-buffer; 10 mM EDTA (pH 7.4) buffer and were disrupted in a high-pressure homogenizer (Avestin EmulsiFlex; Canada) at 1400 ±100 bar in 6 passages. For chemical fixation 0.2% Glutaraldehyde (G7776; Sigma Aldrich) was added dropwise to the re-suspended pellet and incubated 1 h at 4°C prior to homogenization.

2.6 Fluorescence stain

To discriminate cell debris from inclusion bodies the homogenized cell pellet (5 min 5000 g) was re-suspended and incubated 30min in a 1xPBS solution containing 1% BSA and 2.2 mg/l of a product specific biotinylated primary antibody (courtesy of Sandoz GmbH). After washing with 1x PBS 1% BSA once, the pellet was re-suspended and incubated for 30 min in 1x PBS 1% BSA containing 10 µg/ml secondary IgG antibody labelled with Alexa 488 (AT11001;

Invitrogen Life Technologies). Prior to measurement the suspension was washed and re-suspended in 1xPBS.

2.7 Nano particle tracking (NTA)

An NS500 (Malvern, UK) software release (Nano Sight 3.0) equipped with a 488 nm laser and a CMOS camera (Hamamatsu Photonics, Japan) was used for the conducted NTA measurements. Most of the software settings are proprietary and are not known to the authors. The measurement chamber was primed prior to each measurement with 1x PBS to minimized particle drift. In-between measurements the chamber was flushed twice to avoid sample carryover. All samples were sonicated 1 min prior to measurement and diluted 1:10 in PBS. The focus level was set automatically a standardized camera level of 16 was used in combination with a detection threshold of 20. 6 replicates, 90 sec each were conducted with a 5 sec time delay at a controlled temperature of 25°.

2.8 Titer quantification

Product titer was measured using RP-HPLC after solubilizing the washed pellet of disrupted cells in guanidine hydrochloride. Biomass concentrations were gravimetrically quantified after drying at 105°C for min. 72 h. Therefore 2 mL of culture broth were centrifuged (4500g, 10 min, 4°C) in a pre-weighted glass tube and the pellet was washed once with 5 mL RO water. The determination was done in duplicates. After drying in the drying oven the biomass dry weight was measured on a scale.

Statistical data analysis

Data were subjected to statistical analysis using Datalab Version 3.5 (distributed by Epina http://datalab.epina.at/). Based on an $\alpha = 0.05$ the significance of the correlation was evaluated at hand of the p-value.

3. Results

3.1 TEM+HPF as gentle method for IB visualization

Currently there is no verified method available to quantitatively size a representative number of IBs. This circumstance hinders the establishment of a high throughput method for quantitative IB sizing. Consequently, method assessment can only be based on relative confirmation by comparing result of two otherwise orthogonal methods (Figure 1). Since an absolute measure of the size of a distribution of biologic nanoparticles is not available, a relative verification is targeted by comparing the relative area [%] derived from TEM and the hydrodynamic diameter [nm] derived from NTA. To avoid measuring artefacts it is of utmost importance to minimize the impact of sample preparation in order to preserve the most native IB form. While TEM is capable of visualizing IBs even in the cytosol, NTA can only measure particles in suspension. For this reason HPF has been used as fixation approach for TEM, owned to the gentle fixation properties. Supplemental 1 illustrates the conservation of cellular structures for different induction time points of two representative experiments. In contrast, sample preparation for NTA requires cell homogenization and a consequent FL stain in order to facilitate IB analysis even in the background of cell debris.

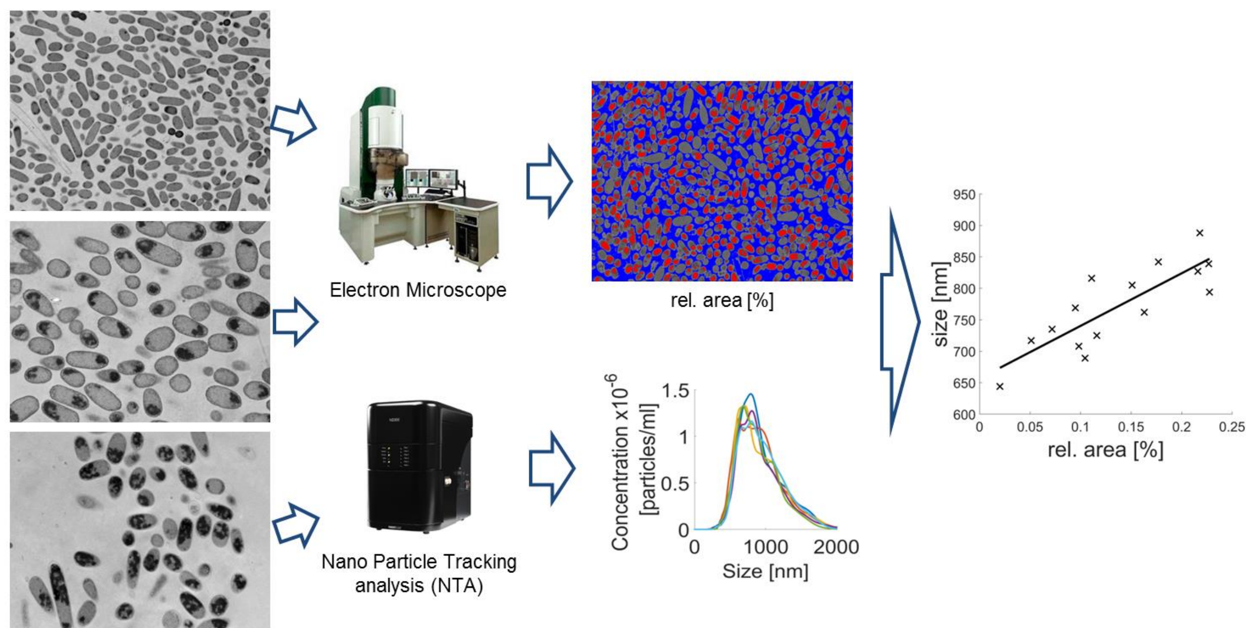


Figure 1 Method assessment by relative method verification: *Flow chart of the targeted relative verification of NTA and TEM derived quantification of IB size, Independent IB process samples are analyzed by NTA as well as by TEM, NTA yields the hydrodynamic particle diameter distribution which corresponds to the IB size [nm] owned to the utilized specific FL stain, TEM derived images of process samples are segmented according to the grey scale;*

3.2 Grey scale image segmentation for quantitative IB sizing not significantly operator dependent

Qualitative IB growth over time can be deducted from the images from Supplemental 1. But for a quantitative assessment of IB size/growth over time, a standardized approach of IB sizing is necessary. Using grey scale image segmentation from TEM images the relative IB size was quantified as IB area per cell [%]. Basal grey values of TEM images have been found to be highly variable owned to background particles. This impairs a uniform background definition and correction. This circumstance impairs fully automated image segmentation. Targeting a sound science method to reproducibly quantify cytosolic IB size 17 independent samples were analyzed by image segmentation (Figure 2). Using this software based approach, operated by individuals, the time per operator and image decreased below 10 sec. For each sample 3-6 TEM images were recorded and segmented by the individual operator in random order at least 3 times. In Figure 2 A-C the variance induced by the different operators is indicated. The respective results are not statistically significant operator dependent, rendering the method transferable for IB sizing and suitable for method verification. Using HPF as gentle fixation method for IBs as biologic nano particles the software aided image segmentation approach was evaluated as suitable orthogonal verification method.

In order to minimize artefacts sample preparation was simplified as far as possible. While mere homogenization and direct NTA measurement did not lead to satisfying results the implementation of a FL stain increased sensitivity of the method (data not shown). Consequently, the samples were measured post homogenization and FL stain without any fixative (native). To illustrate the data basis for a size measurement by NTA, Figure 2 A displays a histogram of tracked and sized particles of an exemplary fluorescence stained sample. Although the particles were tracked in high number, NTA results and the relative IB areas from TEM-HPF were not significantly correlated Figure 2 D. Judging from the TEM-HPF images (Supplemental 1) as well as from the image segmentation (Figure 2 D) a significant difference/ growth over time in size of IBs can be observed. Nevertheless, this trend was not represented by NTA results. Additional test indicated that standard FL beads were identified with high precision even in the background of stained homogenate (data not shown).

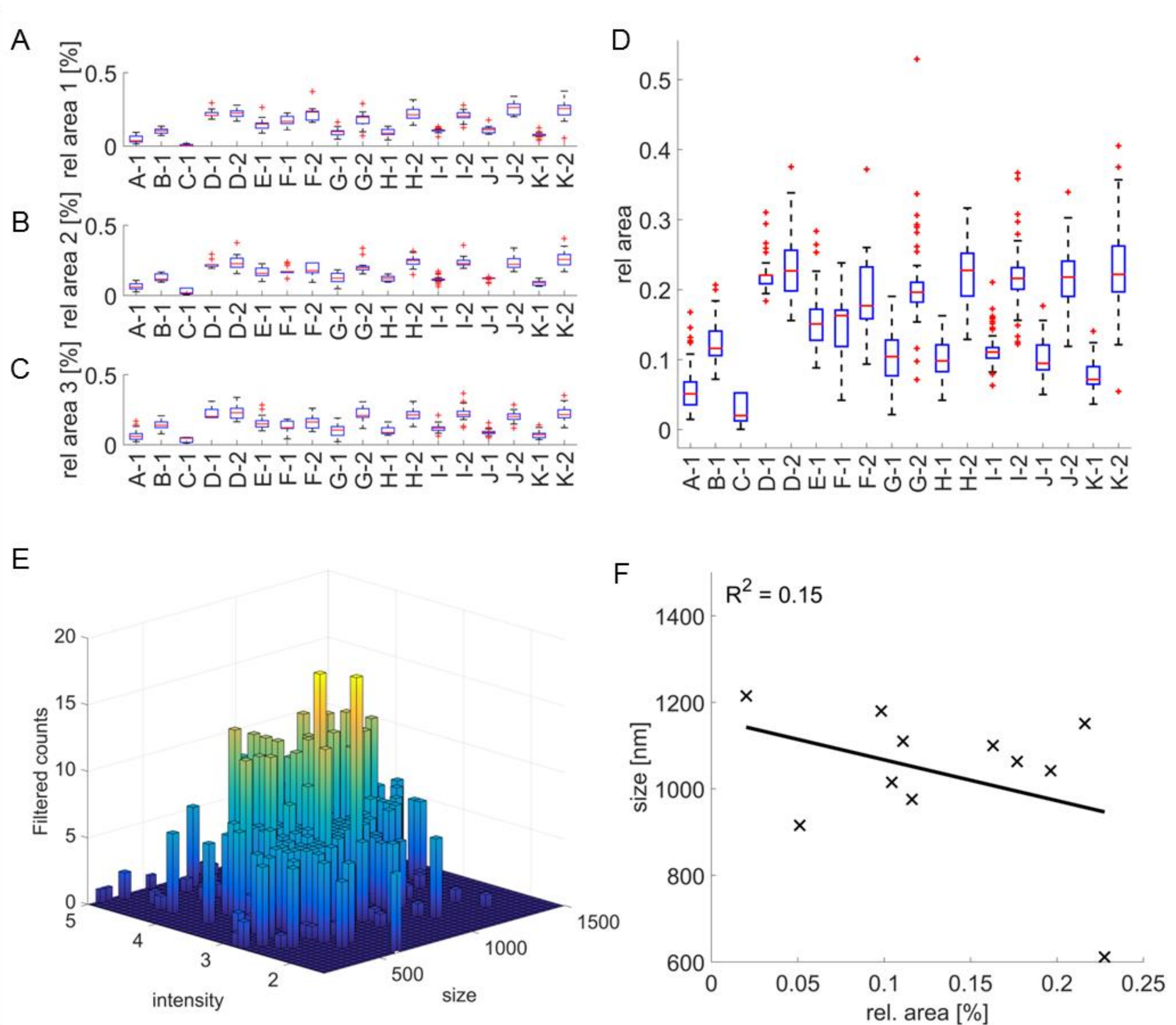


Figure 2 grey scale image segmentation of TEM images for IB sizing is not significantly operator dependent:

The relative area (rel. area [%]) corresponds to the area covered by IBs per cell background, 17 induced samples and one negative sample (C-1), of each sample 3-6 individual TEM images were segmented in random order (n>3), each letter corresponds to an individual experiment, for samples with the same letter (2) indicates a later time point than (1); (A-C) Grey scale image segmentation results operator specific, each subpanel corresponds to one individual operator; (D) all segmentations(> 550) results pooled, whiskers indicate 75% interval; (E) Filtered data of the size distribution of fixated IBs by NTA in the background of cell debris, all tracked particles of one sample measurement including the 6 replicate measurements, filtered by intensity and track length. (F) The correlation of relative IB area [%] to the hydrodynamic diameter derived from NTA [nm], data does not allow was not found to be significant p(t)=0.18;

3.2 Particle fixation for Nano particle tracking analysis (NTA) increases method sensitivity

To investigate the impact of sample preparation on IB sizing by NTA additional tests were conducted using a chemical fixative prior to cell homogenization. The IB sizing results of native IBs and chemical fixated IBs are compared in Figure 3. Figure 3 A/B illustrate the massive improvement of NTA raw data quality by sample fixation prior to homogenization. The standard deviation drops and displays significantly less variance (Figure 3 E). Highly interesting is the observation that fixated samples displayed quantitatively more particles than native samples (Figure 3 F). This is especially interesting in respect to the cross linking properties of glutaraldehyde which should lead to generally bigger and less particles. Presumably, the addition of a fixative prior to homogenization prevents IBs from aggregating post homogenization during sample preparation.

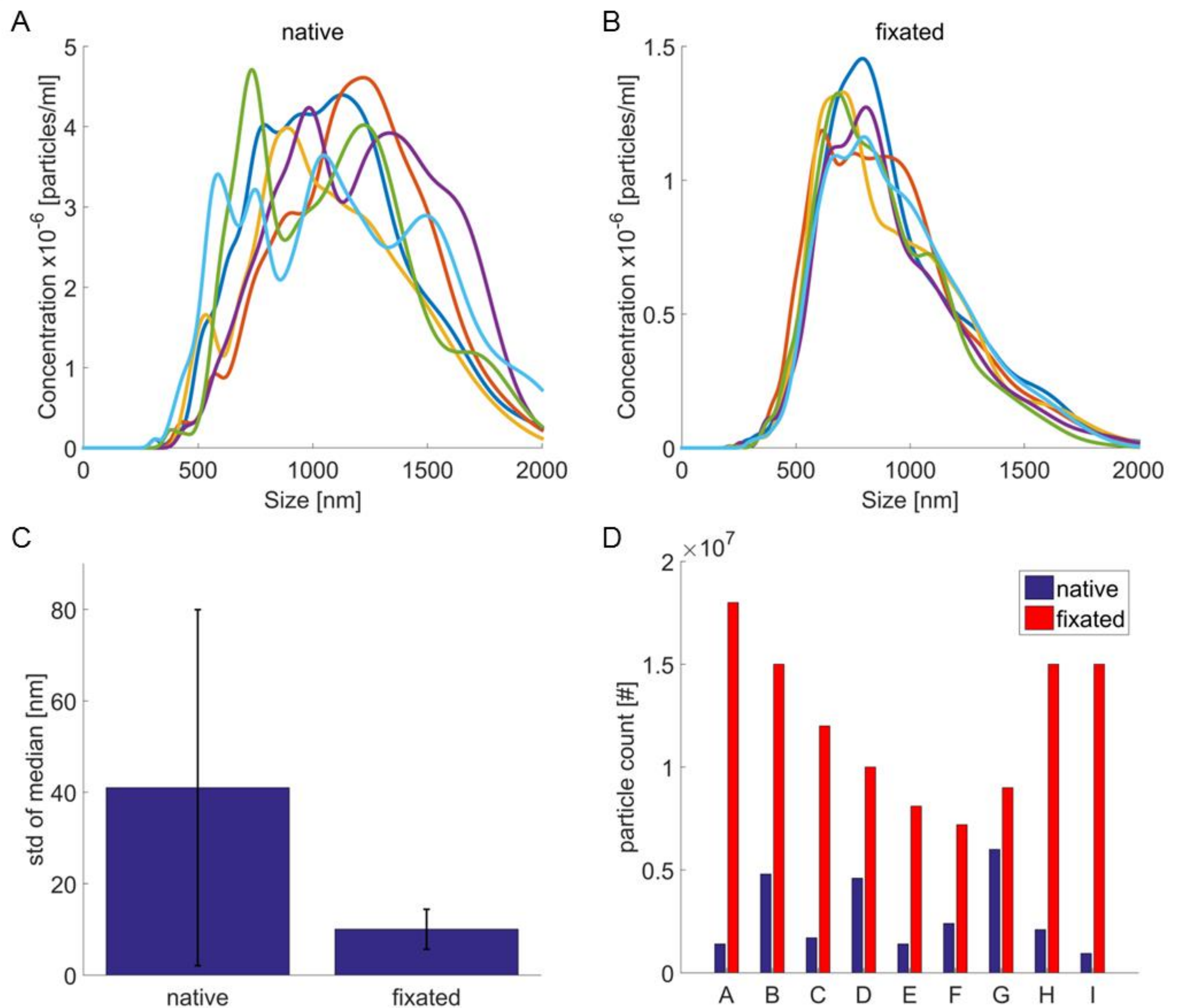


Figure 3 The positive impact of fixation prior to homogenization on NTA particle measurement: *(A) raw data of particle size distribution of FL stained native IBs; (B) raw data of particle size distribution of FL stained fixated IB of the same sample as in A; (C) Fixation prior to homogenization decreases standard deviation of the median size and boosts reproducibility, comparison of the standard deviation of the median of the particle size distribution of the 6 replicate measurements per sample, for not fixated (native) n= 37 and fixated (fixated) n=26 samples, p(t) < 0.001; (D) total particle counts of NTA raw data, A-I correspond to sample names from various fermentations and time points, the observed trajectory in the particle count of fixated samples is presumably of coincidental nature.*

Besides increasing measurement sensitivity, sample fixation impacts the intensity per particle (Figure 4 A). In order to put the IB size obtained from fixated samples into perspective, Figure 4 B illustrates the correlation of all measured TEM and NTA samples. Based on a p value of 0.002 it can be concluded that the rel. IB area [%] and the hydrodynamic diameter [nm] are

correlated. To illustrate that the measure IB size is not a redundant measure of a simple titer quantification the TEM areas as well as the IB diameters derived from NTA are compared to the respective specific titers Figure 4 B/C. Based on the assumption of a uniform particle density the size of IBs should be tightly correlated to the amount of product contained in the particle. Nevertheless, the specific product titer does not display a highly significant correlation to the particle size, neither for particle sizes derived from TEM (Figure 4 C) nor from NTA (Figure 4 D).

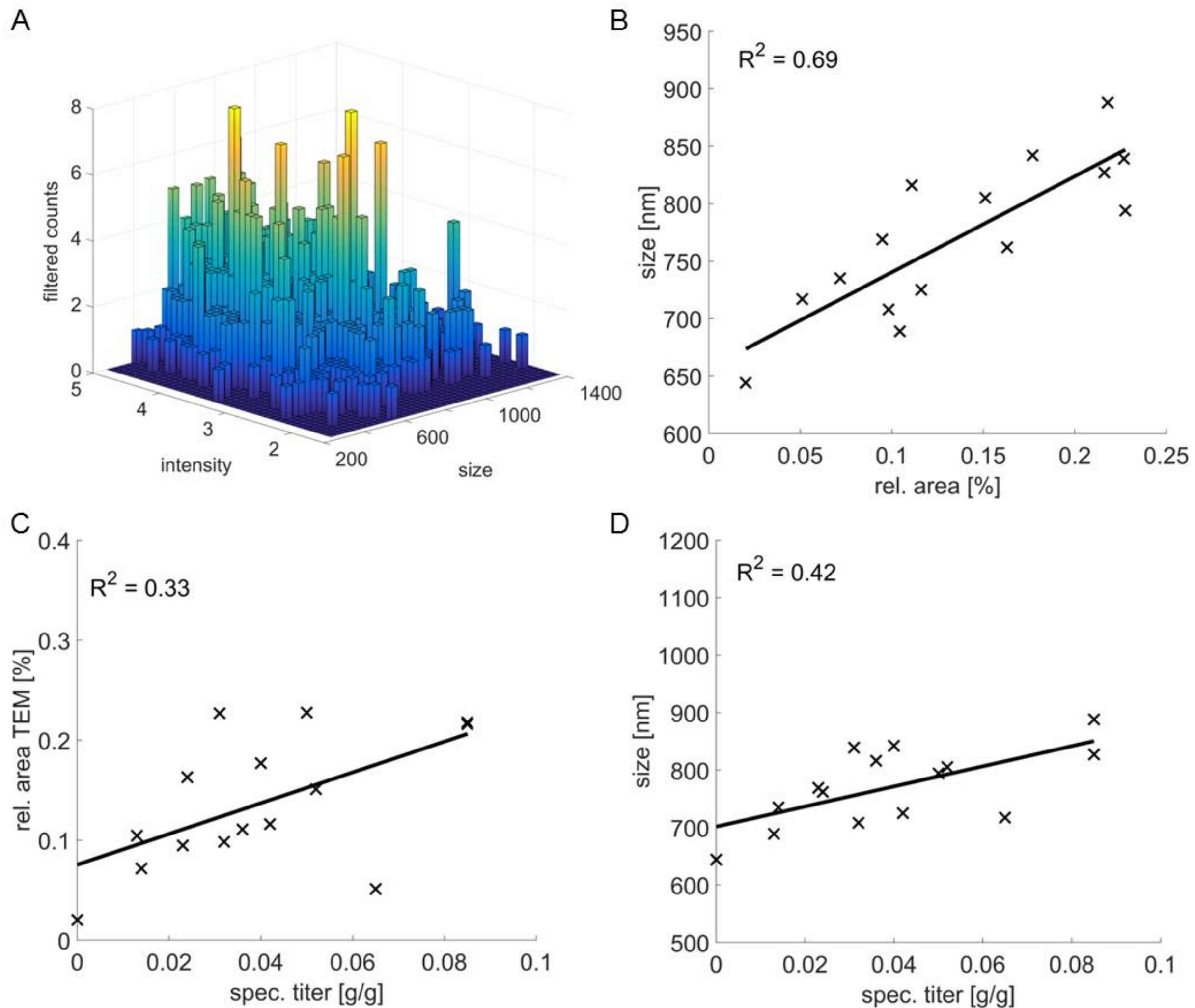
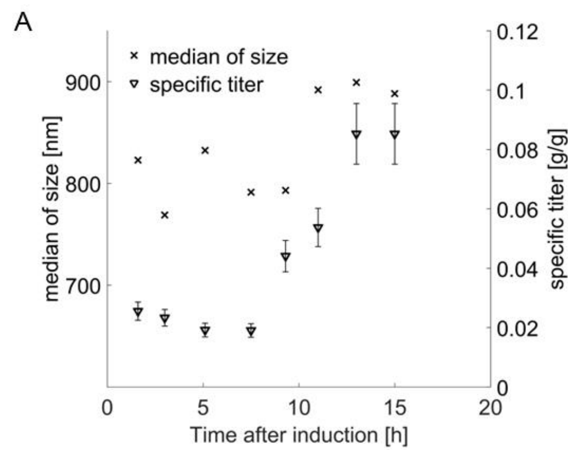


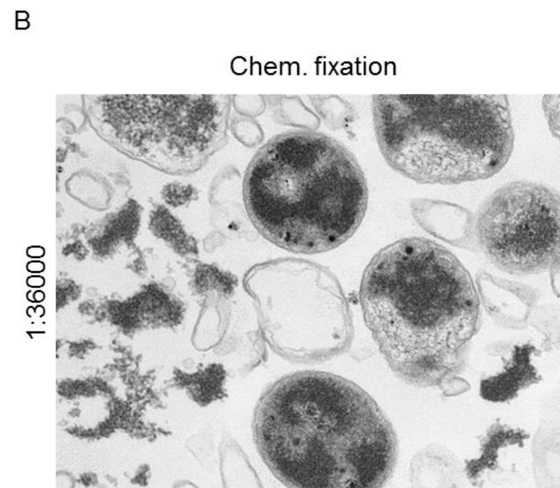
Figure 4 NTA size of fixated samples is significantly correlated to the relative IB area derived from TEM (A) Filtered data of the size distribution of fixated IBs by NTA in the background of cell debris, all particles of one sample measurement including the 6 replicate measurements, filtered by intensity and track length; (B) significant correlation of relative area (rel. area TEM [%]) and particle size derived from NTA (size [NTA]),n=15, R^2=0.69, p(f)= 0.002; (C) correlation of relative area (rel. area TEM [%]) and specific product titer (spec. titer [g/g]), n=15, R^2=0.33, p(f)= 0.026, no serial correlation; (C) correlation of particle size derived from NTA (size [NTA]) and specific product titer (spec. titer [g/g]), n=15, R^2=0.42, p(f)= 0.009, no serial correlation;

The analysis of early and late time points of induction of different sets of experiments increases the observable differences in IB size. In comparison the timely resolution of size over induction time is a greater challenge in regard of method sensitivity. In this respect Figure 5 A illustrates the growth of IB size as well as the progression of specific product titer [g/g] over induction time. In accordance to Figure 4 D the IB size and product titer were not found to be correlated closely. The increase in size (+10-12%) was found to be comparably small given the substantial increase in specific titer (+300-400%).

To investigate the impact of homogenization and correlated sample preparation on sensitivity of the NTA measurement additional samples after homogenization were analyzed Figure 5. It can be observed that in case of HPF (Figure 5 C) the IBs are released into the supernatant while the IBs appear to maintain a more segregated state. The structure of these protein aggregates appeared poriferous and fragile. In contrast the chemical fixation of the same sample prior to homogenization led to denser particles (Figure 5 B). Based on these images it can be inferred that chemical fixation helps to maintain the IB conformation. This conclusion is also in accordance with the previously found positive impact of fixation on NTA sensitivity.



B



C

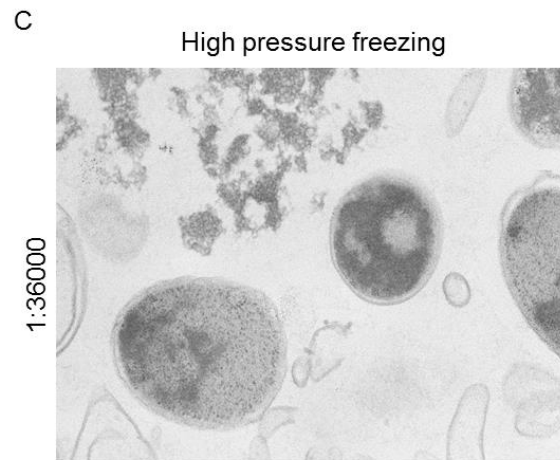


Figure 5 IB sizing by NTA features sufficient sensitivity to resolve IB growth over time;
(A) The specific titer (spec. titer [g/g] and IB size (median of size [nm])over process time since induction (Time after induction [h], as indicated before size and titer are correlated significantly, n=8, p(f)= 0.015, including a serial correlation over time (B) Fixation leads to particle condensation in the homogenate as well as in the cells, TEM image of IBs after homogenization (700 bar, 6 passages) of chemically fixated cells, 1:36 000; (C) TEM image of IBs after homogenization (700 bar, 6 passages) of native cells, 1:36 000;

4. Discussion

To investigation of the interrelation of USP and DSP requires sensitive characteristics of the USP end product– the inclusion bodies. Accordingly, the goal of this contribution was to establish and to verify a method to size of a representative number of IBs with high sensitivity and high efficiency.

4.1 grey scale segmentation of TEM images is a sensitive method for IB characterization

For method verification a second, orthogonal method to assess IB size has been established, since up to date available methods are not particle based, lack efficacy and/or sensitivity to size a representative number of biologic nanoparticles. To minimize the effect of sample preparation and correlated artefacts it was indicated to analyze the IBs in the most native conformation possible. Using only centrifugation prior to HPF, sample preparation was reduced to a minimum. Besides sample preparation the method of sample fixation has been a topic of vivid discussion. A common approach for sample fixation of IBs is chemical fixation [7, 10, 13, 20] or air drying of the specimen on a copper grid [14, 19]. Despite the wide usage of these methods the specimen is altered, subcellular structures are condensed and shrinkage occurs. In contrast to the general opinion regarding the shape of IBs [10], using HPF-AFS the native structure of IBs was found to be far more loose and sensitive than reported before.,

Regardless of the sample preparation, microscopy simplifies the shape of 3D specimen to 2D images. To compensate this drawback of image based solutions a representative number of particles need to be analyzed. Especially since the IBs do not appear to have a fully symmetrical, spherical shape, different orientations need to be accounted for by sizing a larger number of IBs. Peternel et al [20] addressed the problem of the statistical significance and sized 250-350 IBs in order to obtain a histogram of IB size distribution of isolated and washed IBs. In accordance to the latter contribution 160-380 individual cells (containing IBs) were analyzed repetitively by each of the three operators for every sample. But in contrast to the chemical fixation used by Peternel et al [20] AFS was used as highly gentle method, renowned for the ability to preserve cellular substructures. Hereby, we established HPF-AFS-TEM imaging of IBs in the cytosol in combination with grey scale image segmentation as valuable method to reproducibly and independent of an operator quantify a representative number of IBs.

4.2 NTA is a sensitive method to size a representative number of IBs

IB sizing by NTA allows the characterization of a representative number of particles (>1000) per sample. In the context of NTA measurement the chemical fixation of the cells prior to homogenization yielded an increase in sensitivity as well as a substantial increase in total particle count. Although, the total amount of recognized particles was increased (Figure 3 D), the amount of relevant particles was decreased (Figure 3 A/B). This observation might be owned to a certain aggregation tendency of native particles. This tendency could lead to an aggregation of the IBs with cell debris, especially in combination with a disintegration of the IBs triggered by shear stress during homogenization. Cellular debris after high pressure homogenization is about 0.5 µm in size [12]. In combination with IB fragments the resulting size would theoretically overlap with the expected size of native IBs. Fixation impairs this aggregation as well as it putatively decreases the probability of IB disintegration during homogenization, which in turn would lead to an increase in total number of particles but decreases the number of product specific particles.

Despite using a product specific FL stain in combination with a chemical fixation of the cells, background particles were found to bias the sizing of the standard beads in the background of homogenate. The observed strong background signal might be attributed to a bleed through of scattered light through the long pass fluorescence filter, which decreases the method specificity. For future measurements it is advised to circumvent such issues be increasing the distance between excitation wavelength and fluorescence filter. Remaining measurement noise, might be owned to an unspecific FL stain in combination with the possibility of a changing IB density. Particle size calculation in the NTA algorithm assumes a constant or particle density. As a result a potential increase in particle density would lead to an assumed increase in particle diameter in the NTA result.

Besides the methodological advances for sizing IBs by NTA and TEM the results indicate that high pressure homogenization greatly impacts IB properties. Even if different process parameters in USP elicit differences in IB size, it is highly unlikely that these differences are preserved throughout high pressure homogenization. Although chemical fixation impairs the particle aggregation tendency the approach is regarded as not feasible in an industrial production process. Chemical fixation would greatly hinder solubilisation and consequent product recovery. Nevertheless, the ratio of particle count of native and fixated IB samples could potentially be used as measure of stickiness. In this context IB sizing by NTA could help to understand the molecular processes which lead to different aggregation tendencies and in turn impact DSP efficiency. Consequently, NTA could be used to derive an additional response parameter on the basis of which integrated bioprocess development might succeed in investigating the interlink of USP and DSP.

5. Conclusions

The overall goal of this contribution was the establishment and assessment of an simplistic and transferable method for high throughput IB sizing.

- TEM in combination with grey scale image segmentation is a sensitive and reproducible method to quantify the size of native, cytosolic IBs which can be used for method verification.
- NTA is a particle based method allowing to size a great number (>1000) of fluorescence labelled IBs.
- Chemical fixation of IBs prior to homogenization decreases standard deviation and particle count but increases reproducibility of IB sizing with NTA.
- Based on the observed effect of fixation it can be hypothesized that high pressure homogenization annihilates differences in IB size caused by USP. Nevertheless, the ratio in particle count of native homogenate and fixated homogenate offers a measure for IB stickiness.

6. Acknowledgements

We are grateful for the financial support of Sandoz GmbH.

The TEM measurements were performed by the EM Facility of the Vienna Biocenter Core Facilities GmbH (VBCF), member of Vienna Biocenter (VBC), Austria.

7. Conflicts of Interest

The authors declare no conflict of interest.

8. Supplemental

References

[1] Hedhammar, M., Alm, T., Gräslund, T., Hober, S., Single-step recovery and solid-phase refolding of inclusion body proteins using a polycationic purification tag. *Biotechnology Journal* 2006, *1*, 187-196.
[2] Grote, F., Ditz, R., Strube, J., Downstream of downstream processing – Integrated bioprocess development from upstream to downstream. *Chemie Ingenieur Technik* 2009, *81*, 1276-1277.
[3] Freydell, E. J., Ottens, M., Eppink, M., van Dedem, G., van der Wielen, L., Efficient solubilization of inclusion bodies. *Biotechnology Journal* 2007, *2*, 678-684.
[4] Wellhoefer, M., Sprinzl, W., Hahn, R., Jungbauer, A., Continuous processing of recombinant proteins: Integration of inclusion body solubilization and refolding using simulated moving bed size exclusion chromatography with buffer recycling. *Journal of Chromatography A* 2013, *1319*, 107-117.
[5] Wong, H. H., O'Neill, B. K., Middelberg, A. P., Centrifugal processing of cell debris and inclusion bodies from recombinant Escherichia coli. *Bioseparation* 1996, *6*, 361-372.
[6] Strandberg, L., Enfors, S. O., Factors influencing inclusion body formation in the production of a fused protein in Escherichia coli. *Applied and Environmental Microbiology* 1991, *57*, 1669-1674.
[7] Datta, I., Gautam, S., Gupta, M. N., Microwave assisted solubilization of inclusion bodies. *Sustainable Chemical Processes* 2013, *1*, 1-7.
[8] Datar, R. V., Cartwright, T., Rosen, C.-G., Process Economics of Animal Cell and Bacterial Fermentations: A Case Study Analysis of Tissue Plasminogen Activator. *Bio/Technology* 1993, *11*, 349-357.
[9] Dürauer, A., Mayer, S., Sprinzl, W., Jungbauer, A., Hahn, R., High-throughput system for determining dissolution kinetics of inclusion bodies. *Biotechnology Journal* 2009, *4*, 722-729.
[10] Castellanos-Mendoza, A., Castro-Acosta, R. M., Olvera, A., Zavala, G.*, et al.*, Influence of pH control in the formation of inclusion bodies during production of recombinant sphingomyelinase-D in Escherichia coli. *Microbial Cell Factories* 2014, *13*, 1-14.
[11] Taylor, G., Hoare, M., Gray, D. R., Marston, F. A. O., Size and Density of Protein Inclusion Bodies. *Nat Biotech* 1986, *4*, 553-557.
[12] Wong, H. H., O'Neill, B. K., Middelberg, A. P. J., Cumulative sedimentation analysis of Escherichia coli debris size. *Biotechnology and Bioengineering* 1997, *55*, 556-564.
[13] Walther, C., Mayer, S., Sekot, G., Antos, D.*, et al.*, Mechanism and model for solubilization of inclusion bodies. *Chemical Engineering Science* 2013, *101*, 631-641.
[14] Upadhyay, A. K., Murmu, A., Singh, A., Panda, A. K., Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in Escherichia coli. *PloS one* 2012, *7*, e33951.
[15] Luo, J., Leeman, M., Ballagi, A., Elfwing, A.*, et al.*, Size characterization of green fluorescent protein inclusion bodies in E. coli using asymmetrical flow field-flow fractionation–multi-angle light scattering. *Journal of Chromatography A* 2006, *1120*, 158-164.
[16] Elena, G.-F., Joaquín, S.-F., Esther, V., Antonio, V., Tunable geometry of bacterial inclusion bodies as substrate materials for tissue engineering. *Nanotechnology* 2010, *21*, 205101.
[17] Cano-Garrido, O., Rodríguez-Carmona, E., Díez-Gil, C., Vázquez, E.*, et al.*, Supramolecular organization of protein-releasing functional amyloids solved in bacterial inclusion bodies. *Acta Biomaterialia* 2013, *9*, 6134-6142.
[18] Moon, M. H., Lee, S., Sedimentation/steric field-flow fractionation: a powerful technique for obtaining particle size distribution.
[19] Margreiter, G., Messner, P., Caldwell, K. D., Bayer, K., Size characterization of inclusion bodies by sedimentation field-flow fractionation. *Journal of biotechnology* 2008, *138*, 67-73.
[20] Peternel, Š., Jevševar, S., Bele, M., Gaberc-Porekar, V., Menart, V., New properties of inclusion bodies with implications for biotechnology. *Biotechnology and Applied Biochemistry* 2008, *49*, 239-246.
[21] Korz, D. J., Rinas, U., Hellmuth, K., Sanders, E. A., Deckwer, W. D., Simple fed-batch technique for high cell density cultivation of Escherichia coli. *Journal of Biotechnology* 1995, *39*, 59-65.
[22] Lee, S. Y., High cell-density culture of Escherichia coli. *Trends in Biotechnology* 1996, *14*, 98-105.

ii. Biomass sensing

Physiological control strategies rely on a real time biomass estimation which is often based on large amounts of data and challenging model setups [32].The choice of the appropriate real time biomass estimation approach depends on the demands towards the respective field of application. Bioprocess control in production stage can rely on vast amounts of historic process data and process knowledge to establish control strategies. But in order to benefit from physiological process control in production stage, physiological process control has to be realized as early as in process development. In bioprocess development stage process knowledge and historic data is scarce, which constricts applicable real time biomass estimation methods. A comprehensive comparison of generic BM sensing methods in the context of early bioprocess development, for industrially relevant high cell density fermentations is to our knowledge not available. Literature suggests mainly data driven models for real time BM estimation [46, 47]. But the necessary amount of training data is usually not available in early stage bioprocess development [12], which calls for less data dependent hard type sensors and first principle soft sensors.

While the error of the investigated methods did not display a significant correlation to strain or metabolic activity, the combination of the employed methods yielded a surprising reduction of error. Hereby, we were able to illustrate a highly accurate method of real time biomass estimation merely based on first principle mass balances and a hard type sensor.

Accepted Article

Generic biomass estimation methods targeting physiologic process control in induced bacterial cultures

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

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

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

*Corresponding author: christoph.herwig@tuwien.ac.at

Keywords: biomass estimation; biomass sensing; bioprocess development; real time process monitoring; soft-sensor;

Practical application

Within this contribution we have outlined and compared four generic applicable methods for biomass estimation in early bioprocess development. The accuracy and robustness of a hard sensor, soft-sensor and hybrid sensor were discussed based on the coefficient of variation of the root mean squared error (cvRMSE) for two strains and three different levels of metabolic activity. This comparison facilitates a comprehensive overview of appropriate methods for biomass estimation in bioprocess development. Depending on the scope of the planned experiments as well as on the available infrastructure and historic data, the outlined data ease method selection. Hereby, we aim to alleviate physiologic bioprocess development requiring physiological process control which is necessarily based on real time biomass estimation.

Received: 12/16/2015; Revised: 03/19/2016; Accepted: 04/27/2016

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/elsc.201500182.

This article is protected by copyright. All rights reserved.

Accepted Article

Abstract

Advanced bioprocess development strategies focus on the control of physiological entities which rely on accurate real time determination of biomass concentration. Various methods have been proposed in literature but up to this date a comprehensive and differentiated comparison of biomass estimation approaches for early stage bioprocess development is missing. In this contribution, we compared hard sensor, soft-sensor and data-driven approaches for real-time biomass estimation in respect to accuracy, transferability and costs. The outlined methods were tested with two different microbial strains and recombinant products using *E. coli*. To investigate the applicability of the outlined methods, method performance was assessed in correspondence to metabolic activity. Based on statistical descriptors the methods were compared and discussed. The results indicate no significant impact of strain or biomass estimation approach on the measurement quality. The average relative error of 11-13% can be greatly reduced by over 85% combining the outlined methods by the means of weighted average. This approach proved to be highly robust even during highly dynamic process conditions of oscillating specific substrate uptake rates. Concluding, the combination of low cost first principle soft-sensor approaches in combination with a hybrid soft-sensor yields the best information to effort ratio.

Introduction

The governmental regulatory demand regarding biotechnological production of pharmaceuticals has been growing, not only since the Quality by Design (QbD) initiative [1]. Regulatory authorities recommend the use of the latest scientific advances for process control in order to demonstrate process understanding and assure constant product quality [2]. Driven by simplicity, bioprocesses are conventionally developed based on technical process parameters as the volumetric feeding rate [3, 4]. But merely controlling the technical variables, as volumetric feed rates, is ignoring the physiology of the actual product manufacturer – the cells. Consequently, literature has started addressing physiology as by the control of specific rates throughout induction phase [5-7]. Physiological control of cellular factories, independent of the specific rate of interest, relies on the precise real time estimation of one of the most important, yet one of the most challenging variables to measure in real time- the biomass (BM) [8].

The choice of the appropriate real time biomass estimation approach for industrially relevant high cell density fermentations depends on the demands of the respective field of application. Robustness and accuracy are the main demands of production stage towards bioprocess control strategies, while costs and transferability are of limited concern. Commonly, bioprocess control in production stage can rely on vast amounts of historic process data and process knowledge for control purposes. These datasets allow the use of data driven models, as widely discussed and favored in literature [9, 10]. In bioprocess development stage the demands towards the control strategy are transferability, high information to effort ratio and costs. In early bioprocess development stage process knowledge and

This article is protected by copyright. All rights reserved.

Accepted Article

historic data is scarce [11], which constricts real time biomass estimation methods and points towards mechanistic or hard type approaches. These different constraints for real time BM sensing methods in development in comparison to production stage in combination with the limited literature on this topic calls for a comprehensive comparison. In this context, there is a clear need for a comprehensive comparison of real time BM estimation methods for early bioprocess development at hand of: (1) cost, (2) transferability, (3) accuracy, (4) information to effort ratio.

Biomass estimation approaches

In a non-induced state the biomass yield can be regarded as constant. But upon induction the triggered protein overexpression causes alterations in the energy metabolism and consequently variations of the yield coefficients [12, 13]. Therefore, various approaches for real time biomass estimation have been reported [10, 11, 13-17]. These real time approaches can generally be categorized by the underlying principle of biomass estimation: hard type sensor and model based.

Hard type sensor

Probes using light absorption/ scattering are limited to a relatively low biomass concentrations since accuracy is impaired by suspended solids such as cellular debris [18]. Fluorescence probes on the other hand rely on a constant amount of absorbing compounds as NAD(P)H per cell, unless 2D fluorescence spectroscopy is used. Hereby 2D fluorescence spectroscopy requires substantial effort for data evaluation, making the implementation and data interpretation laborious [8]. Similarly near infrared (NIR) measurements can be used for biomass estimation, in case spectra of NIR are recorded and subjected to a model based data evaluation [19]. In case of 2D fluorescence spectroscopy as well as in the case of NIR spectroscopy data evaluation requires substantial training data sets for model building which is hardly available in early stage bioprocess development.

Dielectric spectroscopy has a higher dynamic range of biomass concentration and does not require complex data treatment. The method relies on the fact that viable cells act as capacitors upon polarization in an electric field. The measured signal is a function of volume fraction of the cells and cell size [18]. Since the measurement relies on a transmembrane potential only intact/ living cells are detected [20]. The relative permittivity (pF/cm) from the culture broth can be easily correlated to biomass offering a highly informative online signal for mammalian, yeast and bacterial processes [21-23], although it has been reported to be sensitive to gas holdup [24].

Model based biomass estimation

Mostly mechanistic and data driven models are utilized, although categorization is difficult due to the gradually overlapping definitions [25-27]. To the authors understanding both types generally differ mainly in transparency and the requirement of process data. Mechanistic models feature a higher degree of transparency (white box) in comparison to data driven (black box) models. Mechanistic models are largely based on *a priori* knowledge which is generally applicable and independent of

This article is protected by copyright. All rights reserved.

Accepted Article

strain and product. Necessary parameters are of fixed quantity and measurable e.g. comprise a meaning as material or energy balances but are generally applicable and are hardly dependent on process data.

Data driven or black box models rely mainly on strain specific training process data sets in order to obtain values for the model parameters [26]. Parametrized models are based on a fixed number of parameters (e.g. linear, ordinary differential equations) arising from prior knowledge, while nonparametrized models (e.g. artificial neuronal networks) are based on a variable number of parameters arising from historic process data. Model accuracy consequently depends on the quality and the quantity of available training data sets. Several model algorithms as fuzzy rule systems, multivariate regression, artificial neuronal networks or simple correlations have been proposed in literature for microbial biomass estimation [6, 13, 28-30]. In a setting of early bioprocess development first principle models appear most feasible from the category of mechanistic models, while owned to the limited necessity of training data the use of a hybrid model as representative of data driven models, data appears most promising.

First principle soft-sensor

First principle soft-sensors use transparent elemental balances/ mass balances to derive target variable estimations from real-time measured culture outputs [31]. In case of redundancies, reconciliation procedures can help to detect gross measurement errors [32], also the metabolic state of the culture can be inferred [33]. Owned to the nature of first principle balances these soft-sensors do not rely on historic process data since the underlying principles are generally applicable.

Hybrid model soft-sensor

In the stage of early bioprocess development training data usually is scarce, limiting the options for data driven models. In hybrid soft-sensors, mechanistic *a priori* knowledge of the system facilitates a reduction of model parameters, which in turn minimizes the amount of required training data for parameter estimation of the data driven part. Hybrid models facilitate the use of data driven approaches even in early stage process development. Nevertheless extrapolation over the boundaries of the training data set influences the accuracy negatively.

Goals

This contribution provides a comprehensive comparison of generic methods for high cell density biomass estimation approaches in the context of early bioprocess development for induced *E. coli* cultures for recombinant protein production. Targeting process development, we selected permittivity measurement as a representative hard type sensor and two first principle approaches as well as one hybrid soft-sensor approach as representative of model based approaches in order to facilitate a comprehensive overview.

The application, benefits and limitations of the respective real time biomass estimation methods shall be illustrated and will be analyzed qualitatively and quantitatively.

This article is protected by copyright. All rights reserved.

Accepted Article

Using two industrial relevant *E. coli* strains, yielding a periplasmic and a cytosolic product, the performance of the BM estimation approaches will be illustrated at hand of different levels of metabolic activity. Owned to the representative number of experiments used for quantification of the estimation error this contribution features an unpreceded scope of biomass estimation approaches for bioprocess development.

Materials and methods

Bioreactor system

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

Cultivations

The experiments were based on two industrially relevant *Escherichia coli* strains. Strain A refers to a modified *E. coli* K12 featuring a rhamnose-inducible expression system (rhaBad promoter). A one-time addition of L-rhamnose (1.5 g/L) induced the production of the soluble protein (~24 kDa). The complex media was based on the recipe from Wilms et al. [34] and supplemented with additional components. The limiting C-source was glycerol.

Strain B refers to a recombinant BL21 DE3, producing an intracellular protein (~30 kDa) in form of inclusion bodies, after a one-time induction with IPTG (1 mM). The synthetic media was based on the recipe of Korz, Rinas et al. [35]. Glucose was used as limiting C-source.

Pre-cultures were incubated at 30°C and 170 rpm to an OD600 of approx. 1.5 in 150 mL batch media and 2.5% batch volume aliquots were used for inoculation. Both strains were cultivated at controlled pH (7), dissolved oxygen DO2 (>30%) and temperature (A = 30°C and B = 35°C). The pre-induction feeding strategy was based on an exponential feed forward profile to maintain a predefined growth rate until induction [36]. On attainment of the predefined biomass (~30 g/L) the end of pre-induction fed-batch (EFB) was reached and the cultures were induced after 30 min adaption time (post-induction phase). Stirrer speed was set to 1400 rpm and aeration to 1.4 v/v/m for the whole process. The pH was maintained by adding 12.5% NH4OH, which also served as nitrogen source. The DO2 was kept over 30% by supplementing oxygen to the air.

Metabolic activity

This article is protected by copyright. All rights reserved.

Accepted Article

The post-induction phase featured different levels of metabolic activity realized by control of the biomass specific substrate uptake rate (qs). For real time qs control the feed rate $F_{(t)}$ is dynamically adjusted during the post induction phase. $F_{(t)}$ is calculated via the feed concentration c_s and the biomass $X_{(t)}$ and a defined $qs_{(t)}$ at the corresponding time point (Equation 1).

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

Equation 1: substrate feed rate $F_{(t)}$ [l/h] corresponds to the specific substrate uptake rate qs times actual biomass content $X_{(t)}$ [g] divided by the feed concentration c_s [g/l].

As physiological descriptor th average qs (qs_{mean}) was calculated over a defined amount of cumulatively metabolized substrate (dSn; Equation 2) to categorize the different feeding profiles according to their physiologic characteristics, into low (<0.15 g/g/h), mid (0.15-0.25 g/g/h) and high (>0.25 g/g/h) qs_{mean}. To trigger qs, feed rates where adapted according to the results of biomass estimation in function of the qs set point.

$$dSn_{(t)} = \frac{S_{(t)} - \mathrm{S_{EFB}}}{\mathrm{X_{EFB}}}$$

Equation 2: biomass specific amount of fed substrate since induction dSn, dSn is calculated by the substrate added since EFB ($S_{(t)} - S_{EFB}$), normalized by the biomass content at induction (XEFB); S(t) refers to Substrate in [g] fed since EFB, SEFB refers to the total amount of substrate fed from batch to EFB.

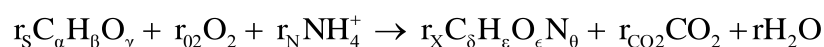
Biomass Sensing

Hardtype sensor: Dielectric spectroscopy

The permittivity sensor (Hamilton Bonaduz, Switzerland) signal was recorded using the Evobox software (Hamilton Bonaduz, Switzerland) in dual frequency mode. Probe and strain specific calibration parameters were used for the online BM estimation (referred to as Perm), which were obtained from training fermentations (strain A n = 4; strain B n = 5).

Model based: First principle soft-sensor

For mass balancing the following stoichiometric equation for an oxidative metabolism was assumed under the prerequisite that no substrate/ by-product accumulation occurs (Equation 3).



$$r_S C_\alpha H_\beta O_\gamma + r_{02} O_2 + r_N NH_4^+ \rightarrow r_X C_\delta H_\varepsilon O_\epsilon N_\theta + r_{CO2} CO_2 + rH_2O$$

This article is protected by copyright. All rights reserved.

Accepted Article

Equation 3: stoichiometric equation for an oxidative metabolism, where a C- and N-source is metabolized under the usage of oxygen and incorporated into biomass. r_S [cmol/h] rate of substrate feed; r_{O2} [cmol/h] and r_{CO2} [cmol/h] are the rates of oxygen/carbon dioxide consumption derived from offgas analysis; r_N [cmol/h] rate of base feed; r_X [cmol/h] rate of biomass growth, r_{H2O} [cmol/h] water as byproduct of oxidation.

The elemental composition (α through θ) of the substrate ($C_\alpha H_\beta O_\gamma$) and the biomass ($C_\delta H_\varepsilon O_\epsilon N_\theta$) are known. The C-balance (referred to as C-bal) Equation 5 and N-balance (referred to as N-bal) was established by considering the C/N- contents and based on to the law of conservation the sum of all rates is zero as long as flows are quantified correctly. For both cases the biomass accumulation rate can be determined independently (Equation 5, Equation 6).

$$r_S + r_x + r_{CO2} = 0$$

Equation 4: sum of C-mole conversion rates [cmol/h], enables the estimation of the unknown biomass formation rate rx.

$$r_N + r_x = 0$$

Equation 5: sum of nitrogen conversion rates [cmol/h], enables the direct calculation of the unknown biomass accumulation rate rx

Model based: Data driven - Hybrid soft-sensor

An exponential function with only two model parameters (k, d) was found sufficient to minimize the distance between process data (n = 6) of cumulative biomass yield over dSn. The biomass estimation in pre-induction fed batch phase was obtained by the first principle soft-sensor (C-bal). This highly simplistic model enabled to describe the yield trajectory with an exponential function independent of the feed dynamics.

$$Y_{xs}(t) = d \cdot e^{-k \cdot dSn(t)}$$

Equation 6: exponential function for the cumulative biomass yield coefficient Yxs in induced cultures described by dSn which is fitted to historic process data to obtain values for the strain and product specific parameters k and d.

Data processing and Data Analysis

Metabolic rates and yield coefficients were calculated with MATLAB R2013 b (Mathworks, Natick, USA). Software was used for the calculation of specific rates and yield coefficients, as we described elsewhere [37].

This article is protected by copyright. All rights reserved.

Accepted Article

For quantitative analysis of estimation accuracy the coefficient of variation of the root mean squared error (cvRMSE) was used [38]. Analysis of variance (Anova) was used to analyze the significance of the differences in the average cvRMSE from experimental sets. Biomass estimations were combined by calculating an average weighted by the error of each methodology [39].

Analytics:

Biomass dry weight (CDW)

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

Substrate conc. and small metabolites

The C-source concentration in the feed media was calculated using the gravimetrically determined density. NH4OH concentration was determined by titration with 1 M HCl, equivalent point was derived by using bromothymol blue as indicator [40]. Acetate concentrations were quantified from the supernatant by enzymatic photometric principle in a robotic system (Cedex BioHT, Roche, Switzerland). The analysis was used as a quality control to exclude possible acetate production due to oxygen limitation or overflow metabolism.

List of Symbols

$C_\alpha H_\beta O_\gamma$	substrate composition
$C_\delta H_\varepsilon O_\epsilon N_\theta$	biomass composition
CDW_{EFB} …	biomass at the end fed batch [g]
c_s …	substrate concentration in feed [g/L]
$dSn_{(t)}$ …	fed substrate normalized by the CDW at the end exp. fed-batch [g/g]
$F_{(t)}$ …	feed flow rate [L/h] after time (t)
$qs_{(t)}$…	biomass specific substrate uptake rate [g/g] at time point (t)
qs_{mean}	average qs within a predefined window of dSn [g/g/h]
r_{CO2}	CER, carbon dioxide evolution rate [mol/h]
r_n	nitrogen consumption rate [c-mol/h]
r_s	substrate consumption rate [c-mol/h]
r_x	biomass conversion rate [c-mol/h]
$S_{(t)}$…	added substrate after time t [g]
S_{EFB} …	fed substrate at the end fed batch (g)
$X_{(t)}$…	CDW [g] at time point (t)
EFB…	time point of induction [h]
t…	process time [h]
DCW…	Dry cell weight

This article is protected by copyright. All rights reserved.

Accepted Article

Results

Dynamic biomass yield decline

At hand of the experiments in FIGURE 1 we want to raise awareness that the stoichiometry of cellular metabolism changes throughout induction time. Whereas feeding profiles based on technical aspects feature a volumetric constant feeding rate (FIGURE 1 A), physiologic feeding profiles target constancy of specific physiologic variables as qs (FIGURE 1 B). Physiologic feeding profiles consequently require the adaptation of the feeding rate according to the growth of biomass. Despite this difference the dynamic decline of Yxs in induction phase can be observed for volumetric constant feeding profiles as well as for physiologic feeding profile. In accordance to literature it can be concluded that with the decline of Yxs in induction phase the stoichiometry of the metabolism is changing.

This article is protected by copyright. All rights reserved.

Accepted Article

Biomass yield prediction based on consumed substrate

Constrained by the availability of training data sets (n=6) the establishment of a fully data driven model did not appear feasible. Constraints in form of *a priori* knowledge were used to simplify the otherwise undirected approach of model establishment, to compensate the lack of comprehensive training data: upon induction the increasing metabolic load term decreases the biomass yield over time [12, 13]. An exponential function featuring only two parameters k and d (Equation 6) was found to depict the trajectory of the cumulative biomass yield versus time after induction (FIGURE 2 A). But commonly, metabolic activity is closer correlated to substrate availability rather than to time. Using dSn as independent variable (FIGURE 2 B) the distance to process data could be further decreased. dSn appears to be a physiological more significant independent variable than time. This can be explained: the cumulative metabolic activity of a high qs experiment in a window of time is significantly different compared to a low qs experiment (Supplemental 1). The established correlation (FIGURE 2 B) facilitates a biomass estimation based on dSn but requires strain and product specific training data sets.

Biomass sensing: Hard-type vs Hybrid vs First principle Soft-sensor

Within the subsequent experiments the general applicability of the outlined biomass sensing approaches is demonstrated and illustrated on a qualitative level (FIGURE 3, Supplemental 5). The comparison was based on a common window of dSn to assure the comparability of the method performance across different levels of metabolic activity. The hard type probe features noise with a low frequency but relatively high amplitude, nevertheless the overall trajectory of biomass grwoth is reproduced. Neverthelexx, the high amplitude noise of estimation is hindering real-time estimation of (specific) rates greatly, since error propagation would lead to substantial error amplification. Although the hard type probe is influenced by gas holdup as well as by electromagnetic fields (stirrer) the approach shows satisfactory results.

The balancing approaches in general display little noise and follow the trajectories smoothly. Inaccuracies of the input data e.g. concentration of base, off gas concentrations impact the overall estimation accuracy. Accumulation of substrate or metabolites also imposes a significant threat to balance based approaches (data not shown). The increasing divergence in case of FIGURE 3 B can be attributed to process dynamics resulting from the high metabolic activity. Small relative errors of the real time data impact the biomass estimation to a greater extend at higher process dynamics. The estimation based on Yxs displayed a good correlation, since the performed experiments did not extrapolate from the training data sets in terms of qs and dSn. Nevertheless for a statistical representative recommendation quantitative analysis of the data is necessary.

This article is protected by copyright. All rights reserved.

Accepted Article

Impacting variables for biomass sensing

The impact of the method of estimation and metabolic activity shall be brought to attention by the means of Table 1 (Supplemental 4). Therefore, the methods are compared based on the cvRMSE of biomass estimation in respect to the offline reference measurement. It should be noted that the total quantification across a uniform window of dSn does not reflect the frequency of the correlated error. The biomass estimation accuracy was not significantly influenced by the method of estimation (strain A $p = 0.115$; strain B $p = 0.116$). Similarly the average specific substrate uptake rate had no significant impact on estimation accuracy. This finding applies for all estimation approaches and is independent of strain and product (strain A $p = 0.597$; strain B $p = 0.722$). The biomass estimation of the hybrid soft-sensor features the highest cvRMSE independent of the strain. In general, the average cvRMSE of biomass estimation is greater for strain B. With an average relative error of 11.7% (strain A) and 13.1% (strain B) the analyzed methods are uniformly applicable.

Significant reduction of noise through combinatory approach

After substantiating that the error of measurement is not correlated to strain or metabolic activity the question of the most accurate method remains elusive. For a clear recommendation the average deviation of the biomass estimation from offline verification measurements was quantified across all levels of metabolic activity. FIGURE 4 A and C illustrate the average error of estimation differentiated according to the methods of estimation.

System perturbations and the noise of measurement can affect the real time estimation of physiological variables; this noise is usually amplified by error propagation. Consequently, an increase in estimation accuracy is highly desirable and obligatory targeting robust physiological process control. FIGURE 4 B/D illustrates the effect of combination of biomass estimation approaches exemplary for one experiment. Hereby i-iii refer to combinations of biomass estimation approaches according to the nature of estimation method. While (i) contains all estimates, (ii) only contains hard and soft-sensor based estimates and (iii) only soft-sensor based estimates. (iv) corresponds to biomass estimation based on the N-balance. The latter combinations i-iv in B/D are based on a weighted average approach using the cvRMSE values of the single BM estimation approaches. The weighted average approach leads to a significant reduction of measurement noise. The combination of all outline biomass estimation methods decreases the cvRMSE significantly by >85%, compared to the estimation approach based on a single elemental balance (N-balance) for both strains.

This article is protected by copyright. All rights reserved.

Accepted Article

Superior robustness of combinatory biomass sensing approaches

To stress the robustness of combinatory methods in dynamic conditions the qs was controlled on oscillating set points in a separate experiment (FIGURE 5 A). In this experiment discreetly different levels of qs were targeted to induce dynamically changing process conditions. Willingly physiological bottlenecks are exceeded leading to an accumulation of substrate in the fermentation broth (post 24 h). Usually undesired substrate accumulation violates a prerequisite of the first principle balancing approach (FIGURE 5 A). In case of accumulation the first principle balancing approaches consequently contribute to a comparably greater error. Therefore, the cvRMSE is decreased by the removal of the accumulation sensitive C-balance (FIGURE 5 C).

Discussion

The pronounced decline of the biomass yield coefficient (FIGURE 1); as also reported elsewhere [41]; highlights the necessity of real time biomass estimation methods. An approach employing a fixed biomass yield as found in literature [11, 42, 43], cannot account for the changes in physiology triggered by induction.

Regarding the hybrid model there is a clear differentiation to hybrid semi-parametric models, which refer to a mixture of parametrized models and nonparamatrized models [26]. Hybrid semi-parametric models usually show a superior accuracy if trained with substantial amounts of training data sets. Concerning our hybrid model the limited number of experiments and variance of metabolic activity presumably affect the quality of fit (FIGURE 2) as well as model performance in general. Model iterations, a greater basis of process data and fewer constraints would have definitely improved the performance of the hybrid model. It can be concluded that unlike in production stage in early bioprocess development data driven approaches are significantly impaired owned to the lack of process data.

The closer correlation of the cumulative biomass yield to dSn in comparison to process time (FIGURE 2) can be attributed to the difference in metabolic activity. The substrate uptake rate determines growth (replication) and often even productivity [44]. The link between substrate uptake and metabolic load has been substantiated in literature [12, 45]. Hereby, especially protein aggregation, as result of recombinant production and/or replicative aging, has been identified as major causes of cellular stress

This article is protected by copyright. All rights reserved.

Accepted Article

and consequent cell death [46, 47]. In this respect data evaluation based on consumed substrate (dSn) appears more feasible in order to enable physiologically valid conclusions from process data.

We have illustrated the performance of three different types of biomass estimation approaches. The criteria for selection and assessment were hereby cost, transferability, accuracy, information/ effort (Table 2). In this respect accuracy was the reason to exclude the otherwise commonly used electron balance based on the degree of reduction (DoR). Since the frequent controller actions necessary in HCD to maintain DO2 would cause substantial noise in real time biomass estimation. Nevertheless, in the context of bioprocess development first principle soft-sensor methods as N-bal and C-bal appear to be the most effective choice.

Unlike the hybrid soft-sensor the first principle softs sensor does not require training data sets for sensor setup and can consequently be directly employed. These methods feature a high degree of transferability, a sufficient accuracy at a low effort in combination with extremely low investment costs. The tested hard type sensor facilitates an overall low cvRMSE but the signal underlies a low frequent noise, which hinders the online estimation of physiological rates by error propagation.

The presented, highly simplistic, combination of biomass estimations based on a weighted average greatly improves signal quality and enables the user to employ advanced physiological process control routines. Regarding error propagation, the observed noise reduction is of great value for the real time estimation of specific rates and yields [22]. Hereby, neither strain nor product displayed a significant impact on estimation accuracy. The transfer of the presented approaches to other organisms depends merely on adequate adaptations to consider the difference in physiology. Consequently, the outlined methodology facilitates already in early process development an accurate real time biomass estimation, which enables development to put the cells in center of attention by physiological bioprocess development.

This article is protected by copyright. All rights reserved.

Accepted Article

Conclusions

The goal of this contribution was the application and the comprehensive comparison of suitable methods for real time biomass estimation in the context of bioprocess development.

- Balancing approaches constitute the most simplistic, transferable and cost efficient method for biomass estimation. Owned to the underlying general applicable principles this approach can be used for other microbial cultures for recombinant protein production. By accounting for all relevant reactions the principle can be even transferred to mammalian cell culture.
- The cumulative consumed substrate (dSn) represents a physiological highly relevant substitute for time on the x-axis. In case of highly limited process data, dSn used in a hybrid model allows predicting the biomass yield accurately.
- Especially in bioprocess development, a combination of biomass estimations approaches greatly benefits robustness and accuracy. We recommend using this approach independent of the employed methods to maximize robustness of biomass sensing.

Acknowledgements

We are grateful for the financial support of Sandoz GmbH.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] FDA, Guidance for Industry Q9 Quality Risk Management, 2006.
[2] FDA, Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, 2004.
[3] Ramalingam, S., Gautam, P., Mukherjee, K. J., and Jayaraman, G., Effects of post-induction feed strategies on secretory production of recombinant streptokinase in Escherichia coli, *Biochemical Engineering Journal,* vol. 33, 2007, pp. 34-41.
[4] Wong, H. H., Kim, Y. C., Lee, S. Y., and Chang, H. N., Effect of post-induction nutrient feeding strategies on the production of bioadhesive protein in Escherichia coli, *Biotechnology and Bioengineering,* vol. 60, 1998, pp. 271-276.
[5] Levisauskas, D., Simutis, R., Borvitz, D., and Lübbert, A., Automatic control of the specific growth rate in fed-batch cultivation processes based on an exhaust gas analysis, *Bioprocess Engineering,* vol. 15, 1996, pp. 145-150.
[6] Gnoth, S., Jenzsch, M., Simutis, R., and Lubbert, A., Control of cultivation processes for recombinant protein production: a review, *Bioprocess Biosyst Eng,* vol. 31, 2008, pp. 21-39.
[7] Levisauskas, D., Inferential control of the specific growth rate in fed-batch cultivation processes, *Biotechnology Letters,* vol. 23, 2001, pp. 1189-1195.

This article is protected by copyright. All rights reserved.

Accepted Article

[8] Olsson, L., and Nielsen, J., On-line and in situ monitoring of biomass in submerged cultivations, *Trends in Biotechnology,* vol. 15, 1997, pp. 517-522.
[9] Jenzsch, M., Simutis, R., Eisbrenner, G., Stuckrath, I. *et al.*, Estimation of biomass concentrations in fermentation processes for recombinant protein production, *Bioprocess Biosyst Eng,* vol. 29, 2006, pp. 19-27.
[10] de Assis, A. J., and Filho, R. M., Soft-sensors development for on-line bioreactor state estimation, *Computers & Chemical Engineering,* vol. 24, 2000, pp. 1099-1103.
[11] Dabros, M., Schuler, M., and Marison, I., Simple control of specific growth rate in biotechnological fed-batch processes based on enhanced online measurements of biomass, *Bioprocess and Biosystems Engineering,* vol. 33, 2010, pp. 1109-1118.
[12] Schaepe, S., Kuprijanov, A., Simutis, R., and Lübbert, A., Avoiding overfeeding in high cell density fed-batch cultures of E. coli during the production of heterologous proteins, *Journal of Biotechnology,* vol. 192, Part A, 2014, pp. 146-153.
[13] Jenzsch, M., Simutis, R., and Luebbert, A., Generic model control of the specific growth rate in recombinant Escherichia coli cultivations, *J Biotechnol,* vol. 122, 2006, pp. 483-93.
[14] Jenzsch, M., Gnoth, S., Beck, M., Kleinschmidt, M. *et al.*, Open-loop control of the biomass concentration within the growth phase of recombinant protein production processes, *J Biotechnol,* vol. 127, 2006, pp. 84-94.
[15] Riesenberg, D., Schulz, V., Knorre, W. A., Pohl, H. D. *et al.*, High cell density cultivation of Escherichia coli at controlled specific growth rate, *Journal of Biotechnology,* vol. 20, 1991, pp. 17-27.
[16] Passrinha, L. A., Bonifacio, M. J., and Queiroz, J. A., Application of a fed-batch bioprocess for the heterologous production of hSCOMT in Escherichia coli, *J Microbiol Biotechnol,* vol. 19, 2009, pp. 972-81.
[17] Jobe, A. M., Herwig, C., Surzyn, M., Walker, B. *et al.*, Generally applicable fed-batch culture concept based on the detection of metabolic state by on-line balancing, *Biotechnol Bioeng,* vol. 82, 2003, pp. 627-39.
[18] Kiviharju, K., Salonen, K., Moilanen, U., and Eerikäinen, T., Biomass measurement online: the performance of in situ measurements and software sensors, *Journal of Industrial Microbiology & Biotechnology,* vol. 35, 2008, pp. 657-665.
[19] Arnold, S. A., Gaensakoo, R., Harvey, L. M., and McNeil, B., Use of at-line and in-situ near-infrared spectroscopy to monitor biomass in an industrial fed-batch Escherichia coli process, *Biotechnology and Bioengineering,* vol. 80, 2002, pp. 405-413.
[20] Markx, G. H., and Davey, C. L., The dielectric properties of biological cells at radiofrequencies: applications in biotechnology, *Enzyme and Microbial Technology,* vol. 25, 1999, pp. 161-171.
[21] Dabros, M., Dennewald, D., Currie, D., Lee, M. *et al.*, Cole–Cole, linear and multivariate modeling of capacitance data for on-line monitoring of biomass, *Bioprocess and Biosystems Engineering,* vol. 32, 2009, pp. 161-173.
[22] Ehgartner, D., Sagmeister, P., Herwig, C., and Wechselberger, P., A novel real-time method to estimate volumetric mass biodensity based on the combination of dielectric spectroscopy and soft-sensors, *Journal of Chemical Technology & Biotechnology,* vol. 90, 2015, pp. 262-272.
[23] Vojinović, V., Cabral, J. M. S., and Fonseca, L. P., Real-time bioprocess monitoring: Part I: In situ sensors, *Sensors and Actuators B: Chemical,* vol. 114, 2006, pp. 1083-1091.
[24] Lübbert, A., "Gas Hold-up," *Encyclopedia of Bioprocess Technology*: John Wiley & Sons, Inc., 2002.
[25] Thompson, M. L., and Kramer, M. A., Modeling chemical processes using prior knowledge and neural networks, *AIChE Journal,* vol. 40, 1994, pp. 1328-1340.
[26] von Stosch, M., Oliveira, R., Peres, J., and Feyo de Azevedo, S., Hybrid semi-parametric modeling in process systems engineering: Past, present and future, *Computers & Chemical Engineering,* vol. 60, 2014, pp. 86-101.
[27] Günther, M., and Velten, K., *Mathematische Modellbildung und Simulation: Eine Einführung für Wissenschaftler, Ingenieure und Ökonomen*: Wiley, 2014.
[28] Shimizu, K., A tutorial review on bioprocess systems engineering, *Computers & Chemical Engineering,* vol. 20, 1996, pp. 915-941.
[29] Komives, C., and Parker, R. S., Bioreactor state estimation and control, *Curr Opin Biotechnol,* vol. 14, 2003, pp. 468-74.
[30] Sagmeister, P., Wechselberger, P., Jazini, M., Meitz, A. *et al.*, Soft-sensor assisted dynamic bioprocess control: Efficient tools for bioprocess development, *Chemical Engineering Science,* vol. 96, 2013, pp. 190-198.

This article is protected by copyright. All rights reserved.

Accepted Article

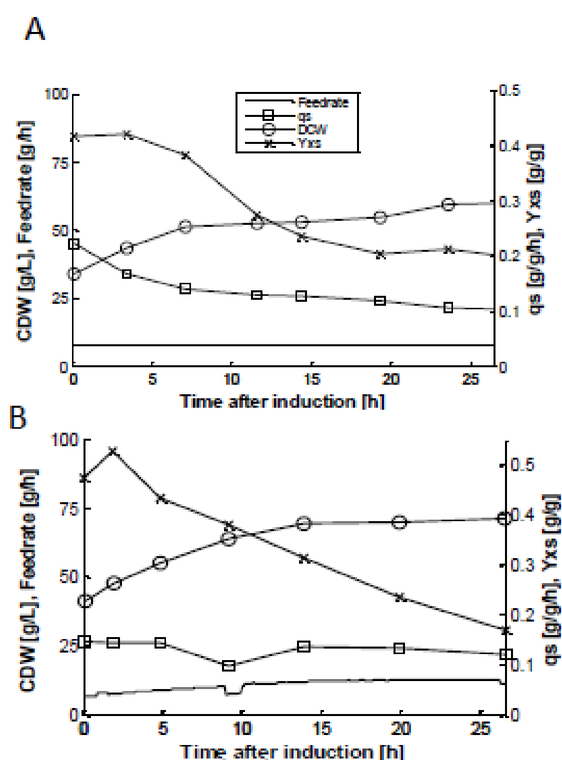
[31] Kadlec, P., Gabrys, B., and Strandt, S., Data-driven Soft-sensors in the process industry, *Computers & Chemical Engineering,* vol. 33, 2009, pp. 795-814.
[32] van der Heijden, R. T. J. M., Heijnen, J. J., Hellinga, C., Romein, B. *et al.*, Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates, *Biotechnology and Bioengineering,* vol. 43, 1994, pp. 3-10.
[33] Herwig, C., Marison, I., and von Stockar, U., On-line stoichiometry and identification of metabolic state under dynamic process conditions, *Biotechnology and Bioengineering,* vol. 75, 2001, pp. 345-354.
[34] Wilms, B., Hauck, A., Reuss, M., Syldatk, C. *et al.*, High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the Escherichia coli rhaBAD promoter, *Biotechnology and Bioengineering,* vol. 73, 2001, pp. 95-103.
[35] Korz, D. J., Rinas, U., Hellmuth, K., Sanders, E. A. *et al.*, Simple fed-batch technique for high cell density cultivation of Escherichia coli, *Journal of Biotechnology,* vol. 39, 1995, pp. 59-65.
[36] Lee, S. Y., High cell-density culture of Escherichia coli, *Trends in Biotechnology,* vol. 14, 1996, pp. 98-105.
[37] Sagmeister, P., Wechselberger, P., and Herwig, C., Information Processing: Rate-Based Investigation of Cell Physiological Changes along Design Space Development, *PDA J Pharm Sci Technol,* vol. 66, 2012, pp. 526-41.
[38] Willmott, C. J., On the validation of models, *Physical geography,* vol. 2, 1981, pp. 184-194.
[39] Aehle, M., Simutis, R., and Lubbert, A., Comparison of viable cell concentration estimation methods for a mammalian cell cultivation process, *Cytotechnology,* vol. 62, 2010, pp. 413-22.
[40] De Meyer, T., Hemelsoet, K., Van Speybroeck, V., and De Clerck, K., Substituent effects on absorption spectra of pH indicators: An experimental and computational study of sulfonphthaleine dyes, *Dyes and Pigments,* vol. 102, 2014, pp. 241-250.
[41] Fan, D. D., Luo, Y., Mi, Y., Ma, X. X. *et al.*, Characteristics of fed-batch cultures of recombinant Escherichia coli containing human-like collagen cDNA at different specific growth rates, *Biotechnol Lett,* vol. 27, 2005, pp. 865-70.
[42] Henes, B., and Sonnleitner, B., Controlled fed-batch by tracking the maximal culture capacity, *Journal of Biotechnology,* vol. 132, 2007, pp. 118-126.
[43] Akesson, M., Karlsson, E. N., Hagander, P., Axelsson, J. P. *et al.*, On-line detection of acetate formation in Escherichia coli cultures using dissolved oxygen responses to feed transients, *Biotechnol Bioeng,* vol. 64, 1999, pp. 590-8.
[44] de Hollander, J. A., Kinetics of microbial product formation and its consequences for the optimization of fermentation processes, *Antonie Van Leeuwenhoek,* vol. 63, 1993, pp. 375-81.
[45] Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M. *et al.*, Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview, *Microb Cell Fact,* vol. 7, 2008, pp. 11.
[46] Lindner, A. B., and Demarez, A., Protein aggregation as a paradigm of aging, *Biochim Biophys Acta,* vol. 1790, 2009, pp. 980-96.
[47] Wang, P., Robert, L., Pelletier, J., Dang, W. L. *et al.*, Robust growth of Escherichia coli, *Curr Biol,* vol. 20, 2010, pp. 1099-103.

This article is protected by copyright. All rights reserved.

Accepted Article

FIGURE 1: Dynamic biomass yield decline substantiates the dynamic metabolic stoichiometry in induction phase:
Experiments using strain A with different substrate feeding profiles (Feedrate [g/h]): (A) constant volumetric feeding rate, (B) constant physiology based on a real time biomass sensing approach. Both experiments share a common average metabolic activity in terms of the specific substrate uptake rate (qs [g/g/h]) and illustrate the decline of the biomass yield (Yxs [g/g]) over time.

FIGURE 1

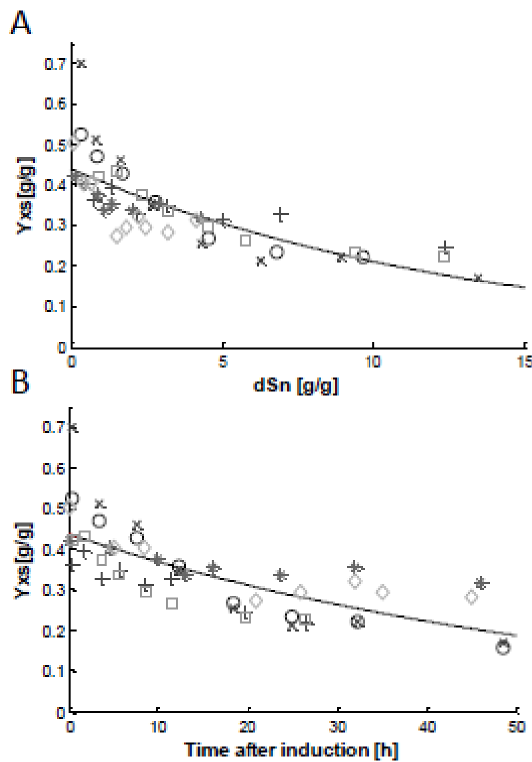


This article is protected by copyright. All rights reserved.

Accepted Article

FIGURE 2: dSn is a more feasible common denominator for biomass yield trajectories than time: Cumulative biomass yield (Yxs) trajectories over time/dSn for 6 independent experiments of strain A (Strain B Supplemental 2); every class of symbols represents an independent experiment (A): Yxs [g/g] over time after induction [h] displays a low quality (R2 = 0.45); (B): Yxs [g/g] over dSn [g/g] displays a better quality of fit (R2 = 0.70)]; Experiments were conducted at different levels of specific substrate uptake rate to maximize the range of contained metabolic activity in terms of qs_{mean}

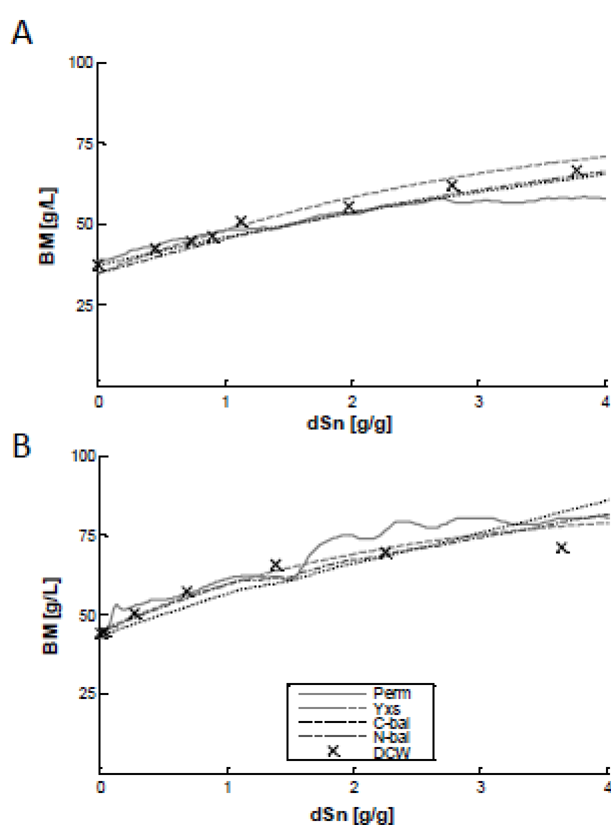
FIGURE 2



This article is protected by copyright. All rights reserved.

FIGURE 3: Performance of biomass sensing methods appears qualitatively impacted by process dynamics:
Data of strain A, further data as well as data on strain B can be found in Supplemental 5; Experiments differ in the physiological descriptor of the physiologic substrate uptake rate qsmean [g/g/h]; (A) low qs < 0.15 [g/g/h]; (B) high qs >0.25 [g/g/h]. The offline biomass quantification is depicted as asterisks (DCW) as point of reference for the comparison of the biomass estimation approaches. The continuous lines are the result of the biomass estimation based on the method as indicated: first principle soft-sensor C balance (C-bal) and N-balance (N-bal), hybrid soft-sensor (Yxs), and hard type sensor (Perm). For the sake of comparability experiments are displayed within a common window of the cumulative and normalized fed substrate (dSn [g/g]).

FIGURE 3

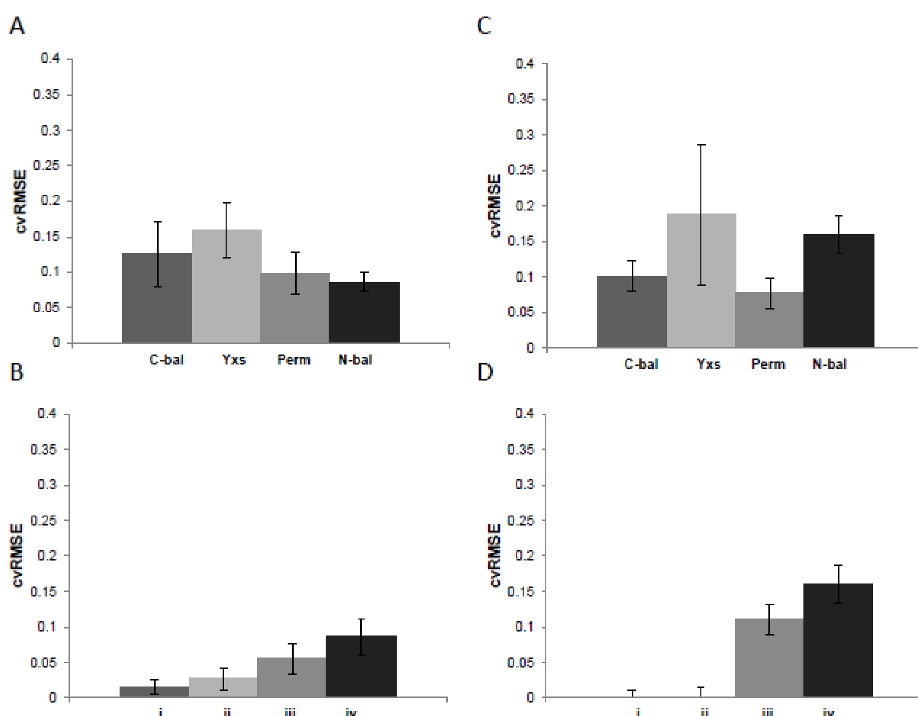


Accepted Article

This article is protected by copyright. All rights reserved.

FIGURE 4: Combinatory approaches for biomass sensing allow an extensive reduction of noise: Error of estimation in respect of the estimation, (A) cvRMSE of biomass estimation methods: first principle soft-sensor C balance features an error of 12.6 % (C-bal) and N-balance features an error of 8.6 % (N-bal), hybrid soft-sensor features an error of 15.9 % (Yxs), and hard type sensor based features an error of 9.8 % (Perm) of strain A, n=9; (B) cvRMSE of biomass estimation methods: first principle soft-sensor C balance features an error of 10.1 % (C-bal) and N-balance features an error of 15.9 % (N-Bal), hybrid soft-sensor features an error of 18.8 % (Yxs), and hard type sensor based features an error of 7.7 % (Perm) of strain B, n=9 ; (B), (D); Combinations of biomass estimation approaches by weighted average according the error of each method for one exemplary experiment n=1 (qs=0,34), all biomass estimation methods (i), soft-sensor and hard type sensor estimation methods (ii), weighted average soft-sensor methods (iii), first principle soft-sensor N-balance (iv); (B) strain A (D) strain B;

FIGURE 4



Accepted Article

This article is protected by copyright. All rights reserved.

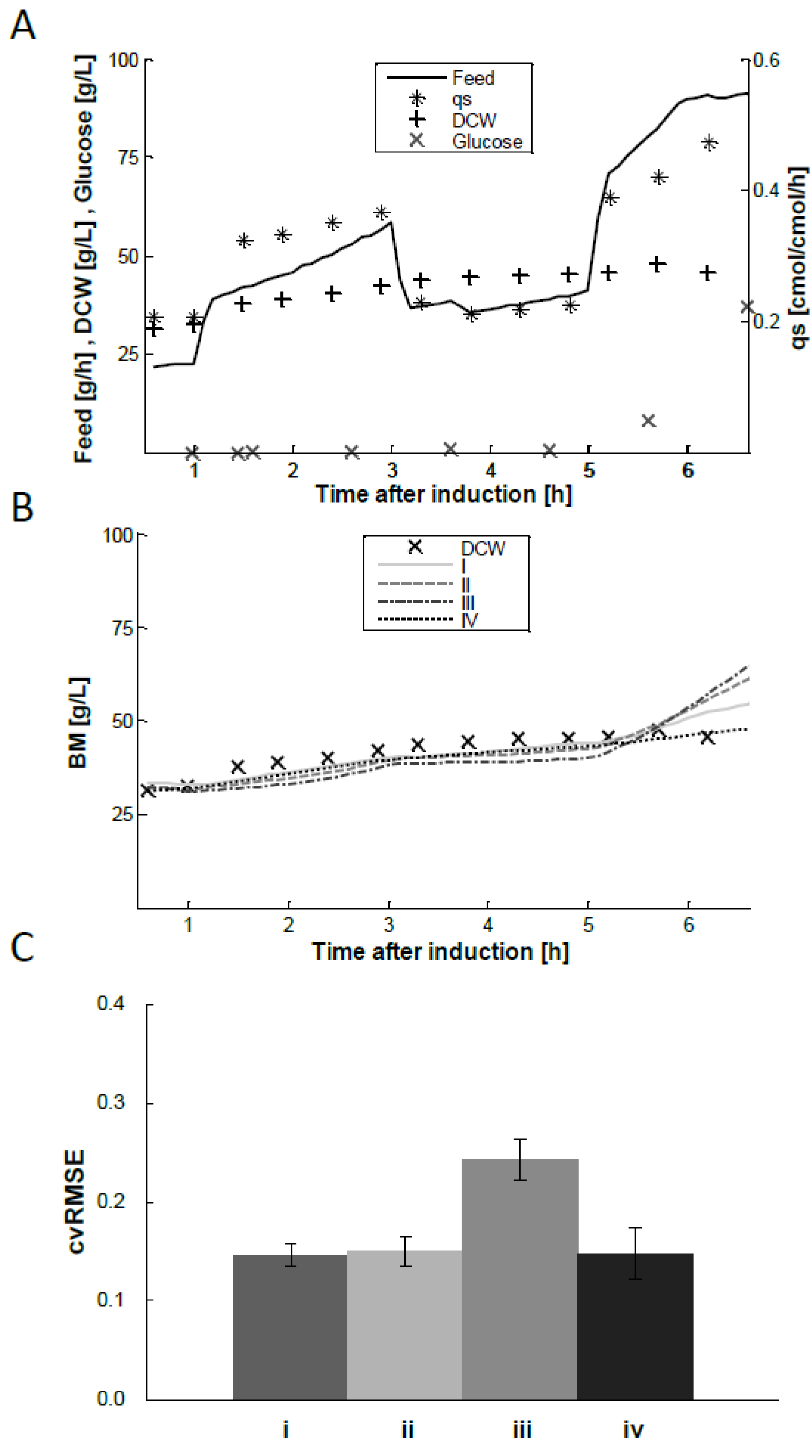
Accepted Article

FIGURE 5: Process dynamics for robustness testing substantiate the robustness of the weighted average approach: All data from strain B; (A) Different levels of qs (0.4 g/g/h for high and 0.2 g/g/h low level) are targeted to challenge the biomass estimation methods by process dynamics. Accumulation at 5.5 h after induction constitutes a significant process event which consequently impacts biomass estimation; (B) Biomass offline verification (DCW [g/l]) and weighted average estimations according to the error of each method, all biomass estimation methods (i), soft-sensor and hard type sensor estimation methods (ii), weighted average soft-sensor methods (iii), first principle soft-sensor N-balance (iv). (C) cvRMSE of weighted average combinations according to the error of each method, all biomass estimation methods (i), soft-sensor and hard type sensor estimation methods (ii), weighted average soft-sensor methods (iii), first principle soft-sensor N-balance (iv);

This article is protected by copyright. All rights reserved.

Accepted Article

FIGURE 5



This article is protected by copyright. All rights reserved.

Table 1 quantification of noise of real time biomass estimation does not indicate a significant impact of strain/metabolic activity. Comparison of cvRMSE of different biomass sensing approaches (n=3): C-balance (C-bal); cumulative biomass yield trajectory (Yxs); Hard Type Sensor (Perm); N-balance (N-bal); Strain A: the impact of qs level (low/mid/high) was insignificant ($p = 0.597$, $\alpha = 0,05$), the method of biomass estimation insignificantly influence estimation accuracy ($p = 0.115$, $\alpha = 0,05$); Strain B:, the impact of qs level (low/mid/high) was insignificant ($p = 0.722$, $\alpha = 0,05$), the method of biomass estimation insignificantly influence estimation accuracy ($p = 0.116$, $\alpha = 0,05$);

	high qs		mid qs		low qs	
	Biomass range [g/L]	cvRMSE [%]	Biomass range [g/L]	cvRMSE [%]	Biomass range [g/L]	cvRMSE [%]
Biomass estimation: Strain A						
C-bal	42.8-85.2	16.9	39.8-76.8	13.0	34.7-57.9	7.8
Yxs		19.0		11.5		17.0
Perm		7.6		8.9		13.2
N-bal		7.4		8.4		10.0
Biomass estimation: Strain B						
C-bal	29.1-76.1	11.8	22.9-49.3	7.8	28.7-49.8	10.9
Yxs		29.9		10.8		15.6
Perm		9.2		8.7		5.3
N-bal		13.2		18.5		16.2

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article

Table 2: Assessment table of the illustrated biomass sensing approaches; Evaluation ranges from not applicable (-) over limited performance (+); good performance (++) to superior performance (+++); Cost of investment for installation (Cost); Effort and necessary adaptions to transfer the approach to another strain (Transferability); accuracy in terms of the average cvRMSE (Accuracy); Ratio of information to effort in terms of method complexity and data treatment (Information/effort); First principle soft-sensor based on the C balance (C-bal); First principle soft-sensor based on the N-balance (N-bal), hybrid soft-sensor (Yxs), hard type sensor (Perm); combination of soft-sensor, hard type and hybrid soft-sensor (i) combination of soft-sensor and hard type sensor estimation (ii), combination of soft-sensor methods (iii)

	Cost	Transferability	Accuracy	Information/effort
C-bal	-	+++	++	++
N-bal	-	+++	++	++
Yxs	-	+	+	+
Perm	+++	++	+	++
i	+++	+	+++	+
ii	+++	++	+++	+++
iii	+	+++	+	++

This article is protected by copyright. All rights reserved.

Discussion

As analytics constitute the backbone of every physiological conclusion, the accuracy and information content of analytical methods used for the quantification of physiology and its phenomena is of utmost importance. In this section, analytical methods have been re-assessed, verified or established to quantify and detect physiological phenomena as protein release, protein aggregation and biomass growth.

In the context of protein quantification we demonstrated a massive error of measurement for a commonly applied method. Regarding information content, an error greater than 200 % renders a method basically worthless. This example highlights the necessity to implement proper controls and to assess method sensitivity and accuracy under conditions as close as possible to the final application (1a).

As second example of the importance of sensitive analytical methods we addressed the analytical topic of sizing cytosolic protein aggregates (inclusion bodies). Using transmission electron microscopy as orthogonal verification method is was possible to substantiate nano particle tracking analysis as method for high throughput IB sizing (1a). Based on the results it was concluded that by using industrial relevant cell disruption technologies, as high pressure homogenization, native inclusion body properties as e.g. size are not preserved. Also in this example the thorough assessment of the capabilities of the analytical method led to far reaching conclusions which actually lead to a revision of the basic hypothesis.

The elucidation of the role of physiology in the context of substrate supply and protein expression requires precise control of physiological variables throughout induction phase. Since physiological variables are normalized by the biomass content, this approach requires real time biomass estimation. In this context a combination of the first principle mass balances and a hard type sensor facilitated a reduction of biomass estimation error by 85 % (1b). This finding is especially relevant for early bioprocess development since this state is mainly characterized by the lack of strain specific knowledge. Hereby, the proposed method meets all needs of early bioprocess development as transferability and simplicity while granting a high level of accuracy.

The demands towards analytical accuracy are higher in the context of physiological bioprocess development as compared to technological bioprocess development. Technological bioprocess development does simply rely on less analytical input data. In contrast physiologic variables comprise by definition higher amounts of information but are in turn based on a greater deal of analytics. Despite the perceived gap of analytical requirements for physiological bioprocess development, established methods can mostly be employed if the sensitivity has been assessed within the final environment of application.

2) Effectivity of physiological control

Various approaches of process control have been illustrated in literature, whereas especially the control of the substrate supply has been proven as highly promising targeting maximum productivity [31, 48-50]. Technological as well as physiological feeding strategies have been widely employed, but a sound science comparison and investigation of the pros and cons is still missing. To foster the adoption of physiological bioprocess development we illustrate the transition from technological to physiological control approaches, investigating the process step by step (Figure 1).

The following section outlines how to quantify physiology within well-defined process phases and investigates physiological process evaluation and its benefit for technologically controlled processes. Consequently, physiologically controlled experiments are directly compared to technologically controlled processes in order to assess the benefit of physiological control. The use of physiological variables as factors for design of experiments requires boundaries called physiological capacities. The quantification of such physiological capacities requires additional strain characterization experiments, which increase the perceived effort for physiological bioprocess development. Nevertheless, substrate accumulation is common in late induction phases, especially in the context of physiologically controlled processes. Since such a, currently unpredictable, event does not represent thorough bioprocess understanding, the physiological reason of substrate accumulation must be investigated. And finally, a novel control approach is introduced to increase robustness and efficiently avoid substrate accumulation within induction phase to substitute laborious strain characterization experiments.

In a nutshell, volumetric constant and physiological feeding profiles are compared by transformation of raw process data into physiological data. Additionally, the impact of substrate supply on productivity as well as the nature of physiological capacities in respect to metabolic activity is analyzed. On the basis of this analysis the effectivity of physiological control approaches and physiological bioprocess development shall be evaluated.

2) Advantages/challenges of physiological process development

Pre/Post induction phase
Independent post induction phase benefits from low substrate supply

Physiological Feeding
Physiological process control boosts titers by >90%

q_{Scrit} – physiological capacity to metabolize substrate
q_{Scrit} is heavily correlated to time and metabolic activity

Closed loop physiological control
Automated qscrit sensing avoids substrate accumulation

Figure 1: Structure of the process control section: The interrelation of the pre and post induction phase is investigated in respect to the substrate supply. In the subsection of Physiological Feeding volumetric constant feeding strategies are compared to physiologically constant approaches. Subsequently the dependency of the physiological capacity to metabolize substrate is discussed. Before the pieces are brought together in a physiological closed loop approach for online sensing of physiological capacities

i. Pre and Post induction Phase

Early process development is characterized by the lack of strain specific knowledge. As a consequence process development is strongly dependent on expert knowledge. The selection of process parameters for investigation is commonly based on a risk assessment approach [51]. The initially unclassified process parameters are consequently categorized into critical and non-critical process parameters [18] by empirical investigation within an DoE. The majority of DoEs is using product related variables as solely response for data evaluation. To facilitate knowledge feedback for subsequent bioprocess development routines it is indicated to include more product independent variables e.g. physiological variables. Physiological variables condensed to a single numeric value within a well-defined process phase, might display a higher degree of transferability. Only a sound science based phase definition for data evaluation would facilitate physiologically transferable conclusions. Comparing various technical (volumetric constant) feeding profiles, within a physiological meaningful phase definition, the approach and a workflow of information feedback is illustrated and the benefits of the approach are discussed.

Bioprocess development workflow: transferable physiological knowledge instead of technological correlations

Reichelt W.N.[a], Haas F.[b], Sagmeister P.[b], Herwig C.[a]

[a] Christian Doppler Laboratory for Mechanistic and Physiological Methods for Improved Bioprocesses, Institute of Chemical Engineering, Vienna University of Technology, Getreidemarkt 9/166, A-1060 Vienna, Austria
[b] Research Division Biochemical Engineering, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorfer Strasse 1A/166-4, 1060 Vienna, Austria;

[*]Corresponding author: christoph.herwig@tuwien.ac.at; Tel.: +43-(0)1-58801-166400

Abstract

Microbial bioprocesses need to be designed to be transferable from lab scale to production scale as well as between setups. Although substantial effort is invested to control technological parameters, usually the only true constant parameter is the actual producer of the product: the cell. Hence, instead of solely controlling technological process parameters the focus should be increasingly laid on physiological parameters. This contribution aims at illustrating a workflow of data life cycle management with special focus on physiology. Information processing condenses data into physiological variables, while information mining condenses the variables further into physiological descriptors. This basis facilitates data analysis for a physiological explanation for observed phenomena in productivity.

Targeting transferability we demonstrate this workflow using an industrially relevant *E. coli* process for recombinant protein production and were able to substantiate the following three points: 1) The post-induction phase is independent in terms of productivity and physiology from the pre- induction variables specific growth rate and specific biomass. 2) The specific substrate uptake rate during induction phase was found to significantly impact the maximum specific product titer. 3) The time point of maximum specific titer can be predicted by an easy accessible physiological variable: While the maximum specific titers were reached at different time points (19.8 +/- 7.6 h), those maxima were reached all within a very narrow window of cumulatively consumed substrate dSn (3.1 +/- 0.3 g/g). Concluding, this contribution provides a workflow on how to gain a physiological view on the process and illustrates potential benefits.

Introduction

Extensive effort is invested into bioprocess development of recombinant protein production in *Escherichia coli*. Owned to *E. coli* inherent attributes of fast growth, simplistic genetic engineering, inexpensive media and the possibility of high cell density cultivations, *E. coli* is one of the most intensively characterized and industrially exploited prokaryotic production platforms for heterologous protein production [1,2].

Progressively expiring patent protection brings forward generic drug industry and heats up competition for the most efficient and robust bioprocesses [3]. While competition is intensifying, the quality by design (QbD) initiative additionally increases the demands concerning bioprocess development [4]. This cumulates in the necessity for highly efficient bioprocess development routines with maximized information to effort ratio. In bioprocess development investigated but insignificant factors, strain specific but hardly transferable correlations as well as sampling in areas other than of the maximum productivity consequently are highly undesirable scenarios. Within this contribution we illustrate a generic applicable workflow for bioprocess development which aims to maximize the information to effort ratio. This workflow is exemplified addressing three topics: (1) The establishment of bioprocess platform knowledge to minimize relevant factors for investigation; (2) A physiological explanation for observed variance to allow greater transferability between setups. (3) Prediction of area of maximum specific titer to minimize sampling effort.

(1) Establishment of bioprocess platform knowledge

Within bioprocess development the number of investigated process parameters is the biggest cost driver in terms of effort for bioprocess development. The selection of eligible factors for further investigation is commonly based on a risk assessment approach based on expert and theoretic literature knowledge [5,6]. Hereby, the establishment of bioprocess platform knowledge would facilitate omitting otherwise routinely tested factors for consecutive products. The establishment of platform knowledge is demanding [7] but facilitates effort reduction in case of product changeover. Based on platform specific prior knowledge the number of factors eligible for investigation can be greatly reduced.

Industrial bioprocesses are commonly subdivided into the phase of biomass accumulation (pre-induction phase) and the phase of product formation (post-induction phase) [8]. In pre-induction phase the most relevant variables from a process technological point of view, are the specific growth rate and biomass concentration at induction [9]. The dependency of the pre-induction phase on the post-induction phase was postulated [10-12], whereas other authors did not find a dependency [9,13]. Since literature remains inconclusive in this point, the investigation of the respective interrelation remains an obligation. By the means of the obligatory investigation of the interrelation of pre- and post-induction phase an exemplary but generally applicable workflow for platform knowledge generation shall be demonstrated. Using In this work, we investigate another product in the same strain background as in the work of Wechselberger et al. [9], which will aid to extend individual findings to platform knowledge [14].

(2) A physiological explanation for observed variance

Bioprocess data evaluation routines conventionally correlate factor set points of technological process parameters as pH or temperature with product related variables as response [15,16]. During process transfer or scale up the reactor

system is usually altered, while the actual cellular producers remain constant. Hereby, the focus on cell physiology promises greater transferability than the focus on technical process parameters. Nevertheless, although data is sufficiently available, data exploitation rarely addresses physiologic interrelations especially in an industrial process development environment.

Literature has outlined an information processing approach within which data is condensed into physiologic variables to describe the physiologic state of the cell [17]. In a context of bioprocesses the term "information mining" defines the approach of describing physiology by single numeric values, corresponding to the average of a time dependent physiologic variable within a defined process phase - physiological descriptors. To increase knowledge transferability, information mining targets the use of physiology descriptors rather than to technological process parameters to explain observed variance in productivity. Data evaluation routine is commonly conducted to finalize a cycle of process development, but within a platform system this routine can be used initially for factor selection. Similarly to upcycling this re-assessment of historic process data may supplement or even replace conventional risk assessments, and highlights the relevance of historic process data for process development.

(3) Prediction of area of maximum specific titer

The process analytical technology (PAT) in the framework of the QbD initiative is calling for timely measurements and consequently time-resolved offline sampling [4]. Time resolved sampling features increased sampling effort but offers deeper insight into physiologic and product formation kinetics. In contrast, end-point sampling bears the risk of missing the maximum product concentration, since product degradation processes during the fermentation are common [18,19]. However, in an industrial process development setting it is hardly feasible to continuously sample in high frequently. Hence, it is necessary to predict the important phases in order to maximize the information per sample ratio, preferably based on an easily accessible physiologic relevant variable.

Roadmap

Investigating the relationship of independent pre- and post-induction phase, we illustrated the establishment of platform knowledge exemplarily. Using the same strain but different product, the comparison of the main conclusions to the work of Wechselberger et al. [9] will extend the scope of findings to platform knowledge. A two factor screening design was conducted to investigate the impact of pre-induction specific growth rate and the biomass concentration at the point of induction on post-induction phase.

The information processing approach condenses data to information. Subsequently information mining condenses the information into physiological descriptors to facilitate in-depth physiological process understanding (Figure 1). The physiological descriptor most significant to explain observed variance in titers is investigated further. Additionally a physiological variable determining the area of maximum titer shall be identified within subsequent data analysis.

Materials and methods

Host

A modified K12 *E. coli* strain (kindly provided by Lonza Ltd., Visp, Switzerland) was used as a model expression system for the project. The strain features a rhamnose-inducible expression system (rhaBAD promoter). The recombinant protein product was a Fab antibody. Since the strain is unable to utilize rhamnose as a C-source, a one-time addition of inducer was sufficient.

Media

For fermentation a predefined media was used [20].

Bioreactor setup

Fed-batch experiments conducted out in a DASGIP multi-bioreactor system consisting of four glass bioreactors with a working volume of 2 l each (Eppendorf; Hamburg, Germany). The reactors are equipped with baffles and three disk impeller stirrers. The DASGIP control software v4.5 revision 230 was used to control the process parameters: pH (Hamilton, Reno, USA) and pO_2 (Mettler Toledo; Greifensee, Switzerland; module DASGIP PH4PO4), temperature and stirrer speed (module DASGIP TC4SC4) and aeration (module DASGIP MX4/4). The pH control was facilitated by 12.5% NH_4OH base addition by the pump module DASGIP MP8. Feed was added using the same pump module. The reactors were sterilized at 121°C for 20 min. CO_2, O_2 concentrations in the off-gas were quantified by a gas analyzer (module DASGIP GA4) using the non-dispersive infrared and zircon dioxide detection principle, respectively. The gas flow was controlled by the gas mixing module DASGIP GA4.

Fermentation parameters

The pre-culture was inoculated in shake flasks from frozen stocks (100 ml in 1 l flasks). After approx. 17 h at 30°C and 200 rpm, a volume equivalent of pre-culture of 2.5% of the batch volume was used to inoculate the batch media. After inoculation of the batch medium (20 g/l C-source), the C-source in the batch medium was consumed within 12 h. The pre-induction feeding strategy was based on an exponential feed forward profile according to Equation 1. It describes the exponential increase of the timely feed flow rate $F_{(t)}$ in dependence of the initial flow rate F_0 and the specific growth rate μ.

Equation 1 Feed profile of pre-induction phase: $F_{(t)}$ [ml/h] is increased exponentially dependent on time t [h] and the specific growth rate μ [1/h].

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

Hereby the start flow rate F_0 depends on the amount of biomass present in the reactor and the targeted specific growth rate. The substrate concentration in the feed c_S and the biomass yield Y_{XS} is used to transform the biomass growth into a feed flow rate F_0.

Equation 2 Initial feed flow rate F_0 [ml/h] at the start of pre-induction phase: depends on biomass ($x_0 * V_0$) present in the reactor the specific growth rate μ [1/h] and the substrate concentration c_S [g/l] as well as on the biomass yield Y_{XS} [g/g]. The biomass yield Y_{XS} [g/g] was derived from dry cell weight measurements at the end of several batch fermentations.

$$F_0 = \frac{\mu * X_0}{c_S * Y_{XS}}$$

After the fed-batch, the culture was induced with rhamnose. Upon induction the volumetric constant feed rate F(t) [ml/h] is scaled based on the initial specific substrate uptake rate qs_{init} [g/g/h] and the amount of biomass at the end of fed batch X_{EFB} [g] (Equation 3). The post-induction feed rate was calculated according to equation 3 at the end of pre-induction [9]. The volumetric feed rate was left constant throughout induction phase. Dissolved oxygen levels (pO_2) were controlled over 25% by supplementing pure oxygen to the air. The pH was kept constant at 7 by adding 12.5% NH_4OH, which also served as nitrogen source. Temperature was set to 35°C for the whole process.

Equation 3 initial substrate uptake rate qs_{init} [g/g/h] calculated at the start post-induction phase: qsinit as setpoint serves as basis for calculating the volumetric constant feed rate $F_{(t)}$ [ml/h]depending on the feed concentration and the total biomass X_{EFB} [g]

$$qs_{init} = \frac{F_{(t)} * c_S}{X_{EFB}}$$

Offline analysis

- Biomass dry cell weight

  Biomass dry cell weight concentrations were gravimetrically quantified after drying for 72 h at 105°C. The initial biomass concentration, which was required for the calculation of F_0 (Eq. 2) was measured by photometric principle (OD 600 nm). Samples were diluted to the linear range of OD measurement < 0.8 and consequently converted to a biomass concentration by the use of an established linear regression. Given the better reproducibility, process data evaluation was based on dry cell weight measurements.

- Titer measurement

  Offline samples were centrifuged (4300 rcf, 10 min) and pellets were washed with distilled water and then stored at -20°C. Frozen pellets were re-suspended in 100 mM Tris, 10 mM Na-EDTA, pH 7.4 to a final volume of 20 ml and homogenized at 1400 ± 100 bar for 6 passages (Avestin EmulsiFlex, Ottawa, Canada).

  The product quantity was determined via a proprietary industrial protein G affinity chromatography method using a pH gradient. Since only correctly folded product is bound the measurement of product quantity is also regarded as a measure of product quality.

Data processing and evaluation

Calculation of metabolic rates and yield coefficients was conducted with Matlab 2012 b (Mathworks, Natick, Massachusetts, USA). Software was used for the information processing in terms of calculation of specific rates (Equations 8 and 9) and yield coefficients, as described elsewhere [9].

Information mining yields a physiological descriptor as single numerical value, which corresponds to the average of the timely variable calculated for all variables and experiments compared within the same window of calculation (t_{EFB} – t)

Equation 4 calculation of physiological desrciptors expemplary for qs. Within a defined window of calculation which is kept constant for all experiments of interest e.g. the average is calculated of a physiological variable of interest to yield a single numerical value – a physiological descriptor.

$$qs = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{t_{EFB}}^{n} qs_{(t)}$$

Statistical analysis

Since common DoE evaluation is based on the set points of the respective factors it does not take usual process deviation into account potentially masking effects triggered by potential set point deviations. To ensure the most realistic response to factor correlation, we used the actual met process variable as input instead of its mere set point.

Variables were tested for co-linearity with Datalab Version 3.5 (distributed by Epina http://datalab.epina.at/). Multi-linear regression models were fitted and analyzed via the statistics software MODDE® (Umetrics, Umeå, Sweden).

Results

Figure 1 outlines the proposed workflow for bioprocess development, aiming to increase the information to effort ratio. Usually, a risk assessment is conducted to select process parameters for further investigation. In our case, targeting the establishment of platform knowledge, the interrelation of pre- and post-induction phase was selected. A two factor screening design was designed to investigate the impact of pre-induction specific growth rate and the biomass concentration at the point of induction on post-induction phase (1a). By Information processing the process data is condensed into information – timely resolved physiological variables (1b). Consequently, information mining (2a) is utilized for further data condensation by calculating physiological descriptors within a given process phase (e.g. constant time window). The resulting descriptors are investigated concerning correlations to process parameters as well as to productivity for hypothesis generation. The identified descriptors can consequently be used as factors for subsequent experiments (2b). Another cycle of information processing (2c) and information mining (3) follows, to deepen the physiological bioprocess understanding and to verify or to refuse generated hypothesis. The following results and discussion will be structured according to the presented approach (Figure 1).

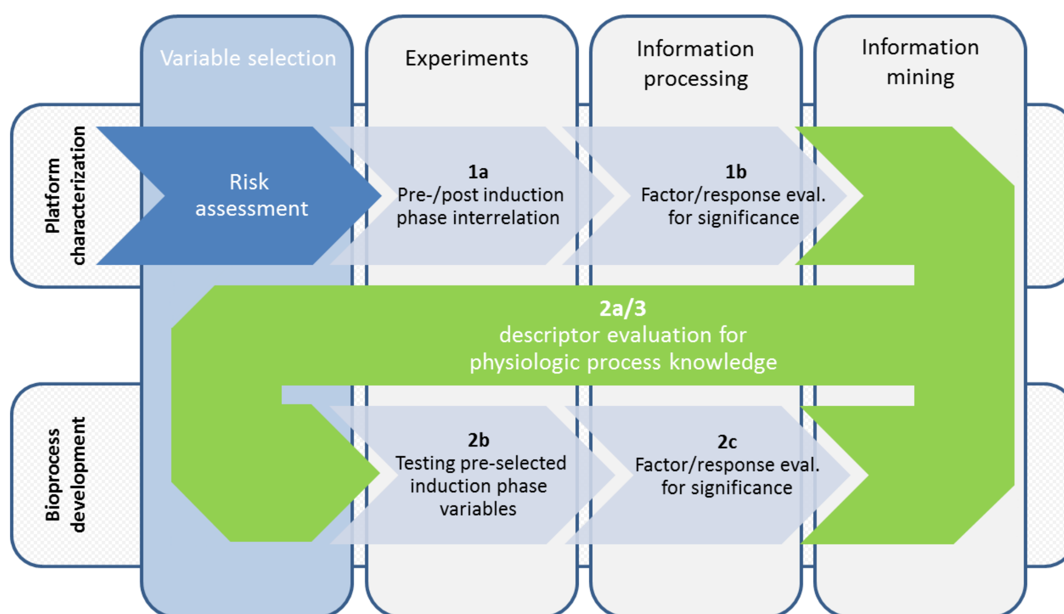


Figure 1: Flow chart of proposed method for efficient bioprocess development: The obtained data set resulting from the obligatory investigation of interrelation of pre- and post-induction phase (1a) is condensed by information processing (1b) and evaluated by information mining (2a). Consequently the most influential physiological descriptors are selected and investigated for their impact on product related variables (2b). In case the physiological descriptor is not directly accessible, for instance the biomass yield, a controllable correlated process parameter should be investigated instead. Established platform knowledge triggers the transition from platform characterization (1a, 1b) via 2a to bioprocess development (2a-2b). Consecutive experimentation within 2b generates data for the information processing approach. Information mining facilitates a continuous information based hypothesis generation to identify variables for consequent design of experiment and hypothesis verification

- ***Step 1a: Investigation of interrelation of pre- and post-induction phase***

To elucidate the impact of the pre-induction on the post-induction phase a DoE-screening design was conducted with a pre-induction specific growth rate of 0.08- 0.16 [1/h] and a biomass at induction of 20.7-44.6 [g/l] as factors (Figure 2 A). The post-induction volumetric constant feeding profile was based on qs_{init} (Equation 3), in all experiments (0.22 +/- 0.027 [g/g/h]).

- ***Step 1b: Information processing***

Using the achieved process values instead of the mere DoE set points as input variables for the evaluation no significant correlation was observed. Although significant variance was observed within the maximum specific titer (Figure 2 B) the observed variance could not be explained by the variance in biomass at induction (Figure 2 C) and pre-induction specific growth rate (Figure 2 D) ($p=0.685$, $\alpha=0.1$).

In accordance with the work of [9] no interrelation of pre- and post-induction was detected for this expression system. This confirmation of previous findings substantiates the prior knowledge. Upon further confirmation with other products, the finding can be regarded platform knowledge within the investigated system.

However, although the original DoE factors are not significant (Figure 2), variance in the maximum specific product titer was observed (0.0189 +/– 0.0037 [g/g]). This raises the question of a physiologic explanation.

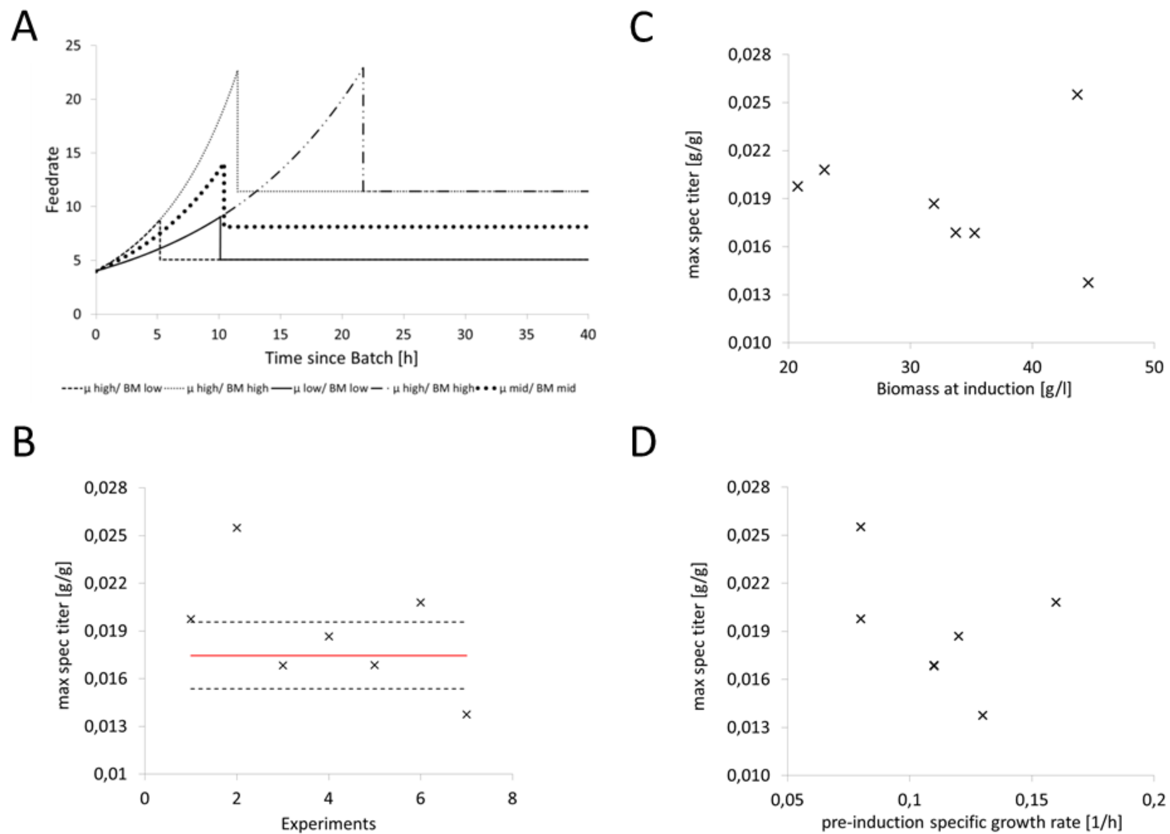


Figure 2 Biomass at induction and pre-induction specific growth rate do not impact maximum specific titers; (A) Exemplary feeding profiles with different pre-induction specific growth rates and different biomass concentrations at point of induction; (B) maximum specific titer [g/g] plotted for each experiment, continuous line indicates mean value (0.0189 g/g), dashed lines indicate the area covered by two times the standard deviation of the center points (+/- 0,001 g/g) ; (C) maximum specific titer [g/g] plotted over biomass concentration [g/l] at induction; (D) maximum specific titer plotted over the pre-induction specific growth rate [1/h], two experiments are highly congruent and optically appear as one experiment

- ***Step 2a: Information mining***

The observed variance in the specific titer is likely to have a physiological cause. An information mining approach is used to identify the physiological trigger for the variance in specific product titer. Using a PCA based on the physiological descriptors the majority of the observed variance (73 %) was analyzed.

Based on the PCA a strong negative correlation of the post-induction specific growth rate (µ) and the specific substrate uptake rate (qs) is inferred.

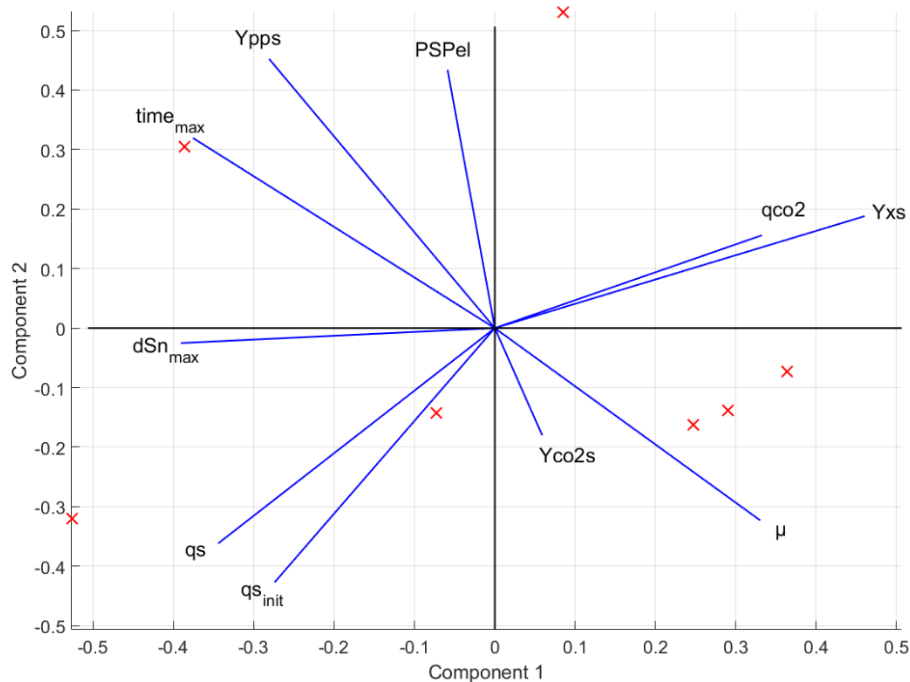


Figure 3 Specific rates as source of variance in the maximum specific product titer; Principal component analysis with scaled and centered values (n=7): μ specific growth rate [1/h]; qs specific substrate uptake rate [g/g/h]; PSPel maximum specific titer [g/g]; $time_{max}$ time until maximum specific titer [h]; dSn_{max} amount of normalized substrate metabolized until maximum specific titer [g/g]; qco2 specific carbon dioxide excretion rate [g/g/h]; Y_{co2s} carbon dioxide yield per substrate [g/g]; Ypps product yield per substrate [g/g]; Yxs biomass yield [g/g];

- ***Step 2b: Data based Design of Experiment***

Aiming to leverage productivity, the information mining approach facilitates factor selection based on empirical data, rather than on a theoretical risk assessment. According to the PCA (Figure 3) Y_{co2s} appears most promising. But of all correlated variables the specific substrate uptake rate is directly accessible via the feeding rate. Nevertheless, an industrial environment usually limits the feeding strategy to a constant volumetric flow rate in favor of simplicity and robustness. As a consequence the feeding rate is calculated based on qs_{init} and kept constant throughout the process. Calculating the appropriate feeding flow rate based on a specific variable requires knowledge of the respective biomass concentration at the point of induction. Given the greater mathematical dependency of μ on biomass measurements, using μ as basis for calculation bears a greater tendency to error propagation, than basing the calculation on qs_{init}. Given the importance of robustness it appears more reasonable to base the feeding strategy on qs than on μ. qs_{init} was investigated in the range of 0.088 – 0.323 [g/g/h] featuring 3 center points.

- ***Step 2c: Information processing***

In correspondence to the workflow (Figure 1) again information processing is applied to evaluate the impact of the physiological DoE factors. As illustrated in Figure 4 , a significant negative correlation of maximal specific titers and qs_{init} ($p(f) = 0.0225$, $\alpha=0.05$) is indicated. This finding also holds true if all 12 conducted experiments are included (Figure 6 B). The highest specific titers were found at the lowest achieved qs_{init} .

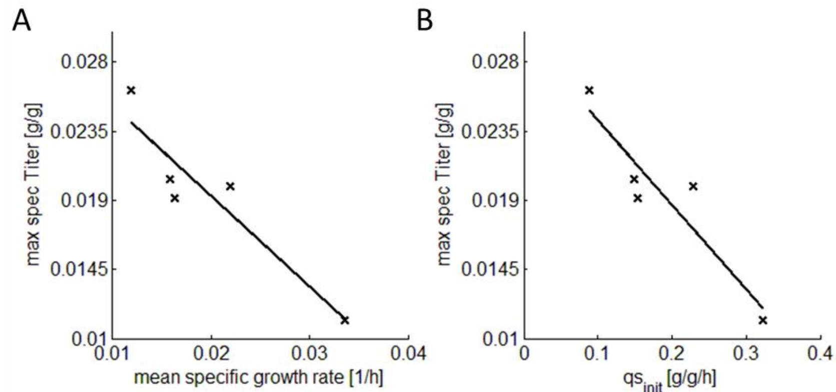


Figure 4 Negative correlation of μ and maximum specific titers; (A) The maximum specific titer [g/g] and mean post-induction specific growth rate [1/h] are highly significantly negatively correlated ($p(f)=0.0214$, $\alpha=0.05$); (B) The maximum specific titer [g/g] and qs_{init} the specific substrate uptake rate at induction [g/g/h] are highly significantly negatively correlated. ($p(f)=0.0225$, $\alpha=0.05$)

- ***Step 3: Information mining***

Following the information mining approach we investigated the influence physiology on productivity in order to find a possible physiological descriptor or predictor to explain the increase in titer.

Plotting the specific titer against time (Figure 5 C) visualizes the difference in the time point of reaching maximum specific titers. Maximum specific titers are reached much later in low qs_{init} experiments than in qs_{init} high experiments.

Hereby, the question is raised whether the substrate or time is the most relevant variable to describe productivity. For this reason the time was substituted with the cumulated amount of fed substrate on the x-axis (Figure 5 B). For the sake of transferability the amount of fed substrate was normalized on the amount of biomass at the end of pre-induction phase (x_{EFB}) yielding the variable dSn [g/g].

Equation 5 dSn corresponds to the integral of the feeding rate rS [g/h] since the end of fed batch t_{EFB} normalized on the biomass concentration at the end of fed batch x_{EFB}

$$\Delta Sn(t) = \frac{\int_{t_{EFB}}^{t} rS\, dt}{x_{\mathrm{EFB}}}$$

Interestingly, plotting the maximum specific titers against dSn (Figure 5 D) nicely aligns the product formation trajectories and the maximum specific titers can be found at similar dSn (3.1 +/- 0.3 g/g). The different physiological relevance is nicely illustrated by the experiment featuring the highest qs. While the decline in specific titer appears irritating over time around 11 h the trajectory appears plausible plotted against dSn, since all experiments appear enter a phase of declining specific titer at high dSn.

While maximum specific titers are reached at significantly different time points (Figure 5 C) (p(f) = 0.0022, α=0.05) they are reached at insignificantly different values of dSn (Figure 5 D) (p(f) = 0.7886, α=0.05). Consequently, insignificantly different amounts of substrate result in different amounts of product, featuring a significant difference in yield of product per substrate [g/g] (p(f)=0.0285, α=0.05) (data not shown). These findings establish dSn as a readily accessible parameter for prediction of maximum specific titers.

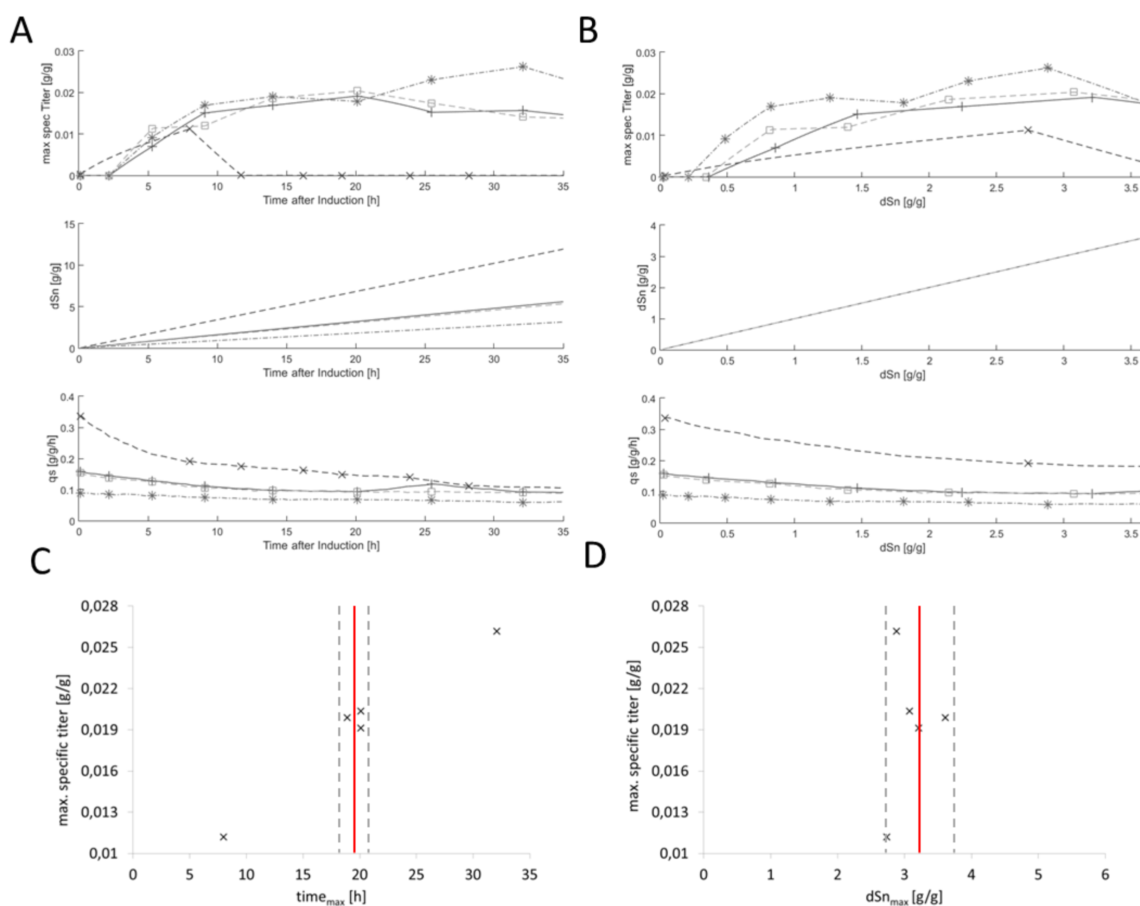


Figure 5 Maximum specific titers are reached insignificantly different dSn. (A) trajectories of specific titer [g/g cumulated fed substrate dSn [g/g] and specific substrate uptake rate qs [g/g/h] plotted against time after induction [h]; **(B)** trajectories of specific titer [g/g], cumulatively metabolized substrate dSn [g/g] and specific substrate uptake rate qs [g/g/h] plotted against dSn cumulated fed substrate normalized on the amount of biomass at induction [g/g]; **(C)** time points of maximum specific titer are significantly different; continuous line indicates mean value, dashed line indicates the area covered by two times the standard deviation of the center points; **(D)** dSn at maximum specific titer are insignificantly different; continuous line indicates mean value, dashed line indicates the area covered by two times the standard deviation of the center points

In terms of information mining all conducted experiments are used to reassess previous findings (Figure 6 A/B). It can be observed that using all available data sets substantiates the hypothesis of the negative correlation of qs_{init} as well as qs and the maximum specific titer. In this context Figure 6 C-E shall bring the difference of evaluating average and initial substrate uptake rates to the reader's attention.

Figure 6 compares the set of experiments for constant qs_{init} with the set of experiments targeting high variance in qs_{init}. Initially the different values of qs_{init} lead to strongly diverging amounts of available substrate per cell. But this difference in $qs_{(t)}$ fades as the trajectories converge increasingly. Due to the increase of biomass the amount of available substrate per cell and time declines ($qs_{(t)}$) (Supplemental 1) owned to the volumetric constant feeding profile (Figure 5 A). Although the initial difference in qs_{init} at (t=0) is substantial, the difference in the physiological descriptor qs over the whole experiment is only minor (Figure 6 C/D). While the process value of qs_{init} is based on a one point calculation at induction qs refers to the physiological descriptor, calculated over the whole process phase (Equation 4). Given the intention to maximize productivity on modulated the most influential variable e.g. qs_{init}. But as a matter of fact the induced variance in the factor qs_{init} (Figure 6 C qs_init variation vs. qs_init constant) lead to a comparably low amount of observed variance in the response of maximum specific titer (Figure 6 E qs_init variation vs. qs_init constant) .

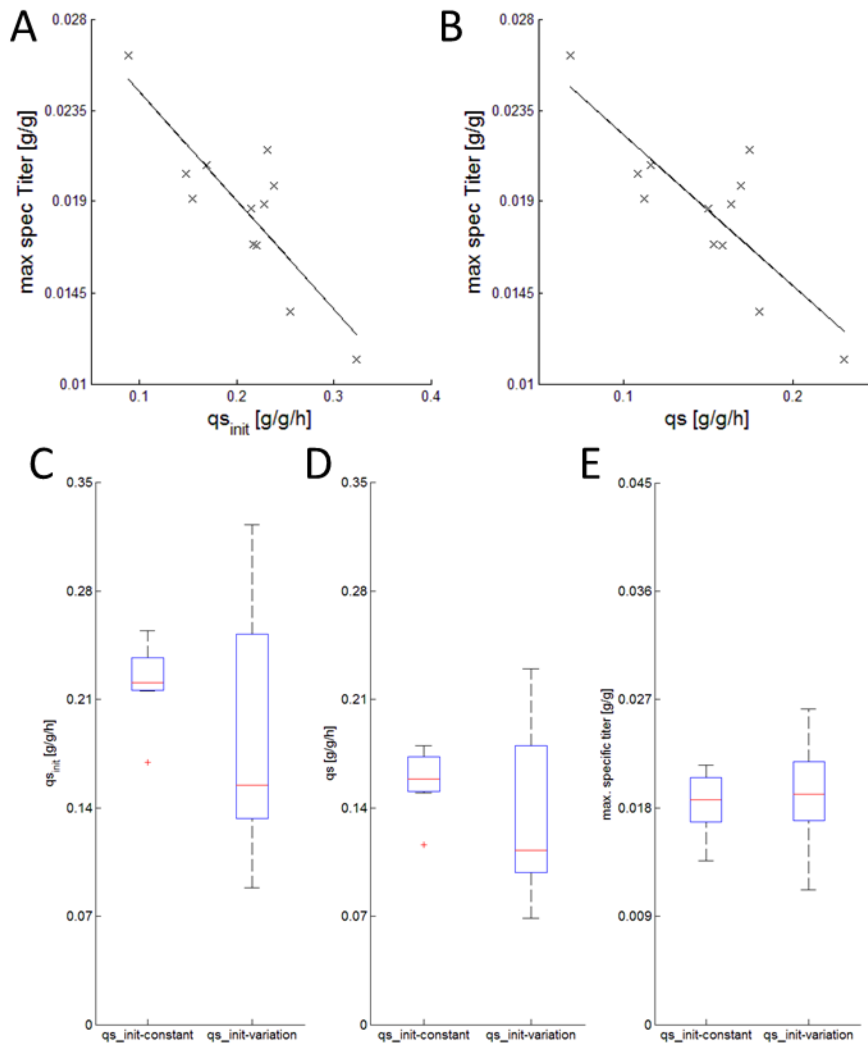


Figure 6 The specific substrate uptake rate impacts maximum specific titers in all experiments significantly; All calculations have been conducted based on a common window of calculation based on dSn. (A) shows the correalation of initial substrate uptake rate and maximum specific titers; (B) shows the correalation of mean substrate uptake rate and maximum specific titers; (C-E) comparison of observed and induced variance within the conducted experiments (n=12) qs_init-constant labels the group of experiments conducted with an equal qs_{init}; qs_init variation labels the group of experiments conducted to investigate the impact of different volumetric constant feeding regimes; (C) shows the induced variance in qs_{init} [g/g/h]; (D) shows the calculated variance in qs [g/g/h]; (E) shows the resulting variance in maximum specific titers [g/g];

Discussion

The goal of this contribution was to outline a workflow on how to expand the scope and to leverage transferability of strain specific findings. In general the workflow emphasizes the necessity of profound data evaluation beyond the conventional DoE factor set point and response evaluation. Data reduction within information processing and information mining is necessary but a delicate step. Single numerical process values descriptors of DoE factors as well as for physiological variables are only comparable if calculated within a defined process phase. Nevertheless, this approach requires a standardized evaluation routine as well as standards for information processing and information mining. It is important to assess experiments as close to cell physiology as possible by focusing on process values but it shall be noted that without quality control of the input data error propagation can significantly impact the final evaluation. Nevertheless, focusing on physiology instead of technological interrelations, the generated process knowledge is better transferable. Using the outlined approach we were able to substantiate the following three findings:

Information mining facilitates a continuous information based hypothesis generation to identify variables for consequent design of experiment and hypothesis verification.

<u>(1) The establishment of bioprocess platform knowledge to minimize relevant factors for investigation</u>

On the basis of the work of Wechselberger et al. [9] the interrelation of pre- and post- induction phase for the same strain but a different product for further investigation (Figure 1 1a-1b) were selected. In accordance to the findings of Wechselberger et al. no significant interrelation of pre- and post-induction phase was found. This fact increases the degree of freedom for optimization and scheduling of equipment occupancy especially in production stage. By repeatedly testing the found interrelations with different products and consequently even with different expression systems, the scope of the initially strain and product specific finding can be extended to platform knowledge.

<u>(2) A physiological explanation for observed variance to allow greater transferability between setups</u>

Despite the insignificance of the investigated factors of pre-induction phase, significant variance was observed in the DoE response, the maximum specific titer. The consequent search for a trigger was based on the condensed information of physiological descriptors and pointed strongly towards an implication of qs. Nevertheless, instead of starting from scratch this circumstance facilitated subsequent factor selection (Figure 1, 2 a-c). The following experiments based on the modulation of qs_{init} successfully increased the variance in titer further. This result substantiates the illustrated approach of factor selection based on historic data. Nevertheless, process development approach based on qs_{init} is strictly limited by the maximum substrate uptake rate especially at the point of induction. Judging from the ratio of induced variance of the factor and observed variance in the response, the modulation of qs appears to facilitate greater variance in the response but requires dynamic feed rate adaptation.

(3) Prediction of area of maximum specific titer to minimize sampling effort

To maximize information to effort ratio sampling should be focused on areas of highest interest – the area of highest specific titers. This substantiates the need for a transferable predictor to identify sampling points of highest relevance, for example concerning time by considering the biological activity [21] or alternatively less time dependent variables, as the consumed substrate. The point of maximum specific titer is reached at different time points but at insignificantly different points of consumed substrate dSn. Since the future of bioprocess development lies within small scale bioreactors [22,23] continuous time resolved sampling is additionally limited by the volume of small scale parallel bioreactors. In this setting sampling point selection is of additional interest and greatly eased by sampling according to dSn.

Although the volumetric constant feeding rate is industrial standard, the growth in biomass quickly decreases the real time qs. Therefore, the direct manipulation of physiology by a real time control of qs for bioprocess development appears far more promising. Recently, novel strategies for the control of the specific substrate uptake rate (qs) were reported [24-26]. The scientific community has been employing the control of physiological variables for bioprocess development purposes in open loop control [27-29] as well as in closed loop control mode [30,31]. Real time control of a major physiologic variable as qs will render this variable a process parameter, which in turn facilitates physiological bioprocess development based on physiological parameters.

Conclusions

The illustrated workflow outlines a roadmap for cost efficient bioprocess development in respect to i) reducing experimental effort in case of product changeover, ii) developing transferable platform knowledge by focusing on physiology and iii) reducing sampling effort. Physiological data interpretation by the means of information processing and information mining makes findings more product and setup-independent and therefore better transferable. Using the illustrated workflow we were able to substantiate the following points:

- A route to platform knowledge has been illustrated by the repeated investigation of the effect of pre-induction on post-induction.
- Based on the data of the insignificant interrelation of pre- and post-induction phase qs_{init} was pinpointed as influential parameter for maximum product titers. This interrelation was substantiated with subsequent experiments. The overall picture of the data strongly points towards a real time control approach to elicit maximum productivity.
- The descriptor of dSn has been found to be powerful predictor for the area of maximum specific titers. Sampling according to dSn focusses the effort on the area of highest interest – the area of greatest titers.

References

1. Lee SY. High cell-density culture of Escherichia coli. *Trends in Biotechnology.* 3// 1996;14(3):98-105.
2. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology.* 2006/09/01 2006;72(2):211-222.
3. Reid B. US biotech prepares to fight generic biologics. *Nat Biotech.* 04//print 2002;20(4):322-322.
4. Rathore AS, Winkle H. Quality by design for biopharmaceuticals. *Nat Biotech.* 01//print 2009;27(1):26-34.
5. Lionberger RA, Lee SL, Lee L, Raw A, Yu LX. Quality by design: concepts for ANDAs. *AAPS J.* Jun 2008;10(2):268-276.
6. FDA. Guidance for Industry Q9 Quality Risk Management. 2006.
7. Greb E, Drakulich A. Platform Technologies: The standardization of upstream and downstream bioprocessing is growing, but several kinks need to be ironed out. *Pharmaceutical Technology.* 2012;36(3):44-50.
8. Sanden AM, Prytz I, Tubulekas I, et al. Limiting factors in Escherichia coli fed-batch production of recombinant proteins. *Biotechnol Bioeng.* Jan 20 2003;81(2):158-166.
9. Wechselberger P, Sagmeister P, Engelking H, Schmidt T, Wenger J, Herwig C. Efficient feeding profile optimization for recombinant protein production using physiological information. *Bioprocess Biosyst Eng.* Nov 2012;35(9):1637-1649.
10. Curless C, Pope J, Tsai L. Effect of preinduction specific growth rate on recombinant alpha consensus interferon synthesis in Escherichia coli. *Biotechnol Prog.* Mar-Apr 1990;6(2):149-152.
11. Norsyahida A, Rahmah N, Ahmad RM. Effects of feeding and induction strategy on the production of BmR1 antigen in recombinant E. coli. *Lett Appl Microbiol.* Nov 2009;49(5):544-550.
12. Shin C, Hong M, Shin H-C, Lee J. High-level production of recombinant human IFN-α2a with co-expression of tRNAArg(AGG/AGA) in high-cell-density cultures ofEscherichia coli. *Biotechnology and Bioprocess Engineering.* 2001/08/01 2001;6(4):301-305.
13. Shin CS, Hong MS, Bae CS, Lee J. Enhanced production of human mini-proinsulin in fed-batch cultures at high cell density of Escherichia coli BL21(DE3)[pET-3aT2M2]. *Biotechnol Prog.* May-Jun 1997;13(3):249-257.
14. Piubelli L, Campa M, Temporini C, et al. Optimizing Escherichia coli as a protein expression platform to produce Mycobacterium tuberculosis immunogenic proteins. *Microbial cell factories.* 2013;12(1):115.
15. Group CB. A-Mab: a Case Study in Bioprocess Development. 12.08.2014 2009.
16. Mandenius C-F, Brundin A. Bioprocess optimization using design-of-experiments methodology. *Biotechnology Progress.* 2008;24(6):1191-1203.
17. Sagmeister P, Wechselberger P, Herwig C. Information Processing: Rate-Based Investigation of Cell Physiological Changes along Design Space Development. *PDA J Pharm Sci Technol.* Nov-Dec 2012;66(6):526-541.
18. Enfors SO. Control of in vivo proteolysis in the production of recombinant proteins. *Trends Biotechnol.* Sep 1992;10(9):310-315.
19. Murby M, Uhlen M, Stahl S. Upstream strategies to minimize proteolytic degradation upon recombinant production in Escherichia coli. *Protein Expr Purif.* Mar 1996;7(2):129-136.
20. Wilms B, Hauck A, Reuss M, et al. High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the Escherichia coli rhaBAD promoter. *Biotechnology and Bioengineering.* 2001;73(2):95-103.
21. Wechselberger P, Sagmeister P, Herwig C. Model-based analysis on the extractability of information from data in dynamic fed-batch experiments. *Biotechnol Prog.* Jan-Feb 2013;29(1):285-296.

22. Betts JI, Baganz F. *Microbial cell factories.* 2006;5(1):21.
23. Bareither R, Pollard D. A review of advanced small-scale parallel bioreactor technology for accelerated process development: Current state and future need. *Biotechnology Progress.* 2011;27(1):2-14.
24. Sagmeister P, Kment M, Wechselberger P, Meitz A, Langemann T, Herwig C. Soft-sensor assisted dynamic investigation of mixed feed bioprocesses. *Process Biochemistry.* 12// 2013;48(12):1839-1847.
25. Zalai D, Dietzsch C, Herwig C, Spadiut O. A dynamic fed batch strategy for a Pichia pastoris mixed feed system to increase process understanding. *Biotechnology Progress.* 2012;28(3):878-886.
26. Dietzsch C, Spadiut O, Herwig C. A dynamic method based on the specific substrate uptake rate to set up a feeding strategy for Pichia pastoris. *Microbial cell factories.* 2011;10:14.
27. Fan DD, Luo Y, Mi Y, Ma XX, Shang L. Characteristics of fed-batch cultures of recombinant Escherichia coli containing human-like collagen cDNA at different specific growth rates. *Biotechnol Lett.* Jun 2005;27(12):865-870.
28. Korz DJ, Rinas U, Hellmuth K, Sanders EA, Deckwer WD. Simple fed-batch technique for high cell density cultivation of Escherichia coli. *Journal of Biotechnology.* 2/21/ 1995;39(1):59-65.
29. Riesenberg D, Schulz V, Knorre WA, et al. High cell density cultivation of Escherichia coli at controlled specific growth rate. *Journal of Biotechnology.* 8// 1991;20(1):17-27.
30. Jenzsch M, Simutis R, Luebbert A. Generic model control of the specific growth rate in recombinant Escherichia coli cultivations. *J Biotechnol.* Apr 20 2006;122(4):483-493.
31. Johnsson O, Andersson J, Lidén G, Johnsson C, Hägglund T. Feed rate control in fed-batch fermentations based on frequency content analysis. *Biotechnology Progress.* 2013;29(3):817-824.

ii. Physiological Feeding

More satisfying than mere physiological data evaluation of volumetric constant feeding profiles is direct physiological process control. Hereby, physiological process control focuses on the control of a biomass specific variable and requires online biomass estimation. In the context of substrate supply the physiological variable of choice would be the specific substrate uptake rate (q_S). For microbial bioprocesses controlling the substrate supply technologically (i.e. by a volumetric constant feeding rate) is state of the art in industry [24]. Technical feed control is a simplistic approach and highly robust but hardly transferable between strains. Technological bioprocess development mainly controls technical variables as the (constant) volumetric substrate feed rate [21, 22, 52]. In contrast physiological bioprocess development controls and investigates physiological variables, as the biomass specific substrate uptake rate (q_s). An increasing number of scientific contributions have been using physiological feeding profiles for the control of specific rates [3-5, 53]. This approach circumvents the impact of underlying trajectories but requires real time biomass estimation. This raises the question whether the increased effort for physiological control in comparison to technical control can be justified in terms of productivity. To address this topic the impact of substrate supply on productivity was studied and compared by conducting technical and physiological feed profiles.

Physiological bioprocess control: the benefit of discriminating the effect of level and trajectory of the specific substrate uptake rate

Reichelt Wieland N.[a], Kager Julian[b], Sagmeister Patrick[b], Herwig Christoph[a,b*]

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

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

***Corresponding author:** christoph.herwig@tuwien.ac.at

KEYWORDS

Physiologic process control; specific substrate uptake rate; dynamic process control; real time biomass estimation; productivity; microbial process development; information to effort ratio;

Abstract

Background

Industrial bioprocesses for pharmaceutical products are engineered to yield high productivity and product quality while being robust and transferable. Especially in microbial bioprocess development, feeding strategies have widely been used to optimize productivity. Hereby, conventional bioprocess development commonly focusses on the maintenance of technical variables as the volumetric substrate feed rate. In contrast, physiologic bioprocess development focusses on the maintenance of physiologic variables, as the biomass specific substrate uptake rate (q_S). Although physiological process development has been discussed before and control strategies are abundantly available, the associated increased effort as for biomass sensing and feed control hasn´t been reviewed critically. The goal of this study was to elucidate whether the additional effort for physiologic bioprocess development is justified by a gain in productivity and process understanding (information to effort ratio).

Results

Using *E.coli* we compared the industrial standard feeding strategy of constant volumetric feeding with a physiological feeding strategy of dynamic feeding. While the industrial standard feeding strategy maintains a technologic variable - the feed rate, the physiologic variable of q_S is in uncontrolled and in free flow. In comparison, the physiological feeding strategy maintains q_S by dynamically adapting the feed rate. Firstly, independent of the feeding strategy we found a generally negative correlation of the average q_S (q_{Smean}) and the maximum specific product titer (+131 %). This finding substantiates q_S as potent candidate based on which maximum specific titers can greatly be increased. But more importantly, in comparison to the industrial standard, physiologic feeding experiments displayed a generally higher product titer (max. +92 %) even for a comparable q_{Smean}.

Conclusions

In this contribution we demonstrated the benefit of physiological feeding and process control by achieving an increase in maximum specific titers in comparison to the more simplistic, technological control approach. Concluding, physiological bioprocess control can be regarded as valuable tool, featuring a desirable information to effort ratio.

Background

The requirements of pharmaceutical industry towards bioprocesses are demanding and include high transferability, high productivity and product quality as well as robustness [1]. In this context, the quality by design (QbD) initiative [2, 3], emphasizes the need for science- and risk based pharmaceutical process development [4, 5]. The main aim of these guidelines is to replace the heuristic approach by more systematic, mechanistic bioprocess understanding [6]. In the lifecycle of a bioprocess setups and scales change frequently, while the cell-line remains constant. Conventional, technological bioprocess development mainly addresses technical, setup specific variables as the volumetric substrate feed rate. In contrast, physiologic bioprocess development controls and investigates physiologic variables, as the biomass specific substrate uptake rate (q_S) or the specific growth rate (µ). Although physiological process development has been discussed before [7-11] and control strategies are abundantly available [12], the associated increased effort as for biomass sensing and feed control hasn´t been reviewed critically. Within this contribution, we address the question if this increase in effort compared to conventional feeding strategies can be justified by the gain in productivity and process understanding.

Post-induction phase is triggered by inducing the culture, which devotes cellular metabolism to product formation. Especially in post-induction phase the substrate feeding rate or the inducer concentration are known as key process parameters regarding the productivity of recombinant microorganisms [13, 14]. Numerous studies reported a clear dependency between productivity and the post induction feed rate [8, 15, 16]. Accordingly, various strategies have been introduced to tailor the feeding rate to the physiological status of the culture while to avoiding detrimental acetate accumulation [17]. Generally, post-induction feed profiles can be categorized by the mode of control of the feeding rate (Table 1). This control can either be technologically controlled focusing on a constant feed rate or physiologically controlled targeting constant physiology.

Table 1 Overview of induction phase feed profiles

Feed profile	control strategy	Advantage	Disadvantage
constant volumetric feed	PID constant flow rate	simple, no extra knowledge/ equipment needed	Scope of conclusions is limited
dynamic volumetric feed	open loop; constant physiologic variable	Constant physiology, scope of conclusions	Increased effort

The simplest post-induction feed profile controls the technological variable of the feed flow rate at a volumetric constant level [18-20] – the “technological feeding strategy”. Offering a simplistic implementation the latter strategy defines today’s industrial standard [21]. Occasionally the post-induction

feed flow rate is tailored to the fed batch phase, e.g. by calculating qsinit at the point of induction [9]. Nevertheless, the growth of biomass leads to an inevitable and uncontrolled decrease of e.g. q_S [20]. While technological the process appears well controlled, physiology underlies great dynamics. The volumetric constant feeding strategy is per definition incapable of reacting to variations in culture physiology [11]. Additionally, this feeding strategy is highly setup specific, since large scale in-homogeneities trigger biomass yield changes which cause differences in culture physiology. In this contribution we refer to this feeding strategy as "q_S free flow".

Substantial effort has been invested in developing generic control approaches to control cell physiology. In this context the scale independent physiological entities like μ [h^{-1}] [10, 11] and q_S [g/g/h] [7-9] have evolved as promising process parameters impacting productivity. Hereby, the physiological variable is maintained by adapting the feed rate in response to the estimated biomass while avoiding substrate accumulation – the "physiological feeding strategy". Concerning the choice of physiologic variable to control literature remains indecisive despite the following two arguments: i) Since the biomass estimation underlies a greater error than substrate feeding rate, the control of μ comprises a greater error [1, 22] than the control of q_S. ii) The specific growth rate can only be altered by modulating the substrate availability. Consequently it appears more feasible to directly control q_S, being an actor on cellular metabolism and in turn on the specific growth rate (Figure 1).

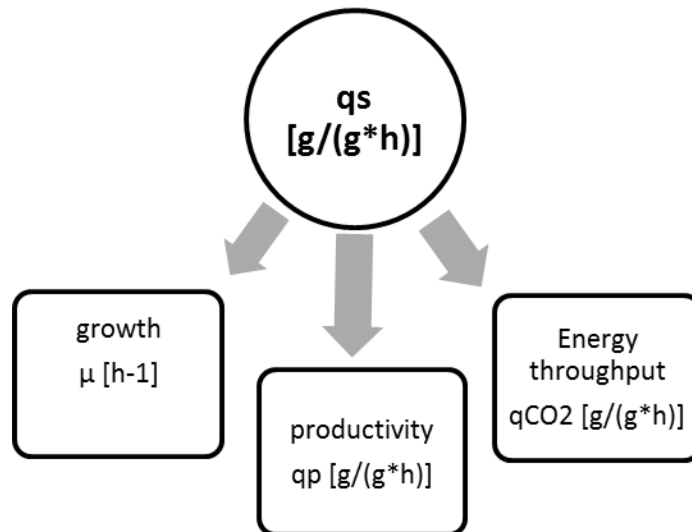


Figure 2 Metabolic relations between qs, µ, qp and qCO2, which represents the energy supply, biomass increase, productivity and cell metabolism of a cell.

Owned to variations of the biomass yield coefficient during induction phase [23, 24] the control of biomass specific variables requires real time biomass estimation. Accordingly, biomass yield independent biomass estimations feature a higher degree of transferability. Various approaches for biomass estimation have been introduced and discussed in literature. Especially in early bioprocess development the biomass estimation is usually obtained from a real time accessible signal using hard type sensors (e.g. permittivity) or soft sensors [12], which estimate biomass based on first principle mass balances [25, 26]. Based on the biomass estimation the trajectory of the physiological variable q_S can be controlled in contrast to the technological feeding strategy. In this contribution the approach of a dynamically adapted feed rate targeting a constant physiology (q_S) along the process is referred to as "q_S control".

To facilitate the comparison of the technologic and physiologic feeding strategy the physiology of the cells has to be put focus throughout data evaluation. The combination of information processing [27] and mining [28] facilitates data condensation and consequent data analysis for transferable, physiological interrelations - for mechanistic knowledge. In accordance with the QbD initiative, the benefit of mechanistic knowledge lies within the transferability of the found interrelations [29], even in case of product changeover [28, 30]. Concluding, only based on mechanistic process knowledge the benefit of physiological bioprocess development and bioprocess control can be identified.

Although physiological bioprocess control is widely discussed in literature [1, 8, 24, 31-34] the applications in industrial-scale remain extremely rare [11]. This circumstance can mainly be attributed to the lack of evidence whether the additional effort for process control is worth its effects. To date, no study was published comparing technological and physiological feeding strategies in terms of effort and benefit.

Goals

Rather than the introduction of another feeding strategy we firstly investigate i) the general impact of substrate availability on productivity to illustrate the differences of the two feeding strategies. Subsequently, for the first time ii) the impact of the q_S trajectory on productivity is analyzed to answer the question whether the increased effort for the physiological feeding strategy can be justified.

Roadmap

The substrate supply has been identified previously [9, 28] as highly influential variable to optimize productivity. On the basis of two sets of experiments "q_S free flow" and "q_S control" at different levels of substrate supply the impact of the feeding strategy is investigated. To quantify the impact of these variables, obtained experimental data is condensed to physiologic variable (information processing) and consequently to physiological descriptors on a common basis of comparison (information mining). Hereby, data evaluation is focused on physiology in order to increase the transferability of findings even in-between rather different sets of experiments. A flow chart of the methodology is depicted in Figure 2.

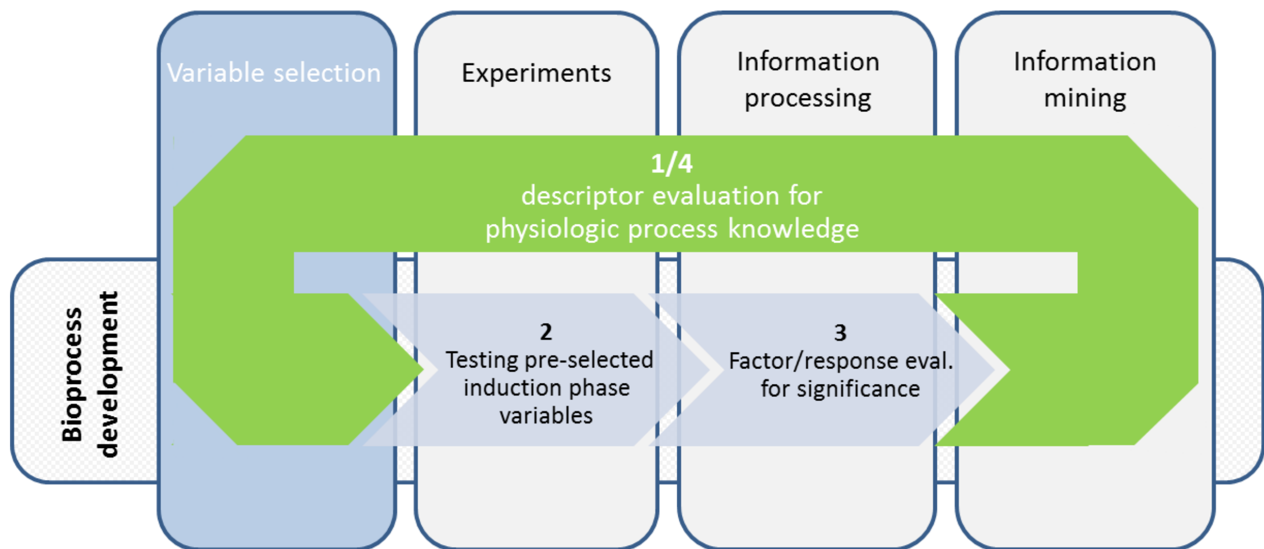


Figure 3: Flow chart of the evaluation methodology with prior process knowledge. The substrate supply has been identified as influential previously [28] and is analyzed within two sets of experiments (1/2); The impact of average substrate supply and the feed strategy shall be analyzed (3). By information mining (4) a physiological conclusion regarding the feeding strategies shall be derived in order to help to replace variable selection within a risk assessment merely based on expert knowledge and to provide the basis for subsequent design of experiments.

Material and Methods

Host

A modified K12 E. coli strain (kindly provided by Lonza Ltd., Visp, Switzerland) was used as a model expression system. The strain features a Rhamnose-inducible expression system (rhaBAD promoter). The recombinant protein product was a Fab antibody.

Media

A modified defined media [14] was used for fermentation.

Bioreactor system

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

Process parameters

The pre-culture was incubated at 30°C and 170 rpm for approx.17h (OD600 = 1.5). A volume equivalent of pre-culture of 2.5% of the Batch volume was used to inoculate the reactors. After depletion of the C-source, after approx. 12h, the pre-induction fed-batch was started. The pre-induction feeding strategy was based on an exponential feed forward profile (equation 4 and 5) to maintain a predefined growth rate.

After reaching the predefined Biomass (~30 g/L) the culture was induced with rhamnose. Temperature was set to 35°C, Stirrer speed to 1400 rpm and aeration to 1.4 vvm for the whole process. The pH was controlled at 7 with NH4OH. The dissolved oxygen (DO2) was kept over 25% by supplementing pure oxygen to the air. As post-induction feeding strategy two methods were compared: i) a constant volumetric feed flow rate was applied referred to as "q_S free flow" (Equation 1) ii) a volumetric dynamic feed rate adapted according to (Equation 2) based on a real time biomass estimation to keep q_S constant, referred to as "controlled q_S".

Equation 1 initial substrate uptake rate qs_{init} [g/g/h]: qs_{init} as setpoint serves as basis for calculating the volumetric constant feed rate $\dot{\mathbf{F}}_{\mathbf{EFB}}$ [L/h] depending on the feed concentration $\mathbf{c_s}$ [g/L] and the total biomass $\mathbf{X_{EFB}}$ [g] at the point of induction

$$qs_{init} = \frac{\dot{F}_{EFB} * \ c_s}{X_{EFB}}$$

Equation 2: substrate uptake rate $q_{S(t)}$ [g/g/h]: q_S as setpoint serves as basis for calculating the volumetric dynamic feed rate $F_{(t)}$ [L/h] depending on the real time estimation/offline measurement of the total biomass X_{EFB} [g]

$$qs_{(t)} = \frac{\dot{F}_{(t)} * \ c_S}{X_{(t)}}$$

Biomass estimation

Real time biomass estimation was based on an in line permittivity sensor. The measured signal is the relative permittivity of the fermentation broth in picofarad per centimeter (pF/cm). The difference of the two frequencies (~ 10 MHz high, ~ 1MHz low - dual frequency mode) correlates linearly with the membrane enclosed volume fraction of cells, which corresponds with the assumption of a constant cell density to the biomass dry weight concentration [35]. The signal to biomass correlation was calibrated during the exponential fed-batch phase by correlating the feed forward biomass profile with the permittivity signal.

Offline analysis

- Biomass dry weight

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

- Product analytics

Two mL of the fresh culture broth was centrifuged (4500 x g, 10 min, 4 °C). The supernatant was analyzed without pre-treatment to determine the CFM titer (cell free media). The pellet samples need to be disrupted and re-buffered before determining the SCF titer (soluble cell fraction). The cell pellets were re-suspended in 20 mL of 0,1 M Tris-buffer (pH 7,4) and were disrupted in a high-pressure homogenizer (Avestin EmulsiFlex; Canada) at 1400 ±100 bar in 6 passages. 500 µL of the homogenate were then applied on gel filtration columns (PD MiniTrap g-25, GE Healthcare, USA) and eluted with 1 mL Eluent A (20mM phosphate-buffer, pH 7,4).

The product titers were measured by a protein G affinity chromatography using a pH gradient on a HPLC system (Thermo Scientific Dionex Ultimate 3000; USA). The column was a HiTrap ProtG (GE Healthcare; USA) with a flow rate of 2 ml/min at 25°C. The Detection was at 390 nm and the elution was forced by changing the pH from 7.4 to 2.5 (20mM phosphate-buffer).

- Substrate conc. and small metabolites

Acetate concentrations were quantified from the supernatant by enzymatic photometric principle in a roboting system (CuBiAn XC; Optocell, Germany). The analysis was used as a quality control to exclude possible acetate production due to oxygen limitation or overflow metabolism. Feed substrate concentration was determined by a density/feed concentration correlation. The feed density was determined gravimetrically.

Data processing & reduction

Calculation of metabolic rates and yield coefficients was conducted with Matlab r2012 b (Mathworks; Natick, Massachusetts, USA). Software was used for the calculation of specific rates and yield coefficients, as we described elsewhere [9]. For further data reduction physiological descriptors where calculated [28] before subsequent statistical data analysis. The utilized window for data comparison was equal for all experiments and determined by the normalized amount of consumed substrate dSn (Equation 3).

Equation 3 dSn corresponds to the integral of the feeding rate rS [g/h] since the end of fed batch t_{EFB} normalized on the biomass concentration at the end of fed batch x_{EFB}

$$\Delta Sn(t) = \frac{\int_{t_{EFB}}^{t} rS\,dt}{x_{\mathrm{EFB}}}$$

Statistical analysis

Since common DoE evaluation is based on the set points of the respective factors it does not take usual process deviation into account potentially masking effects triggered by potential set point deviations. To ensure the most realistic response to factor correlation, we used the averagely met process variable as input instead of its mere set point. Statistical tests were performed with Datalab Version 3.5 (distributed by Epina http://datalab.epina.at)

Results

In order to examine the physiological differences of q_S free flow and q_S control experiments, 15 q_S free flow experiments and 9 q_S control experiments were conducted. To highlight the differences of the two feeding strategies exemplary experiments are displayed in Figure 3. While in q_S free flow experiments the volumetric flow rate remains constant over the whole process, in case of q_S control experiments the feed rate is dynamically adapted to account for the increase in biomass (Figure 3 A, C). Figure 3 F illustrates displays the progression of the biomass yield, emphasizing the necessity of real time biomass estimation. The obviously decreasing biomass yield renders an on fixed yield based q_S control approach unfeasible.

(i) General impact of substrate availability on productivity

To facilitate a conclusion on the benefit of q_S control in comparison to q_S free flow the respective experiments have to be compared by statistical means. Following an information mining approach physiological descriptors must be calculated within a physiologically relevant window of calculation. Especially for dynamically changing rates/ yield/ specific rates (Figure 3 A/F) the window of calculation is of utmost importance. This raises the question of the most relevant dimension for calculation considering physiology and productivity especially in the context of dynamically adapted feeding rates.

To account for the different amounts of fed substrate per biomass in dependence of the feeding strategy a different variable is required to serve physiological relevant common ground for comparison. To serve as physiologically more relevant variable than time we substituted time on the x-axis with the cumulative but normalized consumed substrate since induction: dSn [g/g] (Figure 3 B/D). Irrespective of the feeding strategy the experiments were categorized according to the average q_S achieved within a common window of dSn into q_{Smean} high (> 0.2 g/g/h), mid (0.2-0.12 g/g/h) low (< 0.12 g/g/h).

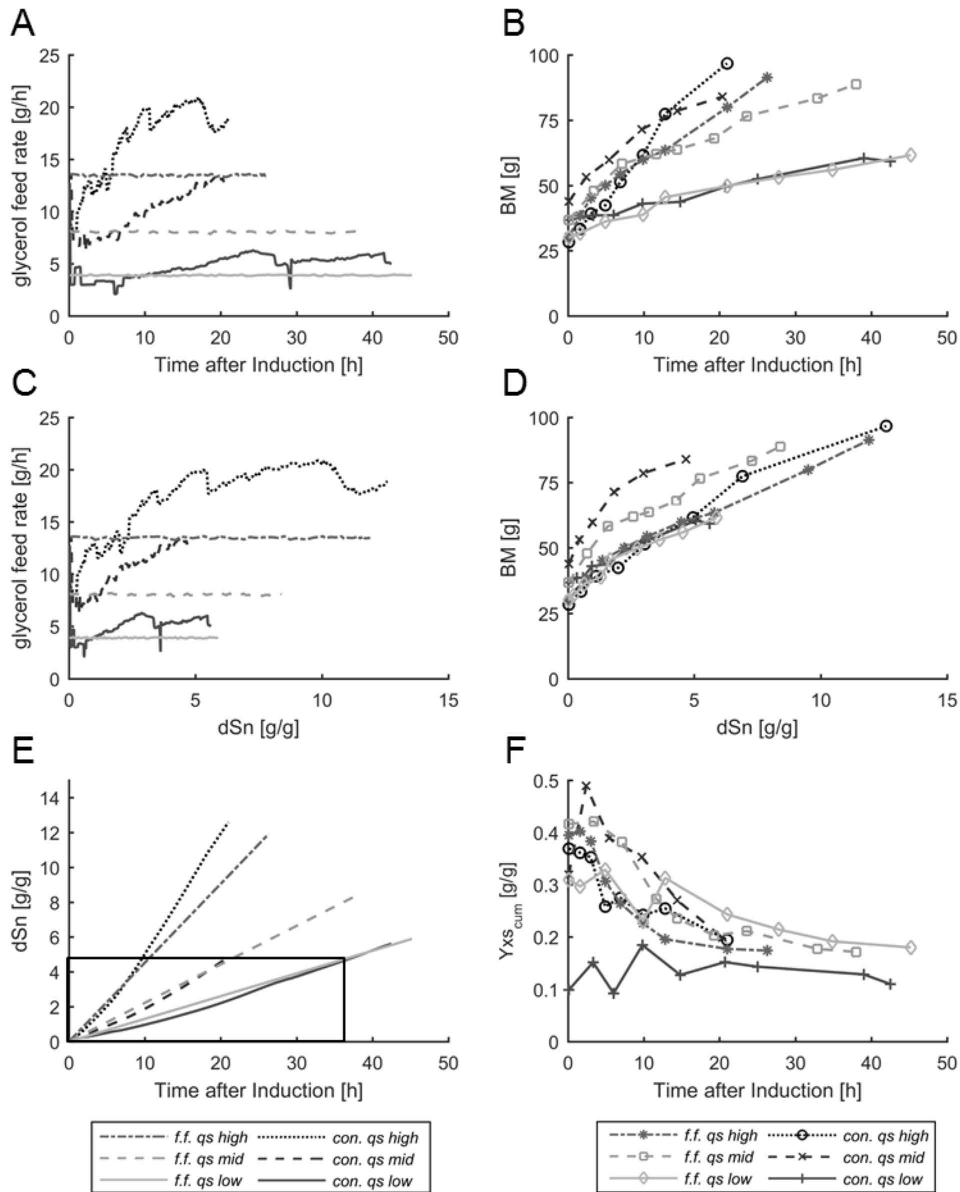


Figure 4: Exemplary experiments of q_S free flow and q_S control differ in substrate supply:

Experiments with comparable levels of q_{Smean} are shown. (A) Glycerol feed rate [g/h] over time after induction [h]; (B) Glycerol feed rate [g/h] over dSn [g/g]; (C) biomass BM [g] over time after induction [h]; (D) biomass BM [g] over dSn [g/g]; (E) cumulative biomass yield Yxs_{cum} [g/g] over time after induction [h], the rectangle highlights the highest common value of dSn shared by all experiments used for phase definition and data evaluation; (F) progression of dSn [g/g] over time after induction [h]; q_S: high (> 0,2 g/g/h), mid (0,2-0,12 g/g/h) low (< 0,12 g/g/h)

Figure 4 illustrates the usefulness of plotting the maximum specific titer over dSn [g/g] in comparison of plotting over time since induction [h] for physiological data interpretation. By aligning the product data over dSn, the induction period can be characterized by three distinct phases. In the first phase (P1) the specific titer in the soluble cell fraction (SCF) is sharply increasing until it reaches the maximum specific

titer until dSn of approx. 2.5 [g/g] (Figure 4 B). The subsequent second phase (P2) is characterized by the increase of specific product concentration in the cell free media (CFM) (Figure 4 D). In this phase product appears to be leaked or transported into the media since the specific intracellular product titer decreases (Figure 4 B). The final stage (P3) is marked by product degradation, in which the absolute amount of product decreases regardless of the localization.

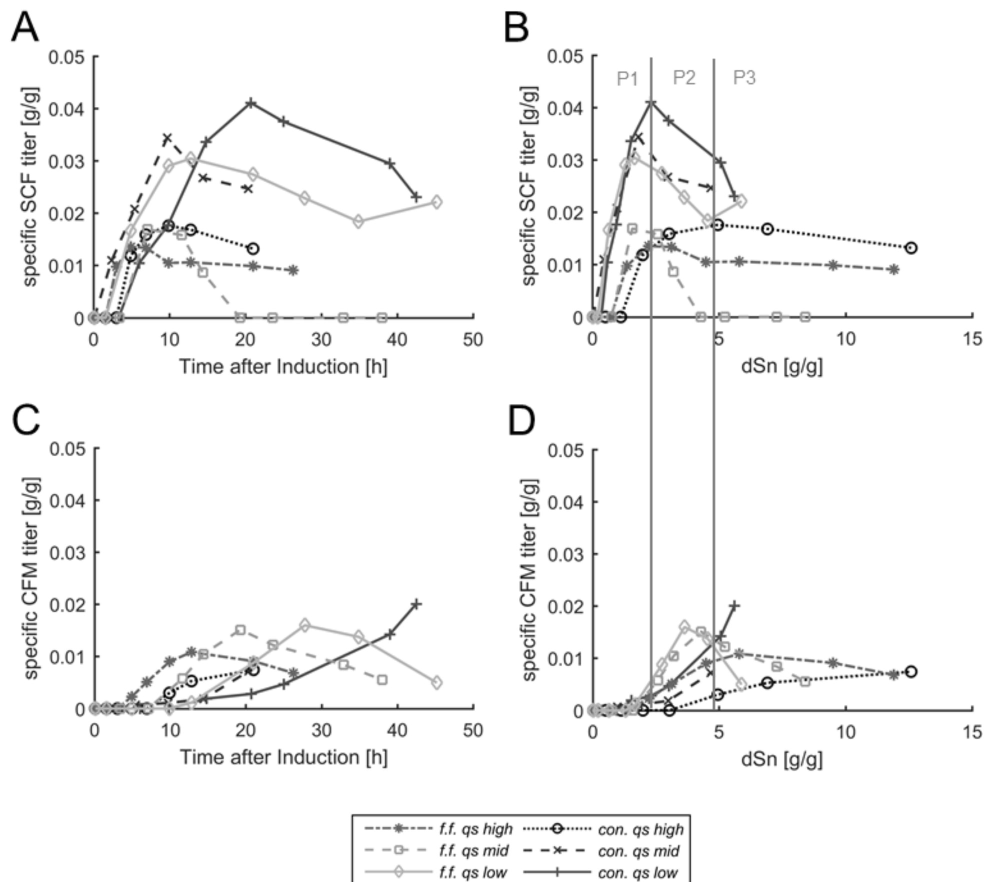


Figure 5: progression of specific product titers over time and consumed substrate.

(A) specific product [g/g] in the soluble cell fraction over time after induction ; (B) specific product [g/g] in the soluble cell fraction over dSn [g/g]; (C) specific product [g/g] in the cell free media over time after induction; D specific product [g/g] in the cell free media over dSn [g/g]; q_S: high (> 0,2 g/g/h), mid (0,2-0,12 g/g/h) low (< 0,12 g/g/h)

Regarding the productivity qp [g/g/h] the same three distinct phases can be observed in the SCF as well as in the CFM (Figure 5). A sharp increase of the qp in the SCF until a dSn of approx. 1 g/g (P1) is followed by a rise of qp in the CFM reaching the maximum value at a dSn of approx. 3.5 g/g (P2). Consequently the overall productivity of CFM starts to decrease, making P3 a phase of less interest from an economic point of view.

This alignment over dSn eases physiological interpretation since it aligns experimental data independent of q_{Smean} and the trajectory of q_S. The greatest common denominator of all conducted experiments was determined at a dSn of 4.5 [g/g]. Consequently physiological descriptors were calculated within an dSn of 0-4.5 [g/g]. This window of comparison includes > 90% of the max specific titers and makes the technologically very different experiments of q_S free flow and q_S control physiologically comparable.

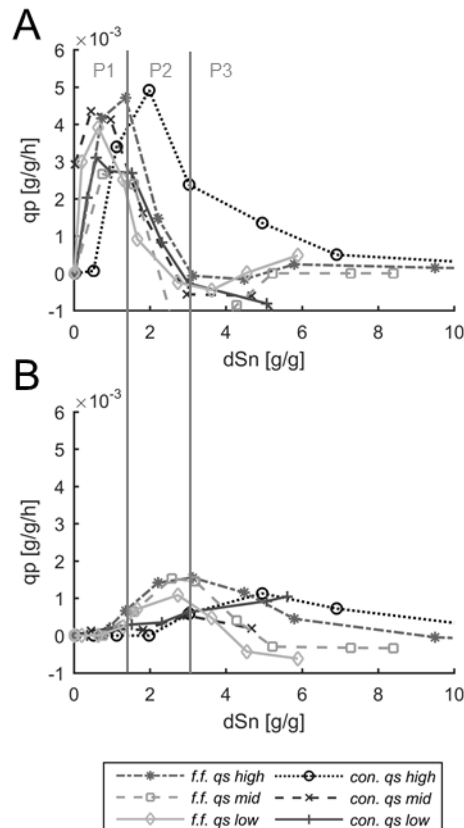


Figure 6: specific productivity over fermentation time in soluble cell fraction (SCF) and cell free media (CFM); (A) specific productivity [g/g/h] in SCF over dSn [g/g]; (B) specific productivity [g/g/h] in the CFM over dSn [g/g]; q_S: high (> 0,2 g/g/h), mid (0,2-0,12 g/g/h) low (< 0,12 g/g/h)

Regarding the feeding strategy within the illustrated experiments different maximum specific were observed. To investigate the role of physiology in the observed variance the maximum specific titers were plotted against the two physiological descriptors q_{Smean} and μ_{mean}. A strong negative correlation can be observed for q_{Smean} (Figure 6 A) as well as for μ_{mean} (Figure 6 B). The phenomena of high titers at low feeding rated could be interpreted as shift in energy distribution from growth to product triggered induction and subsequent substrate availability. The observed correlation of max specific titers and q_{Smean} (Figure 6 A) is statistically significant (R2 = 0.718/Q2 = 0.619) for all 24 conducted experiments.

More importantly, q_S control experiments achieve an overall higher specific titer than q_S free flow experiments. At comparable q_{Smean} levels maximum specific titers in q_S control experiments are higher. Based on this observation it can be concluded that the trajectory of the physiologic state (e.g. q_S) has an overall impact on maximum specific titers.

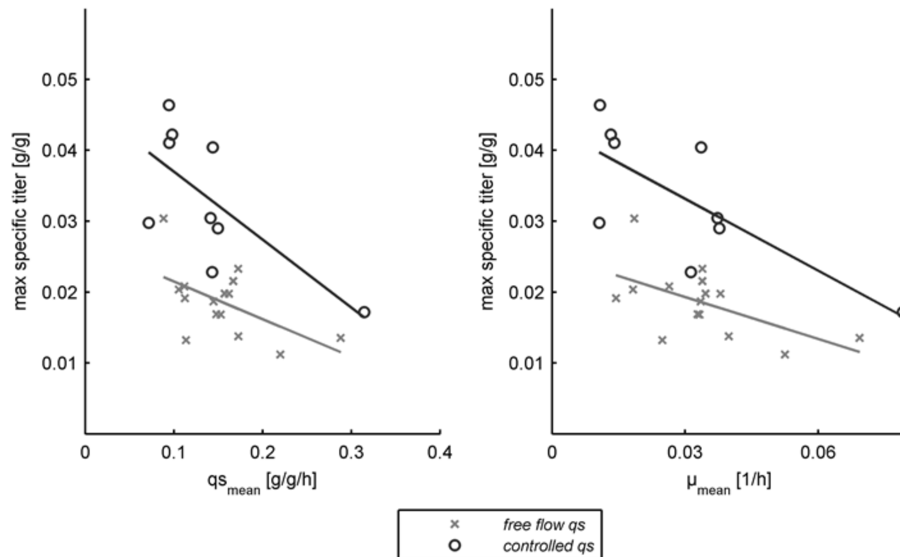


Figure 7 maximum specific titers are negatively correlated to q_{Smean} and μ_{mean}

(A) maximum specific cytosolic titers [g/g] plotted against the average specific substrate uptake rate [g/g/h] calculated in the specified window of dSn; (B) maximum specific cytosolic titers [g/g] plotted against the average specific grwoth rate [1/h] calculated in the specified window of dSn;

(ii) Technological vs physiological control: the impact of trajectories

To investigate the cause for the general tendency of q_S control experiments to display a higher max. specific titer the trajectories of q_S were compared in detail. Within a dSn of 4.5 [g/g], Figure 7 A illustrates the difference in feed rate [g/h] whereas the trajectory of q_S is displayed in Figure 7 B. While q_S control experiments maintain a constant q_S throughout induction, q_S free flow experiments display a steep decline in q_S. This decline in q_S for q_S free flow experiments is owned to the fast increase of biomass as a result of the initially high biomass yield.

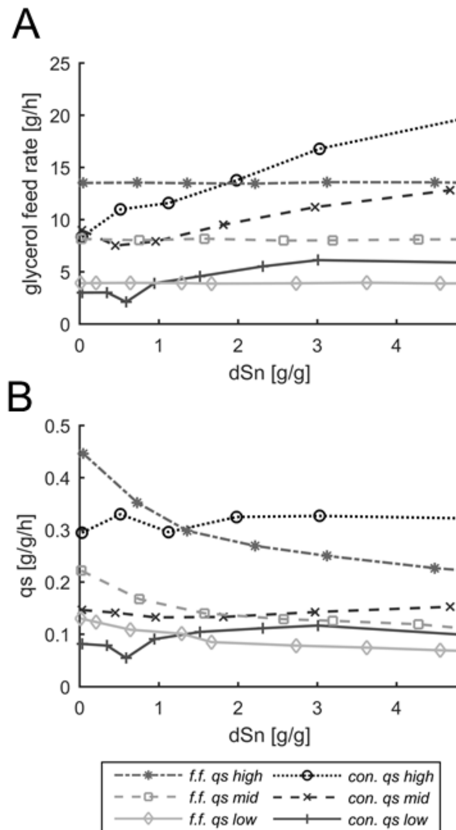


Figure 8 Physiologic feeding requires a dynamic feed profile but keeps q_S constant: (A) substrate feed rate [g/h] over dSn [g/g]; (B) specific substrate uptake rate q_S [g/g/h] over dSn [g/g]; q_S: high (> 0.2 g/g/h), mid (0.2-0.12 g/g/h) low (< 0.12 g/g/h);

To differentiate the impact of q_{Smean} and the trajectory of q_S throughout induction the productivity increase has been evaluated separately (Figure 8). The effect of the feed mode is illustrated by the comparison of q_S free flow and controlled experiments grouped by the achieved q_S mean (Figure 8 A). Generally, low q_{Smean} experiments of q_S free flow experiments display 68% higher maximum specific titers than the maximum specific titer of a q_S free flow experiment with a high q_S mean. Regarding q_S control experiments low q_{Smean} displayed up to 131% more product than high q_{Smean} experiments (Figure 8 A). In comparison to the group of q_S free flow experiments with a comparable q_S mean, the q_S control experiments display an up to 92% higher maximum specific titer. Concluding constant substrate availability and consequently the underlying trajectory of q_S appears to have an additionally beneficial effect independent of the q_S level. In Figure 8 B the feeding strategies have a clear significant effect in addition to the strong negative impact of q_{Smean} on maximum specific titers. This findings constitutes a substantial benefit for physiological bioprocess control since q_S control experiments feature a generally higher titer than q_S free flow experiments.

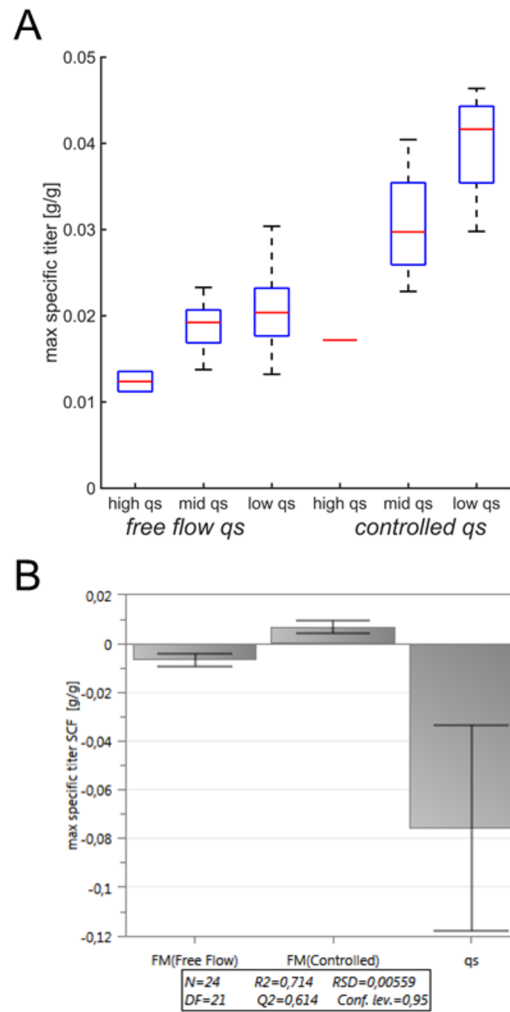


Figure 9 Box Plot comparing the two classes of experiments: (A) q_S free flow (n=15) and controlled q_S (n=9), experiments grouped by level of q_S: high (> 0.2 g/g/h), mid (0.2-0.12 g/g/h) low (< 0.12 g/g/h); (B) multilinear regression model : q_S free flow (FM Free Flow), controlled q_S (FM controlled) and the mean specific substrate uptake rates correlated with the maximum specific titer.

Discussion

Within this contribution we addressed the question if the increase in effort for physiological process control is worth its effects. To answer this question we firstly investigated the overall role of substrate supply on physiology and productivity (i). Secondly, we analyzed the impact of the substrate feed strategy and the correlated trajectory of q_S on productivity (ii).

(i) General impact of substrate availability on productivity

In order to facilitate an objective comparison of q_S free flow and q_S control experiments the data was processed by the means of information processing and information mining. Rather than over time since induction dSn [g/g] (Equation 3) was found to increase comparability. dSn has proven a valuable variable for phase definition of production and degradation phases in SCF as well as in CFM. and to feature a useful predictive power of determining the phase of highest titers (Figure 3).

Numerous publications studied the proportionality of the interrelation of substrate supply and productivity. While some contributions stated a direct proportional relation [8, 11, 36] others stated an indirectly proportionality [16, 37]. Obviously the nature of the interrelation appears to be dependent on physiology and/or product. In this contribution the examined expression system displays a clear negative correlation of q_{Smean} and the maximum specific titer (Figure 8, Figure 6). It can be hypothesized that substrate availability has an impact on the energy distribution. The additional metabolic load of anabolism, imposed by induction, affects the energy household of the host organism severely, leading to a shift in metabolism [38]. This hypothesis goes hand in hand with the observed decrease of the biomass yield coefficient (Figure 3 F). Obviously upon induction of the rhaBAD promotor a major part of metabolic energy is directed from biomass growth to product formation [14]. This might lead to an increasing metabolic load, eventually triggering the observed sharp decline in productivity (Figure 5). The highly complex and non-linear nature of bioprocesses burdens great challenges on bioprocess monitoring and bioprocess control approaches [39] and denies fixed yield approaches for real time biomass estimation.

(ii) Technological vs physiological control: the impact of trajectories

The hallmark difference between q_S free flow and q_S control experiments is the substrate availability for each cell. While in q_S free flow experiments initially un-proportionally much substrate is available for a given q_{Smean}, the specific substrate availability towards the end of fermentation declines severely. Underlying trajectories of investigated physiological variables can severely compromise conclusions. Consequently, the effect of physiological variables and correlated trajectories can only be investigated using physiological control which renders the variable a parameter. Independent of strain or product the accessible range of q_{Smean} is broadened by controlling q_S. To achieve the same q_{Smean} as in a q_S control setting, a significantly higher q_S free flow has to be set due to the decline of q_S throughout induction phase

(Figure 7 B). Under the constraint of avoiding acetate production the latter circumstance grants a greater range of accessible q_{Smean} by dynamically adapting the federate (q_S control).

Besides the scientific arguments we have shown that there is a substantial economic implication in physiological control, since specific titers can greatly benefit from the constant physiological state (Figure 6, Figure 8). Although the technological control approach of a volumetric constant feed rate throughout post-induction [9, 18-20, 34] is very simplistic the gained mechanistic knowledge remains scarce due to underlying trajectories. Physiological feeding strategies on the other hand require higher effort but offer deeper insights into mechanistic interrelations. Understanding the respective bioprocess by comprehending the effects of physiological variables fosters bioprocess understanding and ultimately process robustness. Hereby, the industrial need for robustness overlaps with the demand of the FDA quality by design (QbD) initiative [3, 40] for increased process understanding [4].

Conclusions

In this contribution the benefits of physiological processes control were investigated and discussed in direct comparison to technological controlled processes. The goal of this contribution was to answer the question if the increase in effort for physiological process control is worth its effects.

Methodological achievements

- Based on the calculation of q_{Smean} within a constant window of dSn we established the general correlation of q_{Smean} and maximum specific product titers.
- Only by controlling physiological variables the impact of trajectory and mean level can be discriminated ultimately leading to mechanistic process understanding

Physiological Achievements

- The trajectory of the physiological variable of q_S throughout the process greatly impacts productivity justifying the effort for physiological control
- The working hypothesis concerning a mechanistic explanation points toward a shift of energy distribution from biomass production to product formation, triggered by substrate availability/energy input.

The most surprising finding within this contribution was the great impact of the trajectory of the examined physiological variable (q_S) on the productivity. This great impact of the trajectory does not only justify physiological control, it points out the necessity of controlling physiological variables to gain true mechanistic process understanding and to discriminate the effect of level and trajectory of a given physiologic variable.

Declarations

AUTHORS CONTRIBUTIONS

Experiments and data evaluation were conducted by RW and KJ. The manuscript was drafted be RW. All authors contributed during revision of the manuscript and approved the final version.

AVAILABILITY OF DATA AND MATERIAL

The datasets supporting the conclusions of this article are included within the article.

ACKNOWLEDGEMENT

The authors are very grateful to Lonza Ltd. (Visp, Switzerland) for their support and providing the *E. coli* strain used in this study.

COMPETING INTEREST

The authors declare no conflict of interest.

FUNGDING

We are grateful for the financial support of the Christian Doppler Gesellschaft.

CONSENT FOR PUBLICATION

Not applicable.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

References

1. Dabros M, Schuler M, Marison I: Simple control of specific growth rate in biotechnological fed-batch processes based on enhanced online measurements of biomass. *Bioprocess and Biosystems Engineering* 2010, 33:1109-1118.
2. ICH F: Guidance for Industry Q9 Quality Risk Management. 2006.
3. ICH F: Guidance for Industry Q8 PharmaceuticalDevelopment (R2) 2009.
4. Rathore AS, Winkle H: Quality by design for biopharmaceuticals. *Nat Biotech* 2009, 27:26-34.
5. Lionberger RA, Lee SL, Lee L, Raw A, Yu LX: Quality by design: concepts for ANDAs. *AAPS J* 2008, 10:268-276.
6. Herwig C: Prozess Analytische Technologie in der Biotechnologie. *Chemie Ingenieur Technik* 2010, 82:405-414.
7. Picon A, Teixeira de Mattos MJ, Postma PW: Reducing the glucose uptake rate in Escherichia coli affects growth rate but not protein production. *Biotechnol Bioeng* 2005, 90:191-200.
8. Dietzsch C, Spadiut O, Herwig C: A dynamic method based on the specific substrate uptake rate to set up a feeding strategy for Pichia pastoris. *Microb Cell Fact* 2011, 10:14.

9. Wechselberger P, Sagmeister P, Engelking H, Schmidt T, Wenger J, Herwig C: Efficient feeding profile optimization for recombinant protein production using physiological information. *Bioprocess Biosyst Eng* 2012, 35:1637-1649.
10. Korz DJ, Rinas U, Hellmuth K, Sanders EA, Deckwer WD: Simple fed-batch technique for high cell density cultivation of Escherichia coli. *Journal of Biotechnology* 1995, 39:59-65.
11. Gnoth S, Jenzsch M, Simutis R, Lubbert A: Control of cultivation processes for recombinant protein production: a review. *Bioprocess Biosyst Eng* 2008, 31:21-39.
12. Reichelt WN, Thurrold P, Brillmann M, Kager J, Fricke J, Herwig C: Generic biomass estimation methods targeting physiologic process control in induced bacterial cultures. *Engineering in Life Sciences* 2016:n/a-n/a.
13. Zalai D, Dietzsch C, Herwig C, Spadiut O: A dynamic fed batch strategy for a Pichia pastoris mixed feed system to increase process understanding. *Biotechnology Progress* 2012, 28:878-886.
14. Wilms B, Hauck A, Reuss M, Syldatk C, Mattes R, Siemann M, Altenbuchner J: High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the Escherichia coli rhaBAD promoter. *Biotechnology and Bioengineering* 2001, 73:95-103.
15. Yee L, Blanch HW: Recombinant protein expression in high cell density fed-batch cultures of Escherichia coli. *Biotechnology (N Y)* 1992, 10:1550-1556.
16. Cheng L-C, Hor L-I, Wu J-Y, Chen T-L: Effect of specific growth rate on the production of a recombinant nuclease by Escherichia coli. *Biochemical Engineering Journal* 2003, 14:101-107.
17. Luli GW, Strohl WR: Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations. *Appl Environ Microbiol* 1990, 56:1004-1011.
18. Chen X, Liu L, Li J, Liu J, Du G, Chen J: Optimization of glucose feeding approaches for enhanced glucosamine and N-acetylglucosamine production by an engineered Escherichia coli. *Journal of Industrial Microbiology & Biotechnology* 2012, 39:359-365.
19. Ryan W, Collier P, Loredo L, Pope J, Sachdev R: Growth kinetics of Escherichia coli and expression of a recombinant protein and its isoforms under heat shock conditions. *Biotechnol Prog* 1996, 12:596-601.
20. Sanden AM, Prytz I, Tubulekas I, Forberg C, Le H, Hektor A, Neubauer P, Pragai Z, Harwood C, Ward A, et al: Limiting factors in Escherichia coli fed-batch production of recombinant proteins. *Biotechnol Bioeng* 2003, 81:158-166.
21. Sagmeister P, Jazini M, Klein J, Herwig C: Bacterial Suspension Cultures. 2014:40-93.
22. Rocha I, Veloso AC, Carneiro S, Costa R, Ferreira E: Implementation of a specific rate controller in a fed-batch E. coli fermentation. 2008.
23. Schaepe S, Kuprijanov A, Simutis R, Lübbert A: Avoiding overfeeding in high cell density fed-batch cultures of E. coli during the production of heterologous proteins. *Journal of Biotechnology* 2014, 192, Part A:146-153.
24. Jenzsch M, Simutis R, Luebbert A: Generic model control of the specific growth rate in recombinant Escherichia coli cultivations. *J Biotechnol* 2006, 122:483-493.
25. de Assis AJ, Filho RM: Soft sensors development for on-line bioreactor state estimation. *Computers & Chemical Engineering* 2000, 24:1099-1103.
26. Jobe AM, Herwig C, Surzyn M, Walker B, Marison I, von Stockar U: Generally applicable fed-batch culture concept based on the detection of metabolic state by on-line balancing. *Biotechnol Bioeng* 2003, 82:627-639.
27. Sagmeister P, Wechselberger P, Herwig C: Information Processing: Rate-Based Investigation of Cell Physiological Changes along Design Space Development. *PDA J Pharm Sci Technol* 2012, 66:526-541.
28. Reichelt WN, Haas F, Sagmeister P, Herwig C: Bioprocess development workflow: transferable physiological knowledge instead of technological correlations (under submission). *Biotechnological Progress* 2016.

29. Bolker BM: Ecological Models and Data in R. *Princton UNiversity Press* 2008:396.
30. Glassey J, Gernaey KV, Clemens C, Schulz TW, Oliveira R, Striedner G, Mandenius CF: Process analytical technology (PAT) for biopharmaceuticals. *Biotechnol J* 2011, 6:369-377.
31. Gnoth S, Jenzsch M, Simutis R, Lubbert A: Product formation kinetics in genetically modified E. coli bacteria: inclusion body formation. *Bioprocess Biosyst Eng* 2008, 31:41-46.
32. Levisauskas D: Inferential control of the specific growth rate in fed-batch cultivation processes. *Biotechnol Lett* 2001, 23:1189-1195.
33. Soons ZTA, van Straten G, van der Pol L, van Boxtel AB: Online automatic tuning and control for fed-batch cultivation. *Bioprocess and Biosystems Engineering* 2008, 31:453-467.
34. Wechselberger P, Sagmeister P, Herwig C: Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes. *Bioprocess Biosyst Eng* 2013, 36:1205-1218.
35. Ehgartner D, Sagmeister P, Herwig C, Wechselberger P: A novel real-time method to estimate volumetric mass biodensity based on the combination of dielectric spectroscopy and soft-sensors. *Journal of Chemical Technology & Biotechnology* 2015, 90:262-272.
36. Hellmuth K, Korz DJ, Sanders EA, Deckwer WD: Effect of growth rate on stability and gene expression of recombinant plasmids during continuous and high cell density cultivation of Escherichia coli TG1. *J Biotechnol* 1994, 32:289-298.
37. Jenzsch M, Gnoth S, Beck M, Kleinschmidt M, Simutis R, Lubbert A: Open-loop control of the biomass concentration within the growth phase of recombinant protein production processes. *J Biotechnol* 2006, 127:84-94.
38. Striedner G, Bayer K: An advanced monitoring platform for rational design of recombinant processes. *Adv Biochem Eng Biotechnol* 2013, 132:65-84.
39. Komives C, Parker RS: Bioreactor state estimation and control. *Curr Opin Biotechnol* 2003, 14:468-474.
40. Group CB: A-Mab: a Case Study in Bioprocess Development. 2009.

iii. q_{Scrit} – physiological capacities

Targeting physiological bioprocess development physiological variables can be used as factors in a respective DoE. In this case, the design of experiment calls for a clear definition of physiological constraints (i.e. limits) for each physiological variable. In the context of the definition of the physiological feasible space rather than minima, physiological maxima are of concern. Exceeding physiological capacities leads to unwanted accumulation of substrate or metabolites, which inhibits growth [54] and protein production [55]. Physiological capacities feature two definitions: q_{Smax} defines the total cellular capacity to metabolize substrate and q_{Scrit} is defined as the cellular capacity of anabolism and oxidative catabolism, metabolism without accumulation (Åkesson, Hagander, & Axelsson, 1999).

To quantify maximum physiological capacities, various approaches have been outlined; all aiming to generate a perturbation of C-source availability in an otherwise C-source limited process [5, 33-35]. Targeting a physiological DoE the state of the art approach was followed for quantifying physiological capacities. Within a verification experiment the feasibility of the state of the art is discussed and an alternative method is introduced using an industrial relevant production strain.

Substrate accumulation in induced bioprocesses caused by declining physiological capacities

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

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

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

ABSTRACT

During the cultivation of *E. coli* for recombinant protein production, substrate accumulation is often observed in induction phase. Uncontrolled substrate accumulation leads to difficulties in transferring or scaling processes and even to fail batches. The phenomena of metabolite/substrate accumulation is triggered by exceeding the physiological capacity to metabolize substrate (q_{Scrit}). In contrast to the common understanding of q_{Scrit} as "static" value we hypothesized that q_{Scrit} is of dynamic nature.

Following the state of the art approach of physiological strain characterization, substrate pulse experiments were used to quantify q_{Scrit} in induction phase. The observed, temperature dependent and timely declining q_{Scrit} was expressed through a linear equation to serve as boundary for physiologically controlled experiments. Nevertheless, within a physiologically controlled verification experiment accumulation was observed, although the q_{Scrit} boundary was not exceeded. A second set of experiments was conducted, oscillating q_S physiologically between discrete plateaus physiologically controlled. Hereupon, we were able to deduct a significant interrelation of the metabolic activity and the timely decline of the q_{Scrit}. This finding highlights the necessity of a comprehensive but laborious physiological characterization for each strain or, alternatively, of physiologic feedback control to facilitate real time analysis of q_{Scrit} in order to effectively avoid substrate accumulation.

KEYWORDS

physiological capacity; substrate accumulation; physiologic bioprocess development; dynamic experiments; bioprocess control; balancing approach

Background

Escherichia coli is one of the most exploited organisms for industrial production of recombinant proteins [1, 2]. Although *E. coli* has been characterized comprehensively, substrate accumulation is a frequently observed phenomena in induction phase. This substrate accumulation is counterproductive [3] owned to the associated negative effects on physiology [4]. Moreover, substrate accumulation impairs controllability and even causes fail batches. In literature substrate and metabolite accumulation is commonly regarded as a consequence of exceeding the physiologic capacity to metabolize substrate (q_{Scrit}). Given the importance of q_{Scrit} for experimental design, regardless of the utilized substrate feeding strategy, a quantification of q_{Scrit} appears obligatory. Consequently, various approaches for strain characterization have been outlined to firstly quantify q_{Scrit} [5-8] and consequently constrain the experimental design according to physiologic capacities. While the latter contributions regard q_{Scrit} as a static physiologic parameter, within this contribution the hypothesis that q_{Scrit} is a variable rather than static was tested. Furthermore, the aim was to investigate and to identify correlated physiological variables which trigger the putative dynamics in q_{Scrit}.

Physiology can be described in time dependent manner using physiological variables e.g. (specific) rates and yields, or process phase specific using physiological descriptors [9]. Of all physiological variables, specific rates are of highest interest, since specific rates assure biomass independent comparability and transferability. Consequently, an increasing number of scientific contributions has addressed the topic of controlling specific rates for physiologic bioprocess development [6, 10-12]. Such physiological process development approaches, targeting the control of specific rates e.g. the specific substrate uptake rate (q_S) or specific growth rate (μ), requires a robust control strategy as well as a concise definition of physiological constraints for experimental design.

For accurate control of specific rates during induction phase, real time biomass estimation is obligatory. Especially in industrial relevant high cell density fermentations, real time biomass estimation is challenging. In general, literature favors data driven models or hybrid models for real time biomass estimation [13, 14]. Nevertheless, in the setting of bioprocess development, historic process data is scarce, which impairs the use of data based algorithms. Consequently, hard type sensor or first principle mass balance based approaches are regarded as more feasible, especially in early bioprocess development [15].

To ensure feasibility of the experimental design, experiments are commonly designed within the congruent region of the technical and physiological feasible space. Commonly the technical feasible space is defined by maximum mass/energy transfer but also by temperature/pH stability of the setup. The physiological feasible space is defined by strain specific physiological capacities e.g. to metabolize substrate. While the technical feasible space is setup specific, the physiological feasible space is strain specific and has to be characterized individually for every strain to provide the constraints for the experimental design. Additionally, technical process parameters (e.g. pH, temperature) can be interlinked with physiologic capacities and therefore impact the physiologic feasible space. Consequently, the quantification of physiologic capacities in response to technical process parameters constitutes the basis for the definition of the constraints for experimental design.

In regard of circumventing substrate accumulation physiological maxima, rather than minima, are of concern. Exceeding physiologic capacities leads to accumulation of substrate or metabolites, which impairs reproducibility, transferability, inhibits growth [4] as well as protein production [3]. The physiologic capacity to metabolize substrate

is commonly described by two variables: q_{Smax} defines the total cellular capacity to metabolize substrate and q_{Scrit} is defined as the cellular capacity of anabolism and oxidative catabolism without accumulation of metabolites [5].

To quantify physiologic capacities, various approaches have been outlined, all aiming to generate a perturbation of C-source availability in an otherwise C-source limited process [6-8, 16]. Ultimately, the question of the molecular cause of the observed physiologic limitations is of less concern, than the timely resolution of the trajectory and the interrelation with other process parameters. Introduced as process control strategy Åkesson et al. used periodical "up-pulsing" by temporarily modulating the otherwise exponential substrate feed rate to trigger a transient surplus of C-source. On the basis of the response of the DO_2 signal the exponential feed rate was adapted to avoid q_{Scrit} throughout the process [5]. But according to more recent literature the amplitude of up-pulsing was too low and did not yield a saturation of the glucose uptake system [7], which consequently only allows qualitative conclusions towards q_{Scrit}. Still relying on the DO_2 signal and also designed as process control strategy, Lin et al. achieved metabolic saturation using concentrated C-source shots in addition to an otherwise volumetric constant substrate feed rate [7]. A recently introduced more explorative approach from Schaepe et al. is based on a highly dynamic feeding profile with periodical up-pulsing of the substrate feed rate. Also relying on the DO_2 signal this approach illustrated a decline of q_{Scrit} over induction time [17]. In contrast Henes et al. introduced "down-pulsing" as control strategy, an exponential feed strategy with a fixed yield, which is intermitted periodically. If the DO_2 signal does not react simultaneously to the substrate feed intermittence, q_{Scrit} has been exceeded, which in turn triggers a controller action. [6].

Summarizing, published work on the quantification of physiologic capacities relies widely on the highly sensitive DO_2 signal as response. However, the DO_2 signal is highly sensitive given the absolute low solubility of oxygen in aqueous media. Additionally the DO_2 level is commonly regarded as critical process parameter and has to be kept above a certain threshold. If the DO_2 signal is used as response to quantify q_{Scrit} as well as for DO_2 controller action, definite signal interpretation becomes challenging. Concluding, given the importance of q_{Scrit} an alternative experimental approach to quantify q_{Scrit} independent of the DO_2 signal is required.

In accordance to the state of the art, the first approach was based on highly concentrated, repetitive substrate pulses to quantify the q_{Scrit} in response to the technical process parameters temperature as well as time. Having quantified the q_{Scrit} trajectory in pulse experiments, we attempted and failed to verify the maximum physiologic capacity by a physiologically controlled experiment. Consequently, in a third set of experiments, we investigated the physiological feasible space and its dependency on the average physiologic activity (q_{Smean}) by physiologically controlled oscillations of q_S. Based on the obtained results the hypothesis of a variable q_{Scrit} and its interrelations to physiologic descriptors shall be evaluated to answer the question of the physiologic root cause of substrate accumulation in late process phases.

Materials and methods

2.1 Cultivations

2.1.1 Bioreactor system

Fermentations were conducted in a DASGIP multi-bioreactor system with four parallel reactors with 2 L of working volume each (Eppendorf; Hamburg, Germany). The reactors were equipped with baffles and three disk impeller stirrers. The DASGIP control software v4.5 revision 230 was used for control: pH (Hamilton, Reno, USA), pO_2 (Mettler Toledo; Greifensee, Switzerland; module DASGIP PH4PO4), temperature and stirrer speed (module DASGIP TC4SC4), aeration (module DASGIP MX4/4) and pH (module DASGIP MP8). CO_2, O_2 concentrations in the off-gas were quantified by a gas analyzer (module DASGIP GA4) using the non-dispersive infrared and zircon dioxide detection principle, respectively.

2.1.2 Strain & media

A recombinant BL21 DE3 *E. coli* strain was cultivated, producing an intracellular protein (~30 kDa) in form of inclusion bodies. The synthetic media was based on the recipe of Korz, Rinas et al. [18], where the limiting C-source was glucose.

2.1.3 Process parameters

Pre-cultures were incubated at 30°C and 170 rpm to an OD_{600} of approx. 1.5 in 150 mL batch media and 2.5% batch volume aliquots were used for inoculation. After depletion of the C-source in an initial batch phase, the pre-induction fed-batch was started. The pre-induction feeding strategy was based on an exponential feed forward profile to maintain a predefined growth rate. On attainment of the predefined biomass the cultures were induced after 30 min adaption time. Stirrer speed was set to 1400 rpm and aeration to 1.4 v/v/m for the whole process. The pH was maintained by adding 12.5% NH_4OH, which also served as nitrogen source. The dissolved oxygen (DO_2) was kept over 30% by supplementing oxygen to the air.

2.1.4 Pre induction: exponential feed forward profile

The starting feed rate in L/h (F_0) was calculated using a gravimetrical biomass yield in g/g ($Y_{X/S,g}$), the starting biomass in g (X_0), the concentration of the feed solution in g/L ($c_{S,g}$) as well as the specific biomass growth rate as described elsewhere [19].

2.1.5 Post induction feeding strategies

After the depletion of C-source in the batch phase the culture was induced with IPTG (1 mM). Upon induction either one of three types of experiments were performed: Pulse, q_S control and q_S oscillation experiments.

Pulse experiments: Upon induction substrate pulses were applied for the determination of q_{Scrit}, comprising 3 consecutive glucose pulses (20 g/L). The next pulse was applied upon depletion of glucose from the preceding pulses after a 30 min recovery phase. Within this contribution two pulse experiments were performed at different temperatures.

q_S control experiments: For real time q_S control the substrate feed rate is dynamically adjusted during the post induction phase. The timely variable of the feed flow rate is calculated via the feed concentration $c_{s,m}$ and the biomass $X_{g,(t)}$ at the corresponding time point (Equation 1). The necessary real time biomass quantification in g was conducted using a first principle soft-sensor as described elsewhere [20]. Based on the biomass estimation and $q_{S(t)}$ was controlled within a verification experiment on a q_S trajectory to maximize the average metabolic activity.

$$q_{S(t)} = \frac{\dot{F}_{(t)} * c_{S,g}}{X_{g,(t)}}$$

Equation 1: Specific substrate uptake rate at timepoint t

q_S oscillation experiments: Using the latter described control approach, oscillations in post induction q_S were performed at different average q_S levels (q_{Smean}) levels, q_S amplitudes and oscillation frequencies. The q_S mean set points ranged from 0.23 – 0.4 g/g/h. Based on these q_{Smean} values, amplitude levels quantified as changes in q_S of 0.1 – 0.2 g/g/h were used as set points. In simplified terms, for a q_S oscillation with a q_{Smean} of 0.23 g/g/h and an amplitude of 0.2 g/g/h the q_S was timely changed from 0.03 – 0.43 g/g/h throughout the induction phase of an experiment. Additionally different frequencies of 0.25 – 1 1/h for complete q_S oscillations were used. In total nine oscillation experiments were conducted to resolve the dependency of q_{Scrit} on q_{Smean}.

2.2 Process evaluation and data analysis

Metabolic rates and yield coefficients were calculated with Matlab r2013 b (Mathworks; Natick, Massachusetts, USA). The calculation of specific rates and yield coefficients was conducted as described elsewhere [21].

The here used q_{Sbal} is the specific substrate uptake rate in g/g/h calculated by substitution of r_S with the term $r_{O_2} \times RQ + r_X$ in C-mol/h via the DoR-Balance and the carbon balance (Equation 3, further details in supplemental). The biomass formation rate r_X in C-mol/h was derived from DCW measurements, the other necessary rates from off gas measurement. The C-molar molar mass of glucose in g/C-mol was used to obtain suitable units of r_S. Subsequently the division of rs by the offline BM in g $X_{m,(t)}$ q_{Sbal} was obtained in g/g/h.

$$qs_{bal} = \frac{r_{O_2} \times RQ + r_X}{X_{m,(t)}}$$

Equation 3: Molar specific substrate uptake rate from oxygen uptake, respiratory quotient and biomass

2.3 Offline analytics:

2.3.1 Biomass dry weight (DCW)

Biomass concentrations were gravimetrically quantified after drying at 105°C for min. 72 h. Therefore 2 mL of culture broth were centrifuged (4,500 x g, 10 min, 4°C) in a pre-weighted glass tube and the pellet was washed once with 5 mL deionized water. The determination was done in duplicates. After drying in the drying oven, the biomass dry weight was measured on a scale.

2.3.2 Substrate conc. and small metabolites

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

Results

Physiological maxima for physiologic DoE design

To define the physiological feasible space q_{Scrit} has to be quantified in response to the technical process parameters (e.g. T). Representing the experimental state of the art, substrate up-pulsing for q_{Scrit} quantification was used. Following an explorative approach rather than a control approach with the pulse experiments, q_{Scrit} was quantified based on the offline data. FIGURE 1 displays the residual glucose concentration in the supernatant of repetitive pulse experiments as well as the calculated q_{Scrit} derived from the glucose and biomass (DCW) quantification. Regarding a q_{Scrit} trajectory over time the temporal resolution is of great concern. In a system of growing biomass and excess substrate the calculation of q_{Scrit} requires offline analytics of substrate concentration and biomass and can consequently be correlated to analytical error. This necessity for offline quantification of the substrate concentration and biomass constitutes the major bottle neck for temporal resolution and causes noise in q_{Scrit} by error propagation. Nevertheless, an overall dependency of q_{Scrit} on process time can be concluded, judging from the qualitative decline over time. At the start of induction the results indicate a q_{Scrit} of 0.55 [g/g/h] at 35°C (FIGURE 1 A), while for 20°C q_{Scrit} is reached at 0.33 [g/g/h] (FIGURE 1 B). These results indicate a substantial impact of process temperature on q_{Scrit}, which is congruent with the theoretic metabolic point of view. This finding illustrates the interlink of the technological and physiological parameters and highlights the importance of investigating critical physiological parameters in response of technical variables. Summarizing, the maximum physiological capacity of the oxidative metabolism appears to be a function of time and dependent on the technical process parameter - temperature.

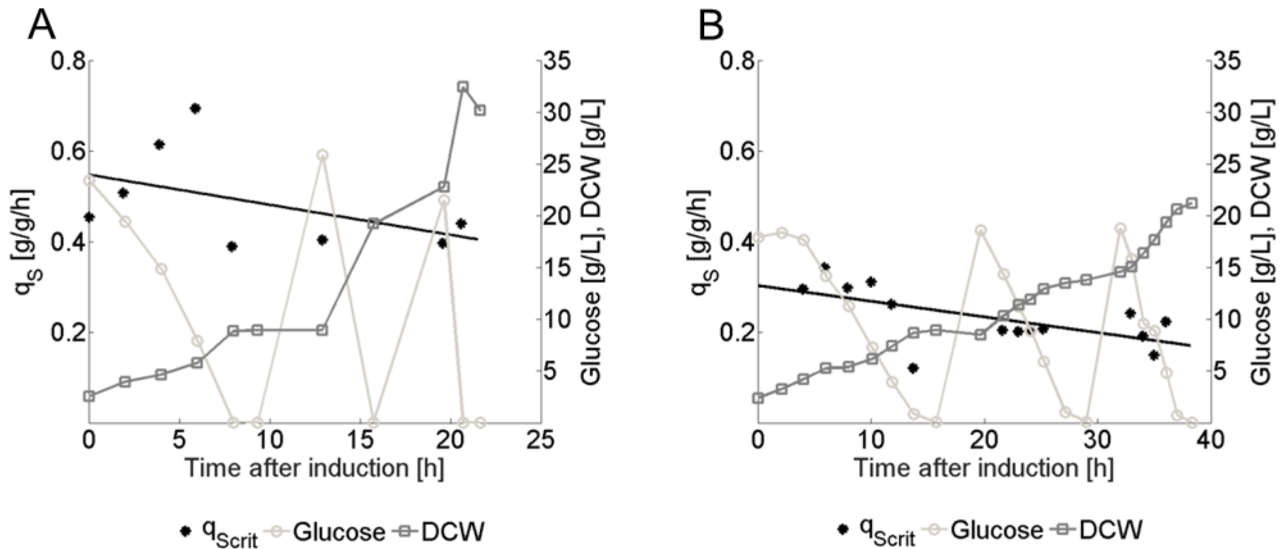


FIGURE 1: q_{Scrit} declines as function of time and temperature q_{Scrit}=f(t,T);
Repetitive substrate pulses (n=3) of 20 g/L were administered to fully saturate the oxidative metabolism and to investigate q_{Scrit} over time; in between pulses a time delay of 30 min was scheduled to facilitate full clearance of accumulated substrate and metabolites; offline samples are indicated by symbols, lines in-between the symbols are meant as guide for the eye. The calculation of q_{Scrit} [g/g/h] requires the rate of the decline in glucose concentration in g/L/h calculated from the timely resolved glucose concentrations in g/L, as well as the offline DCW in g/L measurements as inputs; acetate accumulation was at all times below 0.3 g/L (data not shown); the linear function represents a linear fit of q_{Scrit} values over time. **A**: 35°C (q_{Scrit} = -0.0066*t+0.55) indicating a maximum q_{Scrit} at start of induction of 0.55 [g/g/h]; **B**: 20°C (q_{Scrit} = -0.004*t+0.33) indicating a maximum q_{Scrit} at start of induction of 0.33 [g/g/h]; Hereupon a trajectory for 29°C was deducted mathematically: q_{Scrit}= -0.0056*t+0.46 (not shown)

Sample interval independent data evaluation of oxidative metabolism based on off gas signal

Based on a positive correlation of q_S and productivity (data not shown), the maximization of q_S is of great interest. Given the timely decline of q_{Scrit} (FIGURE 1 A/B) it is indicated to control q_S on a steadily declining trajectory to follow q_{Scrit} without exceeding it. Consequently, the subsequent experiment was designed to control q_S on a predefined trajectory, based on the previously defined q_{Scrit} trajectory (q_{Scrit}= -0.0056*t+0.46), to achieve the highest q_{Smean} possible, while avoiding substrate accumulation. To achieve this goal, real time biomass sensing and dynamic substrate feed adaptation were implemented to realize the predefined q_S trajectory. Despite a carefully designed experiment and a tightly controlled q_S accumulation occurred 6 hours after induction unexpectedly (FIGURE 2 A). This observation challenges the result and the approach illustrated in FIGURE 1, raises the question of transferability and infers additional underlying variables impacting the q_{Scrit} trajectory besides time after induction and T.

The observed substrate accumulation (> 5g/L) renders the supernatant measurements again the bottle neck for high temporal resolution of q_{Scrit}. This becomes obvious comparing the C-balance calculated with (FIGURE 2 B) and without supernatant sampling (FIGURE 2 A). Using the available high resolution off gas data to calculate the rate of oxidative metabolism (r_{Sox}), data processing is less dependent on offline sampling and consequently reduces the noise inflicted by analytical errors (refer to supplemental for further information). Subsequently, the calculation of r_{ox} via the DoR facilitated a higher temporal resolution of q_{Sbal} (Equation 3) as well as a qualitatively better closing mass balances (FIGURE 2 B). Nevertheless, this observation requires statistical analysis before a final conclusion can be drawn.

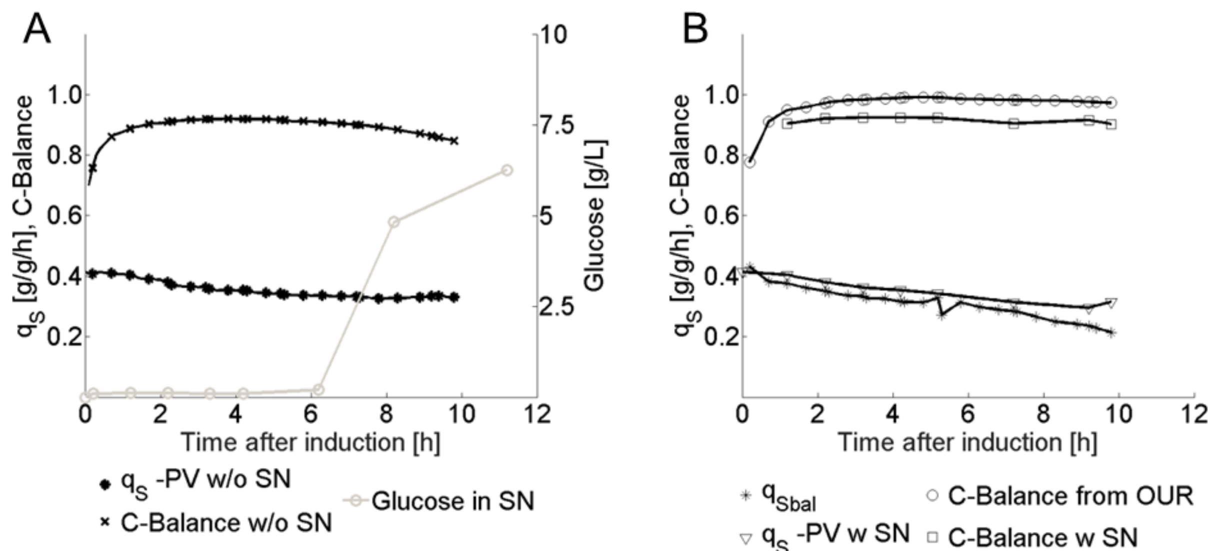


FIGURE 2: Physiological control of qs supposedly beneath the q_{Scrit} displays accumulation;
The appendices w SN and w/o SN indicate the correction (w), or the absence of the correction (w/o) of fed substrate with total glucose measured in the supernatant in g. "from OUR" indicates the use of the balancing approach which uses the C-balance and the DoR-balance as described for the variable q_{Sbal} and explained in more detail in the supplemental. **A**: Physiological control of qs (q_S-PV w/o SN) along a qs set point trajectory: q_S-SP = -0.009*t+0.4 derived from the pulse experiments; q_S – PV refers to the process value of q_S calculated with the feeding rate in g/L/h, the substrate concentration in the feed solution in g/L, the offline DCW in g/L as well as the reactor volume V as inputs. q_S and cumulative carbon balance (C-balance w/o SN) were calculated

without taking supernatant measurements (SN) into account only based on the substrate feed rate; Glucose is accumulating ongoing from 6 hours after induction (Glucose in SN); Acetate accumulation throughout the process was below 0.3 g/L (data not shown); **B**: q_S (q_{Sbal}) and cumulative C-balance are calculated with the balancing approach (C-balance from OUR) in comparison to the offline analytics depending approach taking SN measurements into account (q_S-PV w SN and C-balance w SN); symbols correspond to offline samples while lines in-between are meant as guide for the eye.

Since the q_{Scrit} trajectory did not hold true in the experiment for the maximization of q_{Smean}, the role of q_{Smean} in the decline of q_{Scrit} appears intriguing, but remains to be investigated. To achieve different levels of q_{Smean} and to assess q_{Scrit} controlled while minimize substrate accumulation, alternating phases of excess and limited substrate availability are necessary - controlled q_S oscillations. But in the context of dynamically changing conditions a precise quantification of q_{Scrit} requires high frequency data, which is especially challenging in case of supernatant substrate analytics. Reaching the goal of a robust and temporally resolved quantification of q_{Scrit}, the here introduced q_{Sbal} approach assessed regarding benefit and robustness.

Better closing C-balances substantiate q_{Sbal} approach

Since the analytical data is often correlated to noise, the benefit of reducing the necessity of analytical offline measurements was investigated by comparing the conventional approach of including supernatant measurements to the q_{Sbal} approach. Consequently, the average level as well as the noise of C-balance were analyzed and compared by statistical means. Using the induction phase as window of calculation, the mean level of C-balance (FIGURE 3 A) as well as the average standard deviation (FIGURE 3 B) were calculated and compared. It can be clearly observed that the mean C-balance value of the q_{Sbal} approach (from OUR) is significantly increased ($p(t) = <0.001$), indicating a better defined system. Additionally the standard deviation was significantly decreased ($p(t) = 0.0023$) which indicates a lower level of noise and consequently substantiates the benefit of the q_{Sbal} approach. Although it appears obvious, that by excluding the noise originating from the analytical offline data, data quality is improved, Figure 3 nicely illustrates the benefit of using the q_{Sbal} approach. Nevertheless, it has to be mentioned, that this approach is highly dependent on the quality of the offgas measurement data.

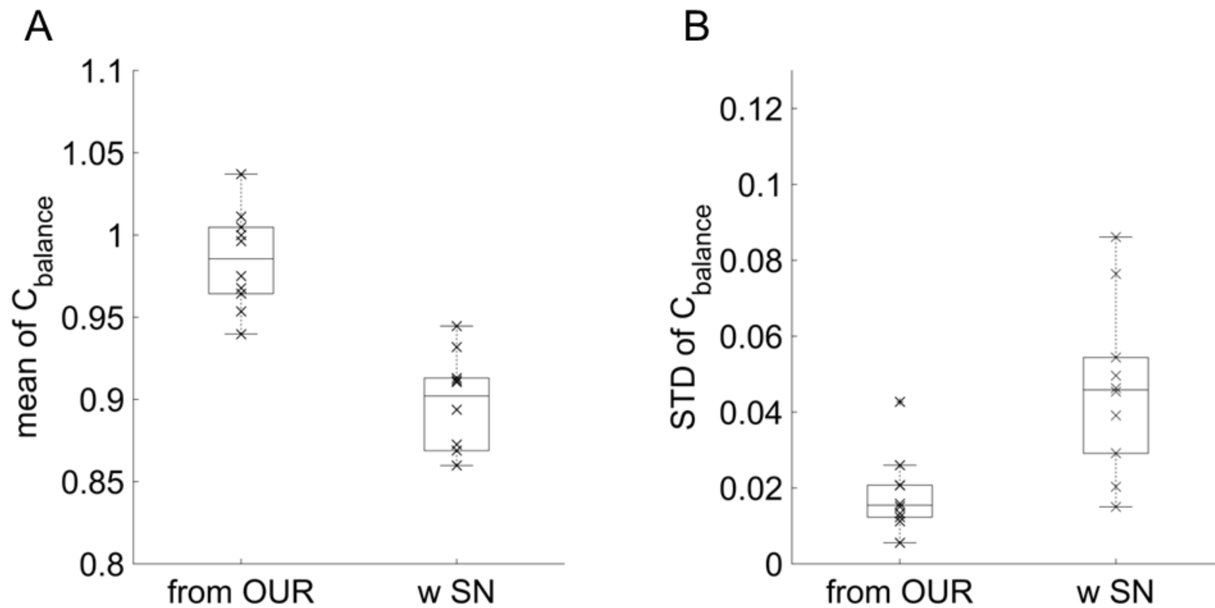


FIGURE 3 Cumulatively mass balance calculation based on q_{Sbal} yields a better defined system;
A: The boxplots show the mean values of the C-balance calculated at every offline BM (n=10); (from OUR) corresponds to the variable derived from the q_{Sbal} approach; (w SN) corresponds to the C-balance including SN measurements. The means for latter is significantly ($p(t) = <0.001$) lower than for the model based approach. This can be explained by not quantified components in the SN (refer to *Supplemental 1*). **B**: Comparison of noise: the average standard deviation of each experiment compared over the two balancing approaches; the averaged STD is significantly ($p(t)=0.0023$) lower for the balancing approach (from OUR) compared to the C-balance including SN measurements (w SN). This can potentially be attributed to analytical error of offline measurements which impact the conventional calculation to a greater extend.

Offgas data quality is critical for q_{Sbal}

Shifting the focus from offline measurements to online offgas data comprises risks as well. To test and to illustrate the correlated risks the consecutive set of experiments was designed to challenge the robustness of the outlined approach (q_{Sbal}) by real time controlling q_S on a steady trajectory in comparison to oscillating q_S between discrete plateaus. The experiments illustrated in FIGURE 4 underline the sensitivity of the q_{Sbal} approach to highly dynamic process conditions. The comparison of a steady feeding profile (FIGURE 4 A) to the oscillatory feeding profile (FIGURE 4 B) shows a substantial difference in noise of the C-balance. This phenomenon can be attributed to controller actions to maintain DO_2 within bounds (FIGURE 4 C/D). The regulation of the oxygen partial pressure leads to spikes in the OUR. This technical cause would hinder the correct estimation of r_{ox} in real time since it leads to increased noise in the DoR and via q_{Sbal} in the C-balance. In case the DO_2 controller actions only affect O_2 partial pressure of gassing, it would be indicated to focus on the CER instead for the q_{Sbal} approach. Nevertheless, the CER would be largely affected by pH fluctuations e.g. in case of overflow metabolism. To minimize such DO_2 controller actions the OTR should outweigh the OUR. This can be achieved by reducing the biomass concentration in the experimental design or by utilizing a system allowing a higher k_La value.

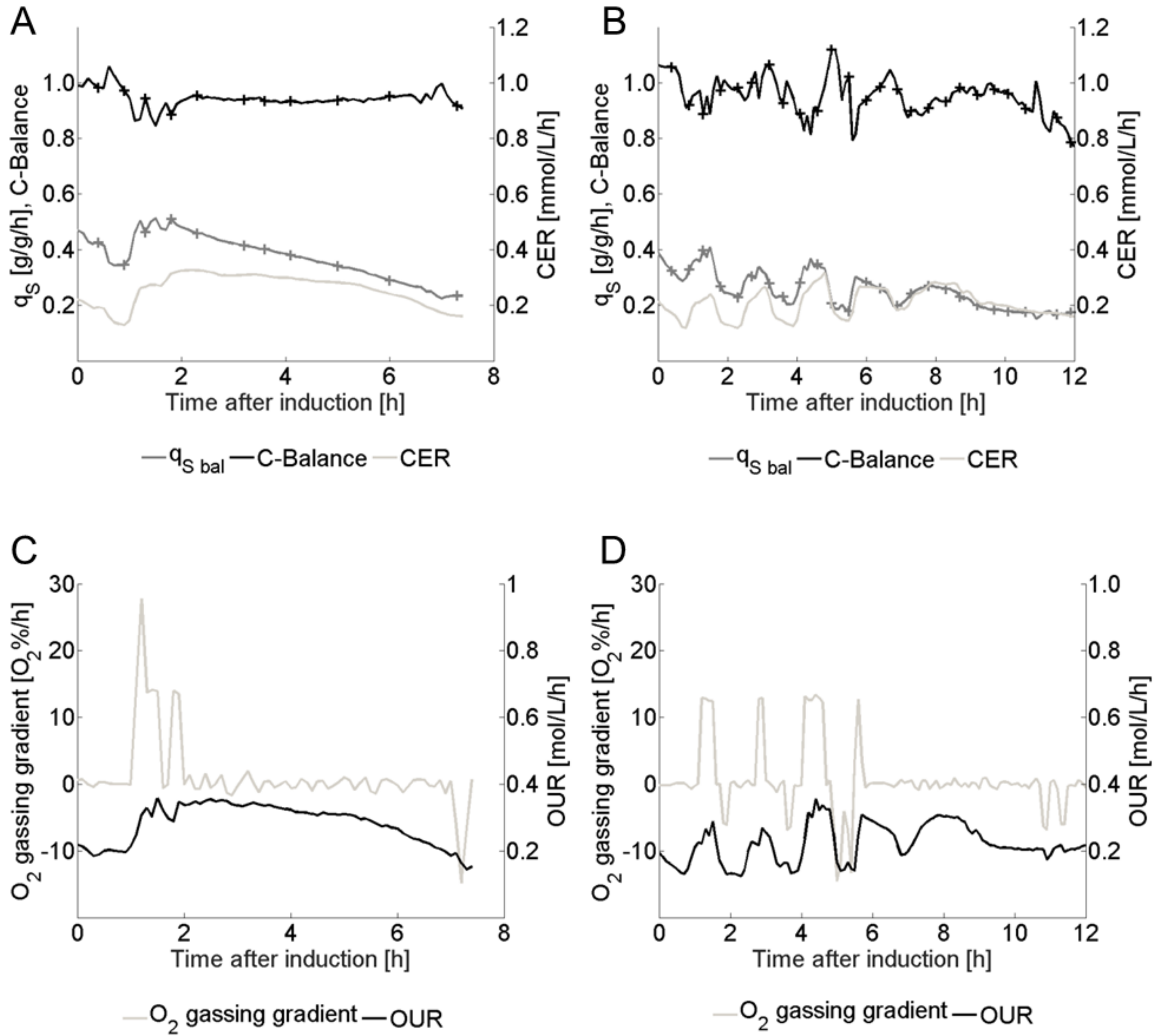


FIGURE 4 Process dynamics calling for frequent DO_2 controller action cause noise in C-balance in real time
A/B: C-balance calculated with balancing approach (left y-axis) in real time, q_{Sbal} and CER (right y-axis) for two different processes showing the limitations of the balancing approach, offline sampling points are indicated as symbols and resemble the verification measurements; C/D: The oxygen uptake rate (OUR) as well as the change in partial pressure of O_2 by the DO_2 step controller (O_2 gassing gradient) is displayed. A/C: q_S profile of low dynamics calls for less controller actions and consequently displays less noise in the C-balance; B/D: A highly dynamic process (oscillatory q_S profile) shows spikes in OUR, that directly affect the real time C-balance. The respective spikes in the C-balance coincident with the step controller actions, further data regarding these experiments is provided in Supplemental 3.

q_{Sbal} can predict accumulation of unknown metabolites

Since the C-balance based on q_{Sbal} displays a better closing C-balance and less correlated noise, the prediction of substrate accumulation in a real time context is of great interest. Besides the reduced need for identification of components, the estimation of the oxidative metabolism allows a higher temporal resolution and eases the quantification of accumulated C-source. Offline metabolite quantification is limited in temporal resolution by the sampling interval of supernatant. In case of supernatant sampling the separation of cells and supernatant is a highly time critical step which can greatly interfere with the data quality. Biomass sampling is less sensitive consequently automation is less challenging which in turn generally allows for a higher sampling frequency than for supernatant.

In this context, estimating the oxidative metabolism using high resolution biomass sampling an offgas data via q_{Sbal} reduces the importance of the supernatant sampling and grants a higher temporal resolution (FIGURE 5 A/B). Within a q_S controlled experiment the use of the q_{Sbal} approach consistently predicted a higher level of accumulation than substantiated by measuring the expected metabolites as acetate and accumulating substrate (FIGURE 5 A). Only further going analysis of the supernatant revealed the release of additional C-sources as organic acids (Supplemental 1). A bias introduced by sample handling in combination with release of unquantified organic compounds are the presumed reasons for the difference of the average level of the C-balances. Owned to the higher level of estimated accumulation r_{ox} is smaller, which in turn affects the yields of biomass and of CO_2 (FIGURE 5 B). To sum it up, using the qSbal approach allows the estimation of accumulated C-source based on offgas data and biomass sampling. Since biomass sampling and quantification is less sensitive than supernatant sampling and C-source quantification data quality is increased.

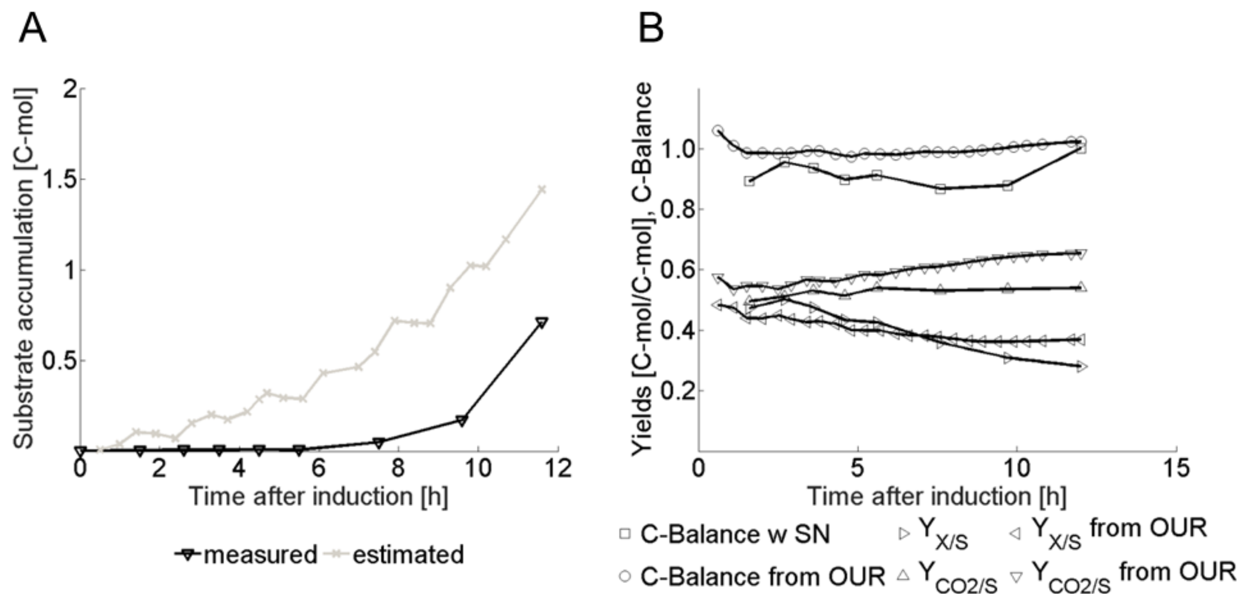


FIGURE 5: Model based data evaluation leads to an overall improvement of data quality and indicates accumulation of a not quantified metabolite

Points indicate the time point of calculation which is based on offgas measurement and biomass quantification and as indicated the metabolite/substrate quantification in the supernatant, lines in-between the symbols are meant as guide for the eye; **A**: Comparison of the theoretical accumulation (estimated) calculated using the cumulative balancing approach and actual measured accumulation (measured) in the SN, symbols depict the offline sampling points; the estimated accumulation in C-mol is calculated as the difference in the metabolized substrate deduced from both, the r_S from DoR- and C-balance and the r_S calculated from the feeding rate but without supernatant measurements. The calculated accumulation is constantly higher than the actual measured total accumulation of C-source; this can be attributed to the non-quantified compounds in the SN (see Supplemental 1). **B**: (from OUR) corresponds to the variable derived from the q_{Sbal} approach; (with SN measurements) corresponds to the C-balance including SN measurements. Cumulative molar yields ($Y_{X/S}$; $Y_{CO2/S}$; $Y_{X/S}$ from OUR; $Y_{CO2/S}$ from OUR) as well as C-balance calculated cumulatively with both methods ($C_{balance}$ w SN; $C_{balance}$ from OUR).

q_{Sbal} displays dependency on metabolic activity

The outlined approach of q_{Sbal} grants a high temporal resolution only dependent on biomass sampling which finally allows to assess the trajectory of q_{Scrit} over time in response to the average level of physiological activity (q_{Smean}) within q_S oscillations experiments. Various physiologically controlled q_S oscillation experiments were conducted (n=9) to investigate the interrelation of q_{Smean} and q_{Sbal}. On the basis of FIGURE 6, the dependency of q_{Scrit} on time and moreover on metabolic activity (q_{Smean}) shall be brought to the reader's attention. Based on the outlined strategy of quantifying the oxidative metabolism, the q_S-SP (set point) can be put in relation to a q_S-PV (process value) (FIGURE 6 A/C). If the PV of q_S does not equal the SP, the metabolic state of the culture is not under control which can have two reasons: i) If the SP is smaller than the PV, accumulated C-source is being oxidized. ii) If the SP is greater than the PV q_{Scrit} is being exceeded. Consequently substrate/metabolites are being accumulated until the SP is reduced in the consequent phase of low qs-SP. By fitting a linear function, a time dependent decline of q_{Scrit} becomes apparent (FIGURE 6 A/C; Supplemental 3). Nevertheless, this time dependent decline does not appear to be transferable in-between experiments (FIGURE 6 A/C), indicating a further significant but underlying variable. FIGURE 6 B illustrates the significant ($p< 0.001$) correlation of the q_{Scrit} slope in dependency of q_{Smean}.

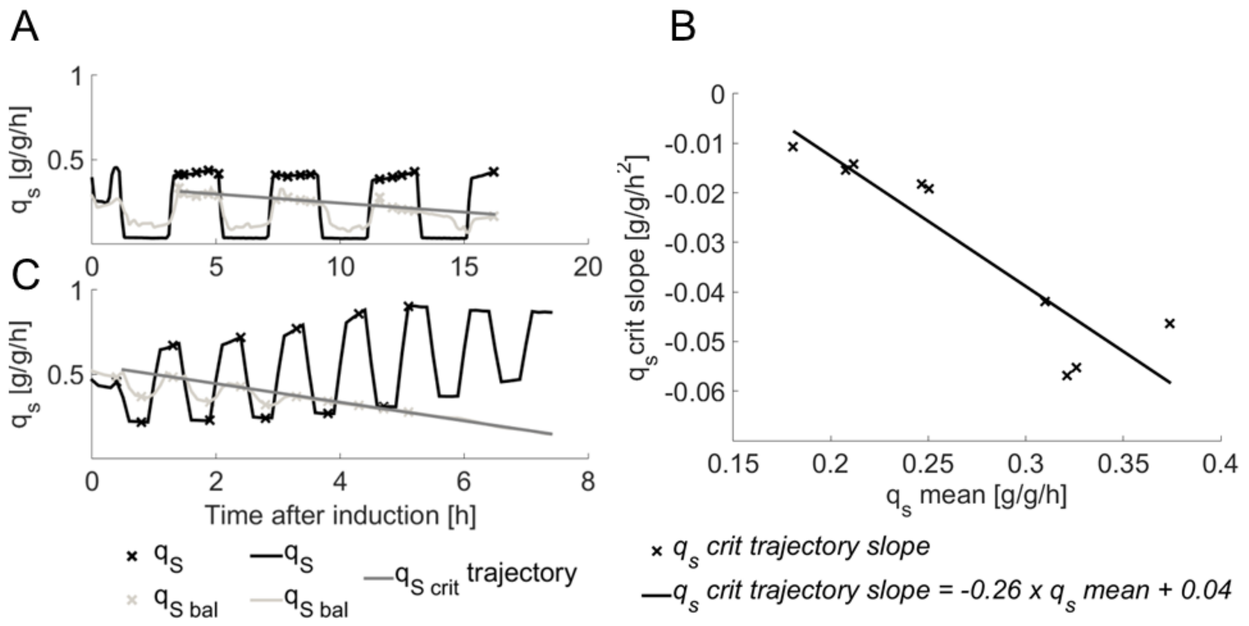


FIGURE 6: The critical specific substrate uptake rate is a function of time and average metabolic activity (q_{Smean})
A & **C**: Two exemplary q_S oscillation profiles of total nine experiments (Supplemental 3) to determine the q_{Scrit} trajectories; the symbols depict the offline samples used for calculation of the q_{Scrit} trajectories. The lines in-between correspond to the real time calculations. q_S calculated from the balancing approach (q_{Sbal}), q_S which was calculated from the substrate feeding rate (q_S). A linear curve was fitted into these data points describing the decline of q_{Scrit} (q_{Scrit} trajectory); **B**: Dependency of the decline of the slope of the q_{Scrit} trajectory on the q_{Smean} for 9 different oscillation experiments (with different q_S means). The slope of the q_{Scrit} trajectories (q_{Scrit} trajectory slope) in [g/g/h^2] is negatively correlated to q_{Smean} [g/g/h];

Discussion

Within this contribution we challenged the common understanding of q_{Scrit} as a “constant” and tested our hypothesis of it being a variable. Furthermore, we investigated the state of the art approach of quantifying q_{Scrit} and aimed to establish a approach to quantify the q_{Scrit} independent of q_{Smean}. On the basis of the exemplary investigation of the physiological feasible space we want to raise awareness concerning the dynamic nature of the physiological capacity of q_{Scrit}.

Maximum productivity is more often correlated to high specific growth rates [22, 23]. High specific growth rates in turn require high substrate supply [24]. So while the maximization of the substrate supply and therefore q_{Smean} is in favor of productivity it increases the risk of exceeding the physiologic capacities and to consequently accumulate substrate. To quantify the q_{Scrit} repetitive up-pulsing were used. In differentiation to literature [5], a higher final pulse concentration of 20 g/L glucose was utilized. Nevertheless the pulse concentration was 50% lower in comparison to Phue et al. [25] to assure saturation while minimizing pulse duration.

The pulse experiments (FIGURE 1) substantiated the impact of time of induction as well as the impact of the process parameter temperature on q_{Scrit}. This correlation of metabolism and temperature is regarded as common knowledge from a theoretic metabolic point of view but has not been quantified up this point. Using repetitive up-pulsing it was possible to deduct a q_{Scrit} trajectory.

For verification of the deducted q_{Scrit} trajectory a verification experiment was designed to maximize q_{Smean} while avoiding q_{Scrit}. Nevertheless, the conducted experiment to real time control q_S on a trajectory following q_{Scrit} has led to substrate accumulation 4 h earlier than predicted by the trajectory. Presumably, the continuously high metabolic activity during the steady q_S experiment (FIGURE 2) led to a faster decline in $q_{Scrit,}$ in comparison to the repetitive pulse experiments, which comprise phases of recovery in-between pulses (30 min). In this context the investigation of correlated process parameters is highly challenging and has been attempted previously. But using a volumetric constant feeding rate the inherent time dependent decline of μ can compromise conclusions regarding the inferred correlation of μ and q_{Scrit} [7].

Periodic up-pulsing of the feeding rate has been used to timely resolve the trajectory of the maximum physiologic capacity of the respective strain, but this approach is technologically highly demanding regarding DO_2 control [17]. The widespread dependency on the DO_2 signal as response for process control and/or data evaluation [5-8, 17] constitutes a major drawback to the opinion of the author. DO_2 levels below 30% have been reported to alter transcription and to impact the metabolism [25]. This fragile metabolic state in-between glucose saturation [5] and oxygen limitation [7] questions the experimental setup of up-pulsing in general as well as relying on DO_2 for control purposes. This problem can be avoided by down pulsing [6] or by decoupling control and data analysis from DO_2.

Controlling the physiological rate of interest, avoids the question whether the observed decline in e.g μ_{crit} is actually correlated to time or actually to an overestimation of produced biomass as a result of the assumption of a constant biomass yield as concluded by Henes et al. [6]. Summarizing, published work on the quantification of physiologic capacities as q_{Scrit} widely underappreciates real time biomass estimation, while relying on the highly sensitive DO_2 signal as response.

Since the state of the art of quantifying q_{Scrit} led to an unsatisfactory result an alternative workflow is required, which includes physiological control and decreases the dependency on DO_2 as response signal. In this context a controlled q_S oscillation experiment, as approach to quantify q_{Scrit} , was introduced. Regarding a workflow of quantifying the physiologic capacities, the subsequent step of data evaluation is crucial for accuracy and highly dependent on data quality. Especially in high cell density (HCD) bioprocesses, transitions from growth to limitation can occur within 15-30 s [17], which emphasizes the necessity of high frequency sampling. In contrast to biomass quantification, the volatility of metabolites and substrates in the supernatant constitutes a significant risk for metabolite/substrate quantification. By basing data evaluation on biomass and offgas data as inputs for the here introduced q_{Sbal} model approach, only verified by analytical data, noise is reduced and overall data quality can be increased (FIGURE 3). Since the model approach provides an estimation of the total accumulation of C-source, the necessity of identification and quantification of each metabolite is reduced.

In a subsequent set of experiments of nine q_S controlled oscillations in-between discrete levels, the correlation of q_{Scrit} decline and q_{Smean} could be illustrated. A positive correlation of q_{Scrit} and μ has been suggested before [7]. However, in the latter contribution, the C-source was provided by a constant feeding rate. Consequently, the growth in biomass leads to an inevitable time correlated decline in μ. Not controlling the physiology in terms of μ hereby impairs the differentiated identification of the cause of q_{Scrit} decline between time and μ. The in this contribution established impact of the physiologic descriptor of q_{Smean} on q_{Scrit} during induction phase could be interpreted as the cumulating impact of the metabolic stress imposed by protein expression. This theory would align with literature reporting the high level of metabolic stress inferred by the IPTG induction system [26, 27]. Upon induction with IPTG the metabolic focus is directed towards target protein expression. Consequently, the average level of metabolic activity described by q_{Smean} translates into energy invested into protein expression and therefore stress. Also, such an cumulating impact has been widely discussed for higher organisms as metabolic memory effect [28]. For microbials the correlation of q_{Smean} and q_{Scrit} infers a descriptive function of q_{Smean} towards a memory effect.

Summarizing, a sound science based identification of the cause requires the control of one of the two variables of interest, as presented in this contribution. This generally applicable principle emphasizes the necessity for physiological process control in the context of physiological process development.

Conclusions

The goal of this paper was to analyze the root causes of observed substrate accumulation during the induction phase for recombinant protein production using *E. coli* as a host. We hypothesized that q_{Scrit} is variable and dependent on time as well as on physiology. Based on the presented data following conclusions were deducted concerning the latter hypothesis

1. The conducted up-pulsing experiments revealed that q_{Scrit} is dependent on induction time and on the process parameter temperature.
2. We found for the first time that the decline of q_{Scrit} is closely correlated to metabolic activity as q_{Smean}, suggesting a cumulating impact of substrate metabolization during induction phase.
3. Based on the impact of q_{Smean} on the decline of q_{Scrit}, we propose to utilize q_{Smean} as descriptor for the metabolic memory effect.

The illustrated time dependency of q_{Scrit} requires a time resolved q_{Scrit} trajectory instead of a static numeric value. However, since this trajectory is additionally dependent on process parameters and on memory effects, a large number of experiments is required in order to avoid accumulation. As an alternative, we propose to utilize the outlined model for a real time feedback control on physiological variables was. Hereby, the model would require real biomass estimation and offgas analysis as input variables and deliver a process value of the current q_S as output. Thereby, a simple step controller would facilitate the feedback control of q_S, while sensing q_{Scrit} in real time. This online detection of q_{Scrit} could eliminate the need for experiments to determine the q_{Scrit} trajectory for each setting of process parameters and strains.

ABBREVIATIONS

CH_2O	c-molar substrate composition
$CH_{1,82}O_{0,5}$	c-molar biomass composition without nitrogen
c_{feed} …	substrate concentration in feed [g/L]
$dSn_{(t)}$ …	fed substrate normalized by the CDW at the end exp. fed-batch [g/g]
$F_{(t)}$ …	feed flow rate [L/h] after time (t)
$F_{S,V}$ …	flow rate of feed solution [L/h]
S_{in}…	substrate inflow [c-mol/h]
q_S…	biomass specific substrate uptake rate [c-mol/c-mol/h] or [g/g/h]
$q_{S(t)}$…	biomass specific substrate uptake rate [g/g/h] at time point (t)
q_{Smean} …	average q_S within a predefined window of dSn or time [g/g/h]
q_{Sbal} …	biomass specific substrate uptake rate calculated with C-balance and DoR-Balance [c-mol/c-mol/h]
q_{Scrit} …	the critical specific substrate uptake rate as defined by Åkesson, Hagander, & Axelsson, 1999 [g/g/h]
q_{Sglc}	specific substrate uptake rate calculated from the glucose concentration gradients in the pulse experiments [g/g/h]

q_{Sglc} fit	linear fit to determine the slope of the decline in q_S glc of the pulse experiments
RQ …	respiratory quotient [mol/mol]
r_{acc} …	rate of accumulating substrate and acetate [c-mol/h]
r_{CO2}	CER, carbon dioxide evolution rate [mol/ h]
r_S	substrate conversion rate [c-mol/h] or [g/h]
r_{O2} …	OUR, oxygen uptake rate [mol/h]
r_X	biomass conversion rate [c-mol/h]
γ_S …	Degree of Reduction (DoR) of substrate [mol/mol]
γ_X …	Degree of Reduction (DoR) of biomass [mol/mol]
μ …	specific biomass growth rate [1/h]
V_0 …	volume at t = 0 [L]
X…	CDW [g] at (0) batch end or after time (t)
X_m…	Biomass [c-mol]
x…	CDW concentration [g/L] at (0) batch end or after time (t)
$Y_{X/S}$…	biomass yield on substrate [g/g] or [c-mol/c-mol]
$Y_{CO2/S}$…	carbon dioxide yield on substrate [c-mol/c-mol]
$Y_{O2/S}$…	oxygen yield on substrate [c-mol/c-mol]
$Y_{H2O/S}$…	water yield on substrate [mol/c-mol]

indices:

g…	gravimetrical entity
m…	c-molar entity
t…	process time [h]
SP …	set point; the intended value of a given process parameter
PV …	process value; the actual value of a given process parameter based on measured quantities
w SN …	addition of w SN indicates values calculated with taking glucose and acetate measurements in supernatant into account
w/o SN …	addition of w/o SN indicates values calculated without taking glucose and acetate measurements in supernatant into account
bal …	denotes values calculated with the balancing approach (Equation 5 - Equation 12)

Declarations

a. Acknowledgements

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

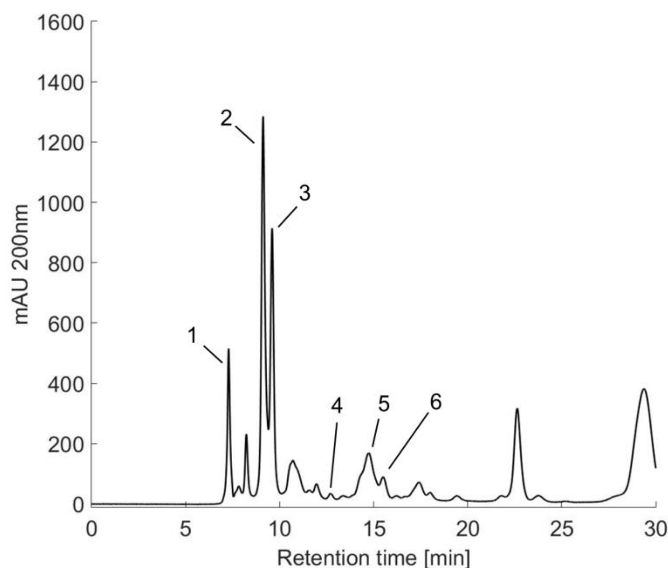
b. Authors' contributions

HC supervised the study. WR and MB conceived the study. All the authors participated in the experimental design. MB, PT, PK performed the cultivations, and analytics. WR and MB analyzed the data and wrote the first manuscript version. WR and JF finalized the manuscript for submission. All authors read and approved the final manuscript.

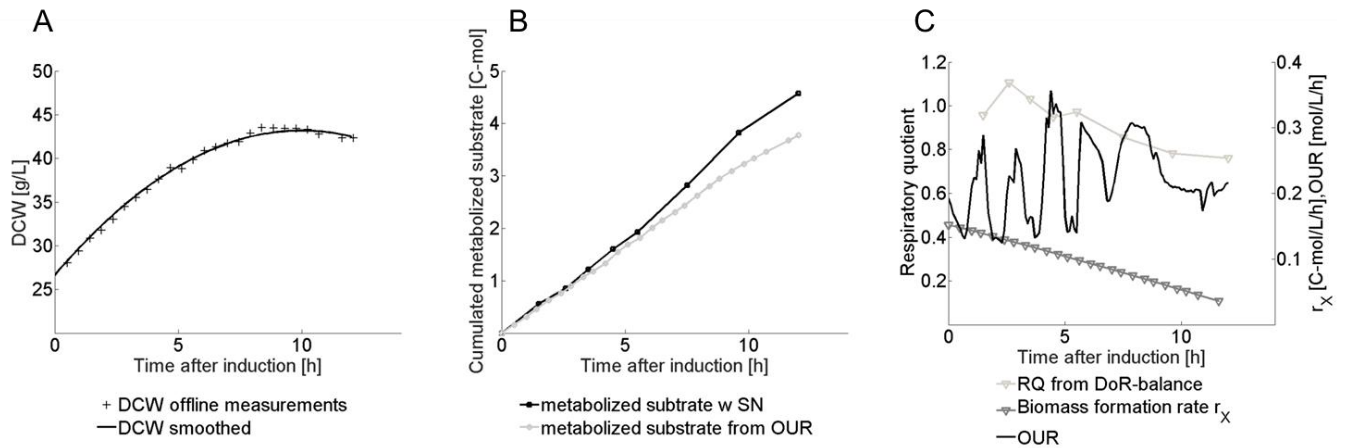
c. Conflicts of Interest

The authors declare no conflict of interest.

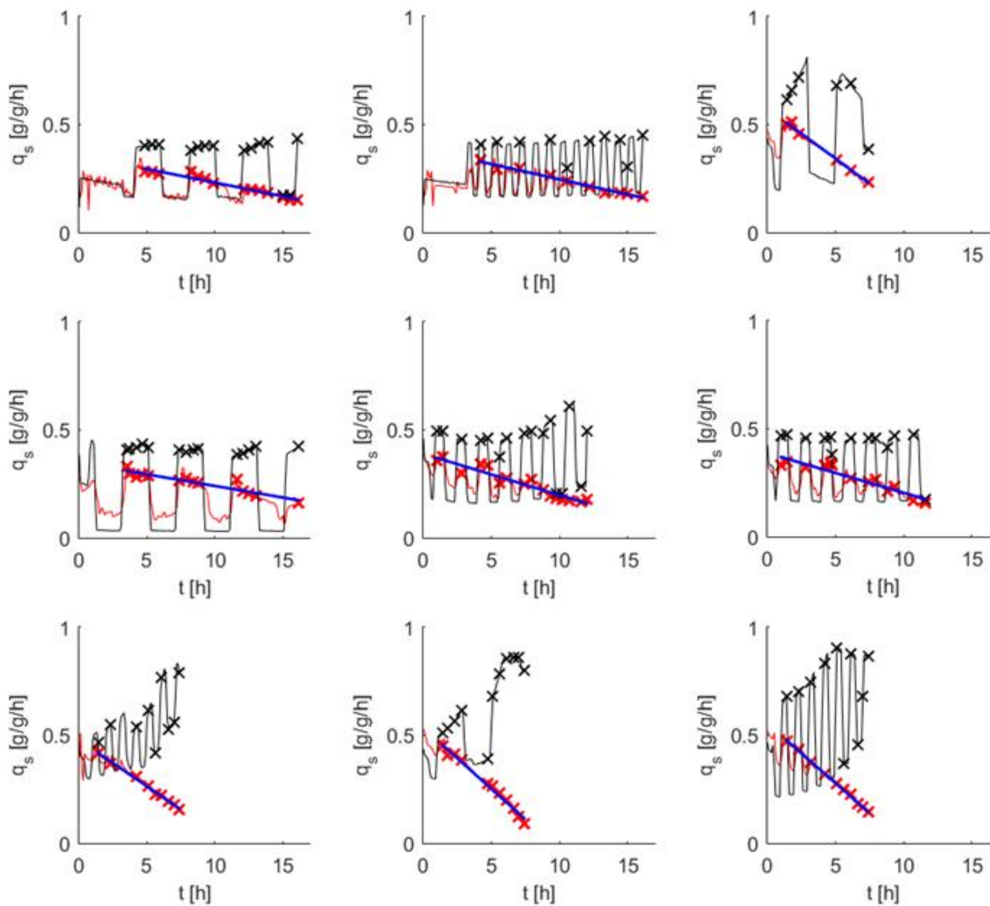
Supplemental



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



Supplemental 2: (A) Biomass dry cell weight (DCW) in g/L; (B) metabolized substrate calculated from the feeding rate and supernatant measurements (black line) and from DoR- and C-Balance (blue dotted line) with marks representing the offline measurement points for biomass (o) and supernatant (+); (C) RQ (red line) calculated with the cumulative biomass yield and cumulative CO2/S Yield, the oxygen uptake rate in mol/L/h (black line) and the biomass formation rate rX in mol/h, marks show the offline measurement points for biomass (triangles) and supernatant (x); The data corresponds to the experiment shown in FIGURE 4 B/D & FIGURE 5 A/B.



Supplemental 3: decline of qScrit over time (all included experiments); The data points correspond to offline sampling points for which the qS , calculated from the qSbal balancing approach (red), is lower than the qS which was calculated from the feeding rate (black). A linear curve was fitted into these data points describing the decline of qScrit (blue) – quality criteria for this fit was R2 > 0.8. B

References

[1] K. Terpe, Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems, Applied Microbiology and Biotechnology 72(2) (2006) 211-222.
[2] G. Walsh, Biopharmaceutical benchmarks 2010, Nat Biotech 28(9) (2010) 917-924.
[3] E.B. Jensen, S. Carlsen, Production of recombinant human growth hormone in Escherichia coli: expression of different precursors and physiological effects of glucose, acetate, and salts, Biotechnol Bioeng 36(1) (1990) 1-11.
[4] G.W. Luli, W.R. Strohl, Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations, Applied and environmental microbiology 56(4) (1990) 1004-11.
[5] M. Akesson, E.N. Karlsson, P. Hagander, J.P. Axelsson, A. Tocaj, On-line detection of acetate formation in Escherichia coli cultures using dissolved oxygen responses to feed transients, Biotechnol Bioeng 64(5) (1999) 590-8.
[6] B. Henes, B. Sonnleitner, Controlled fed-batch by tracking the maximal culture capacity, Journal of Biotechnology 132(2) (2007) 118-126.
[7] H.Y. Lin, B. Mathiszik, B. Xu, S.O. Enfors, P. Neubauer, Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of Escherichia coli, Biotechnol Bioeng 73(5) (2001) 347-57.
[8] M. Åkesson, P. Hagander, J. Axelsson, A probing feeding strategy for Escherichia coli cultures, Biotechnology Techniques 13(8) (1999) 523-528.
[9] W.N. Reichelt, F. Haas, P. Sagmeister, C. Herwig, Bioprocess development workflow: transferable physiological knowledge instead of technological correlations (under submission), Biotechnological Progress (2016).
[10] D. Levisauskas, R. Simutis, D. Borvitz, A. Lübbert, Automatic control of the specific growth rate in fed-batch cultivation processes based on an exhaust gas analysis, Bioprocess Engineering 15(3) (1996) 145-150.
[11] S. Gnoth, M. Jenzsch, R. Simutis, A. Lubbert, Control of cultivation processes for recombinant protein production: a review, Bioprocess Biosyst Eng 31(1) (2008) 21-39.
[12] D. Levisauskas, Inferential control of the specific growth rate in fed-batch cultivation processes, Biotechnology Letters 23(15) (2001) 1189-1195.
[13] M. Jenzsch, R. Simutis, A. Luebbert, Generic model control of the specific growth rate in recombinant Escherichia coli cultivations, J Biotechnol 122(4) (2006) 483-93.
[14] A.J. de Assis, R.M. Filho, Soft sensors development for on-line bioreactor state estimation, Computers & Chemical Engineering 24(2–7) (2000) 1099-1103.
[15] W.N. Reichelt, P. Thurrold, M. Brillmann, J. Kager, J. Fricke, C. Herwig, Generic biomass estimation methods targeting physiologic process control in induced bacterial cultures, Engineering in Life Sciences (2016) n/a-n/a.
[16] I.S. Hunter, H.L. Kornberg, Glucose transport of Escherichia coli growing in glucose-limited continuous culture, The Biochemical journal 178(1) (1979) 97-101.
[17] S. Schaepe, A. Kuprijanov, R. Simutis, A. Lübbert, Avoiding overfeeding in high cell density fed-batch cultures of E. coli during the production of heterologous proteins, Journal of Biotechnology 192, Part A (2014) 146-153.
[18] D.J. Korz, U. Rinas, K. Hellmuth, E.A. Sanders, W.D. Deckwer, Simple fed-batch technique for high cell density cultivation of Escherichia coli, Journal of Biotechnology 39(1) (1995) 59-65.
[19] P. Wechselberger, P. Sagmeister, H. Engelking, T. Schmidt, J. Wenger, C. Herwig, Efficient feeding profile optimization for recombinant protein production using physiological information, Bioprocess Biosyst Eng 35(9) (2012) 1637-49.

[20] P. Sagmeister, P. Wechselberger, M. Jazini, A. Meitz, T. Langemann, C. Herwig, Soft sensor assisted dynamic bioprocess control: Efficient tools for bioprocess development, Chemical Engineering Science 96(0) (2013) 190-198.
[21] P. Sagmeister, P. Wechselberger, C. Herwig, Information Processing: Rate-Based Investigation of Cell Physiological Changes along Design Space Development, PDA J Pharm Sci Technol 66(6) (2012) 526-41.
[22] J.A. de Hollander, Kinetics of microbial product formation and its consequences for the optimization of fermentation processes, Antonie van Leeuwenhoek 63(3-4) (1993) 375-81.
[23] L.-C. Cheng, L.-I. Hor, J.-Y. Wu, T.-L. Chen, Effect of specific growth rate on the production of a recombinant nuclease by Escherichia coli, Biochemical Engineering Journal 14(2) (2003) 101-107.
[24] A. Varma, B.W. Boesch, B.O. Palsson, Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates, Applied and environmental microbiology 59(8) (1993) 2465-73.
[25] J.-N. Phue, J. Shiloach, Impact of dissolved oxygen concentration on acetate accumulation and physiology of E. coli BL21, evaluating transcription levels of key genes at different dissolved oxygen conditions, Metabolic Engineering 7(5–6) (2005) 353-363.
[26] P. Dvorak, L. Chrast, P.I. Nikel, R. Fedr, K. Soucek, M. Sedlackova, R. Chaloupkova, V. de Lorenzo, Z. Prokop, J. Damborsky, Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway, Microbial cell factories 14 (2015) 201.
[27] S. Schlegel, P. Genevaux, J.W. de Gier, De-convoluting the Genetic Adaptations of E. coli C41(DE3) in Real Time Reveals How Alleviating Protein Production Stress Improves Yields, Cell reports (2015).
[28] L. Zhang, B. Chen, L. Tang, Metabolic memory: Mechanisms and implications for diabetic retinopathy, Diabetes Research and Clinical Practice 96(3) (2012) 286-293.

Supplemental

CALCULATIONS

The textbook approach for calculation of q_S (Equation 4) requires the feed flow rate $\dot{F}_{(t)}$(L/h), the substrate concentration in the feed solution $c_{S,m}$(c-mol/L), and the biomass $X_{m,(t)}$ (c-mol) as input.

$$q_{S(t)} = \frac{r_S}{X_{m,(t)}} = \frac{\dot{S}_{in}}{X_{m,(t)}} = \frac{\dot{F}_{(t)} \times c_{S,m}}{X_{m,(t)}}$$

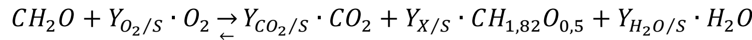
Equation 4: Specific substrate uptake rate from feed flow rate

Equation 4 is based on the assumption that the fed substrate is fully metabolized $\left(\frac{\mathrm{dS}}{\mathrm{dt}} = 0\right)$. In the latter case the substrate conversion rate r_S [C-mol/h] corresponds to the feed flow rate $\dot{S}_{in}$ [C-mol/h] . In case physiological maxima (e.g. q_{Scrit}) are exceeded and accumulation occurs $\left(\frac{\mathrm{dS}}{\mathrm{dt}} \neq 0\right)$ r_s follows Equation 5.

$$\dot{S}_{in} = \mathrm{r}_S + \frac{\mathrm{dS}}{\mathrm{dt}}$$

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

For the offline calculation of q_S and metabolized substrate an approach based on material balancing [27] was used which relies on the availability of high quality offline biomass data. In general, below q_{Scrit} growth of *E. coli* is fully oxidative on glucose as carbon source ($\frac{\mathrm{dS}}{\mathrm{dt}} = 0$) and follows the C-balance specified by Equation 6 [28].



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

Equation 6: E. coli growth equation for carbon

But physiological maxima $q_{Scrit(t)}$ is variable and a function of time and q_{Smean} (FIGURE 6). Exceeding q_{Scrit} consequently constitutes an increasing risk of substrate accumulation ($\frac{\mathrm{dS}}{\mathrm{dt}} \neq 0$) during fermentation. In order to respect $q_{Scrit(t)}$ the variable has to be quantified. But the risk of substrate accumulation close to $q_{Scrit(t)}$ prohibits a direct correlation of the molar substrate (Equation 5). To avoid the necessity of laborious offline supernatant analytics to quantify $\frac{\mathrm{dS}}{\mathrm{dt}}$, r_{CO2} and r_X for the estimation of the substrate conversion rate r_S according to Equation 7 was used. To quantify the oxidative substrate conversion rate r_s only the oxidation to carbon dioxide (r_{CO2}) and biomass growth (r_X) have to be considered, making the quantification of $\frac{\mathrm{dS}}{\mathrm{dt}}$ irrelevant.

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

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

The ratios between biomass growth and carbon dioxide production are defined by the yield coefficients (Y_{xs}/Y_{CO2}). From Equation 7 follows, that the sum of CO_2 ($Y_{CO2/S}$) and biomass yield ($Y_{X/S}$) equals 1(Equation 8).

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

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

So even if the yields are dynamically changing, r_S can be calculated from the sum of r_{CO2} and r_X. Consequently q_{Sbal} is calculated from dividing (Equation 9) by biomass (X_m).

$$q_{Sbal} = \frac{r_{CO_2} + r_X}{X_{m,(t)}}$$

Equation 9: Molar specific substrate uptake rate from balancing approach

As the CO_2 equilibrium between the liquid and the gas phase changes with pH, the r_{CO2} is prone to error under highly dynamic experimental conditions that lead to pH changes. To make the system more robust, r_S is calculated from r_{O2} under consideration of RQ. The RQ (Equation 10) as the ratio between CO_2 production and O_2 consumption and can be expressed by the rates r_{CO2} and r_{O2} as well as by the corresponding molar yields ($Y_{CO2/S}$, $Y_{O2/S}$).

$$RQ \cong \frac{r_{CO_2}}{r_{O_2}} = \frac{Y_{\frac{CO_2}{S},m}}{Y_{\frac{O_2}{S},m}}$$

Equation 10: Respiratory quotient from rates and from yields

As the oxygen yield on substrate in mol/C-mol ($Y_{O2/S}$) is calculated from the DoR balance with the degree of reduction of substrate (γ_S) and biomass (γ_X) the molar biomass yield in C-mol/C-mol ($Y_{X/S}$) (Equation 11) the robustness of the calculation approach against changes in pH is increased. The molar biomass yield is known from offline biomass and supernatant measurements.

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

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

C-Balance w SN / C-balance w/o SN represent the carbon balance (Supplemental Equation 7) where the substrate conversion rate r_S [C-mol/h] is calculated from the feeding rate in L/h using the substrate concentration in the feed solution $c_{S,g}$ [C-mol/L] and the carbon dioxide evolution rate is calculated with offgas analysis data. All carbon balances shown in the figures are normalized by dividing Equation 7 by r_S.

C-balance from OUR is the normalized carbon balance after Equation 7 whereas r_S in C-mol/h is calculated via the DoR-balance (Equation 11) and the carbon balance (Equation 7).

Molar yields in C-mol/C-mol with the appendix "from OUR" used the metabolized substrate in C-mol calculated via DoR-/C-Balance in C-mol as denominator. Yields without the appendix from OUR had the metabolized substrate in C-mol calculated from the cumulated feed solution in L, the concentration of the feed solution in C-mol/L and the total glucose in the supernatant from glucose measurements (g/L) and the reaction volume (L) in g.

iv. Closed loop q_s control

Although, characterization of physiological capacities prior to process development is critical for physiological process development and experimental design, the procedure of quantification is cumbersome. Given the significant decline in q_{Scrit} over time in dependency of the metabolic activity (q_{Smean}) denies a one-time quantification of physiological capacities as being propagated in literature. As a response the physiological feasible space can be further limited which potentially might exclude the optimum, or alternatively more information is required in order to predict the decline of q_{Scrit}. Instead of tighter limitations or additional effort in form of experiments a real time control strategy would be preferable. To effectively avoid substrate accumulation a real time sensing of physiological capacities would be required. In the following contribution the feasibility of real time physiological capacity sensing is assessed by decoupling real time biomass estimation and estimation of the oxidative metabolism. The establishment of a physiological feedback control strategy would eventually complete the development and assessment of physiological process control approaches.

Avoiding substrate accumulation by real time sensing of physiological capacity to metabolize substrate in *E. Coli*

Wieland N. Reichelt[a], Andreas Kaineder[a], Jens Fricke[a], Christoph Herwig[a,b*]

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

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

ABSTRACT

Producing pharmaceutically relevant proteins in microbial bioprocesses, substrate accumulation has to be avoided for the sake of productivity and controllability. But during late induction phase unexpected substrate accumulation is a phenomena often observed in microbial bioprocesses. Despite comprehensive strain characterization and quantification of the physiological capacity to metabolize substrate without accumulation of substrate or metabolites (q_{Scrit}) e.g. by pulse experiments, substrate is found to accumulate in the supernatant. This accumulation is especially pronounced in the context of controlling a physiological variable e.g. the specific substrate uptake rate (q_S).

Recent literature has illustrated a clear dependency of q_{Scrit} not only on induction time but also on the level of metabolic activity. In other words in addition to time after induction the rate of substrate metabolization severely impacts the decline in q_{Scrit}. To effectively avoid substrate accumulation, the dynamics and dependencies of the q_{Scrit} highlight the necessity to sense this physiological capacity in real time. In this contribution a combination of mass balances was used to estimate the current oxidative substrate uptake rate independent of the concurrent biomass estimation, accurately in real time. These biomass yield independent estimations allow the calculation of a real time process value of the physiological variable q_S. In the context of physiological control of q_S the latter approach allowed for the first time physiological feedback control by the comparison of q_S setpoint and q_S process value. Moreover, using a simple algorithm exceedance of q_{Scrit} was detected real time in order to react upon such a breach by q_S setpoint adaptation. By successfully avoiding substrate an metabolite accumulation throughout induction phase of an industrial relevant production process, we were able to illustrate the feasibility of the physiological feedback control.

KEYWORDS

Physiological feedback control; critical physiological capacity; substrate accumulation; real time biomass estimation; oxidative metabolism

Background

Bioprocesses are increasingly employed for the production of pharmaceuticals owned to correlated cost efficacy. Given the simple genetic accessibility, high growth rates and low demands concerning media composition, *Escherichia coli* is one of the most exploited hosts for industrial production of recombinant proteins (Terpe, 2006; Walsh, 2010). To optimize productivity of a given bioprocess, process parameters and their interrelations are investigated during process development within a given range – the characterization space (Rathore & Winkle, 2009). One of the most promising factors used frequently to increase productivity - the substrate feed rate (Babaeipour, Shojaosadati, Khalilzadeh, Maghsoudi, & Tabandeh, 2008; Kavanagh & Barton, 2008; Levisauskas et al., 2003; Ramalingam, Gautam, Mukherjee, & Jayaraman, 2007; P. Sagmeister, Schimek, Meitz, Herwig, & Spadiut, 2014; Sanden et al., 2003) is also the most challenging one. Overfeeding the physiological capacity to metabolize substrate (q_{Scrit}) impairs productivity (Jensen & Carlsen, 1990) and leads to unwanted overflow metabolism or even substrate accumulation. But rather than being a constant q_{Scrit}, is dynamically changing in response to process parameters e.g. pH, temperature and induction time $q_{Scrit}=f(t,pH,T,\ldots)$. This dependency of q_{Scrit} makes the quantification of q_{Scrit} especially laborious but necessary to avoid overflow metabolism and substrate accumulation.

Conventional process development investigates the impact of process parameters on productivity within a design of experiment (DoE). To ensure feasibility of the experiments the DoE is commonly be designed within the technological and physiological feasible space. While technologic constraints (e.g. k_La, heat transfer rate) are setup specific and therefore commonly known, physiological constraints have to be assessed strain and product specific. Especially in respect of the substrate feeding rate the definition of the physiological feasible space is of great concern. Exceeding the physiological feasible space leads to metabolite formation and substrate accumulation, which negatively affect physiology (Luli & Strohl, 1990). Moreover substrate accumulation has been shown to negatively affect productivity (Jensen & Carlsen, 1990) as well as controllability of the respective bioprocess. Consequently, q_{Scrit} constitutes the main constraint to the physiologic feasible space. Given the importance of q_{Scrit} various approaches for the quantification of q_{Scrit} have been discussed in literature (Åkesson, Hagander, & Axelsson, 1999; Akesson, Karlsson, Hagander, Axelsson, & Tocaj, 1999; Henes & Sonnleitner, 2007; Lin, Mathiszik, Xu, Enfors, & Neubauer, 2001).

Other factors investigated within the respective DoE can potentially impact q_{Scrit} e.g. temperature and pH. Additionally, q_{Scrit} has been shown to be dependent on induction time (Reichelt, Brillmann, et al., 2016; Schaepe, Kuprijanov, Simutis, & Lübbert, 2014) as well as on the level of metabolic activity (Reichelt, Brillmann, et al., 2016). Concluding, q_{Scrit} comprises a highly dynamic nature which boosts the necessary effort for strain characterization regardless of the experimental approach utilized for quantification of q_{Scrit}. To reduce dependencies and to increase transferability bioprocess development has increasingly focused on specific physiological rates (Gnoth, Jenzsch, Simutis, & Lubbert, 2008; Henes & Sonnleitner, 2007; Levisauskas, 2001; Levisauskas, Simutis, Borvitz, & Lübbert, 1996) rather than on volumetric feeding rates. To overcome the challenge of a dynamically changing q_{Scrit} and to avoid substrate accumulation a closed loop control approach of a physiological variable is necessary. Focusing on specific physiological rates for process physiological development requires real time biomass estimation and physiological bioprocess control.

Various approaches have been outlined to overcome the challenge of biomass estimation and consequently facilitate the control specific physiological variables. For biomass estimation in general, literature favors data driven models or hybrid models (de Assis & Filho, 2000; Jenzsch, Simutis, & Luebbert, 2006). But for bioprocess development available historic process data is commonly scarce, which restricts the use of data based algorithms. In this context hard type sensor and first principle mass balance based approaches for biomass estimation are regarded as more feasible than data driven approaches. A weighted average based combination of redundant biomass estimations has been shown as highly beneficial to increase accuracy and robustness of the biomass estimation (Reichelt, Thurrold, et al., 2016).

Besides biomass estimation approach physiological control approaches can be discriminated by the controller category.

Two main categories of controllers are employed for physiological process control. While open loop controllers (feed forward) do not measure or estimate the current state of the controlled variable closed loop (feedback) controllers derive the current state of the controlled variable as process value from a direct or indirect measurement. Direct measurements refer to online measurements of e.g. pH, temperature, DO2 and off gas. Indirect measurements or estimations refer to computational values derived from direct measurements. For physiological process control the biggest challenge is the estimation of the variable of interest, since it commonly cannot be measured directly. Consequently the quality of the primary data as basis for subsequent computations is of often underestimated importance. In respect of transferability to industrial scale the number of necessary direct measurements is crucial, since manufactures tend to minimize of measurement device ports to avoid contamination sites.

Open loop controllers are most commonly employed for physiological process control . The high level of simplicity and its robustness concerning measurement errors are the main reasons for the common use.

While technological feedback controllers are widely used e.g. for temperature and pH, examples of closed loop control of physiological variables are extremely scarce. The reason for the limited examples for feedback control is the necessity to determine the variable of interest with sufficient accuracy. Only if the variable of interest can be determined fast enough and with a sufficient signal to noise ratio closed loop control is feasible. Nevertheless, closed loop control is of great interest as real-time quality insurance (Gnoth et al., 2008), owned to its capability to react on process perturbations constitutes the basis of the attributed robustness of closed loop control approaches.

In the stage of bioprocess development strain specific historic process data is commonly scarce the accuracy of multivariate approaches (e.g. artificial neuronal networks) which commonly feature a high degree of accuracy does not appear feasible. An light data driven example, merely using a set of three experiments, Jenzsch et al used an extended kalman filter for biomass estimation in combination with generic model control (Jenzsch et al., 2006) for closed loop control of the specific growth rate. Examples of algorithms independent of apriori information and complex mathematical models have been introduced (Levisauskas, 2001; Rocha, Veloso, Carneiro, Costa, & Ferreira, 2008) although provided experimental data is scarce.

Given the frequent correlation of substrate uptake rate and productivity a control objective can be to run a bioprocess at the highest possible substrate level below q_{Scrit}. These control approaches do not necessarily qualify as closed loop control approaches since the system response used for controller actions is of qualitative nature and not directly

correlated to the controlled variable. Hereby usually a basic open loop control approach is combined with a probing technique e.g. in order to assess the reaction of the culture to sudden substrate starvation or excess. Monitoring the DO2 as response to short substrate up-pulsing it can be determined whether the culture is being overfed or if the open loop feed profile can continue increasing the volumetric feed flow rate (Åkesson et al., 1999; Velut, de Maré, & Hagander, 2007). Based on the same principle also down pulsing by temporary intermittence of the substrate feed rate can be utilized to assess the substrate supply situation in order to maximize the substrate uptake during induction phase (Henes & Sonnleitner, 2007).

More advanced techniques target a quantitative assessment of the metabolic state and substrate accumulation. Since the direct sensing of metabolites and substrate is only possible with substantial analytical effort using an online HPLC or FTIR this approach is commonly not regarded as feasible to pilot or even production scale. Merely based on offgas analysis the respiratory quotient (RQ) can provide valuable insight into the metabolism of the cell (Jobe et al., 2003). But sensing overflow metabolism based on the RQ is only feasible if the metabolite has a different degree of reduction as the substrate. In case of glucose as substrate and acetate as metabolite this approach is consequently not feasible. Nevertheless, using first principle mass balances the approach of Jobe et al. (Jobe et al., 2003) was capable of differentiating between oxidative and oxireductive metabolic states.

We target an independent estimation of biomass and the oxidative metabolism of the culture using first principle mass balances. Since this approach requires off gas analysis besides a volume balance and merely the biomass composition as strain specific information this approach shall feature a high degree of transferability. Using a simple algorithm the controller shall facilitate avoiding substrate accumulation rather than mere sensing of the latter.

Goals

The goal of this contribution is the introduction of a transferable control concept independent of historic process data. Thereby the approach shall be employable even in early bioprocess development. Given the relevance of physiological bioprocess development this approach shall facilitate a reduction in necessary strain characterization experiments, by sensing q_{Scrit} in real time. The feasibility of the introduced approach to effectively avoid substrate accumulation, despite a rapidly declining q_{Scrit} , shall assessed within a fermentation.

Materials and methods

2.1 Cultivations

2.1.1 Bioreactor system

Fermentations were conducted in a DASGIP multi-bioreactor system with 4 parallel reactors with 2L of working volume each (Eppendorf; Hamburg, Germany). The reactors were equipped with baffles and three disk impeller stirrers. The DASGIP control software v4.5 revision 230 was used for control: pH (Hamilton, Reno, USA), pO2 (Mettler Toledo; Greifensee, Switzerland; module DASGIP PH4PO4), temperature and stirrer speed (module DASGIP TC4SC4), aeration (module DASGIP MX4/4) and pH (module DASGIP MP8). CO2, O2 concentrations in

the off-gas were quantified by a gas analyzer (module DASGIP GA4) using the non-dispersive infrared and zircon dioxide detection principle, respectively.

2.1.2 Strain & media

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

2.1.3 Process parameters

Pre-cultures were incubated at 30°C and 170 rpm to an OD_{600} of approx. 1.5 in 150 mL batch media and 2.5% batch volume aliquots were used for inoculation. After depletion of the C-source in an initial batch phase, the pre-induction fed-batch was started. The pre-induction feeding strategy was based on an exponential feed forward profile to maintain a predefined growth rate. On attainment of the predefined biomass the cultures were induced after 30 min adaption time. Stirrer speed was set to 1400 rpm and aeration to 1.4 v/v/m for the whole process. The pH was maintained at 6.9 by adding 12.5% NH_4OH, which also served as nitrogen source. The dissolved oxygen (DO2) was kept over 30% by supplementing oxygen to the air.

2.1.4 Pre induction: exponential feed forward profile

The starting feed rate in L/h (F_0) was calculated using a gravimetrical biomass yield in g/g ($Y_{X/S,g}$), the starting biomass in g (X_0), the concentration of the feed solution in g/L ($c_{S,g}$) as well as the specific biomass growth rate as described elsewhere (Wechselberger et al., 2012)

2.1.4 Post induction feeding strategy:

After the depletion of C-source in the batch phase the culture was induced with IPTG (1 mM). A step controller was used for *real time q_S feedback control* during the post induction phase. Therefore the feed rate is dynamically adjusted every 20 minutes, which is calculated with the base load (calculated with the setpoint) and the adjustment term (calculated with the difference between the setpoint and the process value) to actually reach the given set-point of q_S.

$$\dot{F}_{i+1} = \frac{qsSP_{i+1} \cdot X_i}{c_S} + \frac{(qsSP_i - qsPV_i) \cdot X_i}{c_S}$$

Equation 1: Feed rate for next control interval

Two fermentations with q_SSP = 0.2 g/g/h were performed parallel. In one reactor the setpoint was kept constant, while in the other one the automated setpoint adaption was activated.

2.2 Process evaluation and data analysis

Metabolic rates and yield coefficients were calculated with Matlab r2013 b (Mathworks; Natick, Massachusetts, USA). The calculation of specific rates and yield coefficients was conducted as described elsewhere (P. Sagmeister, Wechselberger, & Herwig, 2012).

2.3 Offline analytics:

2.3.1 Biomass dry weight (CDW)

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

2.3.2 Substrate conc. and small metabolites

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

ABBREVIATIONS

$\widehat{BM}_{mean}$ [g] estimated biomass based on the N-balance, the DoR balance and a permittivity measurement

$\widehat{BM}_{C-Bal}$ [g] estimated biomass based on the assumption that the accumulation term in the C-balance equals zero

rx [g/h] biomass growth rate in g/h and is calculated offline with the biomass dry weight in g/L and the reactor volume in L as inputs

$\widehat{rx}$ [g/h] real-time estimated biomass growth rate

rs [g/h] is calculated offline with the feeding rate in g/h and the glucose accumulation rate in g/h as inputs

$\widehat{rs}$ [g/h] real-time estimated substrate conversion rate

qs [g/g/h] the specific substrate uptake is calculated offline with the biomass dry weight and rs as inputs

$qsSV$ [g/g/h] setpoint of qs

$qsPV$ [g/g/h] the process value of qs, calculated with the estimated substrate conversion rate rs in g/h, the real-time estimated biomass in g as inputs

$\Delta qsPV_i$ [g/g/h] the change of $qsPV$ within the last control interval (20min), calculated with q_SPV_i and with q_SPV_{i-1}

cs [g/L] substrate concentration in feed

F_0 [L/h] starting feed rate in

q_S [c-mol/c-mol/h] or [g/g/h] biomass specific substrate uptake rate

r_{acc} [c-mol/h] rate of accumulating substrate and acetate

r_{CO2} [mol/ h] CER, carbon dioxide evolution rate

rs [c-mol/h] or [g/h] substrate conversion rate in

r_{O2} [mol/h] OUR, oxygen uptake rate

rx [c-mol/h] biomass conversion rate in

μ	[1/h] specific biomass growth rate in
V_0	[L] volume at t = 0
X_i	[g] biomass dry cell weight at t = 0 batch end or at time point t = i
$Y_{X/S}$	[g/g] or [c-mol/c-mol] biomass yield on substrate

Results

Physiological maxima for physiologic DoE design

Rather than controlling a technological variable at a constant level, physiological process control targets maintaining a physiological variable. To illustrate the goal of physiological bioprocess control Figure 1 illustrates biomass growth and the accordingly increased substrate flow rate. The most common approach for physiological control is based on a fixed biomass yield, neglecting the dynamics of the biomass yield for the sake of simplicity. Figure 1 a shows a more advanced control approach, including a real time biomass estimation from a softsensor. Following the control scheme of Figure 1 B, despite real time biomass estimation no estimate of the controlled variable is made. Instead of following a predefined substrate flow profile, the substrate flow rate is calculated incrementally via softsensor biomass estimation.

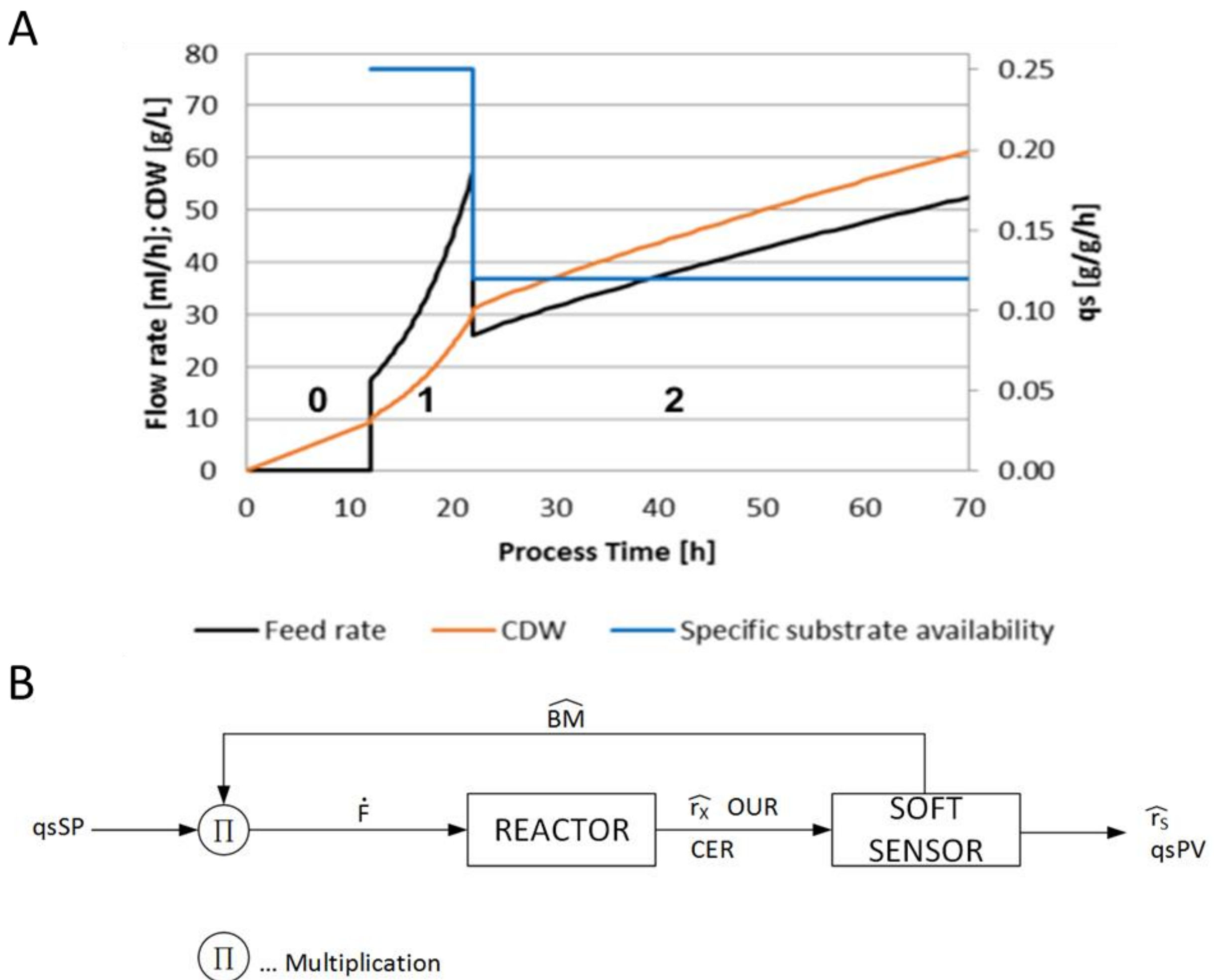


Figure 1 State of the art physiological open loop process control does not provide quantitative real-time information about the controlled variable:
(A) Open loop control including real time biomass estimation based on a softsensor with incremental calculations of the suitable substrate flow rate. (B) open loop controller design for incremental feed forward calculations.

In order to bring the limitations of the previously described controller setup to the reader's attention Figure 2 outlines a fed batch process including substrate accumulation. Upon glucose accumulation due to overfeeding, the biomass estimation of the softsensor clearly deviates from the offline verification data. This is owned to the fact that the used softsensor was based on the assumption that the accumulation term in the C-balance equals zero. Accordingly biomass estimation works well until substrate accumulation.

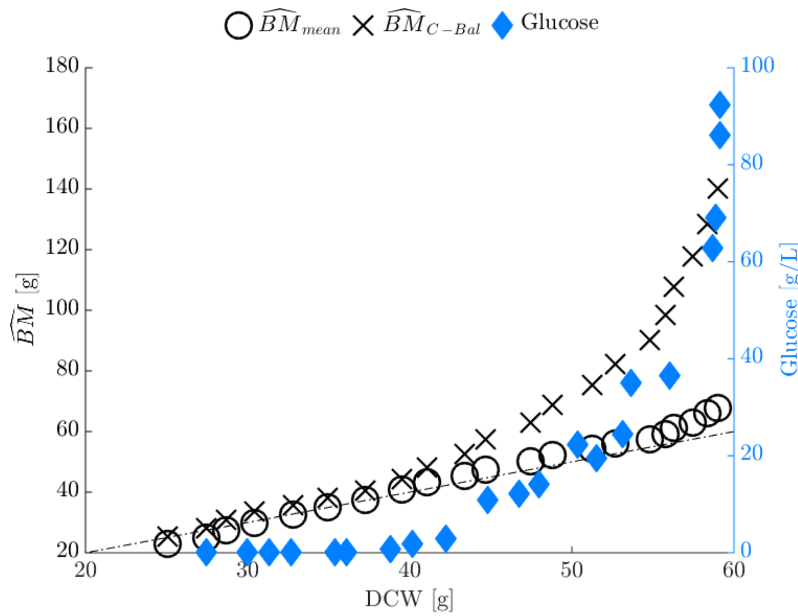


Figure 2 BM estimation based on C-balance is impaired by substrate accumulation;
BM real-time estimations (○,X) and accumulation of glucose (♦). The BM estimation via the C balance (X) shows a increasing error upon glucose accumulation and is not suitable for high q_S fermentations due to the possibility of q_{Scrit} exceedance. The BM estimation via weighted average of permittivity, DOR- & N balance (○) provides a more reliable estimation during glucose accumulation and was therefore chosen for the experiments.

Given the sensitivity of the C-balance based softsensor an alternative approach for biomass estimation is necessary. Altering the softsensor setup can greatly benefit the quality of estimation as well as the robustness of estimation. Figure 3 shows the result of using balancing approaches in combination with a permittivity measurement in order to obtain reliable estimates of the biomass rate (rx) and the oxidative substrate metabolization rate (r_{Sox}). Using a weighted average approach, the biomass estimations based on the N-balance, the DoR balance and a permittivity measurement are combined to make the approach robust against accumulation. Including the DoR balance is in this case straight forward since the primary metabolite (acetate) shares the same degree of reduction as the substrate. Figure 3 illustrates the correlated noise on the estimation of the most crucial rates: the biomass rate (Figure 3 A) and the oxidative substrate uptake rate (Figure 3 B). In case of substrate accumulation (data not shown) the C-balance based estimation quickly deviates from the rate derived from offline references biomass analytics. In comparison to the estimations merely based on the C-balance the noise on the rates of the weighted estimations lies within the order of magnitude of noise observed for the verification data.

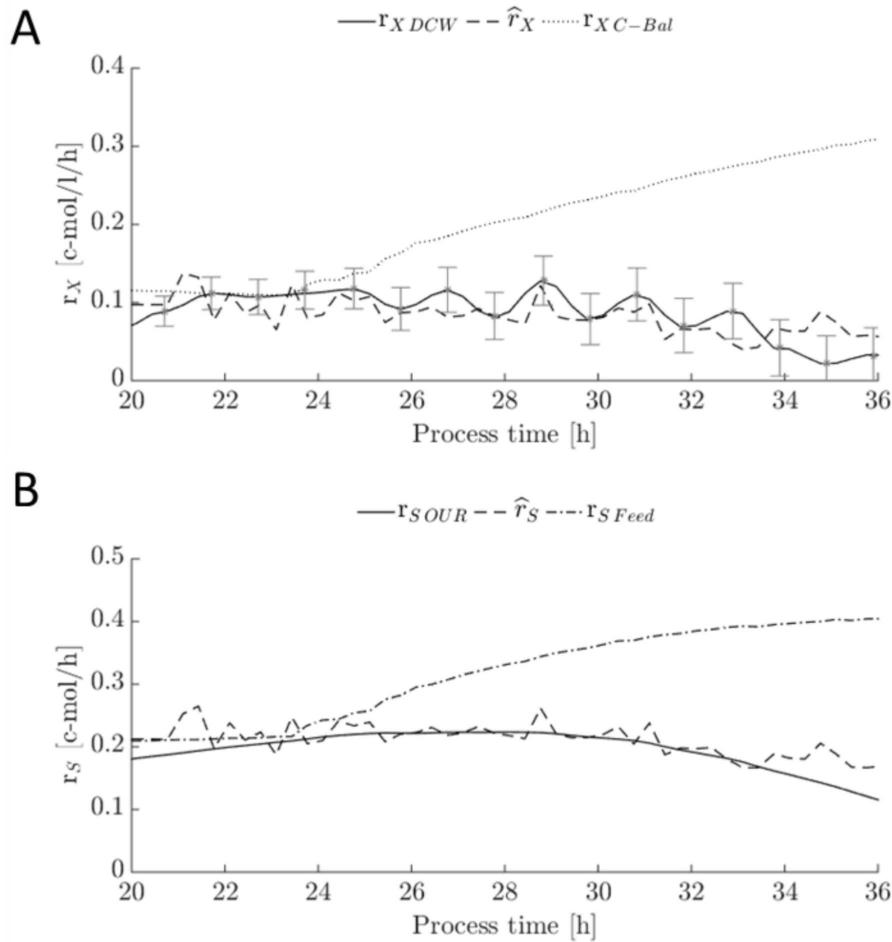
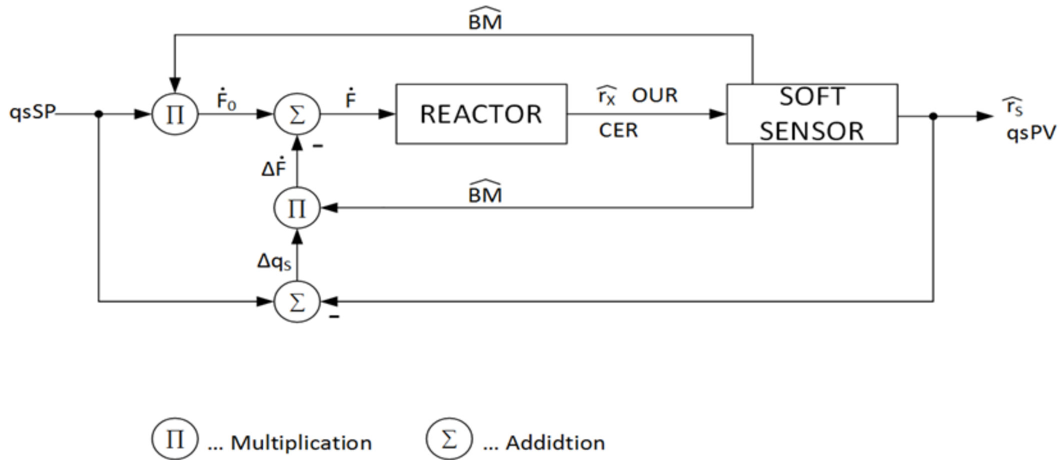


Figure 3 Biomass and substrate uptake rate estimation vs. verification;
(A) Biomass growth rate estimation (dashed line) based on N and DoR balance and permittivity is congruent to offline data (solid line). The BM accumulation rate rX is one of three needed values to estimate the actual substrate conversion rate and furthermore the specific substrate uptake rate q_S. The biomass growth rate based on C balance (dotted line) shows large deviations due to glucose accumulation. (B) Estimation of substrate conversion rate (dashed line) reconciled from rX, CER and OUR shows great congruence to offline data (solid line). The actual substrate conversion rate rS in combination with the biomass is needed to estimate the actual specific substrate uptake rate q_S. The substrate conversion rate based on C balance (dotted line) shows large deviations due to glucose accumulation.

Upon the proof of principle for the real time estimation of rx and rsox the calculation of a process value of q_S, as controlled variable, becomes possible. Figure 4 A explains the controller design to close the loop for the controlled variable q_S by incorporating the obtained values into the previously introduced controller scheme. In contrast now a process value for q_S can be calculated and used for controller action. To increase robustness only the feedback of the process value of q_S impacts the controller action. The basal substrate flow rate is being calculated in time increments based on the biomass estimation, as explained previously (Figure 1 B). Nevertheless, despite closing the loop the q_S setpoint is not maintained and glucose accumulation occurs quickly during induction phase.



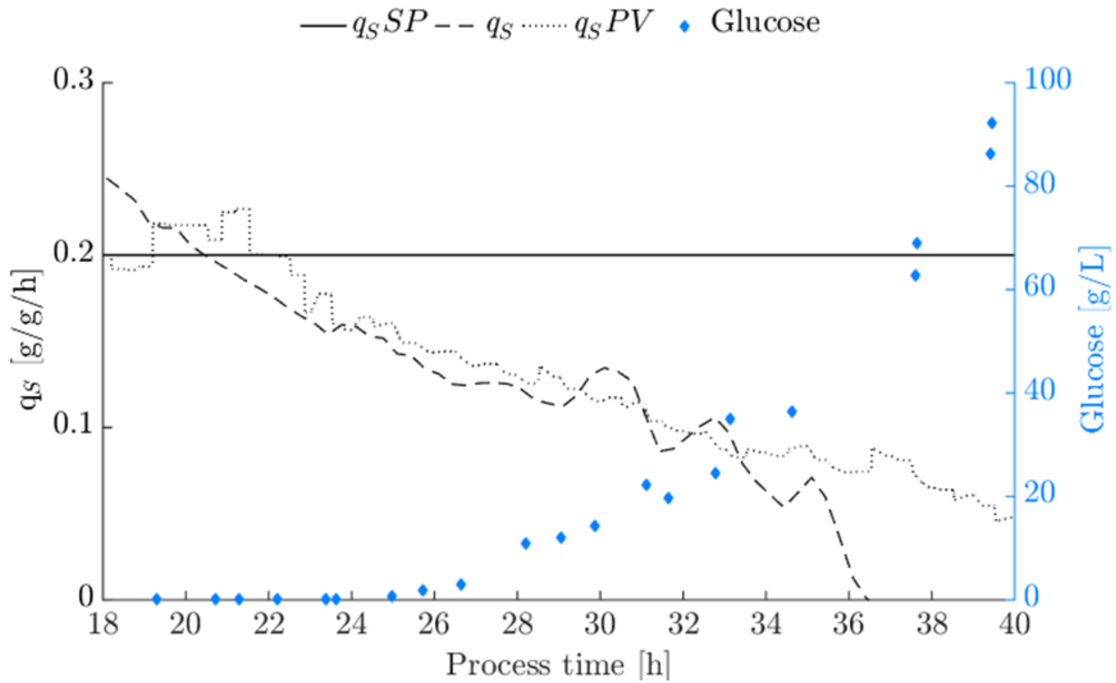


Figure 4 closed loop control approach and its limitations;

(A) Closed loop control provides the possibility of comparing the wanted setpoint with the actually achieved process value. Therefore an interference of the process can improved the quality of the control strategy. (B) Closed loop control of q_S is insufficient without adapting the setpoint, if q_{Scrit} is exceeded. Significant deviation from the q_S setpoint (solid line) due to q_S-max decline. Real time estimation of q_S process value (dotted line) compared to actually achieved q_S PV (dashed line) is within 25% error. High glucose accumulation (blue dots) due to q_{Smax} exceedance.

Owned to the physiological decelline of q_{Scrit} the setpoint of q_S cannot be maintained without substrate accumulation. The cells are simply not capable of metabolizing the amount of substrate defined by the setpoint. As a consequence the setpoint has to be adapted according to the decline in q_{Scrit}. Given the target of reducing the effort for strain characterization, using a predefined q_SSP limitation is not viable. Instead a simple algorithm is required to assess whether q_{Scrit} is reached or not.

Figure 5 illustrates the underlying principle of the q_{Scrit} controlled. By comparing the behavior of the culture upon controller action to previous behavior q_{Scrit} becomes obvious. If the process value of q_S does not increase after an increase in substrate flow the setpoint of q_S needs to be decreased since q_{Scrit} has been reached. Figure 5 A displays the in cooperation of the step controller displayed in Figure 5 B.

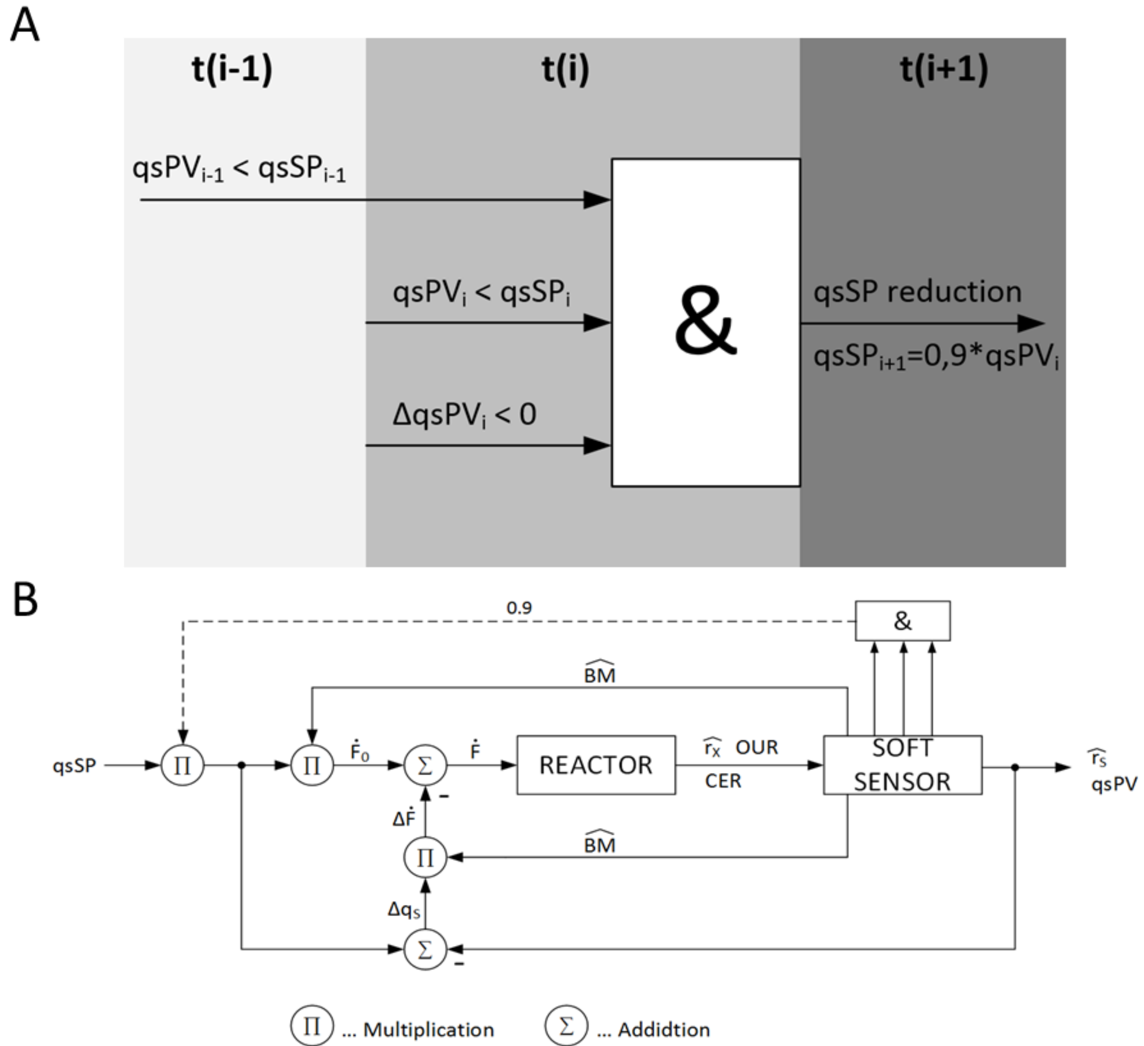


Figure 5 Closed loop control with setpoint adaptation;
(A) A logical query to adapt q_S setpoint (q_SSP), if q_{Scrit} is reached was implemented. Three conditions need to be true to adapt the q_SSP: i) The q_SSP from time point t(i-1) 20 minutes ago was larger than last q_SPV, ii) The q_SSP from time point t(i) is larger than the current q_SPV, iii) the change of the q_SPV within the last 20 minutes (Δq_SPV) was negative. In case of all three conditions, q_SSP is be reduced by 10%. Since in case of substrate limitation q_SPV should always increase upon a q_SSP (federate) increase, if q_SPV is smaller than q_SSP and q_{Scrit} is not reached, this logical query provides a simple method to detect q_{Scrit}. (B) If the system hits its natural limit, a setpoint higher than this limit cannot be achieved and therefore has to be adapted.

The benefit of the introduced control approach is displayed in Figure 6. Although the same q_S setpoint as in Figure 4 has been used, no accumulation occurs upon decline of q_{Scrit}. The controller effectively avoids substrate accumulation throughout the whole induction phase by reducing the q_S setpoint in case q_{Scrit} is being reached by the culture.

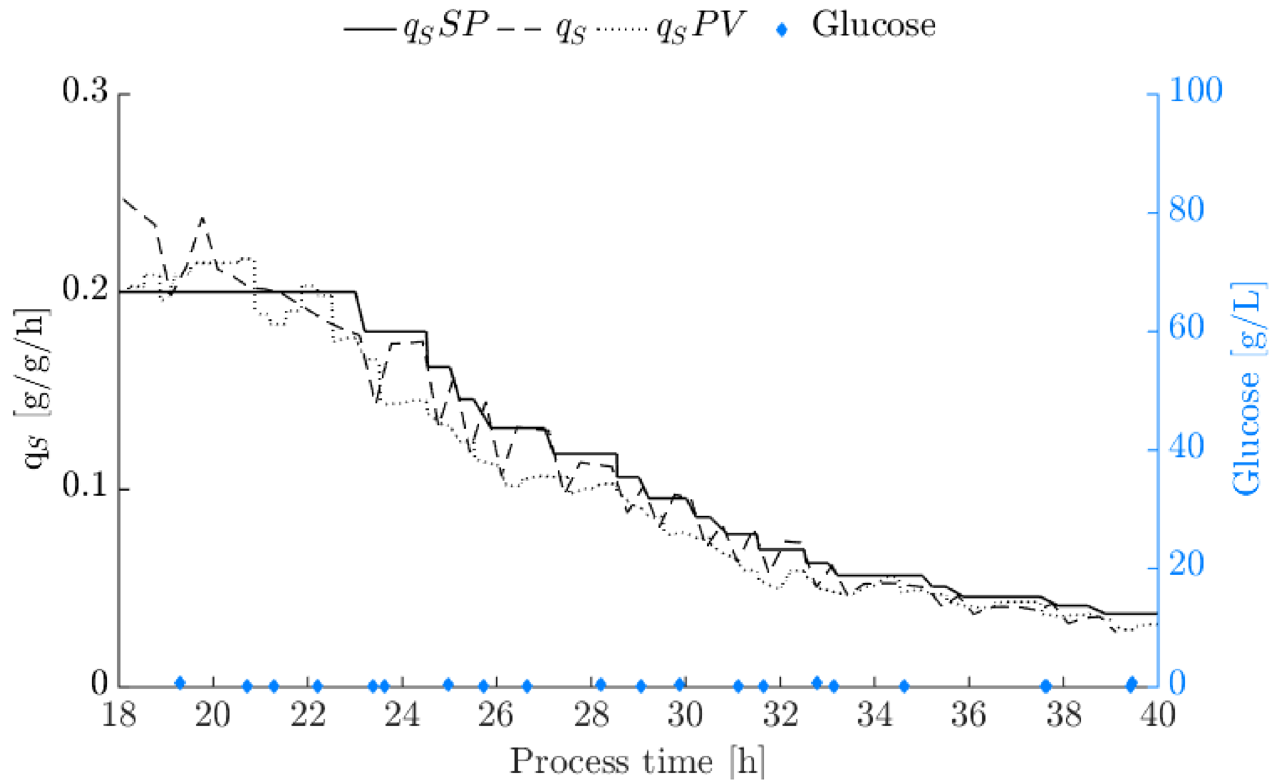


Figure 6 Closed loop control of q_S with adaption of $q_S SP$ (solid line) helps to avoid a breach of q_{Scrit} and facilitates an accurate control of q_S;
Real-time estimation of q_S process value (dotted line) compared to actually achieved q_S process value (dashed line). No significant glucose accumulation (blue dots) due to adaption of $q_S SP$.

Owned to the fact that this controller is merely based on first principle balances and a permittivity probe, the analytical as well as the computational effort is relatively lean. Consequently the introduced concept appears highly transferable even to an industrial environment.

Discussion

Within this contribution we introduced a process control approach capable of physiological feed back control and real time sensing of the physiological capacity to metabolize substrate (q_{Scrit}).

Closed loop control approaches are highly challenging and require a high accuracy of estimation. Only if a viable signal to noise ratio can be obtained a physiological closed loop approach becomes feasible. Nevertheless, various contributions have claimed closed loop control up to this date.

In the contribution of Sagmeister et al., using a first principle softsensor, closed loop control of q_S was claimed. A real time estimation of biomass was used to control the substrate feed rate in order maintain q_S throughout induction by (Patrick Sagmeister et al., 2013). Although real time process data was utilized for the estimation of biomass, no process value of q_S or other physiological variables was obtained. Since actually the feed rate is being controlled this approach potentially qualifies as technical closed loop but lacking the calculation of the process value of any physiological variable not as physiological closed loop control approach.

A very similar concept has been introduced by Jobe et al to sense the metabolic status of the cell in real time. Nevertheless, this approach did not fulfill the requirements of a closed loop control, since no process value of controlled physiological variable was calculated. Calculating an oxidative and an oxireductive metabolic model every 4 min a statistical test was used for the evaluation of the current metabolic state. The substrate feed rate was controlled by an exponential feeding profile, of which the exponent was subjected to controller actions based on the decision concerning the statistical test. Hereby, no process value of the controlled variable μ was calculated and consequently the prerequisites for a closed loop control approach not met. Besides the nomenclature, although the acetic acid was being accumulated the approach of Jobe et al. lacked the sensitivity to take action. The accumulation of acetate was not pronounced enough to trigger controller action; instead the μ controller remained idle. The growth in biomass subsequently decreased the specific growth rate to a level which allowed the uptake of acetate. In contrast the approach presented within this contribution did effectively circumvent the accumulation of acetate and substrate.

Dabros et al (Dabros, Schuler, & Marison, 2010) introduced an algorithm for physiological closed loop control and illustrated its feasibility in non-induced *E.coli* cultures. Using an exponential feed profile a basic substrate feed rate was calculated. The deviation of the process value of μ from μ setpoints triggered an additional PI controller action. Although accumulation was measured using FTIR, the data was only used for reconciliation to improve biomass estimation. If the decline in q_{Scrit} leads to an inevitable decrease in specific growth rate, despite substrate accumulation the controller action would lead to a continuous feed rate increase.

As physiological closed loop controller Jenzsch et al. (Jenzsch et al., 2006) used an extended Kalman filter for biomass estimation in combination with generic model control. Merely using a set of three experiments the model was trained and verified with an additional experiment prior the utilization for process control. This approach was shown to accurately achieve different distinct μsp in induction phase of a microbial bioprocess producing GFP. Despite the good performance the generic model control was based on a constant value for the q_{Scrit} as well as a static

value of the biomass yield. Owned to the underlying hypothesis of a constant biomass Yxs and q_{Scrit} this approach appears sensitive to reaching the physiological capacity q_{Scrit}.

Conclusions

The goal of this paper was the introduction of a transferable control concept capable of effectively avoiding substrate accumulation as well as the illustration of the feasibility of the introduced approach. In respect of the state of the art we were able to establish the following points:

- ➔ Using a combination of first principle mass balances we were able illustrate a transferable concept to independently estimate biomass concentration as well as the rate of oxidative substrate metabolization.
- ➔ Physiological feedback control makes a process value of the physiologic variable accessible but is not sufficient in order to avoid substrate accumulation. This circumstance is owned to the fact that physiologic capacity to metabolize substrate declines over time, making it impossible to maintain the setpoint.
- ➔ Using a simple step controller substrate accumulation can be effectively avoided by setpoint adaptation in response to the violation of physiologic capacities.

The illustrated approach facilitates robust process development without relying on comprehensive strain characterization. This concept is theoretically not limited to the specific substrate uptake rate. Owned to the underlying first principle mass balances this concept appears highly transferable in comparison to data driven alternatives. Furthermore, physiologic feedback control including the introduced controller could be used to complete replace conventional strain characterization in terms of physiological capacities. This will ultimately decrease the effort for process development significantly since not only the effort for strain characterization can be minimized but also number of fail batches owned to substrate accumulation can be greatly reduced.

Declarations

d. Acknowledgements

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

e. Authors' contributions

CH supervised the study. WR, AK and FJ conceived the study and the experimental design. AK performed the cultivations, and analytics. WR and AK processed and analyzed the data. WR wrote the first manuscript version. All authors participated in the final manuscript version, read and approved the final manuscript.

f. Conflicts of Interest

The authors declare no conflict of interest.

References

Åkesson, M., Hagander, P., & Axelsson, J. (1999). A probing feeding strategy for Escherichia coli cultures. *Biotechnology Techniques, 13*(8), 523-528. doi: 10.1023/A:1008906517033

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

Babaeipour, V., Shojaosadati, Seyed A., Khalilzadeh, R., Maghsoudi, N., & Tabandeh, F. (2008). A proposed feeding strategy for the overproduction of recombinant proteins in Escherichia coli. *Biotechnology and Applied Biochemistry, 49*(2), 141. doi: 10.1042/ba20070089

Chen, X., Liu, L., Li, J., Liu, J., Du, G., & Chen, J. (2011). Optimization of glucose feeding approaches for enhanced glucosamine and N-acetylglucosamine production by an engineered Escherichia coli. *Journal of Industrial Microbiology & Biotechnology, 39*(2), 359-365. doi: 10.1007/s10295-011-1046-0

Dabros, M., Schuler, M., & Marison, I. (2010). Simple control of specific growth rate in biotechnological fed-batch processes based on enhanced online measurements of biomass. *Bioprocess and Biosystems Engineering, 33*(9), 1109-1118. doi: 10.1007/s00449-010-0438-2

de Assis, A. J., & Filho, R. M. (2000). Soft sensors development for on-line bioreactor state estimation. *Computers & Chemical Engineering, 24*(2–7), 1099-1103. doi: http://dx.doi.org/10.1016/S0098-1354(00)00489-0

Gnoth, S., Jenzsch, M., Simutis, R., & Lubbert, A. (2008). Control of cultivation processes for recombinant protein production: a review. [Research Support, Non-U.S. Gov't Review]. *Bioprocess Biosyst Eng, 31*(1), 21-39. doi: 10.1007/s00449-007-0163-7

Henes, B., & Sonnleitner, B. (2007). Controlled fed-batch by tracking the maximal culture capacity. *Journal of Biotechnology, 132*(2), 118-126. doi: http://dx.doi.org/10.1016/j.jbiotec.2007.04.021

Jensen, E. B., & Carlsen, S. (1990). Production of recombinant human growth hormone in Escherichia coli: expression of different precursors and physiological effects of glucose, acetate, and salts. *Biotechnol Bioeng, 36*(1), 1-11. doi: 10.1002/bit.260360102

Jenzsch, M., Simutis, R., & Luebbert, A. (2006). Generic model control of the specific growth rate in recombinant Escherichia coli cultivations. *J Biotechnol, 122*(4), 483-493. doi: 10.1016/j.jbiotec.2005.09.013

Jobe, A. M., Herwig, C., Surzyn, M., Walker, B., Marison, I., & von Stockar, U. (2003). Generally applicable fed-batch culture concept based on the detection of metabolic state by on-line balancing. *Biotechnol Bioeng, 82*(6), 627-639. doi: 10.1002/bit.10610

Kavanagh, J. M., & Barton, G. W. (2008). Productivity improvement of recombinant Escherichia coli fermentation via robust optimization. *Bioprocess Biosyst Eng, 31*(2), 137-143. doi: 10.1007/s00449-007-0156-6

Korz, D. J., Rinas, U., Hellmuth, K., Sanders, E. A., & Deckwer, W. D. (1995). Simple fed-batch technique for high cell density cultivation of Escherichia coli. *Journal of Biotechnology, 39*(1), 59-65. doi: http://dx.doi.org/10.1016/0168-1656(94)00143-Z

Levisauskas, D. (2001). Inferential control of the specific growth rate in fed-batch cultivation processes. *Biotechnology Letters, 23*(15), 1189-1195. doi: 10.1023/A:1010528915228

Levisauskas, D., Galvanauskas, V., Henrich, S., Wilhelm, K., Volk, N., & Lübbert, A. (2003). Model-based optimization of viral capsid protein production in fed-batch culture of recombinant Escherichia coli. *Bioprocess and Biosystems Engineering, 25*(4), 255-262. doi: 10.1007/s00449-002-0305-x

Levisauskas, D., Simutis, R., Borvitz, D., & Lübbert, A. (1996). Automatic control of the specific growth rate in fed-batch cultivation processes based on an exhaust gas analysis. *Bioprocess Engineering, 15*(3), 145-150. doi: 10.1007/BF00369618

Lin, H. Y., Mathiszik, B., Xu, B., Enfors, S. O., & Neubauer, P. (2001). Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of Escherichia coli. *Biotechnol Bioeng, 73*(5), 347-357.

Luli, G. W., & Strohl, W. R. (1990). Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations. *Appl Environ Microbiol, 56*(4), 1004-1011.

Ramalingam, S., Gautam, P., Mukherjee, K. J., & Jayaraman, G. (2007). Effects of post-induction feed strategies on secretory production of recombinant streptokinase in Escherichia coli. *Biochemical Engineering Journal, 33*(1), 34-41. doi: http://dx.doi.org/10.1016/j.bej.2006.09.019

Rathore, A. S., & Winkle, H. (2009). Quality by design for biopharmaceuticals. [10.1038/nbt0109-26]. *Nat Biotech, 27*(1), 26-34.

Reichelt, W. N., Brillmann, M., Thurrold, P., Keil, P., Fricke, J., & Herwig, C. (2016). Physiologic root cause analysis of substrate accumulation in induced bioprocesses (under submission). *Microbial cell factories*.

Reichelt, W. N., Thurrold, P., Brillmann, M., Kager, J., Fricke, J., & Herwig, C. (2016). Generic biomass estimation methods targeting physiologic process control in induced bacterial cultures (under submission). *Engineering in Life Sciences*.

Rocha, I., Veloso, A. C. A., Carneiro, S., Costa, R., & Ferreira, E. C. (2008). Implementation of a Specific Rate Controller in a Fed-Batch E. Coli Fermentation. *IFAC Proceedings Volumes, 41*(2), 15565-15570. doi: 10.3182/20080706-5-kr-1001.02632

Sagmeister, P., Schimek, C., Meitz, A., Herwig, C., & Spadiut, O. (2014). Tunable recombinant protein expression with E. coli in a mixed-feed environment. *Appl Microbiol Biotechnol, 98*(7), 2937-2945. doi: 10.1007/s00253-013-5445-1

Sagmeister, P., Wechselberger, P., & Herwig, C. (2012). Information Processing: Rate-Based Investigation of Cell Physiological Changes along Design Space Development. *PDA J Pharm Sci Technol, 66*(6), 526-541. doi: 10.5731/pdajpst.2012.00889

Sagmeister, P., Wechselberger, P., Jazini, M., Meitz, A., Langemann, T., & Herwig, C. (2013). Soft sensor assisted dynamic bioprocess control: Efficient tools for bioprocess development. *Chemical Engineering Science, 96*(0), 190-198. doi: http://dx.doi.org/10.1016/j.ces.2013.02.069

Sanden, A. M., Prytz, I., Tubulekas, I., Forberg, C., Le, H., Hektor, A., . . . Larsson, G. (2003). Limiting factors in Escherichia coli fed-batch production of recombinant proteins. *Biotechnol Bioeng, 81*(2), 158-166. doi: 10.1002/bit.10457

Schaepe, S., Kuprijanov, A., Simutis, R., & Lübbert, A. (2014). Avoiding overfeeding in high cell density fed-batch cultures of E. coli during the production of heterologous proteins. *Journal of Biotechnology, 192, Part A*, 146-153. doi: http://dx.doi.org/10.1016/j.jbiotec.2014.09.002

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology, 72*(2), 211-222. doi: 10.1007/s00253-006-0465-8

Velut, S., de Maré, L., & Hagander, P. (2007). Bioreactor control using a probing feeding strategy and mid-ranging control. *Control Engineering Practice, 15*(2), 135-147. doi: http://dx.doi.org/10.1016/j.conengprac.2006.05.009

Walsh, G. (2010). Biopharmaceutical benchmarks 2010. [10.1038/nbt0910-917]. *Nat Biotech, 28*(9), 917-924.

Wechselberger, P., Sagmeister, P., Engelking, H., Schmidt, T., Wenger, J., & Herwig, C. (2012). Efficient feeding profile optimization for recombinant protein production using physiological information. *Bioprocess Biosyst Eng, 35*(9), 1637-1649. doi: 10.1007/s00449-012-0754-9

Discussion

The goal of this section was the analysis and discussion of the efficiency of physiological bioprocess control approaches in the context of bioprocess development. For physiological process development, physiology must be made available as single numerical value which can be utilized as response for DoE evaluation. Such a workflow of data processing requires a clear and strict but transferable definition of process phases of high physiological relevance. We introduced variable of the cumulative normalized substrate (dSn) as substitute for time with a higher physiological relevance for process phase definition. Based on dSn various physiological descriptors (e.g. q_{Smean}) can be derived within information mining to increase comparability of experiments and consequently ease physiological conclusions (2a). Additionally, by repeatedly testing the involvement of qs within the same strain but with different products provided great evidence concerning the transferability of the impact of qs (2b).
The introduced workflow consequently facilitates the direct comparison of physiological and technological feeding strategies. In this context it was of great importance to assess whether the employment of physiological feeding strategies is justifiable in terms of productivity. Therefore, the impact of substrate supply on productivity was studied and compared by conducting technical and physiological feeding profiles. Hereby, a substantial increase in titer was found for all levels of substrate supply investigated employing the physiological feeding strategy (2c). This finding highlights the potential of physiological process control and substantiates the physiological process development approach.
After substantiating the benefit of physiologic feeding strategies, physiological bioprocess development was demonstrated from scratch. Physiological process development and correlated physiological process control require a clear definition of the physiological design space in terms of q_{Scrit}. In contrast to literature q_{Scrit} was found to be dependent on time as well as on physiologic activity (q_{Smean}). This finding was only possible owned to the high degree of information mining. Without the great amount of accurate analytical data this finding would not have been possible. Although, data evaluation consumed a substantial amount of time, the relevance of the finding is high and presumably transferable to other products and even strains. This observed substantial decline in q_{Scrit} in dependence of time and q_{Smean} constitutes a substantial challenge for process control since substrate accumulation needs to be avoided (2d).
In the context of a dynamically declining q_{Scrit}, using a combination of first principle soft sensors and a hard type sensor we established a physiological feedback control loop. Based on a simple algorithm the q_s set point was adapted in response to a violation of q_{Scrit}. Using this approach to avoid substrate accumulation strain characterization experiments become obsolete since the introduced control approach is merely based on the biomass composition as strain specific process parameter (2e).

Overall scientific Novelty and Author contributions

Table 2: Review contribution summarizing the state of the art of the scientific disciplines in focus

Review contributions	Topics discussed and reviewed	Author contribution
Ex-situ on-line monitoring and data processing [56]	State of the art of ex-situ on-line monitoring for process design-, analysis- and control purposes	Review of the state of the art Online-HPLC approaches; text revision

Table 3: Scientific contributions addressing (process) analytical approaches.

Analytical Method	Main Findings	Author contribution (WR)
Protein quantification [57]	The error of protein quantification by BCA in complex sample matrices exceeds 200% but can be reduced significantly using internal standard addition.	WR conceptualized the study and conducted it in cooperation with the co-authors. WR supervised Data analysis and drafted the manuscript.
Inclusion body analytics [58]	TEM image segmentation is established as quantitative method for IB sizing in the cytosol. The results are used for verification of NTA sizing results. By comparing the size of fixated to non-fixated IBs the impact of high pressure homogenization is demonstrated	WR conducted parts of the analytics, conceptualized the paper and drafted the manuscript.
Real time Biomass sensing [59]	Methods for real time biomass sensing are compared for different levels of metabolic activity. While the single methods do not differ in cvRMSE the combination of methods can reduce estimation error by over 85%.	WR conducted parts of the experimental work and the data analysis. He established the underlying database and drafted the manuscript.
CedexBioHT for high throughput analysis	The use of CedexHT for hight throughput enzymatic metabolite quantification is demonstrated and assessed for dynamic *E.coli* processes.	WR conducted the experiments and reviewed the manuscript.

Table 4: Scientific contributions addressing process controls

Process control Method	Main Findings	Author contribution (WR)
Improved fermentation process (Patent)	The use of low feeding rates, surprisingly led to an significant increase in specific product titer.	WR largely drafted and iterated the patent draft with industrial partners.
Pre and Post Induction Phase [60]	We were able to substantiate the cumulative consumed substrate dSn as highly relevant scale for the investigation of physiological effects. And substantiated the benefitial effect of low qs for productivity.	WR conducted the experimental work in cooperation. WR conducted the data analysis and drafted the manuscript.
Physiological Feeding [61]	By controlling the physiological substrate uptake rate on a constant level, productivities were significantly increased. Independent of the qs level physiological control positively impacted productivity.	WR conducted parts of the experimental work and the data analysis. Additionally he drafted the manuscript.
qscrit – physiological capacities [62]	Addressing physiology in bioprocess development the physiological capacities have to be known. Here we show that the qscrit is not only a timely variable but moreover dependent on the level of metabolic activity.	WR planned and organized the experiments and drafted the manuscript.
Closed loop qs control [63]	The variability of the physiological capacity to metabolize substrate highlights the necessity to sense this capacity in real time. We illustrate the feasibility of this approach by avoiding substrate accumulation throughout induction phase by closed loop physiological control.	WR planned and supervised the experiments and helped setting up the control algorithm. Moreover, WR drafted the manuscript.

Discussion and Conclusions

Given the current situation in biopharmaceutical industry, bioprocess development is facing seemingly contradicting demands. While regulatory authorities are calling for increased process understanding, industry is mainly interested in productivity and cost efficiency. To meet both demands bioprocess development should put physiology into the center of attention. The underlying hypothesis of this statement and this contribution was that physiological bioprocess development features a better information to effort ratio. For industrial application of physiological bioprocess development two gaps and correlated challenges have been identified within the introduction and discussed in detail in the previous sections.

Analytics for physiological process development

In technological process development, process analytical methods e.g. offline biomass measurement or total protein quantification merely serves a minor role and is hardly used as response. In contrast, physiological process variables are based on a greater deal of analytical data (e.g. q_S). In the context of error propagation it appears logical that the relevance of analytical data is of greater concern for physiological bioprocess development. Although analytical methods are commonly widely available and being employed for decades, it is important to establish and to assess method performance under condition as close as possible to the application reality. Although this conclusion appears obvious, the relevance of orthogonal methods for method verification if often underestimated. Only using orthogonal methods we have been able to quantify the correlated error of measurement, to elucidate the impact of high pressure homogenization on IB size and illustrate the substantial benefit of redundant biomass estimation.

Process Control

But the bottleneck is not exclusively of analytical nature, as general shortcoming data evaluation can be identified. Although especially established contract manufacturers possess process data in abundance, they often lack generalized process evaluation routines. Lacking general applicable routines or definitions makes data comparison across laboratories and even across operators risky and prone to error. In this context, a general applicable software for process data management and process data analytics solution would greatly foster comparability as well as interpretability. Although various solutions have been introduced into the market, they often fail to offer an intuitive operator interface and an easy to use data management solution. This heterogenic approach of data evaluation is not only an industrial phenomenon but also an academic, since data evaluation routines are hardly published or discussed in combination with experimental data. Although most of the know-how can be considered textbook knowledge, the abundancy of possibilities to define systems and reactions constitutes the challenge. Only by uniform system and phase definitions data evaluations become comparable across operators and laboratories.

Physiological bioprocess development clearly fosters transferability of findings (interrelation of qs) and has been shown in this contribution to boost productivity. Only by using physiological process control the effect of level and trajectory of substrate supply could be distinguished, This

finding introduced a novel degree of freedom for bioprocess development. Although physiological bioprocess development is associated with an increase in effort for data analysis as well as fundamental analytics, the information content is higher, which in turn makes conclusions better transferable. As a consequence, the information to effort ratio is improved by the fact that the discussed physiological control approaches operate widely product and even strain independent and can consequently be readily employed .

Throughout induction phase, owned to the growth in biomass, volumetric constant feeding profiles hardly ever lead to overfeeding by time. In contrast, the dynamic feed rate adaption for physiological feeding profiles bear a greater risk of substrate accumulation by exceeding a timely declining q_{Scrit}. Despite this risk physiological feeding strategies offer accessibility to a greater area of physiological activity in terms of q_{Smean}. Although, this physiological activity is limited by the decline in q_{Scrit} over time, the state of continuously high metabolic activity would not be accessible with a volumetric constant feeding rate. Hereby, physiologic process control is challenging in terms of robustness especially in respect of the decline in q_{Scrit}. On the other hand the elucidation of the decline in the q_{Scrit} is a dedicated benefit of physiological bioprocess development. This interrelation of metabolic activity and metabolic capacity is a critical piece of information regarding DoE, but on the other hand makes DoE much more complex. Various contributions in literature have investigated similar effects based on volumetric substrate addition, which in turn did not facilitate this far-reaching revelation. Despite the increase in complexity for the DoE the introduction of an appropriate closed loop control approach was demonstrated to be a viable approach to effectively circumvent substrate accumulation. In this context, physiological closed loop systems grant a real time reaction upon changes in physiology, making physiological process control more robust. This approach could eventually help to overcome scale-up issues such as e.g. inhomogeneity which lead to an altered biomass yield and consequently to a different physiological status of the culture. Since the employed control approaches are merely based on first principle soft sensors and a hard type sensor, the physiological bioprocess development approach becomes independent of fixed strain specific variables besides the elemental biomass composition. Thereby the necessary effort for bioprocess development is greatly reduced since no additional strain characterization experiments are necessary. This reduces the decision concerning the feeding profile optimization merely to a decision of favorable high or low metabolic activity.

To put it in a nutshell, physiologic process development grants deeper insights into relevant physiological processes and fosters the general understanding of the behavior of the production strain regardless of an associated titer increase. It grants additional degrees of freedom and thereby theoretically allows for higher time space yields. Physiological bioprocess control approaches, as the backbone of physiological bioprocess development are even able to substitute strain characterization experiments. Physiologic bioprocess development grants a deeper insight into physiology, which resembles a significant gain of information but requires the setup of more complex physiological control approaches. Nevertheless, this effort is conserved since the illustrated approaches are widely product and strain independent. Especially the discussed methods for bioprocess control can be employed for every strain fulfilling the underlying metabolic hypothesis. Within this contribution the benefit of physiological bioprocess development has been demonstrated merely for *E.coli*. Nevertheless, the underlying principles (e.g. mass balances) are independent of the organism of interest. But for

more complex organisms e.g. fungi, mammalians the control strategies would require a higher degree of complexity.

It can be concluded that physiological bioprocess development asks for a one time effort investment for the establishment of sensitive analytics and physiological control approaches but rewards the effort with a significant and continuous increase information to effort ratio. Although further proof is necessary this comprehensive analysis of physiological bioprocess development already provides clear evidence to initiate a change in paradigm: putting the actual producers into the center of attention – the cells.

References which formed this cumulative thesis

Contributions as first Author

- "Bioprocess monitoring: Minimizing sample matrix effects for total protein quantification with bicinchoninic acid assay" [57]

  Journal of Industrial Microbiology & Biotechnology (accepted 2016)

- "High throughput inclusion body sizing: Nano particle tracking analysis" [58]

  Biotechnology Journal (submitted 2016)

- "Generic biomass estimation methods targeting physiological process control in induced bacterial cultures" [59]

  Engineering in Life Sciences (accepted 2016)

- "Bioprocess development workflow: transferable physiological knowledge instead of technological correlations" [60]

  Biotechnology Progress (submitted and under revision 2016)

- "The benefit of physiological bioprocess control -discriminating the effect of level and trajectory of the specific substrate uptake rate" [61]

  Microbial Cell Factories (submitted 2016)

- "Substrate accumulation in induced bioprocesses caused by declining physiological capacities" [62]

  Process Biochemistry (submitted 2016)

- "Avoiding substrate accumulation by real time sensing of physiological capacity to metabolize substrate in induction phase of pharmaceutical bioprocesses" [63]

  Biotechnology and Bioengineering (Manuscript 2016)

- "Improved fermentation process"

  EP16158294.5 (2016) see Appendix

- "The Ca2+/Mn2+ ion-pump PMR1 links elevation of cytosolic Ca(2+) levels to alpha-synuclein toxicity in Parkinson's disease models" [64]

  Cell Death Differentiation (2013)

Contributions as Co-Author

- "Ex situ online monitoring: application, challenges and opportunities for biopharmaceuticals processes" [56]

  Pharmaceutical Bioprocessing (2014) see Appendix

- "Cedex Bio HT Analyzer as a tool for high-throughput analysis for fast physiological characterization of microbial bioprocesses"

  Application Note Roche Diagnostics (2014) see Appendix

- "Nucleocytosolic depletion of the energy metabolite acetyl-coenzyme a stimulates autophagy and prolongs lifespan" [65]

  Cell Metabolism (2014) see Appendix

- "Comprehensive method assessment of Inclusion bodies analytics"

  Manuscript

References

1. Terpe, K., *Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems.* Applied Microbiology and Biotechnology, 2006. **72**(2): p. 211-222.
2. Walsh, G., *Biopharmaceutical benchmarks 2010.* Nat Biotech, 2010. **28**(9): p. 917-924.
3. Levisauskas, D., et al., *Automatic control of the specific growth rate in fed-batch cultivation processes based on an exhaust gas analysis.* Bioprocess Engineering, 1996. **15**(3): p. 145-150.
4. Levisauskas, D., *Inferential control of the specific growth rate in fed-batch cultivation processes.* Biotechnology Letters, 2001. **23**(15): p. 1189-1195.
5. Henes, B. and B. Sonnleitner, *Controlled fed-batch by tracking the maximal culture capacity.* Journal of Biotechnology, 2007. **132**(2): p. 118-126.
6. Gnoth, S., et al., *Product formation kinetics in genetically modified E. coli bacteria: inclusion body formation.* Bioprocess Biosyst Eng, 2008. **31**(1): p. 41-6.
7. Johnson, B.H. and M.H. Hecht, *Cells by Repeated Cycles of Freezing and Thawing.* Bio/Technology, 1994. **12**: p. 1357.
8. Walsh, G., *Biopharmaceuticals and biotechnology medicines: an issue of nomenclature.* Eur J Pharm Sci, 2002. **15**(2): p. 135-8.
9. Lundkvist, J., et al., *The battle of Alzheimer's Disease - the beginning of the future Unleashing the potential of academic discoveries.* Front Pharmacol, 2014. **5**: p. 102.
10. Simpson, A. and O. Caballero, *Monoclonal antibodies for the therapy of cancer.* BMC Proceedings, 2014. **8**(Suppl 4): p. O6.
11. Reid, B., *US biotech prepares to fight generic biologics.* Nat Biotech, 2002. **20**(4): p. 322-322.
12. Dabros, M., M. Schuler, and I. Marison, *Simple control of specific growth rate in biotechnological fed-batch processes based on enhanced online measurements of biomass.* Bioprocess and Biosystems Engineering, 2010. **33**(9): p. 1109-1118.
13. ICH, F., *Guidance for Industry Q8 PharmaceuticalDevelopment (R2)* 2009.
14. Group, C.B., *A-Mab: a Case Study in Bioprocess Development.* 2009.
15. Lee, S.Y., *High cell-density culture of Escherichia coli.* Trends in Biotechnology, 1996. **14**(3): p. 98-105.
16. Ukkonen, K., et al., *Effect of culture medium, host strain and oxygen transfer on recombinant Fab antibody fragment yield and leakage to medium in shaken E. coli cultures.* Microb Cell Fact, 2013. **12**: p. 73.
17. ICH, F., *Guidance for Industry Q9 Quality Risk Management.* 2006.
18. Rathore, A.S., et al., *Case study and application of process analytical technology (PAT) towards bioprocessing: II. Use of ultra-performance liquid chromatography (UPLC) for making real-time pooling decisions for process chromatography.* Biotechnol Bioeng, 2008. **101**(6): p. 1366-74.
19. Bolker, B.M., *Ecological Models and Data in R.* Princton UNiversity Press 2008(508 edition): p. 396.
20. Glassey, J., et al., *Process analytical technology (PAT) for biopharmaceuticals.* Biotechnol J, 2011. **6**(4): p. 369-77.
21. Bostrom, M. and G. Larsson, *Introduction of the carbohydrate-activated promoter P(malK) for recombinant protein production.* Appl Microbiol Biotechnol, 2002. **59**(2-3): p. 231-8.
22. Wechselberger, P., et al., *Efficient feeding profile optimization for recombinant protein production using physiological information.* Bioprocess Biosyst Eng, 2012. **35**(9): p. 1637-49.
23. Greb, E. and A. Drakulich, *Platform Technologies: The standardization of upstream and downstream bioprocessing is growing, but several kinks need to be ironed out.* Pharmaceutical Technology, 2012. **36**(3): p. 44-50.
24. Sagmeister, P., et al., *Bacterial Suspension Cultures.* 2014: p. 40-93.

25. Chen, H.Y. and E. White, *Role of autophagy in cancer prevention.* Cancer Prev Res (Phila), 2011. **4**(7): p. 973-83.
26. Ramalingam, S., et al., *Effects of post-induction feed strategies on secretory production of recombinant streptokinase in Escherichia coli.* Biochemical Engineering Journal, 2007. **33**(1): p. 34-41.
27. Sagmeister, P., et al., *Tunable recombinant protein expression with E. coli in a mixed-feed environment.* Appl Microbiol Biotechnol, 2014. **98**(7): p. 2937-45.
28. Kroemer, G., G. Marino, and B. Levine, *Autophagy and the integrated stress response.* Mol Cell, 2010. **40**(2): p. 280-93.
29. Gelino, S. and M. Hansen, *Autophagy - An Emerging Anti-Aging Mechanism.* J Clin Exp Pathol, 2012. **Suppl 4**.
30. Kavanagh, J.M. and G.W. Barton, *Productivity improvement of recombinant Escherichia coli fermentation via robust optimization.* Bioprocess Biosyst Eng, 2008. **31**(2): p. 137-43.
31. Sanden, A.M., et al., *Limiting factors in Escherichia coli fed-batch production of recombinant proteins.* Biotechnol Bioeng, 2003. **81**(2): p. 158-66.
32. Jenzsch, M., R. Simutis, and A. Luebbert, *Generic model control of the specific growth rate in recombinant Escherichia coli cultivations.* Journal of Biotechnology, 2006. **122**(4): p. 483-493.
33. Gnoth, S., et al., *Product formation kinetics in genetically modified E. coli bacteria: inclusion body formation.* Bioprocess and Biosystems Engineering, 2008. **31**(1): p. 41-46.
34. Rocha, I., et al., *Implementation of a specific rate controller in a fed-batch E. coli fermentation.* 2008.
35. Sagmeister, P., et al., *Soft-sensor assisted dynamic investigation of mixed feed bioprocesses.* Process Biochemistry, 2013. **48**(12): p. 1839-1847.
36. Dietzsch, C., O. Spadiut, and C. Herwig, *A dynamic method based on the specific substrate uptake rate to set up a feeding strategy for Pichia pastoris.* Microb Cell Fact, 2011. **10**: p. 14.
37. Spadiut, O., et al., *Quantitative comparison of dynamic physiological feeding profiles for recombinant protein production with Pichia pastoris.* Bioprocess and Biosystems Engineering, 2014. **37**(6): p. 1163-72.
38. Schlegel, S., P. Genevaux, and J.-W. de Gier, *De-convoluting the Genetic Adaptations of E. coli C41(DE3) in Real Time Reveals How Alleviating Protein Production Stress Improves Yields.* Cell Reports, 2015. **10**(10): p. 1758-1766.
39. Taylor, G., et al., *Size and Density of Protein Inclusion Bodies.* Nat Biotech, 1986. **4**(6): p. 553-557.
40. Wong, H.H., B.K. O'Neill, and A.P.J. Middelberg, *Cumulative sedimentation analysis of Escherichia coli debris size.* Biotechnology and Bioengineering, 1997. **55**(3): p. 556-564.
41. Datta, I., S. Gautam, and M.N. Gupta, *Microwave assisted solubilization of inclusion bodies.* Sustainable Chemical Processes, 2013. **1**(1): p. 1-7.
42. Castellanos-Mendoza, A., et al., *Influence of pH control in the formation of inclusion bodies during production of recombinant sphingomyelinase-D in Escherichia coli.* Microbial Cell Factories, 2014. **13**(1): p. 1-14.
43. Margreiter, G., et al., *Size characterization of inclusion bodies by sedimentation field-flow fractionation.* Journal of Biotechnology, 2008. **138**(3–4): p. 67-73.
44. Luo, J., et al., *Size characterization of green fluorescent protein inclusion bodies in E. coli using asymmetrical flow field-flow fractionation-multi-angle light scattering.* J Chromatogr A, 2006. **1120**(1-2): p. 158-64.
45. Moon, M.H. and S. Lee, *Sedimentation/steric field-flow fractionation: a powerful technique for obtaining particle size distribution.*
46. Jenzsch, M., et al., *Estimation of biomass concentrations in fermentation processes for recombinant protein production.* Bioprocess Biosyst Eng, 2006. **29**(1): p. 19-27.
47. de Assis, A.J. and R.M. Filho, *Soft sensors development for on-line bioreactor state estimation.* Computers & Chemical Engineering, 2000. **24**(2–7): p. 1099-1103.

48. Cheng, L.-C., et al., *Effect of specific growth rate on the production of a recombinant nuclease by Escherichia coli.* Biochemical Engineering Journal, 2003. **14**(2): p. 101-107.
49. Yee, L. and H.W. Blanch, *Recombinant protein expression in high cell density fed-batch cultures of Escherichia coli.* Biotechnology (N Y), 1992. **10**(12): p. 1550-6.
50. Wilms, B., et al., *High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the Escherichia coli rhaBAD promoter.* Biotechnology and Bioengineering, 2001. **73**(2): p. 95-103.
51. Lionberger, R.A., et al., *Quality by design: concepts for ANDAs.* AAPS J, 2008. **10**(2): p. 268-76.
52. Liu, S., *Bioprocess Engineering: Kinetics, Biosystems, Sustainability and Reactor Design.* 2013: p. 666-667.
53. Gnoth, S., et al., *Control of cultivation processes for recombinant protein production: a review.* Bioprocess Biosyst Eng, 2008. **31**(1): p. 21-39.
54. Luli, G.W. and W.R. Strohl, *Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations.* Appl Environ Microbiol, 1990. **56**(4): p. 1004-11.
55. Jensen, E.B. and S. Carlsen, *Production of recombinant human growth hormone in Escherichia coli: expression of different precursors and physiological effects of glucose, acetate, and salts.* Biotechnol Bioeng, 1990. **36**(1): p. 1-11.
56. Kroll, P., et al., *Ex situonline monitoring: application, challenges and opportunities for biopharmaceuticals processes.* Pharmaceutical Bioprocessing, 2014. **2**(3): p. 285-300.
57. Reichelt, W.N., et al., *Bioprocess monitoring: Minimizing sample matrix effects for total protein quantification with bicinchoninic acid assay (accepted).* Journal of Industrial Microbiology and Biotechnology (accepted), 2016.
58. Reichelt, W.N., et al., *Inclusion body size: transmission electron microscopy and nano particle tracking analysis (manuscript).* Biotechnology Journal, 2016.
59. Reichelt, W.N., et al., *Generic biomass estimation methods targeting physiologic process control in induced bacterial cultures (accepted).* Engineering in Life Sciences, 2016.
60. Reichelt, W.N., et al., *Bioprocess development workflow: transferable physiological knowledge instead of technological correlations (submitted).* Biotechnological Progress, 2016.
61. Reichelt, W.N., et al., *The benefit of physiological bioprocess control – discriminating the effect of level and trajectory of the specific substrate uptake rate (submitted).* Applied Microbiology and Biotechnology, 2016.
62. Reichelt, W.N., et al., *Physiologic root cause analysis of substrate accumulation in induced bioprocesses (submitted).* Microbial Cell Factories, 2016.
63. Reichelt, W.N., et al., *Avoiding substrate accumulation by real time sensing of physiological capacity to metabolize substrate in induction phase of pharmaceutical bioprocesses (manuscript).* 2016.
64. Buttner, S., et al., *The Ca2+/Mn2+ ion-pump PMR1 links elevation of cytosolic Ca(2+) levels to alpha-synuclein toxicity in Parkinson's disease models.* Cell Death Differ, 2013. **20**(3): p. 465-77.
65. Eisenberg, T., et al., *Nucleocytosolic depletion of the energy metabolite acetyl-coenzyme a stimulates autophagy and prolongs lifespan.* Cell Metab, 2014. **19**(3): p. 431-44.

Curriculum Vitae

Personal Information

Last name/ first name(s)	**Reichelt Wieland Norbert**
Address	Thaliastrasse, A-1160, Wien,
Telephone	0043/6508403015
E-mail	wieland.reichelt@tuwien.ac.at
Nationality	German
Date of birth	10.03.1987

Work experience

Since	March 2016
Position	CEO and Co-founder
Primary duties/ activities	Management and Development
Name and address of employer	***Evologic Technologies GmbH,*** *Gumpendorferstr. 1A, 1060 Vienna*
Field of business activity	Bioprocess Development for Production of Mycorrhizal Fungi
From - until	January 2013-May 2016
Position	PhD Position
Primary duties/ activities	Physiological Bioprocess Control for Optimized Bioprocesses
Name and address of employer	***BioVT,*** *Gumpendorferstr. 1A, 1060 Vienna*
Field of Research	Biochemical Engineering
From - until	Oktober 2012-Dezember 2012
Position	Scientific Assistant
Primary duties/ activities	Development of Methodical Approaches for Bioprocess Optimization
Name and address of employer	***BioVT,*** *Gumpendorferstr. 1A, 1060 Vienna*
Field of Research	Biochemical Engineering
From - until	September 2011 – Oktober 2012
Position	PhD Student
Primary duties/ activities	Establishing a Metabolomics Platform, Metabolite Extraction, Data Mining
Name and address of employer	***IMB***, Humboldstraße 50, 8010 Graz
Field of Research	Programmed Cell Death and Aging
From - until	March 2010 –May 2011
Position	Diploma student
Primary duties/ activities	Alpha Synuclein mediated toxicity in yeast with special focus on cytosolic Calcium levels
Name and address of employer	***IMB***, Humboldstraße 50, 8010 Graz
Field of Research	Programmed Cell Death and Aging

From - until	August - October 2009
Position	Internship
Primary duties/ activities	Expression, purification and functional-structural analysis of a recombinant protein (FPLC)
Name and address of employer	***Baxter AG***, Uferstr. 15, 2304 Orth
Field of business activity	Pharmaceuticals
From - until	August - October 2008
Position	Internship
Primary duties/ activities	Expression, purification and functional-structural analysis of a recombinant protein (FPLC)
Name and address of employer	***Baxter AG***, Uferstr. 15, 2304 Orth
Field of business activity	Pharmaceuticals
From - until	September – October 2007
Position	Internship
Primary duties/ activities	Measurement of glucose and lactate levels of blood samples, Analysis of viscosity of Zeolith pastes
Name and address of employer	***Roche Diagnostics Graz GmbH,*** Kratkystraße 2, A-8020 Graz
Field of business activity	Pharmaceuticals, Diagnostics
From - until	August - October 2006
Position	Working student
Primary duties/ activities	Iron making, monitoring of the smelting process
Name and address of employer	***Voest Alpine AG,*** Voestalpine-Straße *1, A*-4020 Linz
Field of business activity	Steelmaking
From - until	August 2005
Position	Internship
Primary duties/ activities	Impact resistance and puncture tests of plastic foil (PE/PP)
Name and address of employer	***Borealis*** **Polyolefine** ***GmbH***, St.-Peter-Straße 25, A-4021 Linz
Field of business activity	Petrochemistry
From - until	July 2004
Position	Internship
Primary duties/ activities	Impact resistance and puncture tests of plastic foil (PE/PP) for customers
Name and address of employer	***Borealis*** **Polyolefine** ***GmbH***, St.-Peter-Straße 25, A-4021 Linz
Field of business activity	Petrochemistry
From - until	August 2003
Position	Internship
Primary duties/ activities	Authorship of a series of articles explaining science to pupils
Name and address of employer	**Helmholz Zentrum Dresden Rossendorf** , Bautzner Landstraße 400, D-01328 Dresden
Field of business activity	Cancer and Nuclear Safety Research, basic and application-oriented research

Education

From - until	Since September 20011
Acquired qualification	PhD in programmed Cell death and Aging
Name and address of institution of education	Karl Franzens Universität Graz
From - until	September 2008 – June 2011
Acquired qualification	Chemistry, Branch Biochemistry (with distinction)
Name and address of institution of education	Karl Franzens Universität Graz
From - until	September 2005 – June 2008
Acquired qualification	Chemistry, undergraduate
Name and address of institution of education	Karl Franzens Universität Graz
From – until	September 1997 – June 2005
Acquired qualification	General qualification for university entrance (major in Chemistry, Biology)
Name and address of institution of education	NAWI branch of BRG Hamerling, Hamerlingstraße 18, A - 4020 Linz
From – until	September 1994 – June 1997
Acquired qualification	Completion of elementary school
Name and address of institution of education	2nd Year elementary school in Dresden (DE), 3rd and 4th year in Neuhofen/ Krems (AUT)
From – until	September 1993 – June 1994
Acquired qualification	Completion of the first year of elementary school
Name and address of institution of education	Elementary school Santa Barbara, California, USA

Awards/Sponsorship

Date	2009, 2011 and 2012
Type of award	Merit grant
Name and address of institution of education	Karl Franzens Universität Graz
Date	2015
Type of award	Start Up realization grant
Name and address of institution of education	Inits

Date	2015
Type of award	Pre-Seed Grant
Name and address of institution of education	Austrian Wirtschaftsservice GmbH

Personal abilities and competences

Mother tongue	German
Additional languages	English, French, Russian
Self-assessment	English competent (C2), French independent (B1), Russian elementary (A1) Classification according to „Gemeinsamer Europäischer Referenzrahmen“
Other	Matlab, Modde, PIMS Trained Soft skills due to numerous internships and stays abroad Experience in leading small groups of students over two years
Additional details	I love athletic challenges of endurance sport (ski hiking, mountaineering) as well as the on the spot focus of acrobatics and skiing but also find the relaxation of scubadiving and even fishing highly enjoyable.
Conference/ Workshops	Bioprocess Engineering; 21-27.09.2014; Croatia Workshop; European Society of Biochemical Engineering Sciences Industrial bioprocess development academy; 15-17.04.2015; Germany Workshop; BASF Scale-up and scale-down of bioprocesses; 11-13.05.2015; Germany Conference Poster; DECHEMA Advanced Methods in Bioprocess Development; 10-11.06.2015; Austria Conference Talk; Technical University of Vienna Euro Biotechnology Congress; 18-20.08.2015; Germany Conference Talk; OMICS International Process Analysis & Automation Congress; 28-30.09.2015; USA Conference Talk; International Society for BioProcess Technology

Date

Signature