Model-based Strategies for Scale-down Studies in Fed-batch Cultivation of *Escherichia coli* Expressing Proinsulin

vorgelegt von
MScEng
Emmanuel Anane
ORCID: 0000-0001-8143-7551

von der Fakultät III-Prozesswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
-Dr. Ing.-
genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Juri Rappsilber
Gutachter:	Prof. Dr. Massimo Morbidelli
Gutachter:	Prof. Frank Delvigne
Gutachter:	Prof. Dr. Peter Neubauer

Tag der wissenschaftlichen Aussprache: 8. März 2019

Berlin 2019

Abstract

A fundamental discrepancy in process performance (yields) exists between laboratory and industrial scale fermentation processes. This discrepancy results from heterogeneous environments in large-scale bioreactors due to the longer mixing times associated with the increasing volumes at industrial scale, compared to laboratory scale bioreactors. The effects of such process inhomogeneity or the so-called scale-up effects on microbial physiology and recombinant product quantity have attracted much attention in the bioprocess research community due to the negative impacts of the heterogeneities on process efficiency. In the past three decades, various forms of single compartment and multi-compartment scale-down bioreactors have been developed to study scale-up effects in fermentation processes. In the same period, there has been a phenomenal increase in the use of high throughput (HT) miniaturized bioreactor systems for strain screening and bioprocess development, which has significantly helped to reduce the times required for early bioprocess development. Yet another (unrelated) field of much progress within the same period is the use of mechanistic models for bioprocess development and control. These mathematical tools have greatly contributed to our understanding of the interactions between the organism and the constraints of growth in bioreactors, as well as the elucidation of otherwise obscure intracellular processes. However, these tools and concepts—scale-down bioreactors, HT minibioreactors and mathematical models have mainly developed in parallel, with little or no interaction among them. In effect, scale-down bioreactors are still operated as standalone, low throughput devices; and the benefits of mathematical models are not fully exploited in both scale-down and HT systems.

In this work, we combine a mechanistic model with a parallel minibioreactor system to form a high throughput platform for studying the response of strains to scale-up effects at the screening phase. The mechanistic model is partly built on the operation mechanism of a two-compartment scale-down bioreactor and partly on advanced physiology of the expression system, *Escherichia coli*. During operation, the model is used to simulate the anticipated gradient profiles in large-scale bioreactors, and these are implemented in the minibioreactors using robotic liquid handling stations. As a demonstration, the platform was used to study the influence of glucose and dissolved oxygen gradients on both cell physiology and the misincorporation of non-canonical amino acids (ncBCAA) into recombinant proinsulin, in 24 parallel fed-batch cultivations of *E. coli*. The results show that in cultivations where the cells were subjected to model-derived glucose and dissolved oxygen gradients, there was a marked increase in the production of side metabolites, with reduced recombinant product formation rates and reduced yields. Additionally, the induced gradients resulted in more than 50-fold misincorporation of the ncBCAA norvaline into the recombinant proinsulin compared to the reference cultivation, which significantly undermines the product quality. The integration of the parallel minibioreactor system and mathematical tools provide an opportunity to perfom scale-down experiments in a high throughput manner. Therefore, screening of multiple strains can be combined with scale-down studies to select the most robust strains for bioprocess scale-up. Furthermore, HT cultivation in the developed platform can be used to generate large amounts of physiological data under defined stress conditions for advanced modelling and strain characterisation purposes.

Zusammenfassung

Eine grundlegende Diskrepanz besteht in den Prozessausbeuten im Laborbioreaktor und in der industriellen Produktion. Diese resultieren aus längeren Mischzeiten mit steigenden Volumina, die in Großbioreaktoren Heterogenitäten bedingen, die so in Laborbioreaktoren nicht auftreten. Die Auswirkungen solcher Prozessinhomogenitäten auf die mikrobielle Physiologie sowie auf die Ausbeute und Qualität rekombinanter Proteine und anderer Produkte sind ein wichtiger Bestandteil der aktuellen Forschung. In den vergangenen Jahrzehnten wurden verschiedene Formen von ein- und mehrstufigen Scale-Down-Bioreaktoren entwickelt, um Scale-Up-Effekte in Fermentationsprozessen im Labor zu untersuchen. Gleichzeitig haben sich, mit dem Ziel einer schnelleren Bioprozessentwicklung, miniaturisierte parallelisierteBioreaktorsysteme für das Stammscreening und die Bioprozessentwicklung etabliert. Parallel zu den instrumentellen Entwicklungen wurden in den vergangenen Jahren mechanistische Modelle für die Entwicklung und Kontrolle von Bioprozessen verstärkt eingesetzt. Solche mathematischen Modelle können zu einem besseren Verständnis der Wechselwirkungen zwischen Organismus und den Limitationen in Bioreaktoren sowie zur Aufklärung intrazellulärer Prozesse beitragen.

Diese Werkzeuge und Konzepte - Scale-down-Bioreaktoren, HT-Minibioreaktoren und mathematische Modelle - haben sich jedoch bisher weitgehend parallel entwickelt. In vielen Fällen werden Scale-Down-Bioreaktoren als eigenständige Geräte mit geringem Durchsatz betrieben, und die Vorteile mathematischer Modelle werden sowohl in Scale-Down-Systemen als auch in HT-Systemen kaum genutzt.

In dieser Arbeit wird ein mechanistisches Modell mit einem parallelen Minibioreaktorsystem zu einer Hochdurchsatzplattform für die Untersuchung von Stämmen auf Scale-up-Effekte in der Screening-Phase kombiniert. Das mechanistische Modell beschreibt die Funktion eines Zwei-Kompartment-Scale-Down-Bioreaktors und die Physiologie des zentralen Kohlenstoffmetabolismus von *E. coli*. Beim Screening wird das Modell verwendet, um die in großen Bioreaktoren erwarteten Gradientenprofile zu simulieren, die in den Minibioreaktoren über eine puls-basierte Medienzugabe mit robotergestützten *Liquid Handling*-Stationen umgesetzt werden. In der aktuellen Studie wurde die Plattform genutzt, um den Einfluss von Glukose- und Sauerstoffgradienten auf die Zellphysiologie und den Fehleinbau von nicht-kanonischen Aminosäuren zu untersuchen. Hierzu wurde Proinsulin als rekombinantes Modellprotein in 24 parallelen Kultivierungen mit dem Zulaufverfahren (Fed-batch) in *E. coli* produziert. Die Ergebnisse zeigen, dass in Kulturen, in denen die Zellen Glukose- bzw. Sauerstoff-Oszillationen ausgesetzt sind, die Produktion von Seitenmetaboliten wie z.B. Essigsäure deutlich zunimmt, und parallel die Produktbildungsrate sowie die finale Produktausbeute reduziert ist. Zusätzlich führt die puls-basierte Fütterung im Vergleich zur Referenzkultivierung zu einem 50-fach höheren Fehleinbau der nichtkanonischen Aminosäure Norvalin in das rekombinante Produkt.

Die Integration des parallelen Minibioreaktorsystems und mathematischer Werkzeuge bieten die Möglichkeit, Scale-Down-Experimente mit hohem Durchsatz durchzuführen. Daher kann das Screening mehrerer Stämme mit Scale-down-Studien kombiniert werden, um robuste Stämme für das Scale-up auszuwählen. Darüber hinaus kann die entwickelte Plattform genutzt werden, um große Mengen an physiologischen Daten unter definierten Stressbedingungen für weiterführende und Stamm-Charakterisierungen zu generieren.

List of Contents

Acknowledgements

I would like to express my sincere gratitude to a a number of people who contributed in various ways to the success of this work.

Firstly, I would like to thank my supervisor, Prof Dr Peter Neubauer for giving me the opportunity to do my PhD in his laboratory at the TU-Berlin. His supervision and ideas have really shaped my perceptions about the bioprocess industry and the immense potential of biotechnology. My sincere thanks also go to my PhD adviser, Dr. Ing. M Nicolas Cruz Bournazou: thank you for your patience with me and the opportunity to learn all those great modelling techniques from you.

Secondly, I would like to thank all the technical staff who made sure that all equipment were up and running: without your input, none of the experiments in this work would have been possible. These people are Dipl. Ing. Thomas Högl, Brigitte Buckhardt, Irmgard Maue-Mohn, Frank Ludwig (Sanofi Aventis Deutschland GmbH) and Dr. Manuel Krewinkel (Sanofi Aventis Deutschland GmbH). I am grateful to Dr. Ing. Sebastian Riedel for his support in enzymatic analytical methods and equipment.

There is one person I would like to dedicate one full paragraph of my acknowledgements to: Klaus Pellicer Alborch. I am very much grateful for all the social and administrative support you have given me in the last three years of my life in Berlin, and beyond. Thank you *'boss'*.

I am also grateful to the European Commission Horizon 2020 MSCA Actions for financial support. I am grateful to Dr. Ing. Stefan Junne for coordinating and handling all things relating to funding, important scientific inputs and for the nice time during training weeks. In this vein, I also express my gratitude to the Biorapid Consortium and the project coordinator, Prof Jarka Glassey (Newcastle University, UK) for the nice times we spent together during this project.

My next thanks go to all the scientists who contributed immensely to the development of the concepts in this thesis by co-authoring the various publications: Dr. Ing Diana C. López C., Associate Prof Gürkan Sin (DTU, Denmark), Prof Krist V. Gernaey (DTU, Denmark), Dr. Ing. Tilmann Barz (AIT, Vienna, Austria), Dr Peter Hauptman (Sanofi Aventis Deutschland GmbH, Frankfurt aM, Germany), Annina Sawatzki, Benjamin Haby, Sebastian Hans, Florian Glauche, Ángel Córcoles Garcia … I also acknowledge the help of Dr. Ing. Christian Retiz, whose work was foundational and paved the way for my work.

Finally, I would like to thank my beloved wife, Florence Appiah Anane for her love, patience and support during the hard times of the PhD life. Thank you for the always being there and for the encouragement.

I thank God for the gift of Life and Love

List of Abbreviations

2-CR	Two-compartment Bioreactor
cGMP	Current good manufacturing practices
CFD	Computational fluid dynamics
CRD	Cellular reaction dynamics
DOT	Dissolved oxygen tension
DoE	Design of experiments
GMoP	Good modelling practices
QbD	Quality by Design
PBM	Population balance modelling
HTPD	High throughput process development
PFR	Plug flow reactor
STR	Stirred tank reactor
ncBCAA	Non-canonical branched chain amino acid
PAT	Process analytical technology

List of Symbols

K_{ap}	= Monod-type saturation constant, intracellular acetate prod. (g g^{-1} h^{-1})
K_{sa}	= Affinity constant, acetate consumption (g L^{-1})
K_s	= Affinity constant, glucose consumption (g L^{-1})
K_{ia}	= Inhibition constant, inhibition of cellular growth by extracellular acetate(g L^{-1})
K_{is}	= Inhibition constant, inhibition of acetate uptake by glucose (g L^{-1})
p_{Amax}	= Maximum specific acetate production rate (g g^{-1} h^{-1})
q_{Amax}	= Maximum specific acetate consumption rate (g g^{-1} h^{-1})
q_m	= Specific maintenance coefficient (g g^{-1} h^{-1})
q_{Smax}	= Maximum specific glucose uptake rate (g g^{-1} h^{-1})
Y_{as}	= Yield of acetate on substrate (g g^{-1})
Y_{oa}	= Specific oxygen used per gram of acetate metabolized (g g^{-1})
Y_{xa}	= Yield of biomass on acetate (g g^{-1})
Y_{em}	= Yield of biomass on glucose, excluding maintenance (g g^{-1})
Y_{os}	= Oxygen used per gram of glucose metabolized per gram biomass (g g^{-1})
Y_{xsof}	= Yield of biomass other products of overflow routes, excluding acetate (g g^{-1})
C	= Carbon content of (s) substrate, (x) biomass

List of publications

I. **Anane E**, López C DC, Neubauer P, Cruz Bournazou MN. *Modelling overflow metabolism in Escherichia coli by acetate cycling.* Biochem Eng J. 2017; 125:23—30, https://doi.org/10.1016/j.bej.2017.05.013

II. **Anane E**, López C DC, Barz T, Sin G, Gernaey KV, Neubauer P, Cruz Bournazou MN. *Output uncertainty of dynamic growth models: effect of uncertain parameter estimates on model reliability.* Biochem Eng J. (2019). Submitted (*8th February 2019*).

III. **Anane E**, Sawatzki A, Neubauer P, Cruz Bournazou MN. *Modelling concentration gradients in fed-batch cultivations of E. coli —towards the flexible design of scale-down experiments.* Journal of Chem. Technol. Biotechnol. 2019; 94(2):516-526, https://doi.org/10.1002/jctb.5798

IV. **Anane E**, Garcia AC, Haby B, Hans S, Krausch N, Glauche F, Hauptmann P, Neubauer P, Cruz-Bournazou MN. *Scaling down further: model-based scale-down studies in parallel minibioreactors for accelerated phenotyping.* (In preparation).

Author Contributions to the Papers

Paper	Co-author	Contribution
I	Emmanuel Anane	Writing the paper*, carried out all experimental work, and helped in parameter estimation.
	Diana C. Lopez C	Parameter estimation
	Peter Neubauer	Research Conceptualization and Supervision
	M. Nicolas Cruz Bournazou	Supervision
II	Emmanuel Anane	Writing the paper, carried out all experimental work and parameter estimations, paper conceptualization
	Diana C. Lopez C, Tilmann Barz	Assisted with development of mathematical methods of singular value decomposition and regularization
	Gurkan Sin, Krist Gernaey	Assisted with mathematical methods of uncertainty analysis, Supervision
	M. Nicolas Cruz Bournazou, Peter Neubauer	Supervision, Project conceptualization
III	Emmanuel Anane	Writing the paper, all experimental work, Parameter estimation, simulations of 2-CR
	Annina Sawatzki	Helped in writing the code for the 2-CR simulations
	Peter Neubauer	Project conceptualization and supervision, major input in the storyline/framework of the paper
	M. Nicolas Cruz Bournazou	Supervision, simulation of gradient profiles of 2-CR
IV	Emmanuel Anane	Writing the paper, experimental work, Parameter estimation
	Benjamin Haby, Sebastian Hans	Helped in experimental work and anlytics, ensuring the operation and functionality of all the robotic facilities.
	Florian Glauche, Niels Krausch	Helped in data analysis (protein quantification)
	Angel Corcoles Garcia, Peter Hauptmann	Data Analysis and interpretation of results for misincorporation of ncBCAAs in the recombinant product
	M. Nicolas Cruz Bournazou, Peter Neubauer	Research Conceptualization and Supervision. Interpretation of results.

* All co-authors read the papers and did minor corrections to the write-up.

1. Introduction

Following process development in laboratory scale bioreactors, a potential bioprocess must be scaled up to production scale to meet demand and for economic purposes. However, the transition of a bioprocess from laboratory to industrial scale is not a trivial task. Traditionally, scale-up has been carried out mostly based on operator experience, with guiding engineering principles such as maintaining constant oxygen transfer capacity, broth turbulence (shear forces), mixing times, power per unit volume or gassing rates between the two scales. Preference is usually given to the scale-up criterion that is most relevant for the given bioprocess i.e. impeller tip speed for shear sensitive cultures such as filamentous fungi or mammalian cell culture, K_La for cultures that are sensitive to anoxic conditions, P/V for power requirements or economic reasons, etc. Once a specific scale-up criterion is selected, the other factors are allowed to change uncontrollably in the larger scale, which, more often than not, leads to unpredictable metabolic and physiological behaviour with concomitant loss of yields and other undesirable consequences. A study on the expression of interferon-α1 in recombinant *E. coli* showed a 33 % decrease in biomass yield and over 50% reduction in recombinant protein yield upon scale-up (Riesenberg et al., 1990). Generally, up to 30% loss in recombinant protein productivity, yields and biomass concentrations upon scale-up are common in the bioprocess industry (Takors, 2012). These drawbacks are partly triggered by the formation of concentration gradients in larger bioreactors (Larsson et al., 1996), caused by increased mixing times due to larger broth volumes (Alvaro R Lara et al., 2006). In such processes, as fluid packets move between feeding zones and starvation zones in inefficiently mixed bioreactors, cells are constantly exposed to oscillating concentrations of substrate, metabolites, dissolved oxygen and carbon dioxide(Bylund et al., 1999, 1998; Enfors et al., 2001; Larsson et al., 1996).

In order to understand and better characterise the behaviour of the culture in the presence of these heterogeneities, the scale-down bioreactor concept was developed. Scale-down bioreactors are laboratory scale-bioreactors that are operated or designed to mimic the heterogeneous environments that are thought to exist in large-scale bioreactors (Neubauer and Junne, 2016a). The most common form of scale-down bioreactors consist of multiple compartments, usually a stirred tank bioreactor connected to a plug flow reactor or to another stirred tank bioreactor. During operation, a stress inducing agent is injected into one compartment where the part of the culture in this compartment is exposed to the concentration gradient for a given period of time, until the stress is nullified by mixing with the bulk of the broth in the bigger compartment. Scale-down simulators based on pulsed inputs (pulse addition of substrates, e.g. glucose, dissolved oxygen or an acid or base) to achieve the concentration gradients experienced in heterogeneous conditions are also reported (Peter Neubauer et al., 1995; Sunya et al., 2013). These operation methods lead to exposure of the culture to varying gradients, which typically leads to the undesirable responses that are also seen in large-scale bioreactors. In the past three decades, scale-down bioreactors have been successfully used to study the response of microbial (Neubauer and Junne, 2016b) and mammalian (Nienow et al., 2013)] cultivations to environmental stresses in large-scale bioreactors. The most recent advances in the development of scale-down bioreactors include the mathematical description of both the heterogeneous environments and populations subgroups as they evolve under concentration gradients. In terms of detailed description of mixing patterns in larger bioreactors, computational fluid

dynamics models (CFD) have been developed to determine the mixing patterns of large-scale bioreactors, given the geometric configuration of the bioreactor and other physico-chemical properties of the broth. Furthermore, the incorporation of cellular reaction dynamics (CRD) concepts, such as substrate uptake rates and specific growth rates into the CFD models allows a detailed review of the physiological response of the culture to the simulated mixing patterns in the bioreactor. In terms of cellular responses to the induced stresses, population balance models have also been developed to characterise the various population sub-groups that emerge from cultures exposed to concentration gradients (Baert et al., 2016; Brognaux et al., 2014). Additional developments in rapid sampling technologies (Visser et al., 2002) and *'omics'* technologies (proteomics, metabolomics, transcriptomics) have been used to investigate molecular level responses of strains to concentration gradients induced in scale-down bioreactors (Buchholz et al., 2014; Simen et al., 2017). These advancements allow a more detailed and systematic characterisation of the strain's response to the given stresses.

Despite the success and advancements in the application of scale-down bioreactors for fermentation development, a number of challenges remain. Firstly, scale-down bioreactors are laborious and difficult to operate. This makes them low throughput devices, requiring several months to characterise the response of one strain to scale-up effects. Secondly, scale-down studies are usually carried out only in the latter stages of bioprocess development, or sometimes even after a potential bioprocess fails in industrial scale. It would be much more desirable to apply scale-down bioreactors at the very early stages of bioprocess development, viz. in the screening phase to help select the most robust strains for scale-up. In this way, strain optimization can simultaneously involve scale-up optimisation (Alvaro R Lara et al., 2006), without carrying out actual large-scale cultivations. This second option would require that scale-down bioreactors are designed to be high throughput devices, since at the screening phase a lot of strains are usually involved.

This thesis covers the work done on further advancement of scale-down bioreactors for the study of microbial responses to scale-up stresses at the screening phase, using physiologically descriptive mechanistic modelling and a high throughput, minibioreactor-based bioprocess development platform. The work delves into the application of mathematical modelling to bioprocess development. A strain of *Escherichia coli* expressing recombinant proinsulin is used as the model organism. Firstly, a model of scale-down bioreactors and a detailed mechanistic model of the physiology of the strain under both homogeneous and heterogeneous fed-batch cultivation conditions are developed. Secondly, these models are used to calculate concentration gradients in the form of varying glucose feed and dissolved oxygen pulses, that are implemented in the high throughput parallel minibioreactor system as a scale-down cultivation platform. We use this cultivation platform to show that under conditions of oscillating glucose and dissolved oxygen concentrations in fed-batch culture, there is a significant loss in product and biomass yields, and most importantly, there is misincorporation of non-canonical amino acids into the recombinant protein, which negatively influences the product quality.

2. Scientific Background

The current trends of development in the production of materials and energy for human use are gravitating towards sustainable practices. Notable companies such as BASF, Dow and DuPont in the chemicals and materials industry have all launched industrial biotechnology departments, in a bid to switch from traditional chemical manufacturing to bio-based material supply routes. Additionally, the application of industrial enzymes in the process industries have significantly helped to reduce the carbon footprint of many resource-intensive industries (laundry, baking, brewing, leather and textile processing, agriculture and food processing), towards sustainable development (Schmid et al., 2002). In the pharmaceutical industry, biopharmaceuticals have taken a centre stage in research and development. Here, the main driver is not only sustainability, but also the unique ability of biological molecules to treat unconventional diseases such as cancer and genetic disorders, which, hitherto, was not achievable with traditional chemical molecules of pharmaceutical origin. For instance, the monoclonal antibody (mAb) and personalized medicine market share in the pharmaceutical industry has grown exponentially in the past decade and is poised to grow further as new biological molecules are discovered and developed every year (Seamans et al., 2008).
It is important to note, at this point, that the bioreactor is central to almost all biotechnological processes; be it in the production of renewable energy from energy crops, in the production of biopharmaceuticals or in the biological production of specific enzymes for industrial applications. We acknowledge the existence of a few cell-free synthesis platforms, which may not use conventional bioreactors but these are in the minority without any industrial applications, as yet (Ogonah et al., 2017; Zemella et al., 2018). Therefore, with the wide acceptance of bioprocesses across many diverse industries, the efficient and economically prudent operation of the bioreactor, at all scales (from development to production), is important for the success of industrial biotechnology. Secondly, there is the need to significantly improve the slow pace of development of industrial bioprocesses (currently, 60—180months) (Neubauer et al., 2017), compared to chemical processes (6 to 18 months). The following sections cover a general discussion of the state-of-the-art in bioprocess scale-up to meet industrial demand, and how the application of two enabling tools, namely mechanistic modelling and high throughput cultivation systems can help to better understand cellular behaviour in bioreactors at different scales. Such knowledge would help to fast-track bioprocess development and ensure more robust fermentation scale-up to meet the needs of the bio-economy.

2.1. The Challenge of Bioprocess Scale-up

Although a few cell-free synthesis platforms are reported (Ogonah et al., 2017), the majority of economically viable bioprocesses use cell factories cultivated in submerged fermentation systems. Since the bioreactor is the heart of every bioprocess, it is important to understand how differences in the internal bioreactor environment affect the growth of cells, especially when there is a change of scale (scale-up). Bioprocess development usually begins with the screening of large strain libraries to select targets that are dominant in growth and recombinant protein production or some other target. This is usually carried out in shaken cultures in parallel microwell plates where only endpoint measurements are possible (Lattermann and Büchs, 2015), or recently in parallel minibioreactors (Back et al., 2016). Following screening, the process is developed further in laboratory scale bioreactors, usually with operating volumes of 1—20 L where control strategies and process characterisation are

done. In the specific case of biopharmaceuticals, samples for clinical trials may be produced as early as in the laboratory development phase, or in early pilot plant phase under cGMP guidelines. Once regulatory clearance is obtained for commercial use of the product, the process is then scaled-up to large-scale bioreactors for mass production. Two important points must be considered in this mass production process: (i) the product quality and attributes that affect the efficacy of the biopharmaceutical candidate must not change upon mass production (scale-up), when compared to the version used in clinical trials and (ii) the process must be economically feasible, so as to yield the necessary returns on investment whilst being appropriately priced for the general market. The second point has to do with the yields (Y_{px}, Y_{xs}, energy consumed per unit of product produced, etc) that are achieved in the large-scale bioreactor. Since the product quality (plus any related attributes) and the yields of fermentation processes have been shown to directly depend on the process conditions prevalent in the bioreactor (Alvaro R Lara et al., 2006), the possibility to meet the above criteria upon scale-up would mean that the process conditions in the large-scale are the same as (or are acceptably close to) the process conditions at development scale. Unfortunately, the increase in scale in bioreactors has many implications for the process conditions inside the bioreactor. Even when geometric similarities are maintained, it is not possible to scale-up a fermentation process to maintain all relevant criteria within acceptable limits. That is, if an 80 L bioreactor is scale-up to 10,000 L whilst maintaining a constant power input per unit volume, the mixing time increases 3 times, the impeller tip speed doubles, and the shear forces increase almost 10 times (Alvaro R Lara et al., 2006). Oldshue showed that a scale-up design to satisfy mass transfer (constant K_La criterion) from a 75 L pilot scale process to a 95,000 L production scale would increase the shear rate by 180 %, whereas maintaining a constant shear rate between the two scales could only produce 40% of the mass transfer requirements of the culture in the large scale (Oldshue, 1966). Therefore, the accurate translation of laboratory scale processes into industrial-scale processes is difficult, and during scale-up, the most important parameter that affects the culture is selected to be maintained constant. The most common consequence of this reality is an inevitable increase in mixing times of up to 200 seconds in larger-scale bioreactors (since scale-up is mostly based on K_La, P/V, impeller tip speed) (Alvaro R Lara et al., 2006). In addition to the increased mixing times, fed-batch processes are fed with concentrated substrates at localized input points, which are mechanically fixed. The longer mixing times and the localized addition of highly concentrated viscous substrates lead to the formation of concentration gradients in the bioreactor (Enfors et al., 2001; Larsson et al., 1996). Cells that are traversing these gradients respond in many ways, e.g. by expressing certain stress response genes (Simen et al., 2017), by shifting metabolic routes and by developing other physiological coping mechanisms (e.g. development of flagella in *E. coli*), which all affect the fermentation efficiency in terms of yields and overall process robustness.

2.1.1. Concentration Gradients and Fermentation Efficiency

When the characteristic time of relevant cellular processes (translation, cell division) is close to the mixing time in large-scale bioreactors, there is measurable influence of gradients on the growth and metabolic behaviour of the culture (Alvaro R Lara et al., 2006; Sweere et al., 1987). The inefficient mixing in large-scale bioreactors leads to the creation of spatial concentration pockets of relevant process parameters, such as substrate (glucose), dissolved oxygen, acidity and temperature. A specific challenge of GMP manufacturing processes is the rigidity of the process due to legislation requirements, which prevents bioreactor retrofitting that would allow practical measurement of

concentration gradients in industrial fermenters. For non-GMP processes such as brewing and biogas processes, multi-position and multi-parameter sensors that are mounted on movable lances within the bioreactor have been used to measure spatial concentrations at specific points in the bioreactor (Bockisch et al., 2014). For GMP processes, techniques such as CFD modelling and regime analysis based on mixing time calculations are used to determine the existence of gradients in large-scale bioreactors (Nørregaard et al., 2018). Recently, movable, free floating balls with embedded sensors have been developed to enhance the measurement of process parameters in different zones of large-scale bioreactors (Lauterbach et al., 2017). The general influence of these spatial gradients on overall fermentation efficiency, and studies undertaken in connection to each gradient type are reviewed below.

2.1.1.1. Substrate Gradients

In fed-batch culture, the idea is to control the growth of the organism by limited substrate supply. Therefore, the existence of excess substrate zones in the broth defeats the purpose of this tight control for the fraction of the culture that comes into contact with these zones. The exposure of the culture to zones of higher substrate concentrations has direct consequences on the uptake capacities of the cells for this substrate (Brand et al., 2018; Lin et al., 2001b). In effect, the excess substrate zones may cause the cells to grow at the maximum specific growth rate, which may plunge organisms such as *E. coli* and *S. cerevisiae* into overflow metabolic states. Several studies to investigate the influence of substrate gradients on fermentation processes involving many expression systems have been reported. For instance, in studies involving *E. coli*, substrate gradients that were mainly induced in scale-down bioreactors have been shown to: (i) reduce growth rate and biomass yields (Bylund et al., 1998; Neubauer et al., 1995), (ii) significantly increase the maximum specific uptake capacities for both substrate and dissolved oxygen (Brand et al., 2018; Lin et al., 2001a) and (iii) contribute to the build-up of both intracellular and extracellular metabolite pools (Bylund et al., 1998; Soini et al., 2011a; B. Xu et al., 1999). Additionally, Simen et al. characterized the genes that respond to perturbations in limiting glucose supply to *E. coli* cultures, and concluded that the expression of genes related to uptake and assimilation are up-regulated under non-regular glucose supply conditions. This molecular level response to substrate gradients resulted in about 15% increase in the maintenance energy requirements for the culture in the presence of substrate fluctuations (Simen et al., 2017). These results were similar to the results of Schweder et al. who monitored the genes that respond to substrate oscillations and oxygen limitation in large-scale bioreactors and observed elevations in mRNA levels in *E. coli*, within a few seconds after induction of glucose gradients (Schweder et al., 1999). Other studies of substrate gradients involving the industrially relevant organism *Corynebacterium glutamicum* have also been reported. Lemoine et al. investigated the influence of glucose gradients in *C. glutamicum* growing in complex medium and observed much stringent and more degenerative responses in the yield of cadaverine (Lemoine et al., 2016), compared to similar scale-down cultivations in defined medium where the product was L-lysine (Lemoine et al., 2015). Further studies using *C. glutamicum* have shown similar responses in medium acidification and shifts in uptake capacities (Käß et al., 2014b), as also observed in *E. coli*. Recently, the studies of Marba-Ardebol et al. on substrate gradients in the yeast *S. cerevisiae* using in-situ microscopy to monitor cell size distribution has broadened the scope of such studies. The authors observed significant morphological responses to the induced substrate gradients, in terms of single cell size distribution, as well as the accumulation of side

metabolites (A. M. Marbà-Ardébol et al., 2018). Lastly, a recent study of substrate gradients in *Penicillium chrysogenum* cultivations using intermittent substrate supply and 2-CR techniques revealed the build-up of metabolites from TCA cycle intermediates, the up-regulation of glucose/hexose uptake genes (capacities) and a general decrease in the specific product formation rate (Wang et al., 2018). Thus, for all the oragnisms studied (*E. coli*, *C. glutamicum, S. cerevisiae, P. chrysogenum*), there is a general agreement/similarity in both the physiological and metabolic responses to substrate gradients, with an overall conclusion that substrate gradients generally lead to low efficiency in fermentation processes. However, some organisms (e.g *C. glutamicum*) seem to be more robust in the presence of concentration gradients than others, e.g. *E. coli*, even though such studies comparing the response of different organisms to the same gradients is not reported.

2.1.1.2. Dissolved Oxygen Gradients

Dissolved oxygen gradients may arise in large industrial-scale bioreactors from a combination of three limitations: (i) inherently low solubility of oxygen in fermentation broths. At atmospheric pressure, the solubility of oxygen in water is about 7mmol/L, which is far less than the oxygen uptake capacity of common bacteria such as *E. coli* at average cell densities. This solubility is further reduced by the dissolution of several components to make the growth media. In effect, typical solubilities of oxygen ca. 5 mmol/L are common in fermentation broths (Enfors, 2011); (ii) Economic considerations and practical limitations in industrial mixing. The volumetric mass transfer coefficient for oxygen transfer from the gas phase to the liquid phase is directly dependent on the power input per unit volume, which, in turn, is dependent on the agitation rate. The agitation rate also determines the energy consumption of the process. Therefore, the K_La has a direct link to energy consumption. Thus, the economic viability of the process puts a constraint on how much oxygen transfer capacity can be achieved in an industrial process; (iii) shear sensitivity of the culture. Although free suspension culture cultivations are no longer considered to be sensitive to agitation-related shear stresses (Nienow, 2006)), processes involving cell agglomerates (fungal cultivations) may be sensitive to shear stresses (Veiter et al., 2018). Therefore the practical limitations in such processes would dictate a lower agitation rate. Oosterhuis and Kossen used actual measurements in a 19 m^3 production bioreactor to show DOT gradients of up to 22% air saturation between stagnant and highly turbulent zones in the bioreactor (Oosterhuis and Kossen, 1984). The implication of such gradients for a fed-batch process is that cells in the low DOT zones will switch to anaerobic metabolism, which may be difficult to reverse for Crabtree positive yeasts like *S. cerevisiae* (due to a strong physiological memory) even when the cells move back to homogeneous zones in the bioreactor (Sanchez-Gonzalez et al., 2009). In cultures where the air is enriched with pure oxygen due to high cell densities, the ocurence of gradients with highly concentrated oxygen in the richer zones can promote the leakage of reactive oxygen species (ROS) from metabolic routes (osmotic shock), which may impair both DNA synthesis and the functioning of important intracellular enzymes, leading to poor growth (Baez and Shiloach, 2014). Several studies in investigating the response of strains to dissolved oxygen gradients are also reported in the literature. These usually go hand-in-hand with studies of substrate gradients, since high metabolic acitivity in the vicinity of excess substrate leads to depletion of dissolved oxygen. Scale-down studies of oxygen gradients in the *E. coli* system show a rapid accumulation of formate, lactate and acetate as products of mixed acid fermentation (Bylund et al., 1999; B. Xu et al., 1999), reduced recombinant product yields, at the example of valinomycin production (Li et al., 2015) and a general decline in growth (Sandoval-Basurto et al., 2005). Further studies involving the exposure of *C.*

glutamicum to dissolved oxygen gradients have been reported (Käß et al., 2014a, 2014b, Lemoine et al., 2016, 2015). A perculiar observation of this organism in response to dissolved gradients is its ability to quickly adapt to the heterogeneous environtments (Lemoine et al., 2015), and therefore does not seem to show significant negative influence on the process yields and robustness (Käß et al., 2014a).

2.1.1.3. Temperature, pH and Carbon Dioxide Gradients

Temperature gradients are among the least studied scale-up effects in bioprocess development, although heat effects are thought to be highly detrimental for a fermentation process (Alvaro R Lara et al., 2006). This is probably due to the difficulty in setting up experimental platforms to achieve differential temperature profiles whilst ensuring enough mixing and mass transfer for aerobic growth in a bioreactor (Lethanh et al., 2005). Higher temperature zones can induce heat shock response in microbial cultivation, and lead to pre-mature accumulation of recombinant proteins in cultures that are induced with a temperature shock (Lethanh et al., 2005). Additionally, cells in the vicinity of higher temperature zones will tend to exhibit increased metabolic activity, which may lead to the aggregation of not only soluble recombinant proteins, but also native proteins of the strain. pH gradients are recently gaining attention in the bioprocess research community. Simen et al investigated the effect of ammonia pulses (shifts in pH) in *E. coli* cultivations and observed a higher maintenance energy and the activation of over 400 genes in response to the pH gradients (Simen et al., 2017). In microbial cultivations, a recent study of CO_2/HCO_3^- gradients in *Corynebacterium glutamicum* showed no significant impact of these stresses in the physiological response of the organism, although there was a marked increase in the expression of certain genes, upon genomic analysis (Buchholz et al., 2014). In a recent report, *E. coli* cells exposed to CO_2 levels above 70 millibar CO_2 partial pressure in the inlet gas led to reduced biomass yields and rapid accumulation of acetate, even under non-overflow and fully aerobic conditions. A detailed study of the genomic response of the strain in the study revealed a novel acid response in the cells due to the acidification of the cytoplasm (Baez et al., 2011, 2009). In mammalian cultivations, CO_2/HCO_3^- gradients are expected to have significant impact on the fermentation process due to the sensitivity of these cells to elevated CO_2 concentrations (Spadiut et al., 2013).

A notable phenomenon in the study of the influence of concentration gradients on fermentation performance was recently reported by Limberg et al. 2017. Whereas previous studies (as discussed above) show that *C. glutamicum* is quite robust to glucose, dissolved oxygen and pH gradients (Buchholz et al., 2014; Käß et al., 2014b; Lemoine et al., 2016), the robustness seems to be exhibited when the stresses are applied in isolation or in certain specific combinations. The results of Limberg et al. 2017 show that when pH gradients are coupled to oxygen limitation, *C. glutamicum* loses it robustness against dissolved oxygen fluctuations (Limberg et al., 2017), leading to yield losses of up to 40%. This implies that the study of concentration gradients in fermentation should be conducted in a multi-faceted manner, to consider all possible gradients and the necessary combinations among them to arrive at a more holistic conclusion for each strain.

2.1.2. Influence of Concentration Gradients on *E. coli* Processes

The existence of concentration gradients in large-scale cultivations of *E. coli* means the cells are intermittently exposed to excess glucose—limiting oxygen or limiting glucose—excess oxygen conditions in a cyclic or oscillating manner. *E. coli* responds in a similar manner to excess glucose environments as Crabtree positive yeasts. In the presence of excess glucose, the uptake and processing of glucose in the upper glycolytic routes in quantities that exceed the fluxes that can be processed through the tricarboxylic acid (TCA) cycle lead to the accumulation of pyruvate. Thus, when *E. coli* comes into contact with concentration gradients, there is a build-up of intracellular pyruvate in the cell (Soini et al., 2011a; Vemuri et al., 2006). In the absence of oxygen or in limiting oxygen conditions, the excess pyruvate is readily converted to acetic acid, formic acid, succinate (from PEP), lactic acid and ethanol in a process known as mixed-acid fermentation (Figure 2.1).

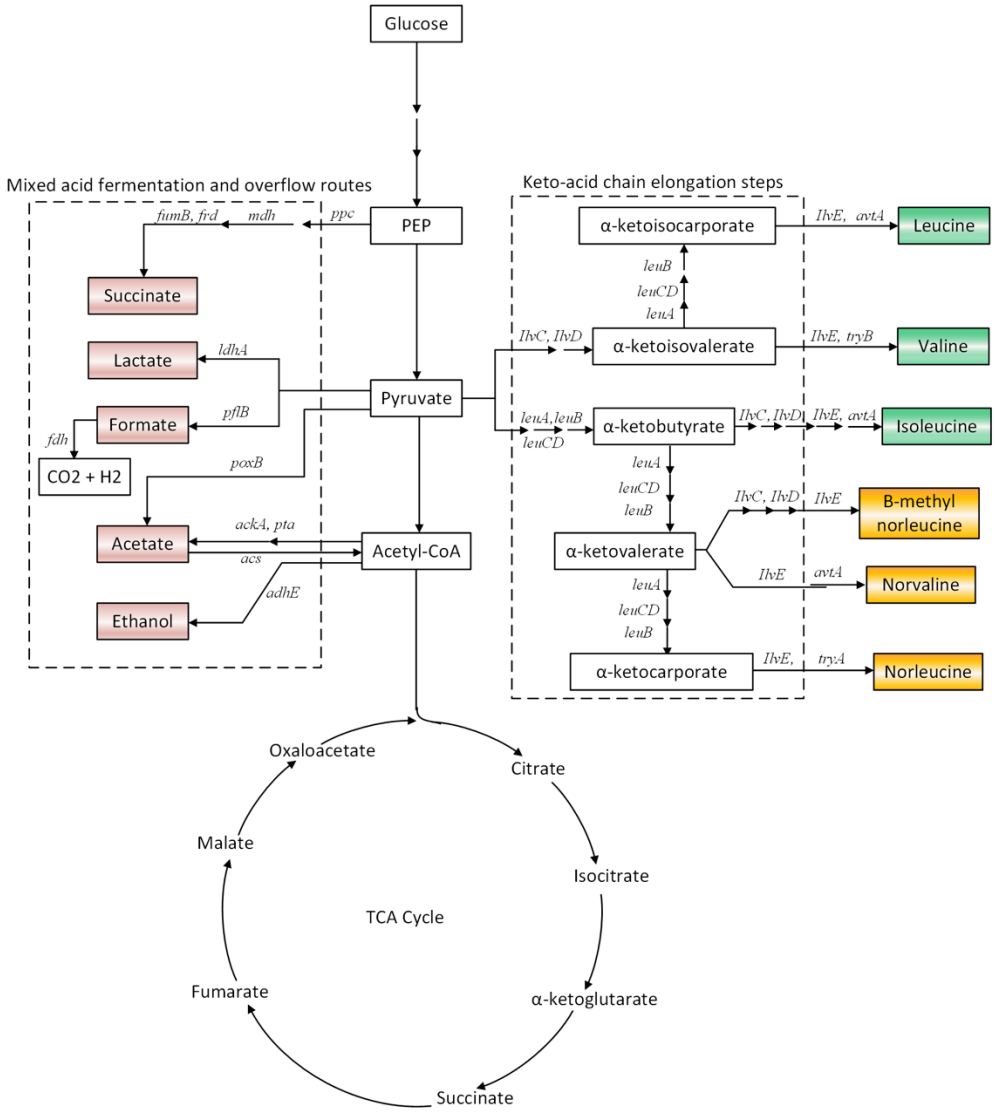


Figure 2.1 A simplified representation of upper glycolytic chain of *E. coli* showing the mixed acid fermentation routes leading to the formation of short-chain organic acids and ethanol (*brown boxes*). Also shown is the keto-acid elongation pathways leading to the formation of the branched chain amino acids (*green boxes*) and their non-canonical analogues (*orange boxes*). The gene clusters/enzymes catalysing the various reactions are also shown in italics. Multiple arrow heads on a pathway shows the number of intermediates on the given metabolic route. Adapted from Apostol et al. 1997, Soini et al. 2008 and Reitz et al. 2018.

In severe anaerobic conditions, formic acid may further be converted to CO_2 and H_2, whereas the left wing of the TCA cycle (succinate) may be shunted back to pyruvate (Vemuri et al., 2006). In the presence of oxygen, the excess pyruvate is mainly converted to acetate, as the major product of overflow metabolism. Since excess glucose conditions in large bioreactors usually occur simultaneously with limitations in dissolved oxygen concentrations (due to higher metabolic activity in the vicinity of the high glucose concentrations), the products of mixed acid fermentation and overflow metabolism are common in large scale cultivations of *E. coli* (De Mey et al., 2007; Taymaz-nikerel et al., 2013). This channelling of carbon resources to extracellular carbonic acids is considered wasteful in fermentation processes because it drastically reduces the yields of both biomass (Luli and Strohl, 1990; Valgepea et al., 2011) and the recombinant protein products (Swartz, 2001). Secondly, the accumulated metabolites such as acetate have been shown to be inhibitory to cell growth in *E. coli* culture (Wolfe, 2005).

Another critical consequence of intracellular pyruvate accumulation in *E. coli* is the further processing of pyruvate into the branched chain amino acids (leucine, isoleucine and methionine) and their non-canonical analogues norvaline, β-methyl-norleucine and norleucine (Reitz et al., 2018; Soini et al., 2011a). Pyruvate is an important compound in the metabolism of *E. coli*, and its intracellular levels modulate the activities of several enzymes both downstream and upstream of the pyruvate node (Bernal et al., 2016). In the presence of elevated pyruvate concentrations, several chain elongation and transamination reactions by various enzymes lead to the formation of α-ketoacids (α-ketobutyrate, α-ketovalerate, α-ketoisovalerate) which serve as precursors for the synthesis of the non-canonical branched chain amino acids (ncBCAAs), norvaline, β-methyl-norleucine and norleucine (for detailed synthesis pathways, see (Apostol et al., 1997; Reitz et al., 2018)). The ncBCAAs are naturally non-proteinogenic, and should therefore not be part of translation products in both wild type and recombinant *E. coli* strains under normal cultivation conditions. However, at higher intracellular concentrations of ncBCAAs, there is an increased chance of mis-acylation of tRNA molecules of the canonical amino acids with the non-canonical forms due to the promiscuity of their corresponding aminoacyl-tRNA synthetases for the ncBCAAs (Cvetesic et al., 2014). This leads to the misincorporation of the ncBCAAs into both cellular material and recombinant proteins, where norvaline replaces leucine, β-methyl-norleucine replaces isoleucine and norleucine replaces methionine. This problem is a common occurrence in large industrial-scale *E. coli* processes and also for CHO cells (Wen et al., 2009). In biopharmaceutical settings where strict regulatory guidelines are enforced, a significant variation in the amino acid composition of a therapeutic molecule may have serious economic and regulatory consequences (Harris and Kilby, 2014). The accumulation and misincorporation of ncBCAAs in *E. coli* has been directly linked to cultivation under micro-aerobic conditions, which may result from glucose gradients in large-scale bioreactors (Cvetesic et al., 2016, 2014; Soini et al., 2008).

Additionally, general performance loss in terms of process yields in the presence of oscillating oxygen concentrations (Jaén et al., 2017; Sandoval-Basurto et al., 2005), pH gradients (Simen et al., 2017) and carbon dioxide gradients (Baez et al., 2011) are also reported in the literature for *E. coli* cultivations.

2.2. Scale-down Bioreactors

To understand the impact of concentration gradients and oscillating environmental conditions on the physiology of microorganisms during fermentation, the scale-down bioreactor concept was developed.

A scale-down bioreactor is a smaller bioreactor (than the one being studied) that has been configured to reproduce the environmental conditions of a larger bioreactor to enable the study of the stress effects of the larger bioreactor. The rationale for creating scale-down bioreactors is simple: large industrial scale bioreactors can seldomly be used for experimental investigation due to cost, cGMP, regulatory and time limitations. The smaller scale abstraction of the large bioreactor provides the flexibility of laboratory cultivations with smaller inventory (media and energy costs) in dedicated research facilities. Since their inception in the mid 80s, scale-down bioreactors have been used extensively in the bioprocess research community to investigate the responses of industrially relevant expression systems to scale-up stresses. For a comprehensive review of scale-down systems, see (Delvigne et al., 2017b; Neubauer and Junne, 2016a, 2010; Takors, 2012).

2.2.1. Types and Operation Mechanisms of Scale-Down Bioreactors

Scale-down bioreactors can be broadly categorized into single compartment and multi-compartment reactors. The early scale-down systems developed were mainly as single compartment bioreactors, either shaped in the form of a tubular closed-loop toroid (Gschwend et al., 1983) or single STRs with pulse-based inputs or internal horizontal discs for increased mixing times (Schilling et al., 1999). Later, the multi-compartment scale-down systems were developed (Figure 2.2), comprising either two connected STRs or an STR connected to a plug flow reactor (Limberg et al., 2016). During operation of single compartment scale-down bioreactor with pulse inputs, the stress inducing agent, usually in the form of a highly concentrated substrate feed or a base or acid is intermittently injected into the bioreactor, at specified intervals (P. Neubauer et al., 1995) or randomly (Sunya et al., 2013). In multi-compartment scale-down bioreactors, one of the compartments is usually a perfectly mixed stirred tank reactor (STR), whereas the other STR or PFR is used to induce the required gradients. The culture is circulated between the perfectly mixed zone and the stress inducing zone, at a rate equivalent to a specified residence time. The stress inducing agent is injected into the heterogenous parts (PFR or 2nd STR), from where it is eventually mixed with the bulk of the culture in the other sections. These operation mechanisms of both the single compartment and multi-compartment scale-down systems produce zones that are similar to feeding and starvation zones in large-scale bioreactors. The resulting periodic exposure of the culture to varying stresses produces stress responses that are also observed in large-scale bioreactors (Enfors et al., 2001). Scale-down techniques have been applied for the successful study of the impact of large-scale gradients for most industrially relevant organisms (Neubauer and Junne, 2016b). Recent improvements in the construction of the plug-flow reactor section include the use of static mixing elements, which prevent back mixing and helps to achieve a higher degree of plug flow behaviour upon aeration than previous hollow tube versions (Junne et al., 2011). Additionally, Lemoine et al used a scale-down bioreactor consisting of one STR connected to two PFRs to simultaneously study the influence of excess substrate and oxygen limitation on the metabolic behaviour of *Corynebacterium glutamicum*, which is the first 3-compartment reactor involving two plug-flow reactors and an STR reported in the literature (Lemoine et al., 2016). Their results show that in the three-compartment bioreactor which combines two types of gradients, the extracellular accumulation of short-chain carbonic acids (lactate and succinate) was higher, with a resulting lower lysine yield than in cultivations where a two-compartment scale-down bioreactor was used. Limberg et al. compared the performance of STR-STR and STR-PFR scale-down bioreactors during the study of the responses of L-lysine producing *Corynebacterium glutamicum* to oscillating glucose and dissolved oxygen concentrations (Limberg et al., 2016). Their results showed similar profiles for

product and biomass yields in the two scale-down systems, which were both significantly lower than the yields under homogenous cultivation conditions.

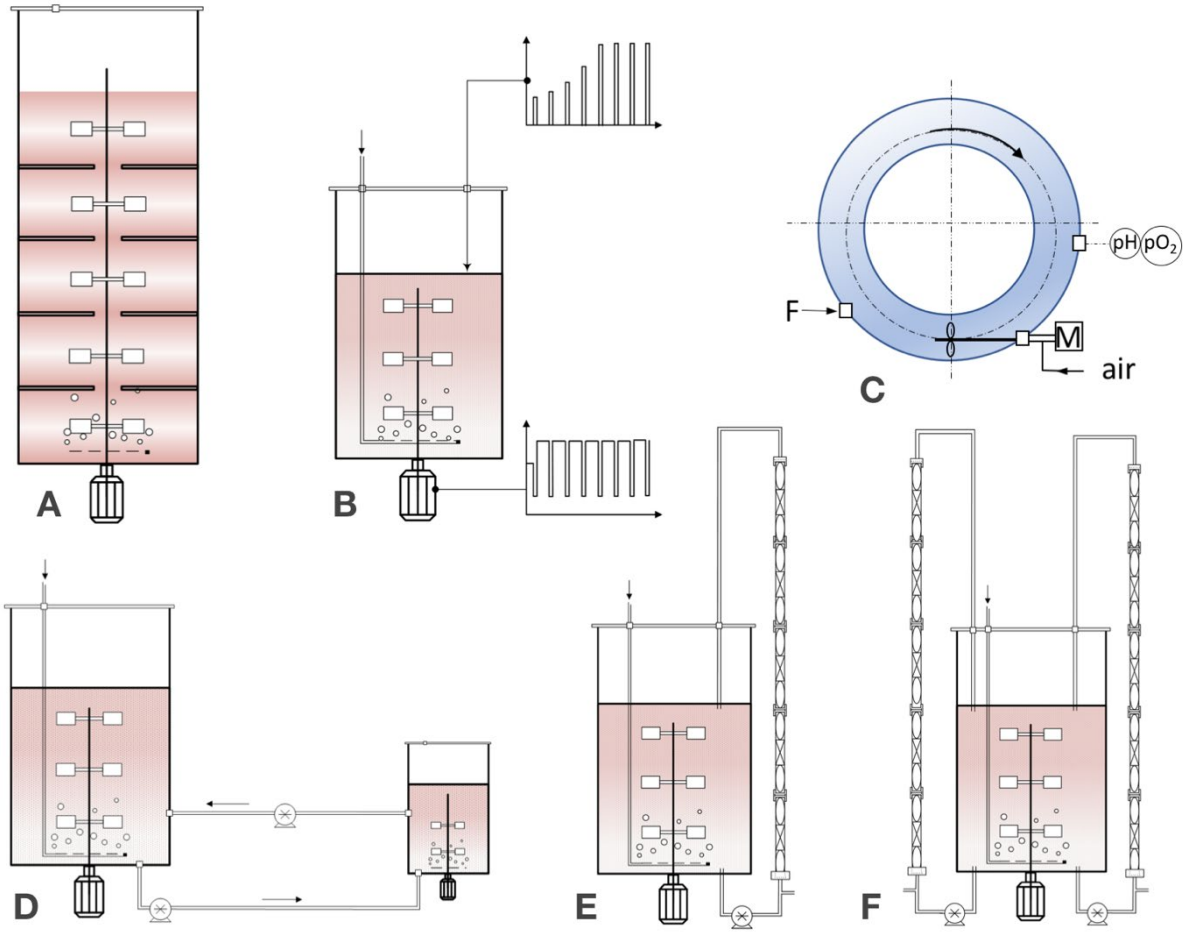


Figure 2.2 Different forms of scale-down bioreactors (A) single compartment bioreactor containing internal discs to increase the overall mixing time (Schilling et al., 1999); (B) Single compartment bioreactor operated with pulse-based inputs to induce scale-up effects; (C) Single compartment toroidal bioreactor to simulate feed addition zones of larger bioreactors (Gschwend et al., 1983); (D)Two-compartment scale-down bioreactor consisting of two stirred tank bioreactors and recirculation pumps; (E) Two-compartment scale-down bioreactor consisting of a stirred tank reactor and a plug flow reactor and (F) Three-compartment scale-down bioreactor comprising two plug-flow bioreactors and a stirred tank bioreactor. Adapted from (Neubauer and Junne, 2016a).

2.2.2. Evolution of Scale-down Bioreactors: Design and Operation

Since the first use of scale-down bioreactors in the mid 1980s, several advancements in technology and our understanding of microbial cultivation processes have shaped how this important class of bioreactors have evolved. The key requirement of scale-down bioreactors is their ability to accurately mimic the heterogeneity of large-scale bioreactors. However, if the conditions in the large-scale are not exactly known, then the scale-down representation may also not be accurate enough or may lead to observations that are not seen in the larger scale (Bylund et al., 2000). Unfortunately, there has been a distinct paucity of information on industrial scale bioreactors in the literature, partly due to proprietary operational conditions in industrial processes and the existence of wide gaps between

research and production facilities in most companies. In spite of these challenges, significant progress has been made in the past three decades on the development of scale-down bioreactors. As summarized in the following sections, some of these advancements concern a better description of the heterogeneity in the larger scale, whereas others span around the use of process analytical technology (PAT) tools (sensors and sampling systems) to better understand the heterogeneity of microbial cultures under oscillating environmental conditions.

2.2.2.1. Computational Fluid Dynamics and Cellular Reaction Dynamics (CFD-CRD) Modelling

As stated earlier, one of the difficulties in the design and operation of scale-down bioreactors is the lack of data on the heterogeneities in large-scale bioreactors (Formenti et al., 2014). This is further attributed to the lack of spatially distributed sensors, as well as the rigidity of industrial scale cultivations/bioreactors in terms of reactor modifications and retrofitting that would enable the measurement of gradient profiles. A laudable solution to this challenge was devised in the mid 1990s when computational fluid dynamics models were used to develop simulations of the mixing patterns in larger stirred tank vessels with known geometric configurations. Measurements from a 30000 L bioreactor were used to validate the simulated mixing patterns in the CFD models (see Neubauer and Junne, 2016a for summary). The results of this work led to a further improvement of the multi-compartment scale-down bioreactor, which was used for further applications to study effects of concentration gradients in *E. coli* and *S. cerevisiae* cultivations (Bylund et al., 1998; George et al., 1993; P. Neubauer et al., 1995). However, despite the success of the CFD modelling, there still existed a gap between the complex hydrodynamic models formulated from the physico-chemical properties of the broth and the biological activity of the cells being cultivated. Recently, the integration of cellular reaction dynamics (CRD), in the form of growth kinetics, substrate uptake rates and product formation rates into CFD models has helped to narrow this gap (Delvigne et al., 2017b; Wang et al., 2015). Furthermore, development in metabolic flux balance models and their coupling to CFD models of large-scale bioreactors have also been realised (Haringa et al., 2018; Wang et al., 2018). Therefore, accurate simulation of mixing patterns and concentration fields in large bioreactors, combined with the dynamics of the cellular response to the simulated gradients allows a better description of the response of the strain to the scale-up stresses (Lapin et al., 2004). Haringa and co-workers used this approach to determine the frequencies and magnitude of substrate gradients that should be achieved in a scale-down version of a 22000 L bioreactor of *S. cerevisiae* (Haringa et al., 2017). The same authors used the CFD-CRD approach to scale down a 54000 L production scale bioreactor of *Penicillium chrysogenum* fermentation (Haringa et al., 2016). In both studies, they used the CFD-CRD tools to determine the response kinetics of the strains, which they termed organism lifelines, in the presence of the heterogeneities. Based on the organism lifelines, new and improved proposals were made to better design scale-down bioreactors for the studied systems. In their most recent study, a 9-pool metabolic flux model of *P. chrysogenum* was coupled with the CFD model of the 54000 L bioreactor. The model simulations were then used to design a 3 L single compartment scale-down bioreactor to study the effect of glucose gradients on product formation rate in *P. chrysogenum*, with the result that the q_P in the scale-down bioreactor dropped by 50 %, as predicted by the model (Haringa et al., 2018). Most interestingly, this drop in product formation rate was significantly reduced when the CFD simulations were used to optimize the glucose feed addition points in the larger bioreactor. Although a lot of simplifying assumptions about the rheology of the broth were made in this study, the results are very promising and open up a new dimension in scale-down studies. Further practical applications

of the proposed methods of Haringa and co-workers based on these CFD-CRD models can be extended to other microorganisms and other reactor geometries, for better scale-down studies.

2.2.2.2. Population Balance Modelling (PBM) and Advanced PAT Tools

The evolution of population subsets in response to induced stresses during scale-down investigations was not taken into account until the previous decade (Delvigne et al., 2017b; Gernaey and Gani, 2010). In multi-compartment scale-down bioreactors, only a fraction of the culture is exposed to the stress-inducing agent at any given time. This fraction is determined by the ratio of the STR to PFR volumes (Wang et al., 2018). In single-compartment pulse-based systems, the response of the culture to the induced stresses is synchronized, without the existence of population sub-sets. However, within the response time of the pulses, the evolution of the population in response to the stresses represents a special type of population heterogeneity that can be followed with advanced PAT tools, such as high resolution online microscopy (A.-M. Marbà-Ardébol et al., 2018) or automated flow cytometry (Delvigne and Goffin, 2014). Therefore, since these population effects also occur in large industrial-scale bioreactors, the concept of population balance modelling was developed to take into account the probabilities of stress responses in various population groups. Morchain and co-workers coupled a PBM with CFD-CRD models in a 70 L scale-down bioreactor to study mixing effects in a 70000 L bioreactor. These techniques were used to identify zone-specific growth rates and localized substrate uptake rates within the larger bioreactor, for stratified population subsets at the single cell level. They concluded that the various local disequilibria between zones and population groups contribute significantly to the poor performance in the larger scale bioreactor (Morchain et al., 2014). The two recent reviews by Gonzalez-Cabaleiro et al. and Delvigne et al. give important bases of the mathematical formulations to describe microbial populations and the tools available to measure/analyse the existence of sub-groups (Delvigne et al., 2017a; González-Cabaleiro et al., 2017). Importantly, the second of these reviews (Delvigne et al., 2017a) highlights the need to separate biological and measurement noises from actual stratifications in the population population response, which, when coupled, can lead to wrong conclusion from PBMs. Further advances in research in the area of PBM within fermentation have considered models of single cell responses to heterogeneities experienced by population subgroups (F Delvigne et al., 2006), stochastic model description of large-scale conditions (F. Delvigne et al., 2006) and a systematic comparison of microbial heterogeneity and bioreactor heterogeneity (Baert et al., 2016), all of which help to elucidate the cell-to-cell variability within the heterogeneous environments in industrial scale bioreactors.

2.2.2.3. Rapid Sampling and 'omics' Technologies

One of the major challenges in the operation of early scale-down bioreactors, and bioreactors in general, is the inability to sample cells at the right physiological state in a multi-compartment or pulse-based system (van Gulik et al., 2013). Accurate sampling within the gradient field of scale-down bioreactors is necessary to capture the responses of strains to the specific stresses being investigated. The introduction of static mixers and spatial sampling points along the length of the PFR module in multi-compartment scale-down bioreactors which have this component greatly improved the sampling domain of the reactor (Junne et al., 2011). A further improvement of the sampling domain was realised in the BioScope, a tubular bypass of the bioreactor that is used for rapid sample withdrawal (van Gulik, 2010). Although this system is technically not a scale-down bioreactor, it was able to mimic some stress

induction zones, such as oxygen depletion (Mey et al., 2010) with residence times of up to 70 seconds during which the culture was exposed to the oxygen limiting conditions. This device and other rapid sampling techniques for general bioreactors (Schädel and Franco-Lara, 2009) allowed the elucidation of *in vivo* kinetic responses of cells to large-scale heterogeneities (Lara et al., 2009; Visser et al., 2002). The rapid sampling also enables the analysis of intracellular metabolite pools by metabolomics techniques, which can be used to detect cellular responses that are not observable in the extracellular space. A typical example in this regard is the work of Carnicer et al. who used quantitative metabolomics to study the effect of dissolved oxygen gradients on free amino acid pools in *P. pastoris* expressing a recombinant protein (Carnicer et al., 2012). Their results show that oxygen limitation in *P. pastoris* leads to a rapid build-up of intracellular free amino acid pools, which, ironically, results in low recombinant protein production rates. Therefore, engineering such strains to free up these pools will significantly improve recombinant product yields. Advanced proteomics and gene expression analysis has enabled complete proteome-wide evaluation of the effects of concentration gradients induced in scale-down bioreactors, as reported in the works of Brognaux et al., 2014; Alvaro R. Lara et al., 2006; Lara et al., 2009 and Simen et al., 2017). The application of these *omics* techniques (metabolomics, transcriptomics, proteomics) to detect the up- and down-regulation of certain genes under stressful cultivation conditions in scale-down bioreactors and the associated elucidation of molecular level response to stresses can help in the genetic engineering of more robust strains to cope with scale-up effects. An example of this application is demonstrated in the work of Schwentner and co-workers, who used metabolomics to improve gene regulatioin in L-valine producing *C. glutamicum* (Schwentner et al., 2018). Once again, the accuracy of the *omics* technologies and the intracellular trends they give depend on the sampling technology employed (van Gulik, 2010). Therefore, to realise the full benefits of this technology in scale-down cultivations, the sample withdrawal and rapid quenching methods developed for general metabolomics need to be incorporated into scale-down bioreactors. The rapid flow-through sampling device with embedded cold quenching developed by Lameiras and co-workers (Lameiras et al., 2015) is typical example that would be suitable for incorporation into multi-compartment scale-down bioreactors.

2.3. Mathematical Models in Bioprocess Development

Mathematical models are becoming increasingly useful in bioprocess development due to advancements in computing power and their ability to help design more informative experiments (Łącki, 2014). Fundamentally, these models can supplement empirical knowledge and give the experimenter a predictive tool to pre-assess the impact of various operational (process) conditions without carrying out actual experiments (Ashyraliyev et al., 2009). Additionally, process variables that are difficult to measure either due to the unavailability of analytical tools or due to the rigidity of the process can be estimated using mathematical models (Herwig et al., 2015; Mears et al., 2016). Thus, process knowledge, which is an essential requirement in the PAT (Process Analytical Technology) and QbD (Quality-by-Design) paradigms, can be generated with mathematical models in a cost-effective manner (Gernaey and Gani, 2010; Junker and Wang, 2006). The use of models in bioprocess development dates back to several decades, but their use has been largely restricted to the study of specific physiological behaviour in fermentation and not so much on process-wide applications (Almquist et al., 2014). For instance, in *E. coli* fed-batch processes, models have been used to

satisfactorily describe biomass growth under limiting substrate conditions (Cockshott and Bogle, 1999), overflow metabolism (Bo Xu et al., 1999) and the mechanisms of inducer transport and inclusion body formation (Calleja et al., 2016; Wang et al., 2001) for specific cases of recombinant protein production.

2.3.1. Classification of Mathematical Models in Bioprocess Engineering

In general, two broad categories of mathematical models are identified in Biochemical Engineering: data-driven (empirical) models and knowledge-based (mechanistic) models. Data-driven models mostly arise from industry, considering the fact that historical data of a particular process that may have been running for decades are usually available. Such data can contain a wealth of information about the process, and tools such as artificial neural networks (ANN) have been developed to extract important information from these data historians. These models are usually trained on the available data based on input-output relations, and used subsequently to predict some future states of the process. On the other hand, a mechanistic model consists of mass balance equations and specific (kinetic) rates of actual biological events which approximate the mechanism of interactions of both extracellular and intracellular components (Almquist et al., 2014). Cell growth and recombinant protein production are underpinned by complex metabolic and physiological processes, as well as the extracellular environment. Mechanistic models attempt to represent these processes by using lumped parameters that characterize the combined effects of intracellular enzymes and transport mechanisms in the cell (Hoffmann et al., 2001). Further classification of the most common mechanistic models from the literature are summarized in Figure 2.3 (Gernaey et al., 2010).

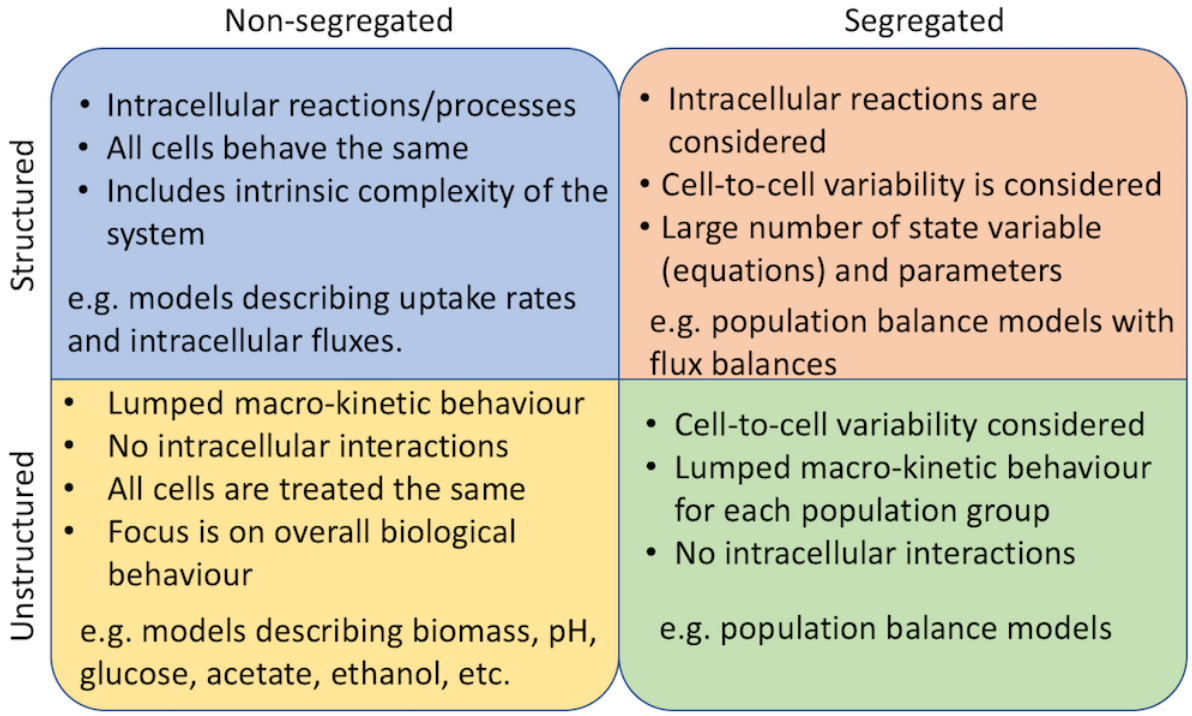


Figure 2.3 Classification and properties of mechanistic models in bioprocess engineering (Gernaey et al., 2010)

Recently, the development of hybrid models has emerged as a promising branch of modelling in biochemical engineering. Hybrid models combine aspects of mechanistic models with data-driven

models. They are particularly useful where some parts of the underlying biological mechanisms of a process are not properly understood, and data is available for ANN training. For instance, in modelling the anaerobic growth of *E. coli*, Kim and co-workers combined mechanistic description of intracellular flux balances with cybernetic networks of external fluxes to obtain a hybrid model that had a better prediction power than either of its component models (Kim et al., 2008). Hybrid models may see more applications in the modelling of recombinant product quality in both microbial and mammalian cultivations, which is usually a multivariate response parameter without clear mechanistic links to process variables. The rest of the discussion in this thesis is restricted to mechanistic models only.

2.3.2. Identifiability, Uncertainty and Sensitivity Analysis of Mechanistic Models

Generally, mechanistic models of biological systems are built with large numbers of parameters due to the complexity of the physiological and metabolic processes they describe. Unfortunately, in the model building process, less consideration is given to the numerical robustness of the model and the mathematical implications of a particular model structure, which usually leads to overparametrization and parameter correlations. This problem is further complicated by the complexity and practical limitations in carrying out biological experiments. The consequence is that during model fitting, the parameter Estimation (PE) can become ill-conditioned, such that it may not be possible to assign unique numerical values to all of the parameters, a condition called non-identifiability (Raue et al., 2009; Saccomani, 2013). The non-identifiable parameters tend to have wide confidence intervals (parameter estimation errors), and there exist various parameter combinations that give the same model fitting (root mean square error of residuals).

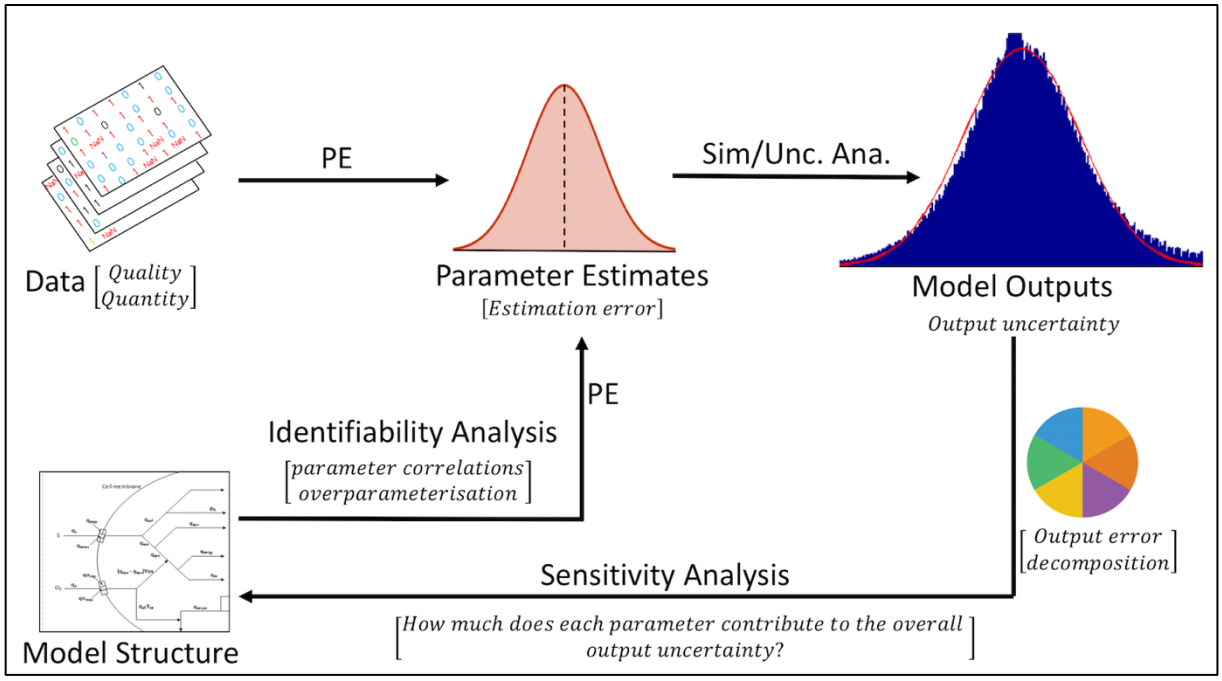


Figure 2. 4 Relationship between data quality, model structure and the uncertainty of model predictions. Also shown are the propagation routes of the error associated with measurements and model structure through the model to the outputs, with the associated model analysis tools for each step. PE—parameter estimation, Sim/Unc. Ana—simulations and output uncertainty analysis.

Identifiability analysis comprises a set of methods to test whether there are inherent problems in the model structure (structural identifiability) and if all the parameters can be uniquely estimated from the given experimental data (practical identifiability). Such analysis is useful in providing information on the amount of data required, the number of experimental replicates and the statistical reliability of the predicted profiles (Sin et al., 2009). Experimental measurements are subject to a number of errors e.g. human error, equipment error, batch to batch differences, etc., which are altogether summarized in the standard deviation of (usually) triplicate measurements of each sampling point. The use of such data to estimate the parameters of a mechanistic model invariably imposes some level of uncertainty (error) on the parameter estimates. Secondly, when the PE problem is ill-conditioned as explained above, the estimated parameters also have a large error (variance) (López C. et al., 2015). It is of importance to determine how the error in the parameter estimates (from these two sources) affect model predictions when the estimates are used to run simulations. Uncertainty analysis is the technique used to determine how the error in the parameter estimates translates into model prediction errors Figure 2. 4. This analysis underpins the reliability and acceptance of a given model for a particular use. Sensitivity analysis, on the other hand, is a similar but reverse concept. Once the uncertainty (or some other model property) in the model outputs is quantified, sensitivity analysis deals with apportioning the output variance into its sources, to quantitatively determine the contribution of each parameter to the model property, in this case the output variance (Saltelli et al., 2004).

The concepts of identifiability, uncertainty and sensitivity analyses have been applied extensively in mechanistic models in physical sciences and engineering (Vajda et al., 1989), but their adoption in models of bioprocess engineering is limited, although model uncertainty has been recognized as a big challenge in biological models (Miskovic et al., 2015). Recently, as part of the so-called Good Modelling Practices (GMoP) guidelines for biotechnology (Sin et al., 2009), these concepts have been promoted, with the emergence of pre-packed software applications (e.g. DAISY) to facilitate advanced analysis of mechanistic models in biology and bioprocess engineering (Bellu et al., 2007; Raue et al., 2009; Villaverde et al., 2015). In the past decade, a number of published models (flux balance models, macro-kinetic models, signal transduction models) of biological systems (*E. coli*, *S. cerevisiae*, Chinese Hamster Ovary cells, *D. melanogaster*, etc) and bioprocesses have been analysed for the identifiability and uncertainty of their parameters (López C. et al., 2013; Quaiser and Mönnigmann, 2009; Sin et al., 2010; Villaverde et al., 2015). A common conclusion from all the models analysed is the general lack of identifiability for biological models, such that even under highly idealized experimental conditions (measuring all state variables at high frequency), some model parameters remain non-identifiable. However, as pointed out by Sin and co-workers in the analysis of the NREL cellulosic fermentation model, once the amount of uncertainty associated with the non-identifiability is known, the model may still be used, but only for certain purposes. In such cases, care must be taken to always consider the level of model reliability that is available to the user. Alternatively, non-identifiable models with large uncertainties can be regularized to reduce the uncertainty in the parameters by incorporating prior knowledge of the system into the parameter estimation problem. The current challenge of the second option is that the mathematical methods available for such regularization (see review by (Chis et al., 2011)) are quite elaborate, and are sometimes beyond the mathematical training of most biotechnologists and bioprocess engineers (Muñoz-Tamayo et al., 2018).

2.3.3. Mechanistic Models of *E. coli*

Undoubtedly, the *E. coli* expression system continues to dominate in bacterial expression of recombinant proteins (Chen, 2012), and is poised to grow further as advances are made in systems biology and computer-based process development (Wang et al., 2009). Of the vast majority of mechanistic models in the literature describing *E. coli* physiology, the following specific mechanistic descriptions are the most common fermentation phenomena encountered in the development of industrial *E. coli* processes.

2.3.3.1. Models for Substrate Uptake in E. coli

The simplest model describing substrate consumption in *E. coli* fermentation is that of Monod kinetics. Although this model is purely empirical, it is versatile in predicting substrate consumption rates in fermentation processes. The maximum specific growth rate has been shown to depend directly on the specific substrate uptake rate, therefore under slow growth conditions, the substrate uptake rate is reduced significantly (Lin et al., 2001b). An extensive mechanistic model describing internal partitioning and usage of glucose was developed by Insel et al (Insel et al., 2007). Depending on the carbon source used in fermentation, an inhibitory term is usually added to the Monod equation to describe inhibitory effects of substrate, product or metabolites (acetate) on substrate uptake (Cockshott and Bogle, 1999). Due to favourable growth conditions during fed-batch culture, cells are assumed to grow in exponential phase, with negligible specific death rate (k_d). This factor is therefore excluded from most *E. coli* models (Horowitz et al., 2010).

2.3.3.2. Models for Overflow Metabolism in E. coli

Perhaps one of the most prominent physiological concepts in *E. coli* fermentation is overflow metabolism and the closely associated mixed-acid fermentation which lead to the excretion of acetate, formate, ethanol and sometimes lactate into the fermentation broth. The re-direction of the carbon source to these compounds, and their excretion into the extracellular broth is considered wasteful in recombinant protein production processes using *E. coli* (De Mey et al., 2007; Valgepea et al., 2011). Recent research using advanced proteomic analysis (Basan et al., 2015; Peebo et al., 2015) and systems biology approaches (Valgepea et al., 2010) show that overflow metabolism is an attempt by fast-growing cells (*E. coli*) to offset an imbalance between high proteomic demands for energy production and biomass synthesis. The underlying metabolic routes to maintain this biomass-energy balance are complex, and modelling such a system can be a daunting task (Kremling et al., 2014). Nevertheless, models that simply assumed saturation of TCA cycle enzymes and one-directional production of acetate, such as that of Xu and co-workers (Luli and Strohl, 1990; Bo Xu et al., 1999) sufficed in describing acetate profiles during *E. coli* fermentation. Other examples of expressions for specific rates of acetate production in various models in the literature are given in Table 1, with an explanation of the underlying assumptions made in each case. Insel et al. presented a mechanistic model of overflow where the production of acetate was defined as a constant fraction of $Y_{x/s}$, and acetate was only consumed after glucose depletion (Insel et al., 2007). However, as shown by Peebo et al (Peebo et al., 2015) and Valgepea et al (Valgepea et al., 2010), intracellular production and re-assimilation of acetate is a continuous process in *E. coli* metabolism, even in the presence of glucose. Overflow metabolism only occurs when the internal rate of acetate production significantly exceeds the rate of its re-

assimilation due to carbon catabolite repression at higher specific growth rates or under anoxic conditions (Basan et al., 2015). The absence of a diauxic delay during the transition from glucose to acetate in batch cultures of *E. coli* suggests that acetate is consumed in the background along with glucose so the cell does not need time to activate acetate consumption pathways upon glucose exhaustion (Bernal et al., 2016; Enjalbert et al., 2015). Surprisingly, only a few *E. coli* models in the literature consider acetate production as an inherent part of *E. coli* metabolism. As shown in Table 1, the few models that account for acetate production use pure empirical relations to describe the specific rates of acetate production when certain pre-set conditions triggering overflow metabolism are met

Table 1 Specific kinetic (rate) expressions for acetate production in different models of recombinant *E. coli* fermentation.

Specific acetate production rate	Assumptions/Notes	Ref
$q_{Ap} = qs_{of,en} Y_{A/S}$ $q_{Ac} = \frac{q_{Ac,\max} A}{A + K_A}$	Production (q_{Ap}) and consumption (q_{Ac}) of acetate are both represented, although the two processes do not occur simultaneously. The accumulated acetate according to q_{AP} is consumed later through q_{Ac}.	Xu et al. (2000)
$q_{ap} = \alpha Y_{x/s}$	Acetate production is a constant fraction of the substrate conversion efficiency, $Y_{x/s}$.	Insel et al. (2007)
$q_{ap} = r_A$	q_{Ap} is a lumped parameter that is estimated directly from experimental data. No mechanistic derivation of the specific rate of acetate production.	Cockshott et al. (1999)
$q_{ap} = \nu = 0.672 K_1 \left(\mu - 3.571 \frac{K_1}{K_2} \right)$ $K_1 = f(\text{Carbon flux to EMP}, X)$ $K_2 = f(X)$	q_{Ap} is determined by the specific growth rate, the fraction of carbon flux to the EMP pathway and the amount of biomass at any given time. The functional relationship between q_{Ap} and m, X and C-flux is represented by explicit algebraic equations.	Ko et al, (1994)

2.3.3.3. Models for Protein Induction and Intracellular Product Formation

The inception of recombinant product formation in fed-batch culture of *E. coli* depends on the promoter driving the expression of the foreign protein. Although recombinant protein production in *E. coli* can be induced by physical means (e.g. temperature and pH shocks, see review by Choi et al (Choi et al., 2006)), the more developed and widely used *tac*-based promoters (*lac, trc, tac*) use the lactose analog isopropyl1-b-D-thiogalactopyranoside (IPTG) for protein induction. A few models that describe the dependence of the product profile on the induction conditions exist in the literature. The

most important parameter that seems to influence the specific rate of product formation (q_p) in almost all modelled cases is the inducer-to-biomass (I/X_{ind}) ratio at the time of induction (Lee and Ramirez, 1992; Ruiz et al., 2011). This follows the hypothesis that the probability of the inducer to bind to the repressor is enhanced by higher intracellular IPTG concentrations (Calleja et al., 2014) which are achieved by higher I/X_{ind} ratios. Conversely, very high intracellular inducer concentrations may lead to increased metabolic burden on the culture resulting in low product formation rates (Donovan et al., 1996; Ruiz et al., 2013). Thus, the I/X_{ind} ratio plays an important role in determining optimal product formation profiles. For example, Calleja et al modelled the product profile based on the inducer (IPTG) concentration and the biomass at the time of induction and found a sigmoidal dependence of the product profile on the non-metabolized inducer and biomass concentrations (Calleja et al., 2016). Mahadevan and Doyle used extensive parametric models and an on-line optimization technique to determine the optimal concentration of metabolized inducer (arabinose) that would maximise the concentration of recombinant chloramphenicol acetyltransferase in fed-batch cultures of *E. coli* (Mahadevan and Doyle, 2003). Other factors considered in mechanistic models of recombinant product formation in *E. coli* include the time to start protein induction (Levisauskas et al., 2003) and the duration of the induction phase (Fahnert et al., 2004). A few mechanistic models that consider intracellular plasmid copy numbers and mRNA formation/degradation rates are also reported in the literature(Calleja et al., 2014; Kramer et al., 2015; Ruiz et al., 2011). However, these haven't been exploited much in process development due to the lack of data and lengthy analytical methods associated with measuring the intracellular variables.

2.3.3.4. Models Describing the Metabolic Burden Induced by Foreign Product Formation

The growth of *E. coli* under non-production cultivation conditions is significantly different from growth under recombinant protein production conditions due to the so-called metabolic burden induced by the foreign protein (Carneiro et al., 2013; Neubauer et al., 2003). The metabolic burden can be attributed to the re-direction of cellular resources towards foreign protein production and synthesis and intracellular transport of non-native amino acids (Hoffmann and Rinas, 2004). Earlier studies of host-plasmid interactions delved into the quanitification of the effects of foreign plasmids on microbial growth (Bailey et al., 1986; Seo and Bailey, 1985). In terms of mechanistic models, three approaches have been identified on the development of model equations to describe metabolic load in *E. coli*:

(i) Growth inhibition by inducer concentration: such models introduce an inhibitory term into the substrate uptake rate (q_s) equation where the presence of the inducer is assumed to inhibit the uptake of the carbon source in a non-competitive way. An example is the model of Kramer et al who used a distinctive differential equation to model inhibitory effects of the inducer (Kramer et al., 2015). In this model, the level of growth inhibition depends on the inducer concentration.

(ii) Use of dedicated decay terms: such models introduce a decay constant by which the specific growth rate declines upon addition of the inducer. Here, the specific substrate uptake rate is unaffected, but cellular resources are channelled to product formation, resulting in a lower specific growth rate. The model of Ruiz et al uses a specific fraction of the growth rate and an apparent biomass yield to describe the effect of protein production on cell growth (Ruiz et al., 2011).

(iii) Re-estimation of parameter values: in this approach, the model structure remains the same but parameter values are estimated separately under production and non-production conditions. The

basic premise is that the metabolic burden is a broader concept, and affects more parameters than just the specific growth rate or the substrate uptake rate. Thus, a re-estimation of model parameters to determine the inhibition on all uptake capacities allows quantification of the collective effects of protein production on cell growth. The model of Neubauer *et al*. uses this approach to estimate the decrease in substrate and oxygen uptake capacities upon induction, as well as the resulting decrease in specific growth rate and yield coefficients (Neubauer et al., 2003). Furthermore, this model considered the segregation of plasmid-containing and plasmid-free cell populations, with clear differences between the parameter values estimated for the two population sub-groups.

2.4. High Throughput Bioprocess Development

Bioprocess development is labour intensive, requiring several experiments that may take several months, to characterise a particular process. Across both industry and academia, bioprocess development (after screening in microtiter plates and shake flasks) is mostly carried out in laboratory-scale bioreactors, usually ranging from 1 to 20 L bench top bioreactors which are run in several one-at-a-time batches to characterise a process (Neubauer et al., 2013). The slow pace of development in such settings and the need to shorten the time-to-market of important biopharmaceuticals have driven the widespread adoption of parallel, minibioreactor systems over the past decade (Bareither and Pollard, 2011; Hemmerich et al., 2018a; Łącki, 2014). The use of parallel cultivations in minibioreactors does not only save time in development (Neubauer, 2011), but also simultaneously enables the efficient incorporation of design of experiments (DoE) principles (Janakiraman et al., 2015; Neubauer et al., 2017), reduces the cost of materials such as media, ensures the testing of process reproducibility (Rameez et al., 2014) and helps to generate enough data to support cellular function models (Cruz Bournazou et al., 2017; Nickel et al., 2017). These high throughput minibioreactors range from 3mL to about 15 mL stirred, single-use vessels that are gamma pre-sterilized. The operation of minibioreactor systems as shaken cultures in well plates and smaller vessels, in the range from 100 µL to 2 mL are also reported (Back et al., 2016), but these are not covered in this thesis due to practical limitations of sampling such micro-scale reactors for detailed offline analysis of metabolites and recombinant proteins. The most recent advances in using high throughput minibioreactors and an overview of existing technologies for bioprocess development are summarized in the following sections.

2.4.1. Monitoring and Control in Minibioreactors

The usefulness of a bioreactor for bioprocess development largely depends on what can be measured, and sometimes when these measurements are taken (online, offline or at-line). Almost all minibioreactor modules on the market today are equipped with optodes: optical sensors that measure back-scattered light or some form of fluorescence. This measuring technique is successfully applied for the online measurement of pH and dissolved oxygen. However, in aerated systems, gas bubbles interfere with the scattering of light and therefore such measurements are not yet applicable for the measurement of biomass in aerated systems (Rameez et al., 2014). However, other non-invasive techniques such as multi-wave fluorescence spectroscopy and Fourier transform infra-red spectroscopy (FTIR), as well as Raman spectroscopy (Kögler et al., 2018) have been reported as efficient

methods for online measurement of metabolites and biomass, although these methods are still in the early stages of development. In terms of control, some of the minibioreactors (e.g. bioREACTOR from 2mag Ag) are incorporated into robotic liquid handling stations which can be used to administer both acid/base for pH control and for feeding. Lately, the use of micropumps for continuous feeding and pH control in minibioreactors and microtiter plates is reported (Ude et al., 2015). These advancements are promising, as they bring the minibioreactor platforms closer to the level of control achieved in actual bench-top bioreactors, which would enable complete automation and digitalization of minibioreactor cultivation, as discussed by Neubauer and co-workers (Neubauer et al., 2013).

2.4.2. Automation: Liquid Handling Stations (LHS)

Apart from offering a platform for control of traditional parameters such as pH, the incorporation of minibioreactors into liquid handling stations allows the automation of cultivations in these systems. For instance, the use of liquid handling stations that can read a worklist generated from programming codes infinitely expands the type of manipulations that can be done during cultivations in minibioreactors. Repetitive tasks, as well as pipetting of micro-volumes that would otherwise be masked with human error are efficiently carried out with liquid handling robots (Knepper et al., 2014). Additionally, the expansion of minibioreactors to include LHS allows the at-line analysis of metabolites, e.g. with enzymatic assays. With the proper communication interphase between the at-line measuring device and the LHS, further cultivation decisions can be taken based on process measurements. Furthermore, when such a platform is interphased with programmable languages such as Matlab or Python, intelligent experiments can be carried out, where the addition of substrates and media components will be decided by the current state and future predictions of the culture. When this concept is implemented in an adaptive framework, a lot of information can be obtained from one parallel run of minibioreactors, as demonstrated by Barz and co-workers (Barz et al., 2018). Recently, Cruz-Bournazou et al used such an interphased high throughput platform to fit and fully characterise a macro-kinetic growth model of wild type *E. coli* in a timespan of 6 hours (Cruz Bournazou et al., 2017).

2.4.3. High Throughput Downstream Processing (DSP)

Consider a 10-hour microbial cultivation in 24 parallel 15 mL minibioreactors under the control of a robotic liquid handling station which takes a sample from each minibioreactor on hourly basis. Further, suppose each of these samples is to be analysed for the concentration of residual glucose, biomass, acetate, formate and ethanol. This simple scenario leads up to 1200 analysis steps during the 10-hour cultivation. Such practical experiences are quite common in the handling of high throughput minibioreactor systems, and therefore there is the need to incorporate high throughput downstream processing devices and equipment into the whole platform. The lack of high throughput DSP in such cultivations only transfers the bottleneck of bioprocessing from the cultivation stage to the analysis phase. Although the examples given above (glucose, acetate, ethanol, biomass) can be easily analysed with spectrophotometric techniques which are also on the order of high throughput systems, real challenges remain with the quantification of intracellular metabolites, protein analysis by SDS-PAGE, cell lysis, recombinant product quantification (in cases of non-fluorescing product) and amino acid analysis which are still done in the traditional way. Although there are high throughput protein analysis devices, such as the LabChip GX II touch (PerkinElmer) which is capable of analysing up to 480 samples

in a single run on capillary gel electrophoresis chips, these are not easy to incorporate into existing configuration of minibioreactor systems. Additionally, pre-analysis steps such as cell lysis and lengthy sample preparation steps may hamper their use in robotic stations, but it is not an impossibility. Other high throughput devices such as the Cedex Bio HT are suitable for high throughput metabolite analysis using enzymatic assays. In the past few years, much progress has been made in the field of HT DSP and its incorporation into HT cultivation systems (Bhambure et al., 2011). For instance, Baumann et al combined HT chromatographic modelling with micro-scale cultivations in an integrated process development platform for faster characterization (Baumann et al., 2015). Another example in the use of HT systems for DSP was reported by Florian and co-workers, who used a DoE approach to develop a lysis buffer for disrupting microbial cells prior to inclusion body purification (Glauche et al., 2017). For a full review on HT applications in DSP, see (Baumann and Hubbuch, 2017).

Florian's paper on cell lysis

2.4.4. Advanced Data Processing

As in the case of sample processing, HT cultivation platforms can easily generate a lot of unexpected numerical data from a single cultivation. After a single run in a parallel HT system, the data content can easily be 100 fold more than data obtained from a bench-top bioreactor. Therefore special and more advanced tools are required for the detailed analysis of this data. If the data is generated from different equipment (e.g. cultivation platform for online data, analysis platform for offline data), it may be necessary to synchronize all the data in a single database with the correct timestamps. Accurate timestamp labelling can be challenging and tricky if samples are to be transferred manually to other peripheral equipment for analysis. The manual treatment of raw HT data can easily get out of hand, or it can take a long time to process the data and make meaningful visualizations and conclusions from the data, which tends to defeat the fast-track purpose of running HT systems. As stated earlier, without HT automated data processing, the bottleneck in bioprocessing is transferred further to data interpretation, even after fast cultivation and analysis phases. It is therefore desirable that in HT systems, the key performance indices (KPIs) of the process can be readily calculated/derived as and when the data is being generated, to help in rapid decision making and process tuning. An example of upstream process optimisation in HT systems during a process run was presented by Nickel and co-workers, where online data generation and optimisation were combined to determine the next steps in the feed profile of a microbial cultivation (Nickel et al., 2017).

Automated data processing in HT systems requires model-based approaches, such as soft-sensors that can calculate KPIs (μ , qp, qS, Yx/s, Ypx) using the raw data, whilst the process is running. Only a few reports are available in the literature on the application of model-based data analysis in HT systems. Landner and colleagues used principal component analysis of data from backscattered light from *E. coli* cultures to infer the influence of induction strength on the expression efficiency, in the BioLector HT platform (Ladner et al., 2017). Secondly, Yordanov et al used Bayesian inference methods to automatically characterize genetic components and cell growth, in automated calculation procedures using data generated from high throughput platforms (Yordanov et al., 2014). Although these techniques are still being developed, there is a huge potential to include automated data processing in HT platforms, to prevent delayed decision making due to delayed data analysis techniques.

2.4.5. Scalability of Minibioreactor Systems

There is a huge difference between running cultivations at a smaller scale and conducting scale-down studies. The latter has to do with reproducing some operational conditions of larger bioreactors in a smaller scale. For instance, Janakiraman et al used the volumetric aeration rates (vvm) as a scale-down factor between parallel *ambr15*^{TM} cultivations of CHO cells and a 15,000 L production scale bioreactor (Janakiraman et al., 2015), whereas Velez-Suberbie et al. used the power per unit volume as a scale-down criterion to compare *ambr15f* cultivations of *E. coli* with a 20 L bioreactor cultivation (Velez-Suberbie et al., 2018). The direct implication of the possibility to conduct scale-down experiments in minibioreactors implies that these systems can also be used to do the reverse, i.e. to scale-up a bioprocess. Rameez et al developed and scale-up a CHO cell line process producing a monoclonal antibody from *ambr15* (15 mL) to 200 L stirred tank bioreactor (Rameez et al., 2014). These comparisons are possible due to the recent improvements in minibioreactor systems, in terms of vessel and impeller geometry, mass transfer capabilities and the possibility for advanced control of process variables. Therefore the minibioreactors have been proven to be scalable, and their application can only improve with time.

2.4.6. Current Challenges of Minibioreactor Systems

Despite the wide use of HT systems for bioprocess development, there are a number of challenges that the end-user must overcome to get useful process information from such a system. A few of the most important ones are briefly discussed below.

2.4.6.1. Sampling Volumes and Analytic Capabilities

With total broth volumes of around 15 mL, the maximum sample volume that can be withdrawn from HT cultivations without adversely influencing the concentration profiles of the culture is limited. Practically, even 1 mL sample volume from such a minibioreactor constitutes almost 10% of the total broth. Therefore, only a few hundred microliters can be withdrawn at a time. This even becomes difficult for fast growing cultures (e.g. *E. coli*) that must be sampled every hour for at-line and offline analytics. The limitation in sampling volume directly translates into limitations in analysis. For recombinant protein production processes, one sample volume may be divided into several aliquots for protein analysis, metabolites and biomass. Even if biomass can be measured non-invasively in the near future (FTIR, Raman methods), there are still limitations in sample volumes required for further protein analysis, e.g. inclusion body extraction, amino acid analysis and SDS-PAGE. However, techniques such as capillary electrophoresis (e.g. LabChip GX II Touch, PerkinElmer) which need less than 10 µL of sample volume for protein analysis may come to the rescue of minibioreactor systems.

2.4.6.2. Evaporation and Jumps in Concentrations

Depending on the mechanical construction, some minibioreactor systems are prone to liquid loss through evaporation, especially those that are directly aerated with a sparger like in bench-top bioreactors. Evaporation becomes a concern for mammalian cultivations, which may take several days for one complete run. Typical evaporation losses of up to 40% of liquid volume over 15 day cultivations

are reported in the literature (Wiegmann et al., 2018). Evaporation effects do not only result in loss of volume, but the remaining broth becomes more concentrated. The profiles of biomass and extracellular metabolite concentrations may increase sharply, not due to biological activity, but due to evaporation effects. Such higher concentrations may lead to wrong conclusions about the growth patterns of the culture. Recently, volume correction mechanisms have been devised to compensate for liquid loss due to evaporation by supplementing the culture with fresh medium (Wiegmann et al., 2018). This technique, however, dilutes the culture and makes concentration comparisons between runs difficult.

2.4.6.3. Cost of Consumables

Apart from the initial huge capital expenditure to install fully operational minibioreactor systems, almost all current minibioreactor systems use gamma-sterilized disposable vessels (Hemmerich et al., 2018a) as bioreactors. These vessels usually have pre-labelled optical sensor spots for non-invasive pH and DOT measurements, which prevents them from being re-sterilized as these spots are heat labile. For a research environment, the cost of these disposable single-use vessels may significantly add to the operation cost of the laboratory, compared to the use of sterilisable stainless steel or glass vessels for fermentation development. In industry, the cost issue may end up being a trade-off between time savings and cost of consumables.

2.4.6.4. Communication among Hardware Components

One of the most formidable challenges in adding peripheral equipment (e.g. flow cytometer) to robotic liquid handling stations is the use of proprietary software on each device. The communication interface of minibioreactors, robot stations and these peripheral equipment are vendor-specific, and may require special drivers and reconfiguration of communication protocols to connect them. These additional drivers may be expensive since they are usually customized for each client, and there is no further technical support for such ad-hoc devices in cases of failure. Furthermore, the in-house retrofitting of software-hardware protocols to connect other devices may render the warranty of certain equipment void, which is a special difficulty. As discussed by Hemmerich et al., a multiplexed hardware-software integration solution is needed to achieve fully functional cultivation, analysis and data processing tasks in one high throughput platform (Bareither and Pollard, 2011; Hemmerich et al., 2018a). This is only achievable if, in future, all equipment manufacturers adopt standardization of connection protocols, to facilitate the connection of all equipment to a single backbone irrespective of the vendor, where they can communicate efficiently and easily.

3. Research Questions and Aim of the Project

There is much knowledge of scale-down systems, specifically, the two-compartment scale-down bioreactor as well as the physiology of *E. coli* in the literature. The possibility of integrating two enabling technologies for bioprocess development (parallel minibioreactor systems and mechanistic models) towards high throughput screening under real cultivation conditions is explored in this project by leveraging on the already existing knowledge of scale-down bioreactors. The aim of this work was to develop and test a high throughput scale-down system that would be suitable for studying the response of strains to scale-up effects at the screening phase. This is achievable by the combination of the two knowledge-bases (HTPD and Scale-down) into one mechanistic model framework that can be used to calculate the appropriate gradient profiles for scale-down studies. Furthermore, the incorporation of this framework into a robotic liquid handling station would allow the implementation of the calculated gradients in parallel minibioreactors, for high throughput study of the effects of scale-up related gradients on strain performance. In this regard, the following research questions have been formulated to form the bases of the work:

RQ1. *What are the mathematical and physiological requirements of mechanistic models for Bioprocess Development?*

Mathematical models are almost ubiquitous in bioprocess development today. But to what extent should these models be detailed to be useful for the specified purpose? What mathematical properties (such as tractability, continuity of equations, identifiability and uncertainty propagation profiles) should they have? Finally, to what extent should old models be updated with current theoretical research results in order to be useful? The responses to these questions are explored in the early part of this research work, by updating a mechanistic model of *E. coli* from the mid 1990s with recent research data, and applying robust identifiability and uncertainty analysis to the model to ascertain the reliability of model predictions. The goal of this part is to establish a reliable and mathematically sound description of the strain.

RQ2. *How can representative, alternative scale-down set-ups be derived from the already existing knowledge (operation mechanisms, design) of multi-compartment scale-down bioreactors?*

Multi-compartment scale-down bioreactors are excellent at reproducing scale-up effects in the laboratory scale. However, to simplify the scale-down concept and introduce it into smaller scale minibioreactors, alternative approaches such as pulse-based methods need to be developed further. The aim here is to use the already existing knowledge of the two-compartment bioreactor to develop a mechanistic model that can be used to design a pulse-based scale-down system that is equivalent to a 2CR and can simulate actual large-scale bioreactors.

RQ3. *How can screening platforms be operated to test scale-up effects when multiple strains are involved?*

There is a need to increase the throughput of scale-down bioreactors to enable robustness analysis of various strains (and not only the selected few) at the screening phase. High throughput minibioreactor systems are beneficial for rapid development, and incorporation of scale-down concepts into these platforms will ensure the selection of the most robust strains for scale-up. But to achieve this incorporation, special tools and technologies are needed.

RQ4. *Can concentration gradients be induced in milliliter-scale cultivations to study scale-up effects? Can a minibioreactor be used as a scale-down bioreactor?*

Fundamentally, minibioreactors have small volumes with highly efficient mixing, which creates a homogeneous environment for the cells. Conventionally, it doesn't seem feasible to create zones in such small bioreactors to study the effects of concentration gradients. The exploitation of appropriately developed mechanistic models to calculate concentration gradients and implement them in smaller minibioreactors with the help of liquid handling stations is explored in this part of the work. This ensures that gradients are achieved in smaller bioreactors for scale-down studies.

RQ5. *What are the specific responses of E. coli in terms of physiology, metabolic states and recombinant protein quality when subjected to pulse-based dissolved oxygen and glucose gradients in a minibioreactor system?*

Would it be possible to reproduce in the minibioreactor scale-down system the responses of *E. coli* to stresses that are already observed in large-scale bioreactors? What specific metabolic and physiological responses can be detected in such a high throughput scale-down bioreactor system? These questions are investigated by subjecting *E. coli* to model-derived glucose and dissolved oxygen gradients in parallel minibioreactor cultivations. The effect of the concentration gradients on the physiological and metabolic states, as well as on the quality of the produced recombinant proinsulin is further explored.

4. Results

4.1. Development of a Mechanistic Model of *E. coli*

A strong understanding of the physiology of the expression system used for recombinant protein production is required for efficient bioprocess development. The mechanistic model of the physiology of the *E. coli* system used in the current study was developed in the late 1990s, with roots in the substrate usage patterns of from *S. cerevisiae* (Pham et al., 1998), which were previously shown to be equally applicable to the *E. coli* system. This model was further refined by Xu et al, who used it to model overflow metabolism and later by Lin et al., who also used it to determine the maximum uptake capacities of *E. coli* under oscillating cultivation conditions. The model has 7 ordinary differential equations built on basic mass balances describing the volume, biomass, glucose, acetate, dissolved oxygen, feed and intracellular recombinant product in a fed-batch process. The specific rates or growth kinetics of the cell are described by a set of non-linear algebraic equations which describe the physiological phenomena such as Monod kinetics and inhibition kinetics. These are coupled to the mass balance equations to form the complete model, which has 19 unknown parameters. A full description of the model is presented in Anane et al.,2017.

4.1.1. Updating the Physiological Basis of the Model of *E. coli* (Paper I)

The basis of the mechanistic model is the partitioning of glucose taken up by *E. coli* into two major metabolic routes: one for oxidative growth and the other for overflow metabolism in case of higher glucose uptake rates. As shown in Figure 4.1A, these partitioning routes account for the total usage of glucose in the cell, as well as the routes of oxygen use in respiratory pathways.

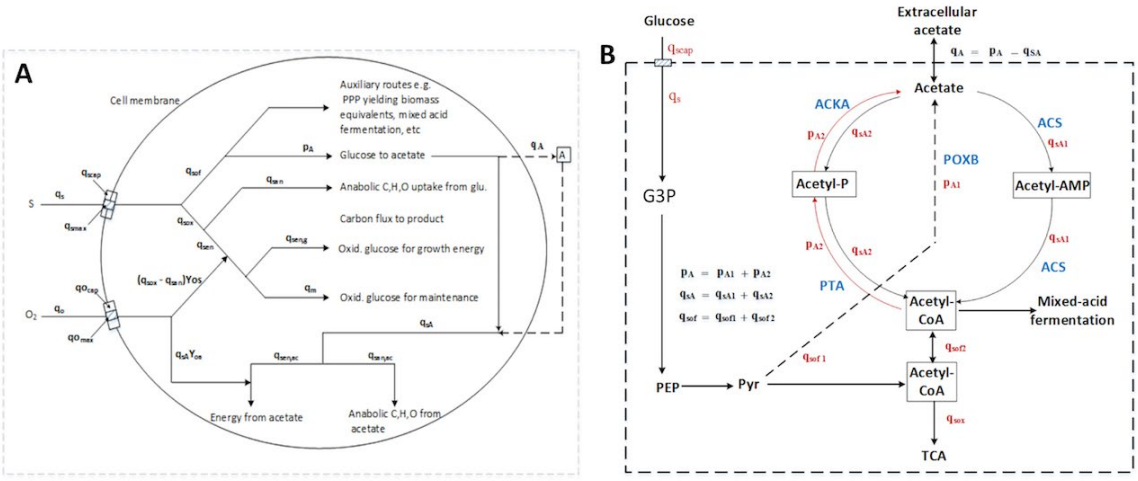


Figure 4.1 (A) Glucose partitioning concept showing metabolic routes of glucose usage under oxidative and overflow metabolic conditions. (B) Acetate cycling concept by which intracellular acetate is constantly being produced and consumed at the same time, which gives the element of continuity to the differential equations used to model the glucose partitioning. Adapted from (Anane et al., 2017) with permission.

In its original form, the branch point of qSox and qSof (oxidative and overflow fluxes) was based on conditional events in the cultivation, such as the availability of oxygen to a certain degree (qO_2 > threshold) or when substrate uptake (qS) was above certain thresholds. Since the time points at which these conditions were fulfilled could not be pre-determined, the mathematical formulation of the model resulted in piecewise functions, some of which were only applicable in certain regimes of the cultivation.

Recent publications on proteomic studies of the enzymes involved in the acetate production and re-assimilation routes, however, showed that the intracellular acetate production and re-assimilation were simultaneous processes. Extracellular acetate is only detected when the intracellular rate of production is greater than the re-assimilation rates. This concept, known as the acetate cycling concept was used in this study to develop the metabolic scheme shown in Figure 4.1B. Based on this scheme, continuous algebraic equations could be developed to describe the overflow metabolism of *E. coli* as the offset of an equilibrium between intracellular acetate production and re-assimilation, which may be triggered by carbon catabolite repression under conditions of high glucose concentration or in limiting dissolved oxygen environments. After model fitting to fed-batch fermentation data (which is described in the next sections), the updated physiological model could be used to develop an advanced overflow profile of the strain, as shown in Figure 4.2.

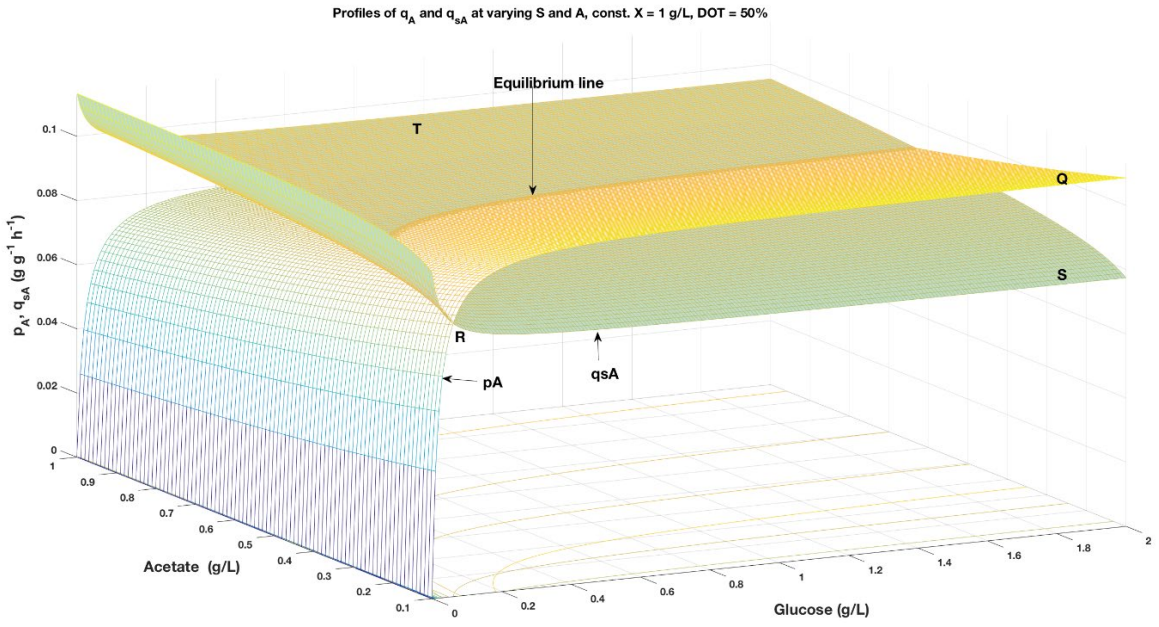


Figure 4.2 Advanced overflow profile of *E. coli* based on the concept of acetate cycling, showing the dynamic equilibrium between intracellular acetate production and assimilation, which is dependent on the residual glucose concentrations. Reproduced from (Anane et al., 2017), with permission.

The 3D representation of overflow metabolism shows the dynamic equilibrium between the two opposing intracellular acetate fluxes (qSA—acetate re-assimilation and pA—acetate production). The equilibrium, however, depends on the amount of glucose present in the extracellular medium. At low glucose concentrations, the rate of acetate consumption (upper surface, point T) is higher than the rate of re-assimilation (surface below point T). As the glucose concentration increases, the intracellular rate of acetate production exceeds the rate of its re-assimilation, which leads to acetate secretion into the extracellular medium.

4.1.2. Model Fitting to Fed-batch Data of Wild Type and Recombinant *E. coli* Strains

The updated model of *E. coli* was fitted to cultivations of two strains of *E. coli*: the wild type W3110 strain and the recombinant BW25113 psW3 lacI^{+} strain expressing a recombinant proinsulin. The two strains were cultivated in fed-batch culture. Details of the model fitting and the resulting parameter values are given in the following sections.

4.1.2.1. Fed-batch cultivations of wild type E. coli W3110 (Paper I)

A fed-batch cultivation of wild-type W3110 was carried out in 3.7 L bench top. The details of the cultivation methodology and analytical methods are presented in (Anane et al., 2017). During the batch phase, a maximum specific growth rate of 0.3 h^{-1} was achieved in the presence of excess glucose, whereas the acetate concentration increased steadily to a maximum of 0.3 g L^{-1} in this phase (Figure 4.3). The model was fitted to the cultivation data by minimisation of the least square function between the measurements and simulations, using the non-linear least square minimisation function (*lsqnonlin*) function in Matlab®. The resulting parameter values for the strain that is not producing any recombinant protein are given in Table 4.1, with results of the parameter uncertainty analysis. The detailed methods of the parameter uncertainty analysis are given in Papers I and II.

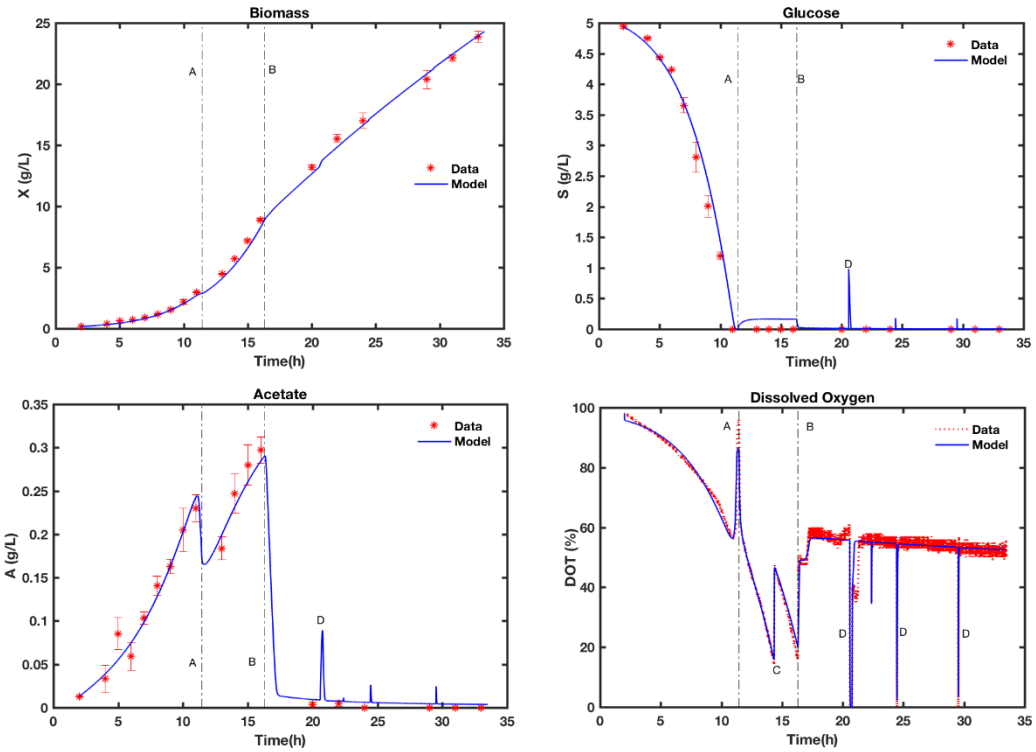


Figure 4.3 Model fitting of fed-batch cultivation of wild type *E. coli* W3110. Point A represents the start of exponential feed fed-batch phase and B the switch to constant feed fed-batch. The peaks (D) represent intermittent glucose pulses that were given to capture the kinetics of acetate production and re-assimilation Reproduced from (Anane et al., 2017), with permission.

Table 4.1 Results of parameter estimation for wild type strain of *E. coli* W3110 (with Tikhonov regularization—see next sections). Also shown is the quantification of the uncertainty associated with the parameter estimators. LB—lower bound, UB—upper bounds, CI—95% confidence interval.

Parameter [+]	Units	Initial guess (literature)*	Estimate	PE uncertainty quantification		
				% σ_θ	LB-CI	UB-CI
K_{ap}	g g^{-1} h^{-1}	0.10	0.5052	15.2	0.3539	0.6565
K_{sa}	g L^{-1}	0.05	0.0134	22.0	0.0076	0.0192
K_o	g L^{-1}	10.0	0.0001	0.0	0.0001	0.0001
K_s	g L^{-1}	0.05	0.0370	8.9	0.0305	0.0435
K_{ia}	g L^{-1}	5.00	1.2399	9.6	1.0062	1.4737
K_{is}	g L^{-1}	10.0	2.1231	27.3	0.9788	3.2673
p_{Amax}	g g^{-1} h^{-1}	0.17	0.2268	6.5	0.1977	0.2558
q_{Amax}	g g^{-1} h^{-1}	0.15	0.1148	6.1	0.1009	0.1287
q_m	g g^{-1} h^{-1}	0.04	0.0129	7.0	0.0111	0.0147
q_{Smax}	g g^{-1} h^{-1}	1.37	0.6356	0.3	0.6320	0.6392
Y_{as}	g g^{-1}	0.80	0.9097	4.5	0.8283	0.9911
Y_{oa}	g g^{-1}	1.06	0.5440	9.5	0.4425	0.6455
Y_{xa}	g g^{-1}	0.70	0.5718	9.9	0.4604	0.6833
Y_{em}	g g^{-1}	0.50	0.5333	2.4	0.5085	0.5580
Y_{os}	g g^{-1}	1.06	1.5620	5.4	1.3941	1.7298
Y_{xsof}	g g^{-1}	0.15	0.2268	12.0	0.1730	0.2807

[+] Parameter descriptions in nomenclature. * References: (Lin et al., 2001b; Martínez-Gómez et al., 2012; Neubauer et al., 2003; Bo Xu et al., 1999)

4.1.2.2. Fed-batch Cultivations of Recombinant E. coli BW25113 and Product Formation Model

The previous cultivation and model fitting did not include the production of any recombinant protein, therefore the version of the model used did not take into account the metabolic burden induced by the expression of a foreign protein in *E. coli*. Thus, further cultivations were done in 15 L BIOSTAT bioreactor (Sartorius Stedim GmbH) at Sanofi Aventis Deutschland GmbH, also in fed-batch mode, for the production of recombinant protein upon induction with IPTG. The details of the cultivation methodology are not presented due to confidentiality reasons. Upon induction of protein formation, there was a steady increase in intracellular proinsulin accumulation in the form of inclusion bodies, up to 3 hours after induction. From 3 hours onwards, the product formation rate declined gently, with a saturation of the total inclusion body fraction of the cell (Figure 4.4). Out of the three recombinant product quantification methods used (scanning densitometry of SDS gels, in-house HPLC at Sanofi and

capillary gel electrophoresis with Labchip GX II touch), the product formation data of the capillary gel electrophoresis (**Error! Reference source not found.**) was used for further model fitting due to consistency in the sequence of product formation data after induction using this method and our inability to publish the HPLC method.

Another feature of the recombinant expression system that was tested in these cultivations was the stability of the plasmid at the end of the product formation phase. Serial dilutions of the broth up to 10^6 times were plated out on both ampicillin-containing and ampicillin-free agar plates, since the recombinant plasmid is under ampicillin selection pressure. The results of the plasmid stability test showed >99.9 % stability for all tested samples. All the individual colonies picked from the diluted sample formed colonies within 1 day on the test plate in the presence of ampicillin.

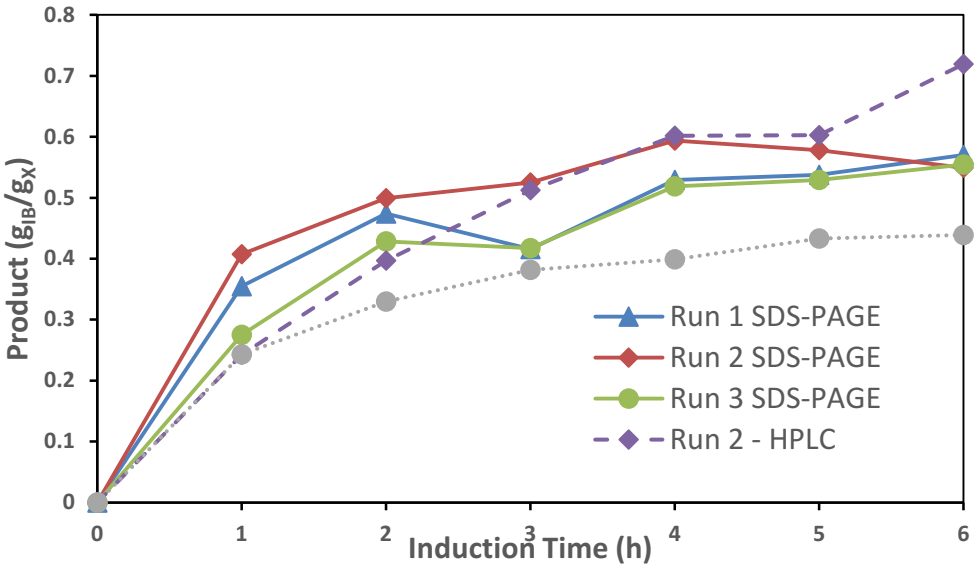


Figure 4.4 Recombinant proinsulin formation rate in fed-batch cultivation of *E. coli* BW25113, quantified using three analytical techniques. SDS-PAGE = standard scanning densitometry of SDS gels using the BIORAD GelDoc image analyser for 3 different cultivations of the same strain; HPLC—high performance liquid chromatography quantification of the product profile of run 2, EPG—quantification of product of run 3 with capillary gel electrophoresis. Due to limitations in availability of equipment, not all methods were applied to all cultivations.

To model the product formation profile, the glucose partitioning model was modified to include a flux under oxidative conditions that supports the formation of inclusion bodies (see modification of Figure 4.1 in Manuscript 5). The accumulation of inclusion bodies in the cell significantly retarded the growth of the culture, as observed in the biomass profiles of the induced cultures compared to the biomass profiles of the wild type *E. coli* (Figure 4.3). The significant flattening of the biomass profile was observed 3 hours after protein induction (ca. 19h, Figure 4.5). This behaviour was consistent for all cultivations that were induced. Since the flattening of the biomass profile was also usually associated with the accumulation of glucose at the end of the cultivation (6 hours after induction), it was hypothesised that the metabolic burden induced by the recombinant protein tends to limit the glucose uptake capacity as the product concentration builds up in the cell, according to the model of Neubauer et al (2003). Therefore, the mechanistic description of substrate uptake according to Monod kinetics (Equation 4.1), which was used to model the wild type cultivations was modified to include an inhibitory term based on the product concentration in the cell, according to Equation 4.2;

$$q_S = \frac{q_{Smax}S}{K_s+S} \qquad \text{Equation 4.1}$$

$$q_S = \frac{q_{Smax}S}{K_s+S} e^{-K_{ip}P} \qquad \text{Equation 4.2}$$

where K_{ip} (g g^{-1}) is the inhibitory constant of product formation on glucose uptake and P is the intracellular inclusion body concentration (g g^{-1}). The exponential decline of qS upon protein induction is consistent with observation of biomass profiles, and is also based on the switch of the glucose feed from exponential to constant feeding in fed-batch cultivation of *E. coli*. The modified model was fitted to the data of cultivations with product formation, and the numerical values of parameters describing the growth of the culture were determined. The fitting of the model outputs to the experimental data, including the recombinant protein and feed rates are shown in Figure 4.5. Under the given cultivation conditions, there was no acetate accumulation. Residual glucose, which was continuously measured online, was only detected in trace amounts (< 0.1 g L^{-1}).

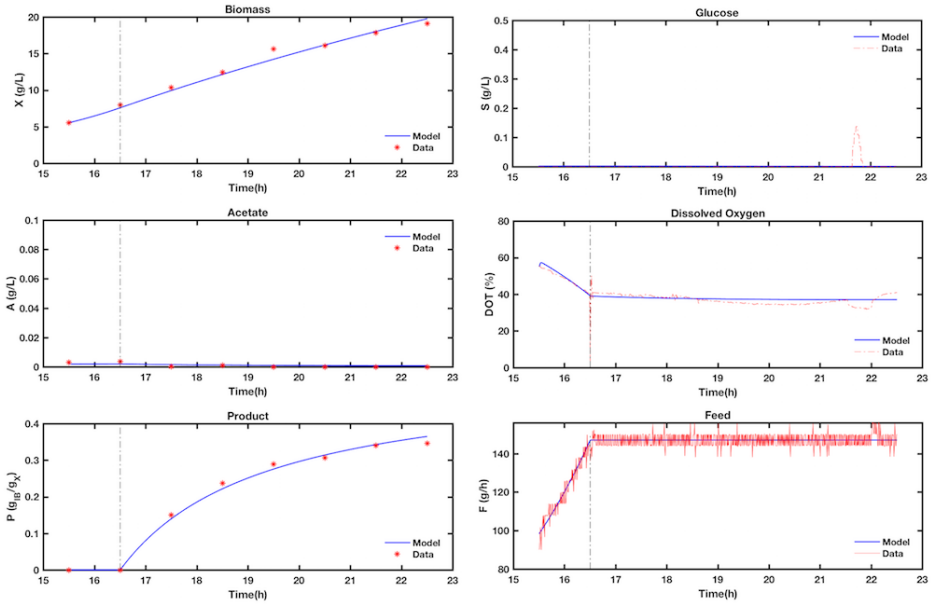


Figure 4.5 Model fitting of recombinant strain of *E. coli* producing proinsulin as inclusion bodies. Notable feature of the growth profile is the sudden flattening of the biomass curve 3-4 hours after protein induction, which may be due to the growth retardation caused by recombinant protein formation.

A comparison of the parameter estimates of the two different strains of *E. coli* is presented in Figure 4.6A. Significant differences in the parameter estimates of the maximum glucose uptake rate between the two strains is noted, where the recombinant strain grows about 3 times faster than the wild type due to the lower glucose uptake capacities for the latter. Additionally, the recombinant strain showed higher acetate production and assimilation rates than the wild type, according to the parameter estimates, which was also confirmed in the cultivations. This unexpected result was probably due to fundamental differences between the two strains. In fact, due to the slow growth of the W3110 strain, further work in this project was carried out only with the fast growing BW25113 strain as a reference.

The parameter estimates of the recombinant strain were used to quantitatively determine the level of inhibition, in terms of decline in substrate uptake rate that was experienced by this strain after protein induction. As shown in Figure 4.6B, there is a sharper decline in specific substrate uptake rate during recombinant protein production (qS_{IB}) than in the absence of inclusion bodies in the cell (qS). The qS

profile without product formation was obtained by setting the product inhibition constant in the model to a value of 0 g g^{-1} (instead of the estimated 1.50 g g^{-1}). Correspondingly, the specific growth rate also declines at a lower rate in the absence of product formation than when inclusion bodies accumulate in the cell. The problem of late fed-batch phase glucose accumulation can be explained by the decline of qS after induction. These results are useful, in the sense that they can be used to determine equivalent declining feed profiles for efficient fed-batch processes of *E. coli*. According to Figure 4.6B, the feed profiles should also decline in accordance with the fall in qS caused by product inhibition, to prevent accumulation of glucose and consequently, also prevent accumulation of acetate, in the latter phases of fed-batch processes of *E. coli*.

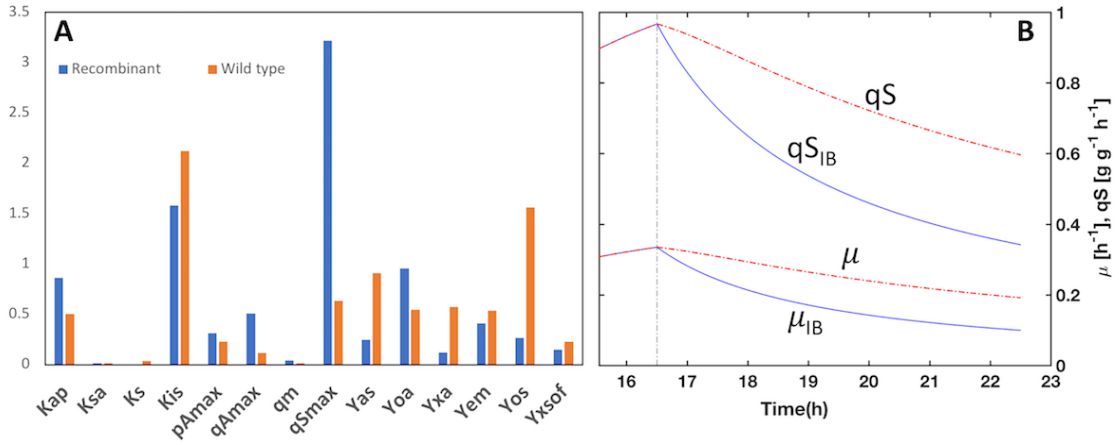


Figure 4.6 (**A**) Comparison of parameter estimates from fed-batch cultivation of wild type *E. coli* W3110 and recombinant *E. coli* BW25113 pSW3 expressing recombinant proinsulin. (**B**) Graphical representation of the inhibition of glucose uptake due to the metabolic burden induced by foreign product formation, and its effect on the specific growth rate in the recombinant strain. Dotted vertical line shows the point of IPTG induction. q_S and μ are the specific substrate uptake rate and specific growth rate respectively, without protein induction; qS_{IB} and μ_{IB} respectively represent the specific substrate uptake rate and specific growth rate under product formation conditions.

4.1.3. Identifiability analysis of the Mechanistic Model of *E. coli* (Paper II)

The mechanistic model of *E. coli* has 15 parameters (for non-protein producing conditions) and 17 parameters when recombinant protein production is considered. As depicted in Figure 4.6 and in Table 4.1, the values of these parameters were estimated from data of laboratory cultivations of *E. coli*, under specified experimental conditions. The identifiability of the parameters was analysed, to ascertain whether each of the parameters could be assigned a unique numerical value during parameter estimation from the available data. The details of the methods for the identifiability analysis and related solution methods are presented in Paper II. At the optimal parameter values (after parameter estimation), the absolute sensitivity matrix of the model was calculated for both cases where data of the recombinant and wild type strains were used. The method of singular value decomposition (SVD) was then used to calculate the rank (r_ϵ) of the sensitivity matrix, which was used to determine the existence of non-identifiable parameters in the parameter estimation problem. For the calculation of the numerical rank, a threshold of 1000 for the condition number and 15 for the collinearity index were used. The results of the SVD with respect to the given thresholds are plotted in Figure 4.7 for the model fitting using data of cultivation of the wild type strain.

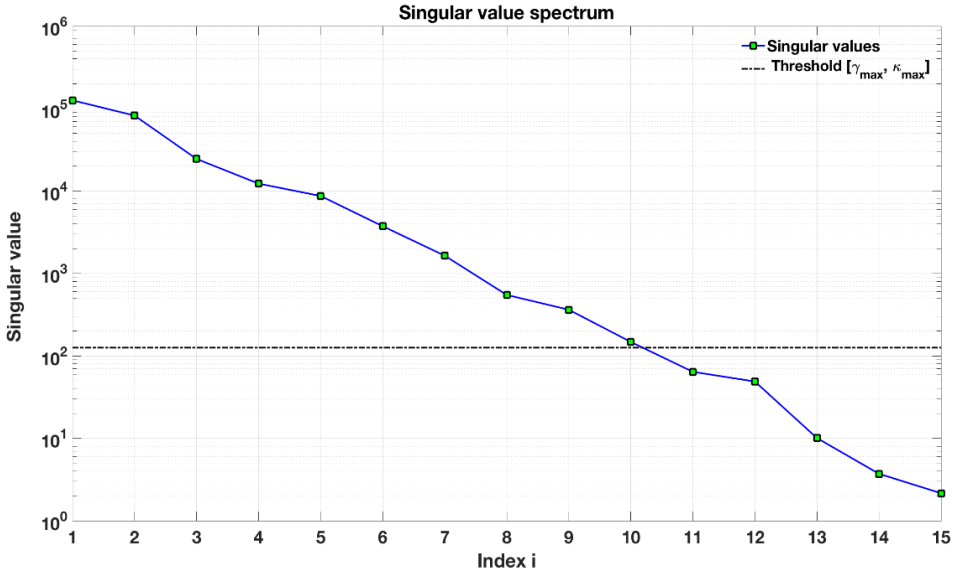


Figure 4.7 Singular value spectrum of the sensitivity matrix ($\tilde{S}$) showing the ill-conditioned parameter estimation problem, based on the thresholds for the numerical rank r_ϵ of $\tilde{S}$. The number of singular values above the threshold line is the numerical rank, in this case $r_\epsilon = 10$, which represents the number of uniquely identifiable parameters in the least square parameter estimation problem. The SVD, however, does not reveal the identity of the uniquely identifiable parameters, so numerical indices are used on the x-axis.

A numerical rank of 10 was determined for the model using the data of the wild type strain, which implies that in the parameter estimation using the minimisation of least squares criterion, only 10 out of the 15 parameters could be uniquely fitted to the data. This is represented by the points above the threshold line in Figure 4.7. However, the SVD does not reveal the identity of these identifiable paramters. The remaining 5 parameters could not be fitted, probably due to non-informative data or the lack of physiological events in the experiments that would trigger the *effects* of those parameters in the cells. Using the data of the recombinant strain producing proinsulin, 14 out of 17 parameters were identifiable after SVD of the sensitivity matrix, using the same thresholds as before. This may be due to the availability of an online glucose sensor for the cultivations at Sanofi, which increased the frequency of data collection for the substrate (glucose data point every 1 min) compared to manual sampling of glucose during the cultivations of the wild type strain (glucose sample every 1 h). The non-identifiability of some of the parameters led to wide confidence intervals for the parameter estimation, as shown in Table 4.2.

Another numerical technique that was used to further investigate the non-identifiability problem is Monte Carlo (MC) parameter estimation. A total of 500 *insilico* datasets were generated from a normal distribution space, defined by the measurements (as mean values) and standard deviations of the wild type cultivation data. For high precision measurements such as DOT, the standard deviation was assumed to be 5% of the measurements. Each of the datasets was then used to carry out parameter estimation, to determine whether the parameters would converge to the same values. The details of the sampling algorithms and parameter estimation are given in Paper II. The probability density functions of the parameters calculated from all 500 parameter estimates are shown in Figure 4.8. Since the measurement errors were assumed to be normally distributed, an identifiable parameter would always converge to the same value in the parameter estimations. As shown in Figure 4.8, some of the parameters showed more broader peaks, signifying their inability to converge to unique values in the

MC parameter estimations. However, although the MC procedure can reveal non-identifiability, it may be erroneous to conclude on the identity of the non-identifiable parameters from the procedure alone, since overall non-identifiability also depends on parameter correlations, which cannot be depicted in probability density functions of Figure 4.8.

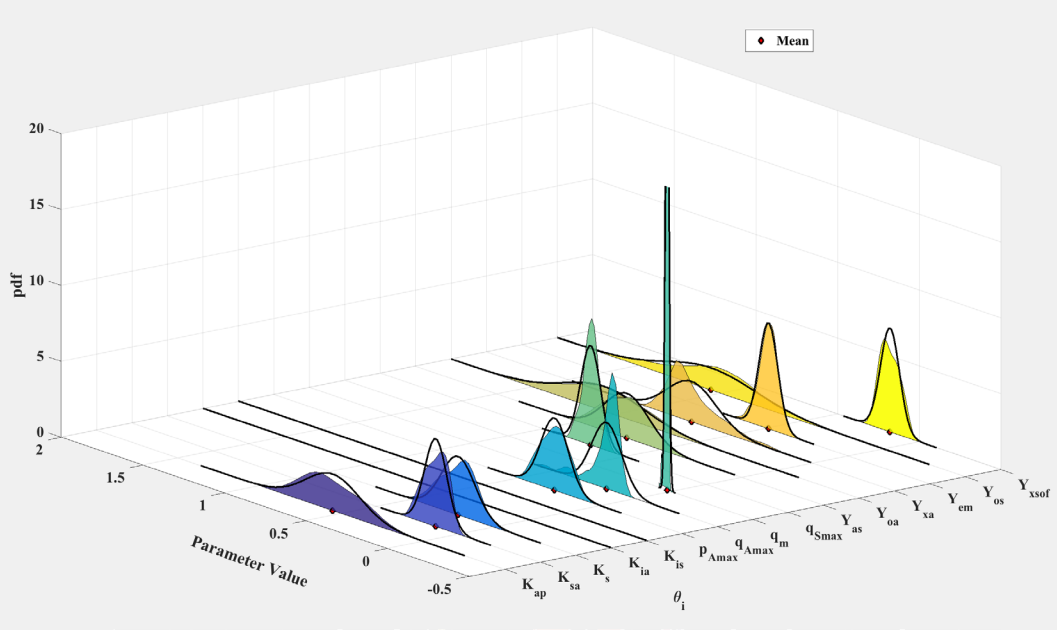


Figure 4.8 Depiction of non-identifiability through Monte-Carlo parameter estimation: distribution of parameter estimates from 500 Monte Carlo data sets. Solid outer lines represent an approximation of the normal distribution for each parameter, whereas the coloured inlay represents the actual distribution of the estimates of each parameter. Typically, the narrower the peak, the more identifiable a parameter would be. However, the spread of the peak alone is not enough to identify the specific non-identifiable parameters as this would also involve the correlation patterns of the parameters.

4.1.3.1. Regularization by Subset Selection

One consequence of the existence of non-identifiable parameters in a parameter estimation problem is the inflation of the variance of the estimated parameters (Table 4.2). During the numerical computation of the estimates, very large numbers are divided by small numbers (ill-conditioning), which leads to numerical instability and inaccurate results. To handle the ill-conditioning problem, the PE problem is made *regular* by the introducing prior knowledge about the system, such as literature values of the parameters. One of the methods of regularization applied in this work was subset selection (Paper II, Supplementary material). In this method, another decomposition of the sensitivity matrix (QR-decomposition) was used to re-order the parameters in descending order of their identifiability (sensitivity of model outputs to the parameters and their correlation with other parameters). Based on the QR-decomposition of the sensitivity matrix for the case where data from wild type strain were used, the parameters were re-ordered as : $\tilde{\theta} =$ [q_{Smax}, p_{Amax}, Y_{em}, Y_{oa}, q_{Amax}, Y_{xa}, Y_{os}, K_s, K_{sa}, K_{ia}, q_m, Y_{as}, K_{ap}, K_{is}, Y_{xsof}]. Therefore, truncating $\tilde{\theta}$ at $r_\epsilon = 10$ (previously determined rank of the system) resulted in the identifiable (active) parameter subset $\tilde{\theta}_{act} =$ [q_{Smax}, p_{Amax}, Y_{em}, Y_{oa}, q_{Amax}, Y_{xa}, Y_{os}, K_s, K_{sa}, K_{ia}], whereas the last 5 elements of $\tilde{\theta}$ were classified as non-identifiable, $\tilde{\theta}_{inact} =$ [q_m, Y_{as}, K_{ap}, K_{is}, Y_{xsof}]. In subsequent parameter estimations, the non-identifiable parameters were given fixed values based on literature (prior knowledge) and taken out of the least squares problem

formulation. This so-called parameter fixing led to more narrow confidence intervals for the identifiable parameters, compared to the case without regularization (Table 4.2).

4.1.3.2. Tikhonov Regularization

Another method of regularizing the ill-conditioned parameter estimation problem that was applied in this work is the Tikhonov regularization. Here, the prior information was added to the objective function of the PE as a penalty term, so that only sensitive and non-correlated parameters are fitted to the data, as described in Paper II. As opposed to regularization by subset selection, preference is not given to some selected parameters when this method is applied, although intrinsically, only some parameters are fitted to the data. However, the regularization by this method has to be tuned to get accurate results, i.e. the contribution of the prior information must be properly balanced with the data at hand, so as not to give too much weight to one side. The tuning parameter was selected from a range of 10^6 to 10^{-3}, using the L-curve method (Paper II). A total of 30 tuning parameters, contained in the vector λ, were generated and tested within the given interval using the geometric series

$$\lambda = a_0 \cdot q^n \qquad \text{Equation 4.3}$$

where q = 0.48, $a_0 = 10^6$ and n = 30. According to the L-curve method, an optimal tuning parameter of 311.6 was selected (corner of the L-curve) for further parameter estimation using the Tikhonov regularization. The results of the parameter estimation by Tikhonov regularization are given in Table 4.1.

4.1.4. Uncertainty Analysis of Model Outputs (Paper II)

As stated earlier, the data used to fit the model and the effect of the ill-conditioning in the parameter estimation lead to some level of variance in the estimated values of the parameters. The level of error in the parameter estimates was quantified as the 95% confidence interval, using Student's t-distribution (Table 4.2). In model applications, however, it is the outputs that are mostly of interest. Therefore, the parameter estimation error, as contained in the confidence intervals, was propagated onto the model outputs by random sampling within the confidence intervals of each parameter, and running model simulations with each of these random samples to determine how the variance of parameter estimates translates into output uncertainty. The sampling and simulation methods, which was done in a Monte Carlo framework are explained further in Paper II.

Table 4.2 Results of parameter estimation with and without regularization by subset selection.

Par		Before subset selection					After subset selection					
			St. Dev		95% CI			St. Dev		95% CI		
	Initial guess	$\widehat{\theta}_{mc1}$	σ	%σ	LB	UB	$\widehat{\theta}_{reg}$	σ	%σ	LB	UB	Fixed value
K_{ap}	0.438	0.508	0.355	69.86	-0.188	1.206	–	–	–	–	–	0.50
K_{sa}	0.016	0.012	0.011	87.11	-0.009	0.035	0.0131	0.0027	20.617	0.0078	0.0184	–
K_{s}	0.035	0.038	0.010	26.25	0.018	0.058	0.0370	0.0032	8.7045	0.0307	0.0434	–
K_{ia}	1.111	1.260	0.338	26.83	0.597	1.923	1.2156	0.0883	7.2662	1.0414	1.3898	–
K_{is}	1.562	1.838	6.130	333.48	-10.17	13.85	–	–	–	–	–	10.0
p_{Amax}	0.203	0.228	0.171	74.99	-0.107	0.564	0.2318	0.0107	4.6036	0.2108	0.2528	–
q_{Amax}	0.106	0.115	0.024	21.25	0.066	0.162	0.1201	0.0049	4.0464	0.1105	0.1297	–
q_{m}	0.013	0.013	0.011	85.14	-0.008	0.035	–	–	–	–	–	0.04
q_{Smax}	0.635	0.635	0.025	3.90	0.586	0.683	0.6361	0.0022	0.3448	0.6318	0.6404	–
Y_{as}	0.827	0.894	0.365	40.89	0.177	1.610	–	–	–	–	–	0.97
Y_{oa}	1.100	0.522	0.101	19.35	0.324	0.720	0.5323	0.0249	4.6768	0.4832	0.5814	–
Y_{xa}	0.611	0.579	0.662	114.37	-0.719	1.878	0.6043	0.0411	6.8045	0.5232	0.6854	–
Y_{em}	0.546	0.532	0.153	28.82	0.231	0.832	0.5432	0.0074	1.3637	0.5286	0.5578	–
Y_{os}	1.100	1.572	0.893	56.80	-0.178	3.322	1.6706	0.0622	3.7221	1.5480	1.7932	–
Y_{xsof}	0.206	0.229	1.034	451.64	-1.798	2.256	–	–	–	–	–	0.15

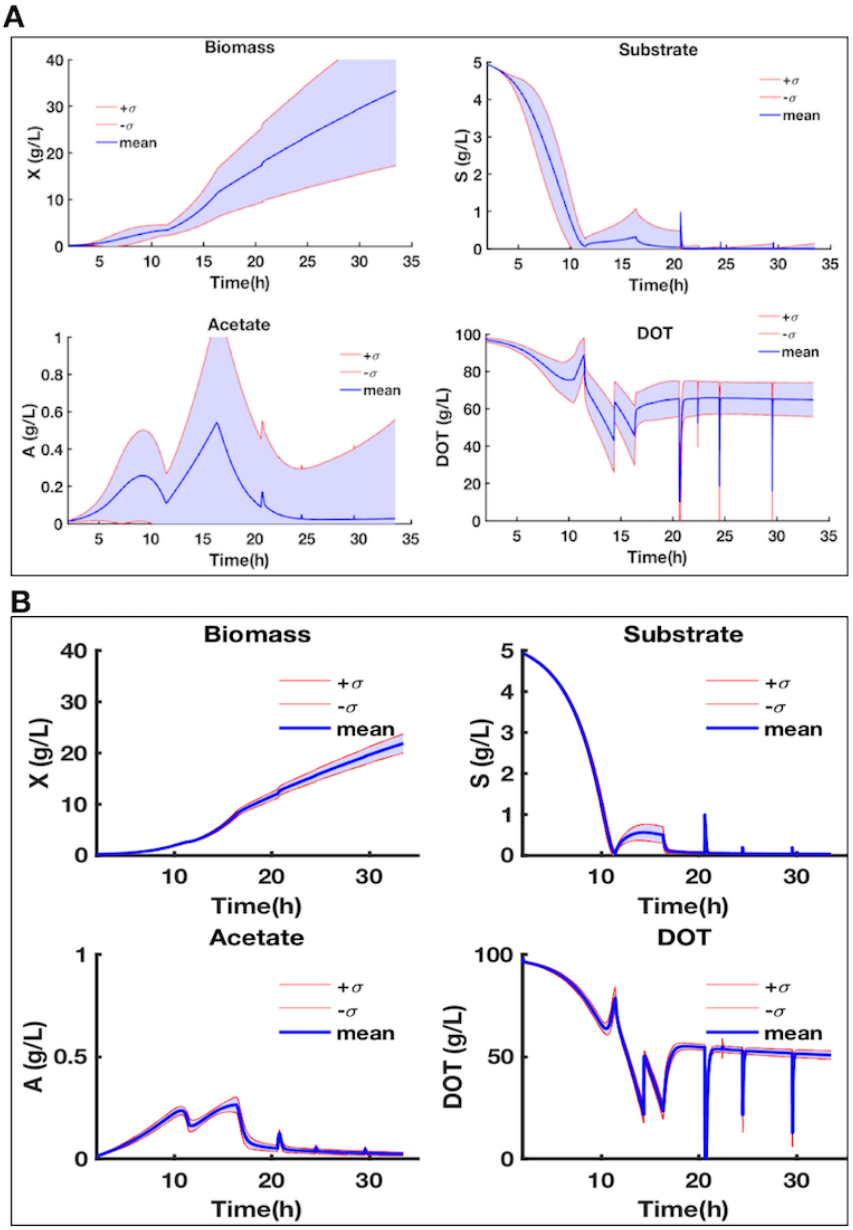


Figure 4.9 Uncertainty analysis of outputs of mechanistic model of *E. coli* (A) before regularization (B) after parameter estimation with Tikhonov regularization.

As shown in Figure 4.9, without regularization of the parameter estimation problem, the output uncertainty analysis led to wide prediction bands for all model outputs for the highly ill-conditioned problem (10 out of 15 parameters identifiable). Since the model fitting using data from the recombinant strain had relatively better identifiability, the PE was not regularized. In the case of model fitting using the data of the wild type strain, the parameter estimation problem was regularized, after which the output uncertainty analysis technique was applied again. The results of uncertainty analysis after regularization are shown in Figure 4.9B. It is apparent that the regularization technique led to more narrow prediction bands for the model outputs by reducing the confidence intervals of the parameter estimates. In the remaining part of the work, regularization by subset selection was mostly applied during parameter estimation.

4.2. Mechanistic Model of Two-compartment Bioreactor and Design of Pulse-based Scale-down Bioreactor (Paper III)

The two-compartment scale-down bioreactor (2CR) has seen many applications, for the study of various large-scale effects (glucose gradients, oxygen limitation, pH gradients) on microbial physiology. Here, a mechanistic model of the two-compartment bioreactor was developed to serve as a bridge between the scale-down concept of the 2CR and a pulse-based scale-down system. It is important to establish the interconvertibility between the two scale-down configurations, because it is easier to implement the pulse-based scale-down system in smaller scales, whereas the 2CR has a stronger footprint in the literature. Thus, according to Figure 4.10, a mechanistic model was established to determine the nature of the concentration gradients that occur in the 2CR, and how such gradient profiles could be used to develop a pulse-based scale-down system in a 3.7 L bench top bioreactor. The details of the model development and simulation conditions are given in Paper III.

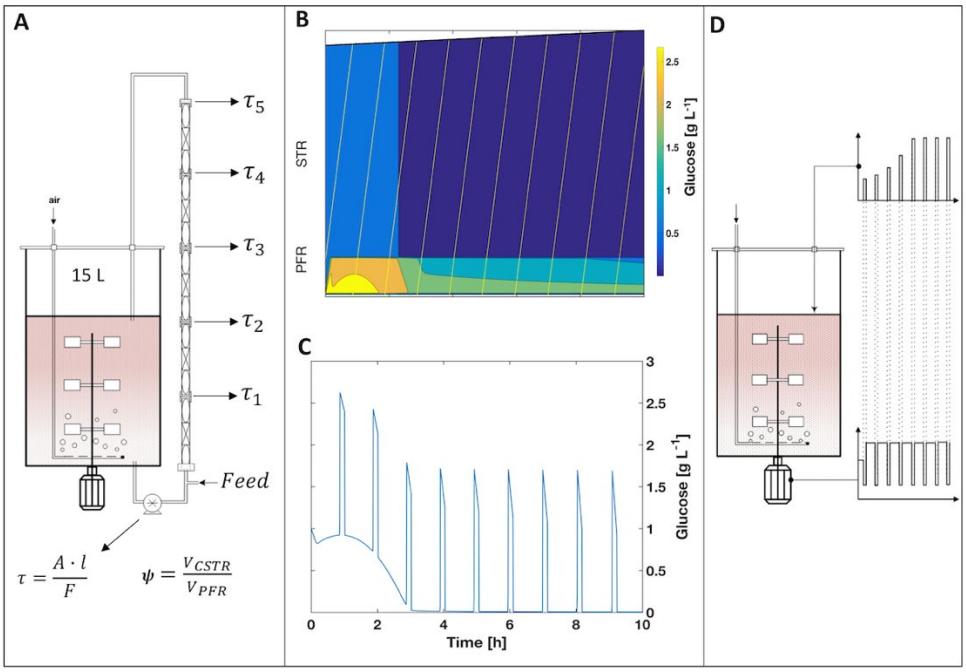


Figure 4.10 Interconvertibility between two-compartment scale-down bioreactor and a bench top pulse-based bioreactor. (A) The exposure time of the culture to the stress inducing agent in the 2CR depends on the flow rate in the PFR section; (B) Glucose concentration gradients experienced by a single cell moving through the perfectly mixed STR and the stressful environment in the PFR. (C) The glucose gradients experienced by the single cell as it circulates between the PFR and STR are similar to repeated pulses in one compartment STR (D), where the pulse frequency and pulse size of the simplified scale-down system in the single compartment STR (D) are derived from the residence times (τ) and volume ratios of the STR to PFR (A). Adapted from (Anane et al., 2018), with permission.

Based on the simulations from the 2CR model, a glucose pulse feeding system was designed for a single compartment STR bioreactor. The calculated exponentially increasing glucose feed was divided into discrete pulses, where the pulse frequency and magnitude were determined from the gradient profiles of the 2CR. Further details of the derivation and calculation of pulses are given in (Anane et al., 2018). The coupling of agitation shifts to coincide with the addition of glucose pulses ensures glucose-rich oxygen-deficient and glucose-limiting oxygen-rich zones in the single-compartment scale-down

system, which is similar to the concentration gradient profiles achieved in the two-compartment reactor (Figure 4.10).

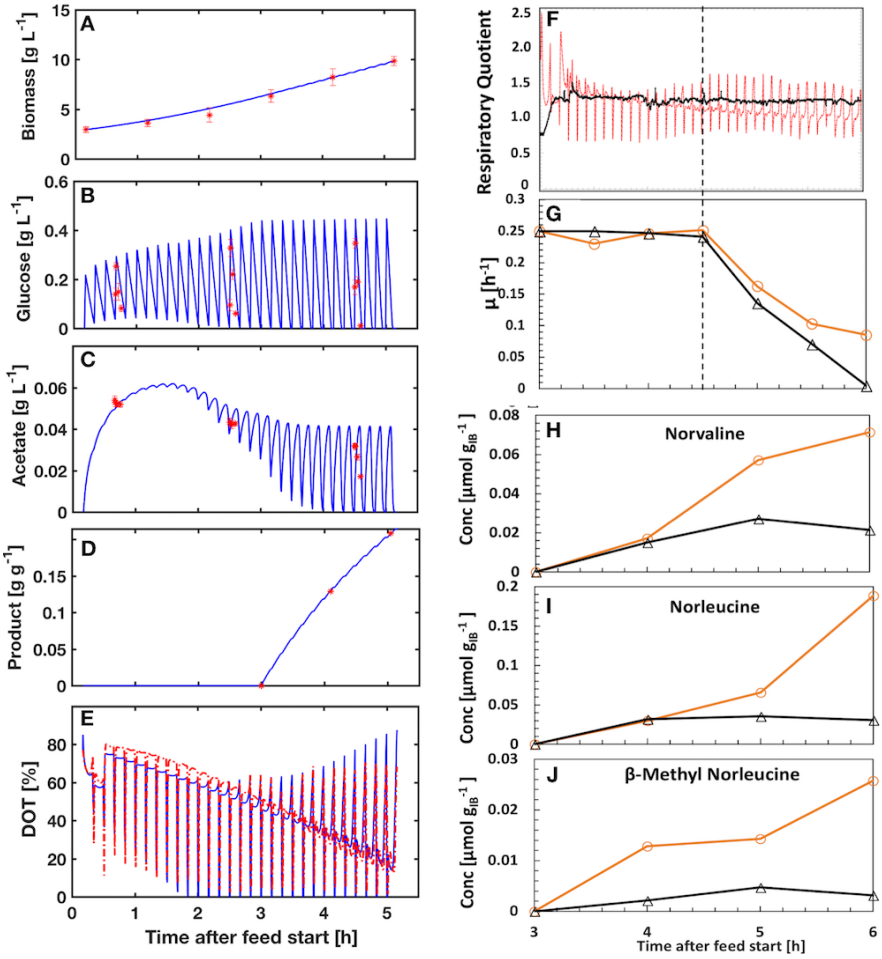


Figure 4.11 Response of *E. coli* to pulse-based scale-down cultivations in 3.7 L bioreactor. (A-E) Growth profiles of *E. coli* showing response to oscillating glucose concentrations. The glucose gradients, which were also accompanied by dissolved oxygen gradients, were induced by the pulse-based fed-batch technique. (F—J) Metabolic and physiological analysis of *E. coli* in the presence of glucose pulses (—O—) and without glucose pulses (—Δ—). In panel F, the oscillating RQ profile shows response to glucose pulses whereas the midline represents RQ of the reference cultivation. The dashed line indicates the point of induction for recombinant protein production. Adapted from Anane et al (2018a), with permission.

The results of the pulse-based cultivation of recombinant *E. coli* in the bench top bioreactor are shown in Figure 4.11, together with the fitting to the macro-kinetic model of *E. coli.* Comparing the pulse-based cultivation to a reference cultivation with smooth glucose feed, there was a significant accumulation of the non-canonical amino acids norvaline, β-methyl norleucine and norleucine in the recombinant protein in response to the pulses. This is in agreement with earlier observations of the accumulation of norvaline in the presence of dissolved oxygen gradients (Soini et al., 2008). However, a key observation of the pulse-based cultivation, as shown in Figure 4.11G, is the slower decline of the specific growth rate when the feed was switched to a constant feed, compared to the reference cultivation. It seems the cells are more robust, with higher glucose uptake capacities in the presence of the pulses (leading to a sustained growth) than when there are no pulses in the cultivation. This is

in agreement of earlier data on the cell viability in large scale bioreactors compared to small scale (Enfors et al., 2001). Overall, the results of the pulse-based scale-down cultivation mostly confirm previous reports about the response of *E. coli* to concentration gradients, with the additional observations of the distinct trends of the concentrations of ncBCAAs in the recombinant proinsulin. Furthermore, the results demonstrate the applicability of mechanistic models for scale-down cultivation design for efficient bioprocess development.

4.3. Scale-down Cultivation in Minibioreactors (Paper IV)

The climax of the current work was the combination of the mechanistic models (model of the strain and the model of the 2CR) with a minibioreactor platform to form a high throughput scale-down system. A full description of the platform and its application to study large-scale effects (glucose and dissolved oxygen gradients) in *E. coli* are given in Paper IV and in the Appendix. The operation of the platform is based on the simulation of the desired concentration gradients in the mechanistic framework, and implementing the gradients with a robotic liquid handling station in fed-batch cultivations. The platform was used to investigate the influence of four different glucose and dissolved oxygen gradient profiles, as well as the influences of a background glucose feeding and induction strength on the physiology and metabolic activity of *E. coli*. Furthermore, the effect of these stresses on the misincorporation of ncBCAAs into the recombinant protein expressed in this strain were also studied. The six specific conditions investigated in the HT scale-down minibioreactor system are given in Table Table 4.3 and summarized below:

i. Influence of mixing time: this was calculated as the frequency of glucose pulses, which were accompanied by corresponding dissolved oxygen pulses whenever a glucose pulse was given. The two frequencies investigated were administering a pulse every 5 min and every 10 min, to represent concentration gradients experienced in larger bioreactors with mixing times in the range of 45 sec to 90 sec.

Table 4.3 Cultivation conditions in 21 parallel scale-down minibioreactors

Bioreactors			Glucose Pulse (min)	IPTG mM	Enzyme-based feeding
B1	B2	B3	-	0.5	linearly to 13 U L^{-1}
C1	C2	C3	5	0.5	-
D1	D2	D3	5	1.0	-
E1	E2	E3	5	0.5	once, 3 U L^{-1}
F1	F2	F3	10	0.5	-
G1	G2	G3	10	1.0	-
H1	H2	H3	10	0.5	once, 3 U L^{-1}

ii. Influence of acute starvation and background glucose supply: here, the effect of the presence or absence of continuous residual glucose in the culture through the enzymatic

glucose release was investigated. The influence of such sustained glucose levels on recombinant product quality and cell physiology is important for strain engineering purposes. This condition was investigated for both 5 min and 10 min glucose pulse cultivations.

iii. Influence of strength of induction: using two different IPTG concentrations, the effects of low and high concentrations of the inducer on microbial growth and recombinant product formation rate was investigated, also in duplicates, with regards to the two different pulsing frequencies.

The results of the scale-down cultivations with respect to biomass growth, recombinant product formation profile and metabolite accumulation in the extracellular medium are shown in Figure 4.13. Generally, the results show a lower biomass production for high frequency gradients/pulses with a higher accumulation of extracellular acetate. Due to lower overall glucose released by the Enbase system in the reference cultivations (B1—B3) (18 ± 0.4 g L^{-1} released compared to 30 g L^{-1} of glucose fed to the other reactors), a lower biomass was recorded in these bioreactors compared to the pulse system. The biomass concentration in the minibioreactors that were subjected to 5 min glucose pulses in the fed-batch phase was 15% less than the biomass in the reference cultivation, *when normalized on the basis of total glucose consumed*. However, despite the relatively lower biomass, the highest concentration of acetate were recorded in the 5 min pulse bioreactors, with about 5 fold higher acetate concentration than in the reference cultivation (OD normalized) and 3 fold higher than that in the 10 min pulse cultivations. On the other hand, the bioreactors that were subjected to 10 min glucose pulses reached higher final biomass concentrations with lower extracellular acetate concentration. Surprisingly, the 10 min pulse cultivations accumulated higher formate than the 5 min, which was the reverse trend for acetate (Figure 4.13). This may be attributed to the higher pulse sizes and higher biomass in the 10 min pulse cultivation, which altogether led to a more rapid decline in the DOT upon pulse addition. The anaerobic conditions thus induced in the low frequency pulse system (DOT in 10 min pulses hit 0 % saturation more often than in 5 min pulses) lead to mixed acid fermentation, with higher formate production rates than the overflow metabolic products observed in the 5 min pulses. The frequency of the pulses did not have a significant influence on the accumulation of inclusion bodies (product) during the cultivations, although the 5 min pulse cultivations had a slightly higher concentration of inclusion per gram of biomass (Figure 4.13).

Possibly due to the slightly higher overall glucose supplied, there was a 6% higher biomass concentration in the cultivations with enzymatic glucose release in the background (E1—E3) than the corresponding cultivations with only glucose pulses (C1—C3). Contrary to expectations, however, the background glucose supply led to about 34% less acetate concentration (Figure 4.13) although the biomass concentration was higher, compared to the same cultivations without the background glucose supply. This may be due to the preservation of a steady pool of intracellular metabolites by the continuous supply of glucose via the enzymatic feeding system, which prevents starvation responses in *E. coli* (Chassagnole et al., 2002; Theobald et al., 1997). In terms of recombinant product formation, there was about 8 % higher inclusion body concentration in the 5 min pulses with enzymatic glucose supply than the corresponding pulse-based feeding without background glucose, whereas the 10 min pulses with background glucose reached 15% higher product quantity than the corresponding cultivation without enzymatic feeding. The strength of induction (IPTG concentration) did not significantly affect both biomass growth and extracellular metabolite accumulation, although there was a slight difference in the product formation profiles.

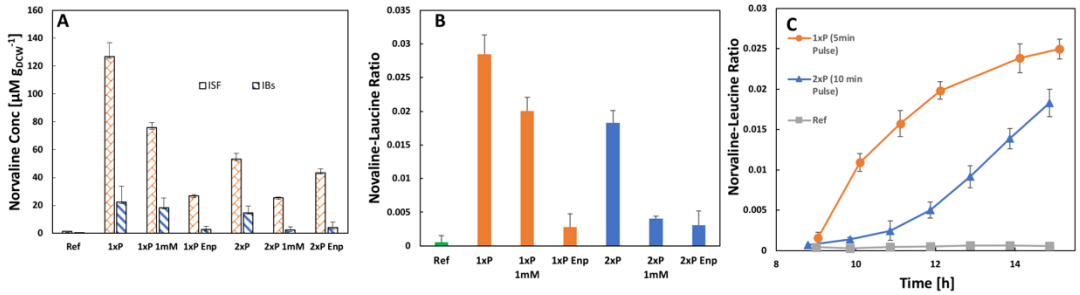


Figure 4.12 (A) Comparison of norvaline in intracellular soluble fraction (ISF) and in purified inclusion bodies (IBs) for the six scale-down cultivation conditions (B) Influence of glucose pulse feeding on overall misincorporation of norvaline for leucine in the purified inclusion bodies of proinsulin (C) Kinetics of misincorporation of norvaline into proinsulin based on different pulse frequencies in the scale-down cultivations.

The influence of the pulse feeding on the recombinant product quality for the various stress conditions are also plotted in Figure Figure 4.12. Since the recombinant protein (proinsulin) used in this study is rich in leucine residues (Reitz et al., 2018), norvaline was the major monitored ncBCAA in the study. Further results concerning the misincorporation of the other ncBCAAs (beta-methyl-norleucine and norleucine) into the recombinant product are given in the Appendix. The higher frequency pulses affected the product quality (misincorporation rate) more adversely than the low frequency gradients. Additionally, the frequency of the pulses influenced the kinetics of misincorporation of ncBCAA into the proinsulin (Figure 4.12C). At higher pulse frequencies, there was a more rapid misincorporation rate, which showed a saturation tendency about 4 hours after protein induction. The low frequency pulses, on the other hand, led to a more gradual misincorporation rate at the beginning, but this seemed not to reach any saturation levels for all the 6 hours that were monitored after protein induction. For the same cultivation conditions (e.g. 5 min pulse, 0.5 mM IPTG), there was 10 times less norvaline misincorporation in the cultivations with background glucose release than the cultivations without the background glucose. In terms of induction strength, the cultures that were induced to a higher inducer concentration had a lower level of misincorporation of ncBCAA than those with lower IPTG concentration.

Generally, there was a higher concentration of the ncBCAA in the intracellular soluble fraction than in the purified inclusion bodies (Figure 4.12A). For instance, in the 5 min and 10 min glucose pulse cultivations (without background glucose feeding), there were respectively 5 fold and 4 fold higher norvaline in the intracellular soluble fraction than in the purified inclusion bodies. The norvaline-norleucine ratio in the inclusion bodies from the 5 min pulse cultivations was about 50 times higher than in the reference cultivation, whereas the 10 min pulse cultivations had about 30 fold higher norvaline misincorporation levels compared to the reference cultivation (Figure 4.12).

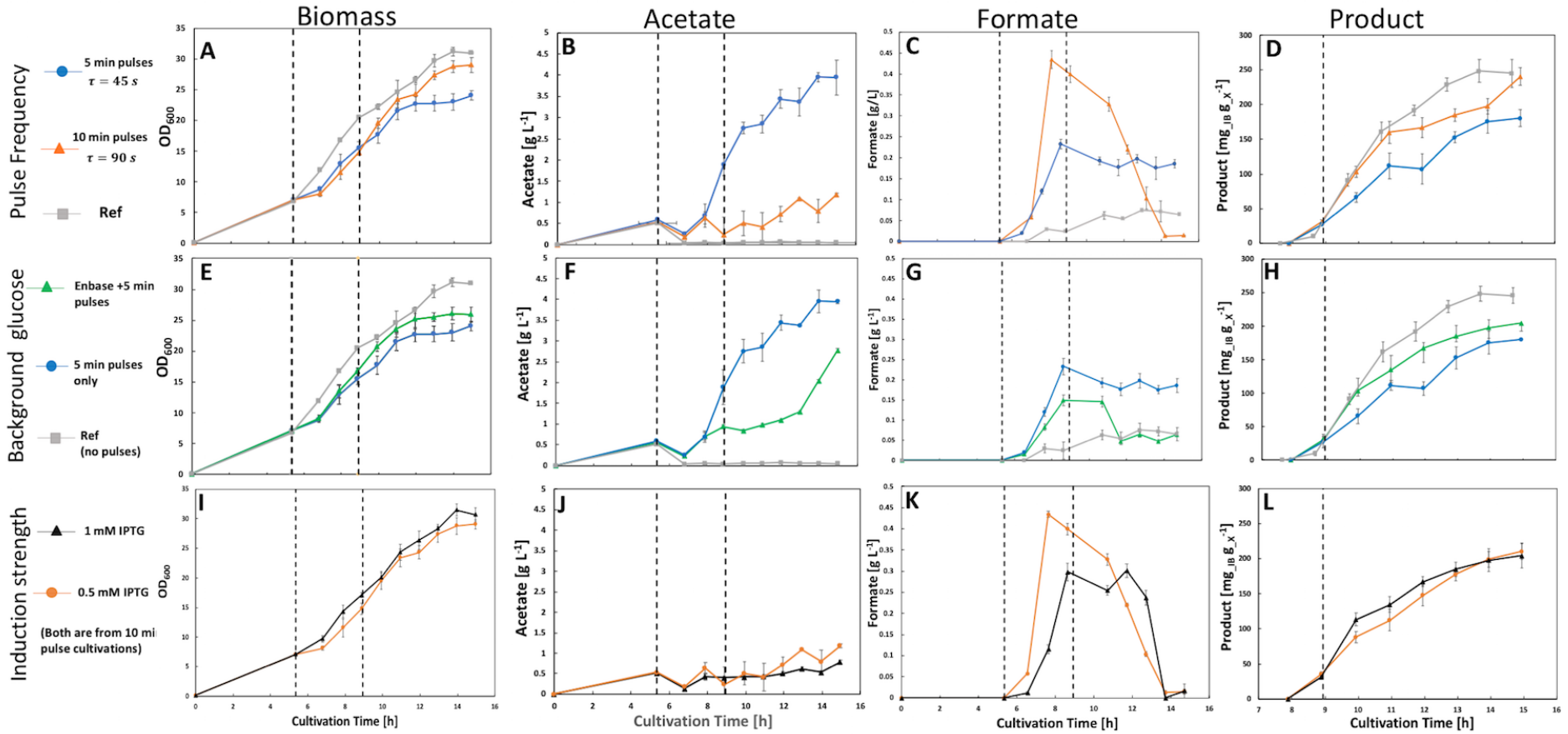
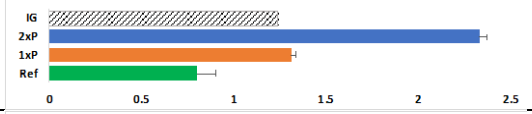
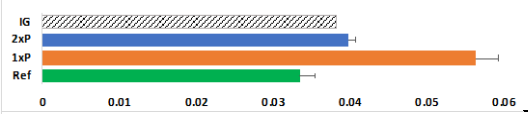
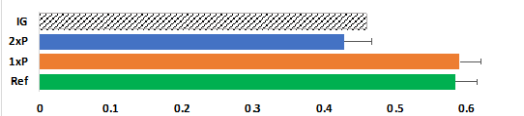
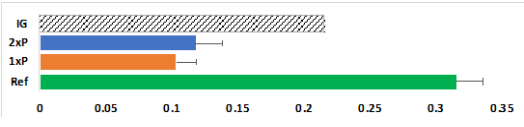
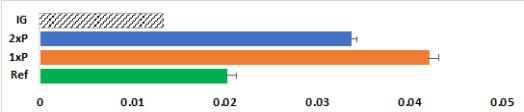
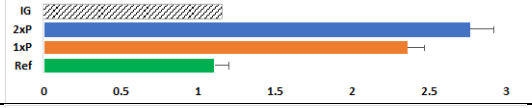
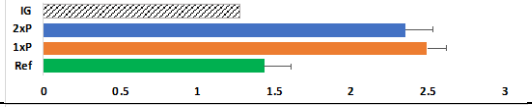
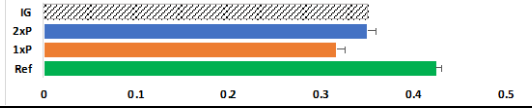


Figure 4.13 Influence of glucose pulse frequency, induction strength and background enzymatic glucose release on growth, metabolite accumulation and the formation of recombinant proinsulin during fed-batch cultivations of *E. coli* BW25113 in minibioreactors. The broken line indicates the beginning and end of exponential feeding fed-batch phase. Recombinant protein production was induced at the end of the exponential feeding phase.

Table 4.4 Results of parameter estimation for scale-down fed-batch cultivation in miniaturized bioreactors, showing the maximum specific rates and yield coefficients of *E. coli* in response to the induced glucose pulses. IG—initial guesses, which were also used to calculate the feed profiles; 2xP—10 min glucose pulse cultivation; 1xP—5 min glucose pulse cultivation; Ref—reference cultivation. Error bars represent the standard deviation for each parameter estimate, which was calculated using the method of *covariance matrix based on sensitivity matrix*, as described by López et al. 2016 [27]

Parameter	Estimated Values	Notes
q_{Smax}		In agreement with the results of [28] and [29], the *E. coli* strain developed a higher capacity for glucose uptake in the presence of glucose pulses, where the adaptation of the maximum uptake capacity seems to be proportional to the pulse size.
K_S		Both pulse-based cultivations resulted in higher K_s values than the reference cultivation, which implies that the *E. coli* strain developed a lower affinity for glucose under the induced concentration gradients.
p_{Amax}		Interesting to note that the specific maximum acetate production rate is the same in both the 5 min pulse-based and reference cultivations. The difference in acetate profiles must therefore be accounted for by other parameters such as differences in its re-assimilation rates
q_{Amax}		Under homogeneous conditions, there is a higher acetate re-assimilation rate, which results in lower extracellular acetate than under pulse-based conditions. Within the pulses, the re-assimilation rate of acetate in the 5 min pulses is slightly lower than in 10 min pulses which is reflected in the acetate profiles.
q_m		Homogeneous cultivation conditions led to lower maintenance energy. Higher frequency glucose pulses divert more energy towards maintenance than low frequency pulses.
Y_{oa}		Both pulse based cultivations used up more oxygen for oxidation of acetate than in the reference. This yield coefficient is directly linked to the specific oxygen uptake rate (qO) in the model.
Y_{os}		Oxygen requirements for glucose oxidation in the pulse-based cultivation was higher than in reference. This may be linked to the higher maintenance energy requirements under heterogeneous cultivation conditions.
Y_{px}		In terms of recombinant product yield, the pulses had a negative effect on product formation. The reference cultivation had the highest mass of inclusion bodies of proinsulin per gram of biomass.

To investigate the influence of the pulses on the intracellular metabolic fluxes, data from each of the cultivation conditions was fitted to the mechanistic model of *E. coli*, and the values of the key parameters describing glucose uptake, maintenance energy, acetate cycling and recombinant product formation are compared and explained in Table 4.4. The details of model fitting for each condition are presented in Paper IV. As detailed in Table 4.4, the use of the mechanistic model of *E. coli* to design the pulse-based scale-down runs and to re-fit the experimental results to the model allows us to derive more information from the experiments than would otherwise be possible. The calculated metabolic fluxes give greater insight into the response of the strain to the various concentration gradients, which can be used to plan further genetic engineering of the strain. Such intracellular responses cannot be easily deduced from the macroscopic environment (metabolite, biomass and substrate measurements) without the use of a model. On the other hand, taking samples to experimentally measure these fluxes with 'omics' technologies is feasible, but this would be quite time consuming and can only deliver results at a later time after the experimental run. The scale-down experiment, as conducted in the minibioreactor system and presented in this chapter are a significant improvement on the throughput of the scale-down bioreactors, which can be used to test the response of various strains to concentration gradients in parallel.

5. Discussion

5.1. High Throughput Scale-down Cultivation in Microbial Processes

High throughput minibioreactor cultivation platforms have come to stay. In the past decade, there has been an exponential increase in the adoption of these systems for early bioprocess development, for obvious reasons (Bareither and Pollard, 2011; Hemmerich et al., 2018b; Rameez et al., 2014). At the same time, due to Quality-by-Design (QbD) requirements, there has been an increasing demand to fully characterise bioprocesses at the development phase, to forestall unforeseen consequences of the final process, upon scale-up (Herwig et al., 2015; Rathore, 2009). This requirement demands that all conditions, including actual large-scale process conditions are considered and tested in the high throughput minibioreactor development phase. Therefore, the question of whether cultivations in minibioreactors are adaptable to mimic concentration gradients that exist in large-scale bioreactors is one of the key points of this thesis. As demonstrated in the results of the high throughput scale-down cultivations, it is possible to achieve such large-scale conditions in miniaturised bioreactors. Particularly, the physiological response of *E. coli* to the induced glucose pulses in the minibioreactors is a proof of concept that gradient profiles that are relevant in industrial-scale cultivations can be reproduced in millilitre scale for scale-down studies. The adoption of two enabling technological methods, namely, robotic liquid handling stations and mechanistic modelling is fundamental for the successful operation of the facility as a scale-down platform. The specific benefits of such a platform for the development of microbial fermentation processes are discussed in the following sections, in light of the previously postulated research questions (RQ).

5.1.1. Large-scale Physiological Data Generation (RQ3)

Many aspects of bioprocess scale-up require detailed physiological data of the organism and full characterisation of the process conditions (Takors, 2012). Typically, this data is generated from standard experimental runs from *DoE* templates (e.g. central composite, Box-Behnken and factorial designs), which are often not enough for data-intensive developmental tasks, such as for training neural network models and for chemometric modelling. Additionally, the recent advances of scale-down cultivations to include CFD-CRD dynamics (Haringa et al., 2017; Wang et al., 2015) and population balance modelling (F Delvigne et al., 2006) require substantial amount of data under specified stressful cultivation conditions to fit these models. The high throughput scale-down minibioreactor system developed and tested in this work is a suitable cultivation platform that can be used to generate such data. By defining the specific stress conditions in the modelling framework, the cultivation data can give detailed insights into the individual organism lifelines as well as the behaviour of selected population groups with respect to a given stress condition. The coupling of the robotic cultivation facilities to a HT protein analysis system (Labchip GX II Touch) and high throughput metabolite analysis platform (Cedex Bio HT) enable rapid processing of samples, as reported for the performed cultivations. Thus, in the current work, the combination of a mechanistic modelling, high throughput parallel cultivation platform, robotics and precision liquid handling, high throughput downstream processing and advanced data analytics through higher level programming languages (e.g. Matlab) enable the collection of large amount of data for efficient strain characterization and

bioprocess development. Until now, data quantity and quality was a major difficulty in bioprocess development (Sin and Gernaey, 2016), but with the adoption of such systems, steady progress can be made towards faster overall process development.

5.1.2. Screening with Scale-up in Mind (RQ4)

Recently, minibioreactor platforms have been applied both in academic and in industrial settings for high throughput screening of yeasts (Back et al., 2016), bacterial (Velez-Suberbie et al., 2018) and mammalian cell lines (Janakiraman et al., 2015). These facilities provide the opportunity to even combine some upstream bioprocess development concepts (e.g. control strategy, scalability) with strain screening (Janakiraman et al., 2015; Rameez et al., 2014). However, there is no report in the literature on adapting these minibioreactor systems to imitate large-scale cultivation conditions during the screening phase. Cultivations in the perfectly mixed millilitre-scale vessels can yield very different results than when the cultivations are performed in large-scale bioreactors where the conditions are heterogeneous, as seen the results of the current work. Hence, performing strain screening under conditions hich imitate large-scale oscillations is necessary to select the most robust strain for further development. The response of different strains in the presence of induced concentration gradients can be used to fine tune basic strain engineering, such as fixing a promoter that is leaky under stressful cultivation conditions and manipulating the expression of specific metabolic pathway enzymes to tailor the response to stresses. These effects may not be observed and dealt with, if the screening is done under homogenous conditions (as is practised currently), with the consequence that an otherwise non-performing strain may be selected for scale-up. Therefore, the developed framework in the current research aims to pioneer a new screening paradigm, where strain screening is done with scale-up in mind, i.e. by considering and dealing with the possible consequences of bioprocess scale-up at the earliest phase. This *screening with scale-up in mind* concept will help to avoid repetitive and laborious fermentation development phases, since all possible production conditions are covered in the early phases.

However, it is worth noting that the large-scale conditions used in the current study (induced gradients by pulse-based feeding) may not be fully representative of the large-scale. The flow patterns that induce concentration gradients in large-scale bioreactors are complex (Haringa et al., 2018), and therefore cannot be *exactly* reproduced in minibioreactor scales. That notwithstanding, it is our opinion that the stresses induced in the presented cultivations are valuable, especially looking at the significant differences in reponse of *E. coli* in the scale-down cultivations. The induced stress conditions can be used to screen strains/processes for their probable robustness in the industrial scale – i.e. from the experiments we can derive useful hypotheses on how a strain would approximately behave in the large scale.

5.1.3. Need for Speed: Faster Bioprocess Development (RQ3)

According to Neubauer et al. the lead times of biotechnological products, especially biopharmaceuticals, from discovery to market, can take up to 15 years (Neubauer et al., 2017). Although other issues such as clinical trials may contribute to this time, bioprocess development and troubleshooting scale-up problems are key contributors to the lengthy lead times. The use of parallel cultivation systems has, undoubtedly, reduced these process development times significantly

(Bareither and Pollard, 2011; Hemmerich et al., 2018b). For instance, seven different cultivation conditions were investigated in the scale-down runs in this work, in which each cultivation lasted for 15 hours. To investigate the response of one *E. coli* strain to all these seven (7) different conditions in a standard two-compartment bioreactor, one would need about three months to finish these runs if only duplicate runs of each condition are performed. However, with the use of the high throughput minibioreactor platform, triplicate runs of each condition were accomplished in a 15-hour period. Prior to screening, the development of strains is nowadays done in a high throughput manner, e.g. with the use of standardized genetic methods (de Lorenzo and Schmidt, 2018) and non-targeted high-throughput strain engineering (Schallmey et al., 2014). Thus, the bottleneck of a faster overall bioprocess development is shifted from strain engineering to screening and cultivation development. The use of parallelized minibioreactor systems for both screening and upstream process development, as demonstrated in the current work, relieves this bottleneck, and ensures that a potential bioprocess reaches production within the earliest possible times. Additionally, the presented framework and methods not only facilitate a rapid bioprocess development, but it also ensures a consistent and efficient cultivation process development by taking into account all the possible cultivation conditions that would be encountered upon process scale-up, as discussed above in Section 5.1.2. Therefore, both speed and efficiency are incorporated into a single cultivation platform for successful bioprocess development.

5.2. Response of *E. coli* to Glucose and Dissolved Oxygen Gradients (RQ5)

Two major types of concentration gradients i.e. substrate and dissolved oxygen gradients, that occur simultaneously (although inversely) in large-scale microbial cultivations were investigated in the current research. Previous research had demonstrated that *E. coli* cells exposed to these gradients respond by showing impaired growth and low yields (Bylund et al., 1998; Han and Lee, 2006; Neubauer et al., 1995). Additionally, under conditions of oscillating concentrations, the recombinant product formation profile is also affected (Soini et al., 2011a), as well as the accuracy of the transcription-translation mechanisms in recombinant protein processes (Biermann et al., 2013; Bylund et al., 2000). All these previous observations are also confirmed in the current research in the presence of oscillating glucose and dissolved oxygen availability. Whereas all the previous observations were made by applying scale-down techniques in multi-compartment or pulse-based cultivations in larger bioreactors (Neubauer et al., 1995; Soini et al., 2008), the gradients and the response effects in the current work are reproduced using millilitre-scale bioreactors, in a high throughput framework. This advancement of scale-down cultivations enables the study of multiple strains under large-scale cultivation conditions.

5.2.1. Physiological Responses to Oxygen and Glucose Gradients

In terms of physiology, *E. coli* showed similar response to glucose pulses in both the bench top and minibioreactor cultivations. Specifically, the repeated gradients of glucose, which were associated with intermittent oxygen limitation, resulted in reduced biomass accumulation rate of the culture compared to the reference cultivations. Consequently, biomass formation, recombinant product

formation and yields (Y_{xs}, Y_{px}) were all lower, whereas maintenance energies (obtained from model fitting) were significantly higher in the pulse-based scale-down cultivations. Despite the lower biomass and lower yield coefficients, the cells in the pulse-based cultivations had higher uptake capacities (qs, qo), with a corresponding higher specific growth rate (μ) in the fed-batch phase than the reference cultivations, especially in the minibioreactor scale-down runs. Thus, previous observations that *E. coli* cells under stressful cultivation conditions are more robust than those growing in homogenous conditions (Brand et al., 2018; Hewitt et al., 2000) were also confirmed in the scale-down runs in this work. The intermittent exposure of the cells to higher concentrations of glucose makes the signalling system or the feedback control loop of the glucose uptake system (phosphotransferase system (PTS)) to remain active, even in the absence of glucose within the short times between pulses. The whole uptake mechanism, including permease proteins and all related enzymes are maintained in more active states under oscillating concentrations (Buhr et al., 1994). This may explain why the cells under scale-down conditions seem to have a higher viability (in terms of μ, qS, qO), although with a higher maintenance energy requirements, as shown in the results of the parameter estimation of the pulse-based cultivations (Paper III and Paper IV). In effect, the higher viability due to the higher uptake capacities under oscillating conditions does not necessarily translate into higher biomass or higher recombinant product formation rates. Therefore, in large-scale fed-batch cultivations when *E. coli* cells circulate from glucose-rich zones to glucose limiting zones, it may lead to a rather higher consumption of glucose due to higher uptake capacities (Brand et al., 2018; Neubauer et al., 1995), but this higher consumption rather pushes more carbon substrate towards products of overflow metabolism, which eventually leads to lower biomass and product yields in recombinant protein production processes.

5.2.2. Metabolic Responses to Oxygen and Glucose Gradients

The macroscopic (physiological) response of the strain to the induced gradients is the sum total of sub-cellular level responses, such as the activation of enzymes that regulate specific fluxes in the metabolic routes under certain stressful conditions. Although metabolomic analysis was not done in the current work, the mechanistic model provided an avenue to simulate the relevant fluxes based on the usage of glucose and acetate production profiles. Therefore, the detailed view of the metabolic response of the strain to the induced concentration gradients was only possible through the model simulations. Such inferential methods have been widely used in the literature, e.g. for the prediction of intracellular pyruvate concentrations in *E. coli* (Zelić et al., 2006), for prediction of acetate concentrations (Ko et al., 1994) and for general metabolic flux balance analysis (Varma and Palsson, 1994). After fitting the mechanistic model to the data of the pulse-based scale-down cultivations and the reference cultivation, the calculated fluxes of overflow routes, maintenance and product formation were compared (Paper IV). Under oscillating conditions, intracellular fluxes of both the glucose partitioning system and acetate cycling routes (Paper I) were significantly higher than in reference cultivations, although the same feed rates were applied (constant μ_{set}). As discussed above, this higher intracellular fluxes led to higher overall physiological activity for the pulse-based cultivations than in the reference cultivations.

5.2.3. Effect of Concentration Gradients on Recombinant Product Quality

One of the most important physiological responses of *E. coli* to the induced concentration gradients in the current work is the integrity of the amino acid composition of the recombinant proinsulin. Earlier

research showed that non-canonical amino acids accumulate in both the cellular material and recombinant protein under rapid shifts in environmental conditions around the cell (Harris and Kilby, 2014; Lara et al., 2006; Soini et al., 2011a).As shown in the results of the pulse-based cultivations in the minibioreactors, norvaline was misincorporated into the recombinant protein, in place of leucine, as reported by other researchers (Fenton et al., 1997; Reitz et al., 2018). An additional observation from the cultivations of the minibioreactors is the influence of the pulse frequency on the kinetics of this misincorporation. At lower frequency, there is an initial slow misincorporation rate, which then increases exponentially at ca. 3 h after induction. This may be attributed to the lower concentrations of intracellular pyruvate at the beginning of the low frequency pulses due to the longer cycle times combined with lower feed rates, which allows pyruvate pools to be depleted before the next glucose pulse is given (Sunya et al., 2013, 2012). On the contrary, in the presence of rapid pulses, there is quick build-up of intracellular pyruvate which is not consumed before the next pulse is given, as can be inferred from the extracellular acetate concentrations in the pulse-based cultivations in the minibioreactors. The ncBCAAs are derived from pyruvate (Soini et al., 2011b). Since pyruvate accumulation occurs under oscillating cultivating conditions (Swartz, 2001), it is not surprising that dissolved oxygen and glucose gradients in large scale *E. coli* cultivations and their corresponding prototypes in the pulse-based scale-down systems resulted in ncBCAAs misincorporation into the proinsulin. This may have negative consequences for the clinical use of therapeutic proteins produced in *E. coli* under conditions of concentration gradients (Wong et al., 2018).

5.2.4. Implications for *E. coli* Process Design, Scale-up and Operation

The overall consequence of concentration gradients on microbial physiology, metabolic states and recombinant protein quantity and quality as reported in the current work means that production scale bioreactors need to be designed and operated with care. The methods employed here mostly create continuous stress-relief (glucose excess-glucose limitation) conditions, where the total microbial population is exposed to the pulse at a given time without creating population sub-sets as may be the case in actual large-scale bioreactors. Nevertheless, the results (yield losses, misincorporation of ncBCAAs, accumulation of metabolites, etc) are similar to previous investigations of scale-up effects in other cultivation set-ups (Junne et al., 2011; Lara et al., 2009; Soini et al., 2011a). This means that the responses of *E. coli* to the induced gradients are quite universal and may also be seen in large industrial-scale cultivations, although actual industrial data from big biopharmaceutical companies showing these effects is not available. The results of the current work can lead to two major directions in further bioprocess development: (i) robust strain engineering to better cope with the induced concentration gradients, as observed by Fuentes et al., (2013). This may include adjusting the sensitivity of the strain to limiting glucose conditions, i.e. modifying Ks to higher values or adjusting qS to consume glucose more slowly even in the presence of excess glucose concentrations or regulating the expression level of genes that code for enzymes responsible for the synthesis of ncBCAAs in *E. coli.* The high throughput scale-down system developed in the current study opens up the possibility to efficiently and quickly test multiple stress effects in such engineered strains to establish the viability of the genetic modifications. (ii) Advanced scale-up studies and optimization of production systems before actual scale-up is done. In that case, specific strain responses to model-derived pulses can be used to define tailored feed profiles, agitation rates and set-points for dissolved oxygen tension for operating the large-scale bioreactor.

5.3. Application of Mechanistic Models in Bioprocess Development (RQ1)

Based on the results of recent research (Enjalbert et al., 2017; Valgepea et al., 2010), an old mechanistic model of *E. coli* was updated to improve its mathematical stability and the continuity of the ODE system over the whole range of cultivation conditions (excess glucose, limiting glucose and starvation conditions). Such improvements were necessary to apply the model to advanced gradient-based methods such as sensitivity and uncertainty analyses. Compared to other models of *E. coli* reported in the literature (Machado et al., 2014), the updated model does not have the hard metabolic switches (e.g. from oxidative to overflow, onset of mixed acid fermentation pathways, etc) that are triggered at some (unknown) time points during the cultivation. But rather, the model has subtle dynamic equilibrium points, by which concentrations of overflow metabolites can change in either direction (towards overflow or oxidative) when the equilibrium points are offset (Anane et al., 2017). The results of fitting the improved model to experimental data showed that the description was more accurate, enabling the 3-dimensional derivation of overflow profile as a continuous process which depends not only on the residual glucose and dissolved oxygen concentrations, but also on intracellular and extracellular acetate concentrations.

5.3.1. Physiological Accuracy and the Parsimony Principle (Ockham's Razor)

Usually, the accuracy of a mechanistic model in describing an observed biological phenomenon is associated with the number of parameters used to build the model. Most of the parameters used in these models are derived from Monod or Michaelis-Menten type equations, which are gradual switching functions that describe various biological states (e.g. growth) under varying environmental conditions. The parameters of these kinetic models may be expanded further by considering concepts such as growth inhibition (inhibition by metabolites, product, substrate) (Andrews, 1968), membrane transport mechanisms (Calleja et al., 2014) and plasmid copy numbers for recombinant products, among others. But to what extent should these models be detailed, in order to be useful? The answer to this question depends on the purpose of the model. As pointed out by Gabor and Banga (2017), the parsimony principle should always be applied in the model building: i.e. the number of parameters should not be more than those required to describe the process in its simplest form (Gábor and Banga, 2015; Rollié et al., 2012). This principle is usually referred to as 'Ockham's razor' in the literature (Omlin and Reichert, 1999). In the *E. coli* model developed and applied in the current work, only 17 parameters were used to describe growth, cellular maintenance, overflow metabolism and the associated acetate production, substrate and dissolved oxygen consumption, mass transfer and recombinant product formation. Thus, the number of parameters were reduced to the minimum possible, required to maintain an accurate description of the physiology of the strain and the process. Based on the results of the model fitting and the follow-up applications in the scale-down design, it can be deduced that the parametrisation of the model was good enough, for the intended purpose of calculating concentration gradients for physiological studies in scale-down cultivations. Thus, the question of the level of detail required in a model (Research Question 1) is directly linked to the purpose of the model. The results also show that constantly updating a model to include the most recent research findings in the field can have enormous benefits for both physiological accuracy and mathematical stability of the model.

5.3.2. Reliability of the Model Predictions

One of the natural consequences of the parameterisation of a model is the level of parameter estimation error associated with the model, which may arise from the quality and quantity of the available data in relation to the number of parameters to be estimated (Guisasola et al., 2006; Lübbert and Simutis, 1994). Despite the parsimony principle applied in the building of the model of *E. coli*, advanced analysis of the identifiability of the model parameters (Paper II) shows that not all parameters could be uniquely estimated from the given fed-batch cultivation data. According to Almquist et al., this non-identifiability of model parameters is one of the major factors hampering the use of models in biotechnology (Almquist et al., 2014). As shown in the results (Section 4.1.4), such models tend to have wide prediction bands and wide confidence intervals for model predictions. Since mechanistic models are specifically built to predict microbial behaviour out of the current experimental space, the uncertainty in predictions due to non-identifiability is a special problem that needs to be addressed (Chowdhury et al., 2015). The principles of regularization applied to the *E. coli* model in the current research are key tools to help increase the prediction fidelity of such non-identifiable models. The special contribution of this work in this regard is the simplification of the otherwise mathematically complex methods, to make regularization techniques usable by non-experts in modelling, such as biotechnologists whose training does not cover such advanced mathematics (Muñoz-Tamayo et al., 2018). Furthermore, the generalized framework for advanced uncertainty and identifiability analysis presented in Paper II summarizes the development steps of a typical mechanistic model in bioprocesses. We hope that the further application of this framework to models of other expression systems (*S. cerevisiae*, CHO cells, *C. glutamicum,* etc) will help narrow the confidence intervals of predicted outputs and make these models more useful in bioprocess development.

5.3.3. Application to Scale-down Design and Operation

Two mechanistic models were pivotal for the operation of the high throughput scale-down system in this work: (i) a relatively accurate and parsimonious model of the strain and (ii) a mechanistic description of the gradient profiles of the 2CR that were used as a guide to design the pulse-based system. The ability to integrate the outputs of these two models into the operation schemes of liquid handling stations to reproduce the heterogenous conditions was key to the success of the scale-down cultivations in the minibioreactors. The success of such experiments is important for further development of model-based approaches in bioprocess development. Until now, most research works involving models in biotechnology have focused on deriving deeper understanding into microbial behaviour in the bioreactor (Almquist et al., 2014), whilst a few extend to monitoring and control of bioprocesses (Mears et al., 2017). However, novel applications such as the methods presented in the current research (model-based stress definition for scale-down studies) and other similar works (model-based online optimization of experiments (Cruz Bournazou et al., 2017)) point in the future direction where mathematical methods will help to design more informative and smart experiments, to move away from the traditional, commonly used design of experiment (DoE) paradigm. The design and implementation of intelligent experiments based on modelling methods is promising, and has the potential to further reduce the development time of fermentation processes when incorporated with high throughput cultivation platforms. This research is a typical example of how the application of mathematical models in the right framework can help to facilitate the digital revolution in the field of industrial biotechnology and bioprocess engineering, as discussed by Neubauer et al. (2017).

6. Conclusions

We successfully developed a high throughput scale-down cultivation platform and showed for the first time that concentration gradients that exist in large industrial-scale bioreactors can be reproduced in millilitre-scale minibioreactors for the purpose of scale-down studies. This platform was achieved by the successful integration of different concepts in bioprocess engineering, namely, mechanistic modelling of strain physiology (growth, overflow, product formation, inhibition), mechanistic modelling of bioreactor operation modes (fed-batch STR, plug-flow and multi-faceted feed systems), high throughput minibioreactor systems and robotic liquid handling stations. This integrated platform is a step in advancing bioprocess development towards digitalization, into the *Industry 4.0* era, which has the potential to shorten the currently long lead times in biotechnology. Furthermore, the digitalization of microbial cultivations will allow more efficient and once-and-for-all (non-repetitive) bioprocess development since more intelligent decisions can be made within one experimental run. The work presented here and similar research reports form a strong bases for this digital transformation in bioprocess engineering.

Parallel millilitre-scale cultivations in the high throughput platform, as well as bench top scale-down cultivations enabled us to further confirm previous observations about the negative influences of dissolved oxygen and glucose gradients on *E. coli* physiology. In the presence of these gradients, our results show a marked reduction in biomass and recombinant protein (product) yields, as well as an increase in the accumulation of extracellular metabolites. Additionally, the induced gradients also had a negative influence on recombinant product quality, measured here as the level of misincorporation of ncBCAAs into the recombinant protein product. These observations, which are also present in actual large-scale *E. coli* cultivations prove that the model-based minibioreactor scale-down system is sufficient for the study of the effects of scale-up related gradients on bioprocess efficiency.

Lastly, the mechanistic modelling techniques applied in this work underscore the importance of models for bioprocess development. The creation of the gradient profiles in the millilitre-scale bioreactors was only possible through the model-based simulations of the physiology of *E. coli* and accurate mathematical description of the whole cultivation system. This, in turn, requires that the models used should meet certain quality criteria, such as establishing the level of uncertainty associated with the model predictions and thoroughly analysing the identifiability of model parameters with regards to the given data. Unfortunately, such detailed model analyses are usually not performed in bioprocess engineering. Therefore, a unique contribution of this work is to open up the detailed model analysis methods to the bioprocess engineering community, to create the necessary awareness about the existence of these methods and to make their application simple for scientists and engineers with limited training in mathematics.

7. Outlook

The high throughput scale-down system developed in this work may be exploited further for the rapid development of bioprocesses under typical large-scale cultivation conditions, taking into account the influence of these conditions on the physiology of the organism and the quality of the recombinant protein product. Future applications of this platform to multiple strains in parallel would allow the selection of not only the best performing strain in terms of growth and productivity, but also the most robust strain for bioprocess scale-up. In such future applications, the following concepts and ideas should be explored for a more thorough bioprocess understanding:

1. The mathematical modelling concepts of the physiology of *E. coli* (e.g. glucose partitioning, overflow through metabolite cycling, metabolic burden of recombinant product, etc.) that were developed in this work should be extended further to cover other relevant microbial expression systems, such as *C. glutamicum* and *S. cerevisiae.* This would enable a more accurate and detailed quantitative description of the strain during growth, which is a pre-requisite for the high throughput scale-down cultivation platform to study scale-up effects in a given organism.
2. The model analysis tools, namely, identifiability and sensitivity analysis techniques that were developed and applied in this work should be applied to already existing mechanistic models in biotechnology. This would ensure more reliable model predictions and a knowledge of the level of uncertainty associated with the outputs of these models, so as to facilitate the use of models in biotechnology, in general.
3. It would be very useful to include advanced proteomics, metabolomics and transcriptomic analyses of samples from the scale-down cultivations to be able to tell the impact of concentration gradients at the molecular level. The methods and conclusions drawn in this work are mostly based on observations in the macroscopic environment (except for recombinant protein quality). However, the detailed analysis of molecular level response will be needed if the results of the parallel scale-down cultivations are to be used for genetic manipulation of expression systems.
4. The calculation of concentration gradients in the scale-down platform using the mechanistic model can be performed in an adaptive manner. The concentration gradients that were induced in the current work were calculated once-off, based on the prevailing conditions (biomass concentration, acetate, growth rate set-point in fed-batch phase, etc.) that were measured at-line, at the beginning of the fed-batch phase. However, in an adaptive framework, new data collected in the course of the cultivation can be fitted to the model, and the gradients re-calculated to suit a pre-determined gradient profile. This has the benefit of reducing strain adaptation to the given gradients, which may have an influence on the response of the strain to the pulses.
5. Simulations from other scale-down models, such as CFD and/or CFD-CRD models can be used to determine more accurate mixing times of larger bioreactors. Furthermore, actual measurements of mixing times in industrial-scale bioreactors should be conducted. The more accurate the mixing time, the better the scale-down design. Therefore, prior to the operation of this developed platform for further scale-down studies, more accurate large-scale data with regards to mixing effects should be provided. This information can then be used in the modelling framework to determine the frequency and magnitude of concentration gradients for scale-down studies.

8. Theses

1. Concentrations gradients affect the physiology of *E. coli*, including the quality of recombinant protein products.

2. Non-canonical branched chain amino acids accumulate in general cellular material *E. coli* in the presence of glucose and dissolved oxygen gradients.

3. Scale-down bioreactors provide an efficient way of investigating concentration gradients.

4. Concentration gradients that occur in large-scale bioreactors can be reproduced in millilitre-scale bioreactors for purposes of scale-down studies.

5. Mechanistic models of biological systems can be highly unreliable in their predictions and should therefore be subjected to appropriate reliability tests.

6. Digitalization and *Industry 4.0* concepts can significantly reduce bioprocess development times.

7. Mechanistic models that were developed in the past should continuously be updated with recent research findings.

8. The incorporation of scale-down studies in early bioprocess development will help to achieve the quality-by-design requirements; where optimal quality of each process step is ensured, and not only the quality of the final product in large-scale.

9. Quantitative modelling can be used to determine the level of growth inhibition induced by the production of foreign proteins in recombinant microorganisms.

10. Non-depletion of metabolite pools by continuous supply of limiting background glucose has positive influence on recombinant protein quality, even in the presence of concentration gradients

9. References

Almquist J, Cvijovic M, Hatzimanikatis V, Nielsen J, Jirstrand M, 2014. Kinetic models in industrial biotechnology – Improving cell factory performance. Metab. Eng. 24, 38–60.

Anane E, López C DC, Neubauer P, Cruz Bournazou MN, 2017. Modelling overflow metabolism in *Escherichia coli* by acetate cycling. Biochem. Eng. J. 125, 23–30.

Anane E, Sawatzki A, Neubauer P, Cruz-Bournazou MN, 2018. Modelling concentration gradients in fed-batch cultivations of *E. coli* -towards the flexible design of scale-down experiments. J. Chem. Technol. Biotechnol.

Andrews JF, 1968. A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates. Biotechnol. Bioeng. 10, 707–723.

Apostol I, Levine J, Leach J, Hess E, Glascock CB, Weickert MJ, Blackmore R, Lippincott J, 1997. Incorporation of Norvaline at Leucine Positions in Recombinant Human Hemoglobin Expressed in *Escherichia coli*. J. Biol. Chem. 272, 28980–28988.

Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG, 2009. Systems biology: Parameter estimation for biochemical models. FEBS J. 276, 886–902.

Back A, Rossignol T, Krier F, Nicaud J-M, Dhulster P, 2016. High-throughput fermentation screening for the yeast Yarrowia lipolytica with real-time monitoring of biomass and lipid production. Microb. Cell Fact. 15, 147.

Baert J, Delepierre A, Telek S, Fickers P, Toye D, Delamotte A, Lara AR, Jaén KE, Gosset G, Jensen PR, Delvigne F, 2016. Microbial population heterogeneity versus bioreactor heterogeneity: Evaluation of Redox Sensor Green as an exogenous metabolic biosensor. Eng. Life Sci. 16, 643–651.

Baez A, Flores N, Bolívar F, Ramírez OT, 2011. Simulation of dissolved CO 2 gradients in a scale-down system: A metabolic and transcriptional study of recombinant *Escherichia coli*. Biotechnol. J. 6, 959–967.

Baez A, Flores N, Bolívar F, Ramírez OT, Bolı F, Ramı OT, 2009. Metabolic and transcriptional response of recombinant *Escherichia coli* to elevated dissolved carbon dioxide concentrations. Biotechnol. Bioeng. 104, 102–110.

Baez A, Shiloach J, 2014. Effect of elevated oxygen concentration on bacteria, yeasts, and cells propagated for production of biological compounds. Microb. Cell Fact. 13, 1–7.

Bailey JE, Da Silva NA, Peretti SW, Seo J-H, Srienc F, 1986. Studies of Host-Plasmid Interactions in Recombinant Microorganisms. Ann. N. Y. Acad. Sci. 469, 194–211.

Bareither R, Pollard D, 2011. A review of advanced small-scale parallel bioreactor technology for accelerated process development: current state and future need. Biotechnol. Prog. 27, 2–14.

Barz T, Sommer A, Wilms T, Neubauer P, Cruz Bournazou MN, 2018. Adaptive optimal operation of a parallel robotic liquid handling station. IFAC-Papers OnLine 51, 765–770.

Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, Hwa T, 2015. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. Nature 528, 99–104.

Baumann P, Hahn T, Hubbuch J, 2015. High-throughput micro-scale cultivations and chromatography modeling: Powerful tools for integrated process development. Biotechnol. Bioeng. 112, 2123–2133.

Baumann P, Hubbuch J, 2017. Downstream process development strategies for effective bioprocesses: Trends, progress, and combinatorial approaches. Eng. Life Sci. 17, 1142–1158.

Bellu G, Saccomani MP, Audoly S, D'Angiò L, 2007. DAISY: A new software tool to test global identifiability of biological and physiological systems. Comput. Methods Programs Biomed. 88, 52–61.

Bernal V, Castaño-Cerezo S, Cánovas M, 2016. Acetate metabolism regulation in *Escherichia coli*: carbon overflow, pathogenicity, and beyond. Appl. Microbiol. Biotechnol. 100, 8985–9001.

Bhambure R, Kumar K, Rathore AS, 2011. High-throughput process development for biopharmaceutical drug substances. Trends Biotechnol. 29, 127–135.

Biermann M, Linnemann J, Knüpfer U, Vollstädt S, Bardl B, Seidel G, Horn U, 2013. Trace element associated reduction of norleucine and norvaline accumulation during oxygen limitation in a recombinant *Escherichia coli* fermentation. Microb. Cell Fact. 12, 116.

Bockisch A, Biering J, Päßler S, Vonau W, Junne S, Neubauer P, 2014. In Situ Investigation of the Liquid Phase in Industrial Yeast Fermentations with Mobile Multiparameter Sensors. Chemie Ing. Tech. 86, 1582–1582.

Brand E, Junne S, Anane E, Cruz-Bournazou MN, Neubauer P, 2018. Importance of the cultivation history for the response of *Escherichia coli* to oscillations in scale-down experiments. Bioprocess Biosyst. Eng. 41, 1305–1313.

Brognaux A, Francis F, Twizere JC, Thonart P, Delvigne F, 2014. Scale-down effect on the extracellular proteome of *Escherichia coli*: Correlation with membrane permeability and modulation according to substrate heterogeneities. Bioprocess Biosyst. Eng. 37, 1469–1485.

Buchholz J, Graf M, Freund A, Busche T, Kalinowski J, Takors R, 2014. CO_2/HCO_3^- perturbations of simulated large scale gradients in a scale-down device cause fast transcriptional responses in *Corynebacterium glutamicum*. Appl. Microbiol. Biotechnol. 98, 8563–8572.

Buhr A, Fliikiger K, Ernis B, 1994. The Glucose Transporter of *Escherichia coli*: Overexpression, Purification and Characterization of Functional Domains. J. Biol. Chem. Chemispry 269, 23437–23443.

Bylund F, Castan A, Mikkola R, Veide A, Larsson G, 2000. Influence of scale-up on the quality of recombinant human growth hormone. Biotechnol. Bioeng. 69, 119–128.

Bylund F, Collet E, Enfors SO, Larsson G, 1998. Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. Bioprocess Eng. 18, 171–180.

Bylund F, Guillard F, Enfors SO, Trägârdh C, Larsson G, 1999. Scale down of recombinant protein production: A comparative study of scaling performance. Bioprocess Eng. 20, 377–389.

Calleja D, Fernández-Castañé A, Pasini M, de Mas C, López-Santín J, 2014. Quantitative modeling of inducer transport in fed-batch cultures of *Escherichia coli*. Biochem. Eng. J. 91, 210–219.

Calleja D, Kavanagh J, de Mas C, López-Santín J, 2016. Simulation and prediction of protein production in fed-batch *E. coli* cultures: An engineering approach. Biotechnol. Bioeng. 113, 772–782.

Carneiro S, Ferreira EC, Rocha I, 2013. Metabolic responses to recombinant bioprocesses in *Escherichia coli*. J. Biotechnol. 164, 396–408.

Carnicer M, Ten Pierick A, van Dam J, Heijnen JJ, Albiol J, van Gulik W, Ferrer P, 2012. Quantitative metabolomics analysis of amino acid metabolism in recombinant Pichia pastoris under different oxygen availability conditions. Microb. Cell Fact. 11, 83.

Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M, 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnol. Bioeng. 79, 53–73.

Chen R, 2012. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. Biotechnol. Adv. 30, 1102–1107.

Chis OT, Banga JR, Balsa-Canto E, 2011. Structural identifiability of systems biology models: A critical comparison of methods. PLoS One 6.

Choi JH, Keum KC, Lee SY, 2006. Production of recombinant proteins by high cell density culture of *Escherichia coli*. Chem. Eng. Sci. 61, 876–885.

Chowdhury A, Khodayari A, Maranas CD, 2015. Improving prediction fidelity of cellular metabolism with kinetic descriptions. Curr. Opin. Biotechnol. 36, 57–64.

Cockshott A, Bogle I, 1999. Modelling the effects of glucose feeding on a recombinant E-coli fermentation. Bioprocess Eng. 20, 83–90.

Cruz Bournazou MN, Barz T, Nickel DB, Lopez Cárdenas DC, Glauche F, Knepper A, Neubauer P, 2017. Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities. Biotechnol. Bioeng. 114, 610–619.

Cvetesic N, Palencia A, Halasz I, Cusack S, Gruic-Sovulj I, 2014. The physiological target for LeuRS translational quality control is norvaline. EMBO J. 33, 1639–1653.

Cvetesic N, Semanjski M, Soufi B, Krug K, Gruic-Sovulj I, MacEk B, 2016. Proteome-wide measurement of non-canonical bacterial mistranslation by quantitative mass spectrometry of protein modifications. Sci. Rep. 6, 1–13.

de Lorenzo V, Schmidt M, 2018. Biological standards for the Knowledge-Based BioEconomy: What is at stake. N. Biotechnol. 40, 170–180.

De Mey M, De Maeseneire S, Soetaert W, Vandamme E, 2007. Minimizing acetate formation in *E. coli* fermentations. J. Ind. Microbiol. Biotechnol. 34, 689–700.

Delvigne F, Baert J, Sassi H, Fickers P, Grünberger A, Dusny C, 2017a. Taking control over microbial populations: Current approaches for exploiting biological noise in bioprocesses. Biotechnol. J. 12.

Delvigne F, Destain J, Thonart P, 2006. A methodology for the design of scale-down bioreactors by the use of mixing and circulation stochastic models. Biochem. Eng. J. 28, 256–268.

Delvigne F, Goffin P, 2014. Microbial heterogeneity affects bioprocess robustness: Dynamic single-cell analysis contributes to understanding of microbial populations. Biotechnol. J. 9, 61–72.

Delvigne F, Lejeune A, Destain J, Thonart P, 2006. Modelling of the substrate heterogeneities experienced by a limited microbial population in scale-down and in large-scale bioreactors. Chem. Eng. J. 120, 157–167.

Delvigne F, Takors R, Mudde R, van Gulik W, Noorman H, 2017b. Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond. Microb. Biotechnol. 10, 1267–1274.

Donovan RS, Robinson CW, Click BR, 1996. Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. J. Ind. Microbiol. 16, 145–154.

Enfors SO, 2011. Fermentation Process Engineering. Hogskoletryckeriet, Stockholm.

Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen B, Krüger E, Schweder T, Hamer G, O'Beirne D, Noisommit-Rizzi N, Reuss M, Boone L, Hewitt C, McFarlane C, Nienow A, Kovacs T, Trägårdh C, Fuchs L, Revstedt J, Friberg PC, Hjertager B, Blomsten G, Skogman H, Hjort S, Hoeks F, Lin HY, Neubauer P, van der Lans R, Luyben K, Vrabel P, Manelius A, 2001. Physiological responses to mixing in large scale bioreactors. J. Biotechnol. 85, 175–85.

Enjalbert B, Cocaign-Bousquet M, Portais JC, Letisse F, 2015. Acetate exposure determines the diauxic behavior of *Escherichia coli* during the glucose-acetate transition. J. Bacteriol. 197, 3173–3181.

Enjalbert B, Millard P, Dinclaux M, Portais J, Létisse F, 2017. Acetate fluxes in *Escherichia coli* are determined by the thermodynamic control of the Pta-AckA pathway. Nat. Publ. Gr. 1–11.

Fahnert B, Lilie H, Neubauer P, 2004. Inclusion bodies: formation and utilisation. Adv. bioc. engi./bio. 89, 93–142.

Fenton D, Lai H, Lu H, Mann M, Tsai L, 1997. Control of norleucine incorporation into recombinant proteins. US Pat. 5,599,690.

Formenti LR, Nørregaard A, Bolic A, Hernandez DQ, Hagemann T, Heins A-L, Larsson H, Mears L, Mauricio-Iglesias M, Krühne U, Gernaey K V., 2014. Challenges in industrial fermentation technology research. Biotechnol. J. 9, 727–738.

Fuentes LG, Lara AR, Martínez LM, Ramírez OT, Martínez A, Bolívar F, Gosset G, 2013. Modification of glucose import capacity in *Escherichia coli*: physiologic consequences and utility for improving DNA vaccine production. Microb. Cell Fact. 12, 42.

Gábor A, Banga JR, 2015. Robust and efficient parameter estimation in dynamic models of biological systems. BMC Syst. Biol. 9.

George S, Larsson G, Enfors SO, 1993. A scale-down two-compartment reactor with controlled substrate oscillations: Metabolic response of *Saccharomyces cerevisiae*. Bioprocess Eng. 9, 249–257.

Gernaey K V., Gani R, 2010. A model-based systems approach to pharmaceutical product-process design and analysis. Chem. Eng. Sci. 65, 5757–5769.

Gernaey K V., Lantz AE, Tufvesson P, Woodley JM, Sin G, 2010. Application of mechanistic models to fermentation and biocatalysis for next-generation processes. Trends Biotechnol. 28, 346–354.

Glauche F, Pilarek M, Bournazou MNC, Grunzel P, Neubauer P, 2017. Design of experiments-based high-throughput strategy for development and optimization of efficient cell disruption protocols. Eng. Life Sci. 17, 1166–1172.

González-Cabaleiro R, Mitchell AM, Smith W, Wipat A, Ofiteru ID, 2017. Heterogeneity in pure microbial systems: Experimental measurements and modeling. Front. Microbiol. 8, 1–8.

Gschwend K, Beyeler W, Fiechter A, 1983. Detection of reactor nonhomogeneities by measuring culture fluorescence. Biotechnol. Bioeng. 25, 2789–2793.

Guisasola a, Baeza J a, Carrera J, Sin G, Vanrolleghem P a, Lafuente J, Matrix FI, 2006. The influence of experimental data quality and quantity on parameter estimation accuracy. Andrews Inhibition Model as a Case Study. Eduacation Chem. Eng. 1, 139–145.

Han M-J, Lee SY, 2006. The *Escherichia coli* Proteome: Past, Present, and Future Prospects, Microbiology and Molecular Biology Reviews.

Haringa C, Deshmukh AT, Mudde RF, Noorman HJ, 2017. Euler-Lagrange analysis towards representative down-scaling of a 22 m3 aerobic *S. cerevisiae* fermentation. Chem. Eng. Sci. 170, 653–669.

Haringa C, Tang W, Deshmukh AT, Xia J, Reuss M, Heijnen JJ, Mudde RF, Noorman HJ, 2016. Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines. Eng. Life Sci. 16, 652–663.

Haringa C, Tang W, Wang G, Deshmukh AT, van Winden WA, Chu J, van Gulik WM, Heijnen JJ, Mudde RF, Noorman HJ, 2018. Computational fluid dynamics simulation of an industrial P. chrysogenum fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization. Chem. Eng. Sci. 175, 12–24.

Harris RP, Kilby PM, 2014. Amino acid misincorporation in recombinant biopharmaceutical products. Curr. Opin. Biotechnol. 30, 45–50.

Hemmerich J, Noack S, Wiechert W, Oldiges M, 2018a. Microbioreactor Systems for Accelerated Bioprocess Development. Biotechnol. J. 13, 1–9.

Hemmerich J, Noack S, Wiechert W, Oldiges M, 2018b. Microbioreactor Systems for Accelerated Bioprocess Development. Biotechnol. J. 1700141, 1–9.

Herwig C, Garcia-Aponte OF, Golabgir A, Rathore AS, 2015. Knowledge management in the QbD paradigm: Manufacturing of biotech therapeutics. Trends Biotechnol. 33, 381–387.

Hewitt CJ, Nebe-von-Caron G, Axelsson B, McFarlane CM, Nienow AW, 2000. Studies related to the scale up of high cell density *E. coli* fed batch fermentations using multiparameter flow cytometry: Effect of a changing microenvironment with respect to glucose and dissolved oxygen concentration. Biotechnol. Bioeng. 70, 381–390.

Hoffmann F, Posten C, Rinas U, 2001. Kinetic model of in vivo folding and inclusion body formation in recombinant *Escherichia coli*. Biotechnol. Bioeng. 72, 315–322.

Hoffmann F, Rinas U, 2004. Stress induced by recombinant protein production in *Escherichia coli*. Adv. Biochem. Eng. Biotechnol. 89, 73–92.

Horowitz J, Normand MD, Corradini MG, Peleg M, 2010. Probabilistic model of microbial cell growth, division, and mortality. Appl. Environ. Microbiol. 76, 230–242.

Insel G, Celikyilmaz G, Ucisik-Akkaya E, Yesiladali K, Cakar ZP, Tamerler C, Orhon D, 2007. Respirometric evaluation and modeling of glucose utilization by *Escherichia coli* under aerobic and mesophilic cultivation conditions. Biotechnol. Bioeng. 96, 94–105.

Jaén KE, Sigala JC, Olivares-Hernández R, Niehaus K, Lara AR, 2017. Heterogeneous oxygen availability affects the titer and topology but not the fidelity of plasmid DNA produced by *Escherichia coli*. BMC Biotechnol. 17, 1–12.

Janakiraman V, Kwiatkowski C, Kshirsagar R, Ryll T, Huang Y-M, 2015. Application of high-throughput mini-bioreactor system for systematic scale-down modeling, process characterization, and control strategy development. Biotechnol. Prog. 31, 1623–1632.

Junker BH, Wang HY, 2006. Bioprocess monitoring and computer control: Key roots of the current PAT initiative. Biotechnol. Bioeng. 95, 226–261.

Junne S, Klingner A, Kabisch J, Schweder T, Neubauer P, 2011. A two-compartment bioreactor system made of commercial parts for bioprocess scale-down studies: Impact of oscillations on Bacillus subtilis fed-batch cultivations. Biotechnol. J. 6, 1009–1017.

Käß F, Hariskos I, Michel A, Brandt H, Spann R, Junne S, Wiechert W, Neubauer P, Oldiges M, 2014a. Assessment of robustness against dissolved oxygen/substrate oscillations for *C. glutamicum* DM1933 in two-compartment bioreactor. Bioprocess Biosyst. Eng. 37, 1151–1162.

Käß F, Junne S, Neubauer P, Wiechert W, Oldiges M, 2014b. Process inhomogeneity leads to rapid side

product turnover in cultivation of *Corynebacterium glutamicum*. Microb. Cell Fact. 13, 6.

Kim J Il, Modes F, Variables C, 2008. A Hybrid Model of Anaerobic *E. coli* GJT001: Combination of Elementary Flux Modes and Cybernetic Variables. Biotechnol. Prog. 24, 993–1006.

Knepper A, Heiser M, Glauche F, Neubauer P, Knepper A, Heiser M, Glauche F, Neubauer P, 2014. Robotic Platform for Parallelized Cultivation and Monitoring of Microbial Growth Parameters in Microwell Plates. J. Lab. Autom.

Ko YF, Bentley WE, Weigand W a, 1994. A metabolic model of cellular energetics and carbon flux during aerobic *Escherichia coli* fermentation. Biotechnol. Bioeng. 43, 847–855.

Kögler M, Paul A, Anane E, Birkholz M, Bunker A, Viitala T, Maiwald M, Junne S, Neubauer P, 2018. Comparison of time-gated surface-enhanced raman spectroscopy (TG-SERS) and classical SERS based monitoring of *Escherichia coli* cultivation samples. Biotechnol. Prog. 1–27.

Kramer B, Tüngler R, Bettenbrock K, Conradi C, 2015. A compact mathematical model of mandelate racemase production and chaperone overexpression in *E. coli*. IFAC-PapersOnLine 48, 23–28.

Kremling A, Geiselmann J, Ropers D, de Jong H, 2014. Understanding carbon catabolite repression in *Escherichia coli* using quantitative models. Trends Microbiol. 23, 99–109.

Łącki KM, 2014. High throughput process development in biomanufacturing. Curr. Opin. Chem. Eng. 6, 25–32.

Ladner T, Mühlmann M, Schulte A, Wandrey G, Büchs J, 2017. Prediction of *Escherichia coli* expression performance in microtiter plates by analyzing only the temporal development of scattered light during culture. J. Biol. Eng. 11, 1–15.

Lameiras F, Heijnen JJ, van Gulik WM, 2015. Development of tools for quantitative intracellular metabolomics of Aspergillus niger chemostat cultures. Metabolomics 11, 1253–1264.

Lapin A, Müller D, Reuss M, 2004. Dynamic Behavior of Microbial Populations in Stirred Bioreactors Simulated with Euler−Lagrange Methods: Traveling along the Lifelines of Single Cells. Ind. Eng. Chem. Res. 43, 4647–4656.

Lara AR, Galindo E, Ramírez OT, Palomares LA, 2006. Living With Heterogeneities in Bioreactors: Understanding the Effects of Environmental Gradients on Cells. Mol. Biotechnol. 34, 355–382.

Lara AR, Leal L, Flores N, Gosset G, Bolívar F, Ramírez OT, 2006. Transcriptional and metabolic response of recombinant*Escherichia coli* to spatial dissolved oxygen tension gradients simulated in a scale-down system. Biotechnol. Bioeng. 93, 372–385.

Lara AR, Taymaz-Nikerel H, Mashego MR, van Gulik WM, Heijnen JJ, Ramírez OT, van Winden W a, 2009. Fast dynamic response of the fermentative metabolism of *Escherichia coli* to aerobic and anaerobic glucose pulses. Biotechnol. Bioeng. 104, 1153–61.

Larsson G, Törnkvist M, Ståhl Wernersson E, Trägårdh C, Noorman H, Enfors SO, 1996. Substrate gradients in bioreactors: Origin and consequences. Bioprocess Eng. 14, 281–289.

Lattermann C, Büchs J, 2015. Microscale and miniscale fermentation and screening. Curr. Opin. Biotechnol. 35, 1–6.

Lauterbach T, Lenk F, Walther T, Int TUD, Grösel M, Lenk S, Gmbh S, Gernandt T, Moll R, Gmbh I, Seidel F, Brunner D, Gmbh Ö, Lüke T, Hedayat C, Büker M, Enas F, Peters A, 2017. Sens-o-Spheres – Mobile , miniaturisierte Sensorplattform für die ortsungebundene Prozessmessung in Reaktionsgefäßen. Dresden.

Lee J, Ramirez WF, 1992. Mathematical modeling of induced foreign protein production by recombinant bacteria. Biotechnol. Bioeng. 39, 635–646.

Lemoine A, Limberg MH, Kästner S, Oldiges M, Neubauer P, Junne S, 2016. Performance loss of *Corynebacterium glutamicum* cultivations under scale-down conditions using complex media. Eng. Life Sci. 16, 620–632.

Lemoine A, Maya Martínez-Iturralde N, Spann R, Neubauer P, Junne S, 2015. Response of *Corynebacterium glutamicum* exposed to oscillating cultivation conditions in a two- and a novel three-compartment scale-down bioreactor. Biotechnol. Bioeng. 112, 1220–1231.

Lethanh H, Neubauer P, Hoffmann F, 2005. The small heat-shock proteins IbpA and IbpB reduce the stress load of recombinant *Escherichia coli* and delay degradation of inclusion bodies. Microb. Cell Fact. 4, 6.

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 Biosyst. Eng. 25, 255–62.

Li J, Jaitzig J, Lu P, Süssmuth RD, Neubauer P, 2015. Scale-up bioprocess development for production of the antibiotic valinomycin in *Escherichia coli* based on consistent fed-batch cultivations. Microb. Cell Fact. 14, 1–13.

Limberg MH, Joachim M, Klein B, Wiechert W, Oldiges M, 2017. pH fluctuations imperil the robustness of *C. glutamicum* to short term oxygen limitation. J. Biotechnol. 259, 248–260.

Limberg MH, Pooth V, Wiechert W, Oldiges M, 2016. Plug flow versus stirred tank reactor flow characteristics in two-compartment scale-down bioreactor: Setup-specific influence on the metabolic phenotype and bioprocess performance of *Corynebacterium glutamicum*. Eng. Life Sci. 16, 610–619.

Lin HY, Mathiszik B, Xu B, Enfors S-OO, Neubauer P, 2001a. Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of *Escherichia coli*. Biotechnol. Bioeng. 73, 347–357.

Lin HY, Mathiszik B, Xu B, Enfors SO, Neubauer P, 2001b. Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of *Escherichia coli*. Biotechnol. Bioeng. 73, 347–357.

López C. DC, Barz T, Körkel S, Wozny G, 2015. Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. Comput. Chem. Eng. 77, 24–42.

López C. DC, Barz T, Penula M, Adriana V, Ochoa S, Wozny G, 2013. Model-Based Identifiable Parameter Determination Applied to a Simultaneous Saccharification and Fermentation Process Model for Bio-Ethanol Production. Biotechnol. Prog. 29, 1064–1082.

Lübbert A, Simutis R, 1994. Using measurement data in bioprocess modelling and control. Trends Biotechnol. 12, 304–311.

Luli GW, Strohl WR, 1990. Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. Appl. Environ. Microbiol. 56, 1004–1011.

Machado D, Rodrigues LR, Rocha I, 2014. A kinetic model for curcumin production in *Escherichia coli*. BioSystems 125, 16–21.

Mahadevan R, Doyle FJ, 2003. On-line optimization of recombinant product in a fed-batch bioreactor. Biotechnol. Prog. 19, 639–646.

Marbà-Ardébol A-M, Emmerich J, Muthig M, Neubauer P, Junne S, 2018. Real-time monitoring of the budding index in *Saccharomyces cerevisiae* batch cultivations with in situ microscopy. Microb. Cell Fact. 17, 73.

Marbà-Ardébol AM, Bockisch A, Neubauer P, Junne S, 2018. Sterol synthesis and cell size distribution under oscillatory growth conditions in *Saccharomyces cerevisiae* scale-down cultivations. Yeast 35, 213–223.

Martínez-Gómez K, Flores N, Castañeda HM, Martínez-Batallar G, Hernández-Chávez G, Ramírez OT, Gosset G, Encarnación S, Bolivar F, 2012. New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. Microb. Cell Fact. 11, 46.

Mears L, Stocks SM, Albaek MO, Sin G, Gernaey K V., 2017. Mechanistic Fermentation Models for Process Design, Monitoring, and Control. Trends Biotechnol.

Mears L, Stocks SM, Albaek MO, Sin G, Gernaey K V., 2016. Application of a mechanistic model as a tool for on-line monitoring of pilot scale filamentous fungal fermentation processes-The importance of evaporation effects. Biotechnol. Bioeng. Submitted, 1–11.

Mey M De, Taymaz-nikerel H, Baart G, Waegeman H, Maertens J, Heijnen JJ, Gulik WM Van, 2010. Catching prompt metabolite dynamics in *Escherichia coli* with the BioScope at oxygen rich conditions. Metab. Eng. 12, 477–487.

Miskovic L, Tokic M, Fengos G, Hatzimanikatis V, 2015. Rites of passage: requirements and standards for building kinetic models of metabolic phenotypes. Curr. Opin. Biotechnol. 36, 146–153.

Morchain J, Gabelle J-C, Cockx A, 2014. A Coupled Population Balance Model and CFD Approach for the Simulation of Mixing Issues in Lab-Scale and Industrial Bioreactors. AIChE J. 60, 27–40.

Muñoz-Tamayo R, Puillet L, Daniel JB, Sauvant D, Martin O, Taghipoor M, Blavy P, 2018. Review: To be or not to be an identifiable model. Is this a relevant question in animal science modelling? animal 12, 701–712.

Neubauer P, 2011. Editorial: Towards faster bioprocess development. Biotechnol. J. 6, 902–903.

Neubauer P, Åhman M, Törnkvist M, Larsson G, Enfors SO, 1995. Response of guanosine tetraphosphate to glucose fluctuations in fed-batch cultivations of *Escherichia coli*. J. Biotechnol. 43, 195–204.

Neubauer P, Cruz N, Glauche F, Junne S, Knepper A, Raven M, 2013. Consistent development of bioprocesses from microliter cultures to the industrial scale. Eng. Life Sci. 13, 224–238.

Neubauer P, Glauche F, Cruz-Bournazou MN, 2017. Editorial: Bioprocess Development in the era of digitalization. Eng. Life Sci. 17, 1140–1141.

Neubauer P, Häggström L, Enfors S-O, 1995. Influence of substrate oscillations on acetate formation and growth yield in *Escherichia coli* glucose limited fed-batch cultivations. Biotechnol. Bioeng. 47, 139–146.

Neubauer P, Junne S, 2016a. Scale-Up and Scale-Down Methodologies for Bioreactors, in: Mandenius, C.-F. (Ed.), Bioreactors: Design, Operation and Novel Applications. Wiley-VCH Verlag GmbH, Weinheim, Germany, pp. 323–354.

Neubauer P, Junne S, 2016b. Scale up and scale down methodologies for bioreactors. Bioreact. Des. Oper. Nov. Appl. 323–354.

Neubauer P, Junne S, 2010. Scale-down simulators for metabolic analysis of large-scale bioprocesses. Curr. Opin. Biotechnol. 21, 114–21.

Neubauer P, Lin HY, Mathiszik B, 2003. Metabolic load of recombinant protein production: Inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in *Escherichia coli*. Biotechnol. Bioeng. 83, 53–64.

Nickel DB, Cruz-Bournazou MN, Wilms T, Neubauer P, Knepper A, 2017. Online bioprocess data generation, analysis, and optimization for parallel fed-batch fermentations in milliliter scale. Eng. Life Sci. 17, 1195–1201.

Nienow AW, 2006. Reactor engineering in large scale animal cell culture. Cytotechnology 50, 9–33.

Nienow AW, Scott WH, Hewitt CJ, Thomas CR, Lewis G, Amanullah A, Kiss R, Meier SJ, 2013. Scale-down studies for assessing the impact of different stress parameters on growth and product quality during animal cell culture. Chem. Eng. Res. Des. 91, 2265–2274.

Nørregaard A, Bach C, Krühne U, Borbjerg U, Gernaey K V., 2018. Hypothesis-driven compartment model for stirred bioreactors utilizing computational fluid dynamics and multiple pH sensors. Chem. Eng. J. 356, 161–169.

Ogonah OW, Polizzi KM, Bracewell DG, 2017. Cell free protein synthesis: a viable option for stratified medicines manufacturing? Curr. Opin. Chem. Eng. 18, 77–83.

Oldshue JY, 1966. Fermentation mixing scale-up techniques. Biotechnol. Bioeng. 8, 3–24.

Omlin M, Reichert P, 1999. A comparison of techniques for the estimation of model prediction uncertainty. Ecol. Modell. 115, 45–59.

Oosterhuis NMG, Kossen NWF, 1984. Dissolved Oxygen Concentration Profiles in a Prod uction-Sea le Bioreactor XXVI, 546–550.

Peebo K, Valgepea K, Maser A, Nahku R, Adamberg K, Vilu R, 2015. Proteome reallocation in *Escherichia coli* with increasing specific growth rate. Mol. Biosyst. 11, 1184–1193.

Pham HTB, Larsson G, Enfors SO, 1998. Growth and energy metabolism in aerobic fed-batch cultures of *Saccharomyces cerevisiae*: Simulation and model verification. Biotechnol. Bioeng. 60, 474–482.

Quaiser T, Mönnigmann M, 2009. Systematic identifiability testing for unambiguous mechanistic modeling--application to JAK-STAT, MAP kinase, and NF-kappaB signaling pathway models. BMC Syst. Biol. 3, 50.

Rameez S, Mostafa SS, Miller C, Shukla AA, 2014. High-throughput miniaturized bioreactors for cell culture process development: Reproducibility, scalability, and control. Biotechnol. Prog. 30, 718–727.

Rathore AS, 2009. Roadmap for implementation of quality by design (QbD) for biotechnology products. Trends Biotechnol. 27, 546–553.

Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J, 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25, 1923–1929.

Reitz C, Fan Q, Neubauer P, 2018. Synthesis of non-canonical branched-chain amino acids in *Escherichia coli* and approaches to avoid their incorporation into recombinant proteins. Curr. Opin. Biotechnol. 53, 248–253.

Riesenberg D, Menzel K, Schulz V, Schumann K, Veith G, Zuber G, Knorre WA, 1990. High cell density fermentation of recombinant *Escherichia coli* expressing human interferon alpha 1. Appl. Microbiol. Biotechnol. 34, 77–82.

Rollié S, Mangold M, Sundmacher K, 2012. Designing biological systems: Systems Engineering meets Synthetic Biology. Chem. Eng. Sci. 69, 1–29.

Ruiz J, Fernández-Castané A, De Mas C, González G, López-Santín J, 2013. From laboratory to pilot plant *E. coli* fed-batch cultures: Optimizing the cellular environment for protein maximization. J. Ind. Microbiol. Biotechnol. 40, 335–343.

Ruiz J, González G, de Mas C, López-Santín J, 2011. A semiempirical model to control the production of a recombinant aldolase in high cell density cultures of *Escherichia coli*. Biochem. Eng. J. 55, 82–91.

Saccomani MP, 2013. Structural vs Practical Identifiability in System Biology. IWBBIO 2013 Proc. 305–313.

Saltelli A, Stefano T, Campolongo F, Ratto M, 2004. Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. John Wiley and Sons Inc, Sussex.

Sanchez-Gonzalez Y, Cameleyre X, Molina-Jouve C, Goma G, Alfenore S, 2009. Dynamic microbial response under ethanol stress to monitor *Saccharomyces cerevisiae* activity in different initial physiological states. Bioprocess Biosyst. Eng. 32, 459–466.

Sandoval-Basurto EA, Gosset G, Bolivar F, Ramírez OT, 2005. Culture of *Escherichia coli* under dissolved oxygen gradients simulated in a two-compartment scale-down system: Metabolic response and production of recombinant protein. Biotechnol. Bioeng. 89, 453–463.

Schädel F, Franco-Lara E, 2009. Rapid sampling devices for metabolic engineering applications. Appl. Microbiol. Biotechnol. 83, 199–208.

Schallmey M, Frunzke J, Eggeling L, Marienhagen J, 2014. Looking for the pick of the bunch: High-throughput screening of producing microorganisms with biosensors. Curr. Opin. Biotechnol. 26, 148–154.

Schilling BM, Pfefferle W, Bachmann B, Leuchtenberger W, Deckwer WD, 1999. A special reactor design for investigations of mixing time effects in a scaled-down industrial L-lysine fed-batch fermentation process. Biotechnol. Bioeng. 64, 599–606.

Schmid A, Hollmann F, Park JB, Bühler B, 2002. The use of enzymes in the chemical industry in Europe. Curr. Opin. Biotechnol. 13, 359–366.

Schweder T, Krüger E, Xu B, Jürgen B, Blomsten G, Enfors SO, Hecker M, 1999. Monitoring of genes that respond to process-related stress in large- scale bioprocesses. Biotechnol. Bioeng. 65, 151–159.

Schwentner A, Feith A, Münch E, Busche T, Rückert C, Kalinowski J, Takors R, Blombach B, 2018. Metabolic engineering to guide evolution – Creating a novel mode for L-valine production with *Corynebacterium glutamicum*. Metab. Eng. 47, 31–41.

Seamans TC, Fries S, Beck A, Wurch T, Chenu S, Chan C, Ushio M, Bailey J, Kejariwal A, Ranucci C, Villani N, Ozuna S, Goetsch L, Corvaia N, Salmon P, Robinson D, Chartrain M, 2008. Cell Cultivation Process Transfer and Scale-Up Part1. Bioprocess Int. 26–36.

Seo J, Bailey JE, 1985. Effects of recombinant plasmid content on growth properties and cloned gene product formation in*Escherichia coli*. Biotechnol. Bioeng. 27, 1668–1674.

Simen JD, Löffler M, Jäger G, Schäferhoff K, Freund A, Matthes J, Müller J, Takors R, Feuer R, von Wulffen J, Lischke J, Ederer M, Knies D, Kunz S, Sawodny O, Riess O, Sprenger G, Trachtmann N, Nieß A, Broicher A, 2017. Transcriptional response of *Escherichia coli* to ammonia and glucose fluctuations. Microb. Biotechnol. 10, 858–872.

Sin G, Gernaey K V., 2016. Data Handling and Parameter Estimation, in: van Loosdrecht, M.C.., Nielsen,

P.H., Lopez-Vazquez, C.M., Brdjanovic, D. (Eds.), Experimental Methods in Wastewater Treatment. IWA Publishing, London, pp. 201–234.

Sin G, Gernaey K V, Lantz AE, 2009. Good modelling practice (GMoP) for PAT applications: Propagation of input uncertainty and sensitivity analysis. Biotechnol. Prog. 25, 1043–1053.

Sin G, Meyer AS, Gernaey K V., 2010. Assessing reliability of cellulose hydrolysis models to support biofuel process design-Identifiability and uncertainty analysis. Comput. Chem. Eng. 34, 1385–1392.

Soini J, Falschlehner C, Liedert C, Bernhardt J, Vuoristo J, Neubauer P, 2008. Norvaline is accumulated after a down-shift of oxygen in *Escherichia coli* W3110. Microb. Cell Fact. 7, 30.

Soini J, Ukkonen K, Neubauer P, 2011a. Accumulation of amino acids deriving from pyruvate in *Escherichia coli* W3110 during fed-batch cultivation in a two-compartment scale-down bioreactor. Adv. Biosci. Biotechnol. 02, 336–339.

Soini J, Ukkonen K, Neubauer P, 2011b. Accumulation of amino acids deriving from pyruvate in <i>*Escherichia coli*</i> W3110 during fed-batch cultivation in a two-compartment scale-down bioreactor. Adv. Biosci. Biotechnol. 02, 336–339.

Spadiut O, Rittmann S, Dietzsch C, Herwig C, 2013. Dynamic process conditions in bioprocess development. Eng. Life Sci. 13, 88–101.

Sunya S, Bideaux C, Molina-Jouve C, Gorret N, 2013. Short-term dynamic behavior of *Escherichia coli* in response to successive glucose pulses on glucose-limited chemostat cultures. J. Biotechnol. 164, 531–542.

Sunya S, Delvigne F, Uribelarrea JL, Molina-Jouve C, Gorret N, 2012. Comparison of the transient responses of *Escherichia coli* to a glucose pulse of various intensities. Appl. Microbiol. Biotechnol. 95, 1021–1034.

Swartz JR, 2001. Advances in *Escherichia coli* production of therapeutic proteins. Curr. Opin. Biotechnol. 12, 195–201.

Sweere a. PJPJ, Luyben KCAM, Kossen NWF, 1987. Regime analysis and scale-down: Tools to investigate the performance of bioreactors. Enzyme Microb. Technol. 9, 386–398.

Takors R, 2012. Scale-up of microbial processes: Impacts, tools and open questions. J. Biotechnol. 160, 3–9.

Taymaz-nikerel H, De Mey M, Baart G, Maertens J, Heijnen JJ, van Gulik W, Mey M De, Baart G, Maertens J, Heijnen JJ, 2013. Changes in substrate availability in *Escherichia coli* lead to rapid metabolite, flux and growth rate responses. Metab. Eng. 16, 115–129.

Theobald U, Mailinger W, Baltes M, Rizzi M, Reuss M, 1997. In Vivo Analysis of Metabolic Dynamics in *Saccharomyces cerevisiae* : I . Experimental Observations. Biotechnol. Bioeng. 55, 305–316.

Ude C, Hentrop T, Lindner P, Lücking TH, Scheper T, Beutel S, 2015. New perspectives in shake flask pH control using a 3D-printed control unit based on pH online measurement. Sensors Actuators B Chem. 221, 1035–1043.

Vajda SVA, Rabitz H, Walter E, Lecourtier Y, 1989. Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. Chem. Eng. Commun. 83, 191–219.

Valgepea K, Adamberg K, Nahku R, Lahtvee P-J, Arike L, Vilu R, 2010. Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. BMC Syst. Biol. 4, 166.

Valgepea K, Adamberg K, Vilu R, 2011. Decrease of energy spilling in *Escherichia coli* continuous cultures with rising specific growth rate and carbon wasting. BMC Syst. Biol. 5, 1–11.

van Gulik WM, 2010. Fast sampling for quantitative microbial metabolomics. Curr. Opin. Biotechnol. 21, 27–34.

van Gulik WM, Canelas AB, Seifar RM, Heijnen JJ, 2013. The Sampling and Sample Preparation Problem in Microbial Metabolomics, in: Metabolomics in Practice. Wiley-Blackwell, pp. 1–19.

Varma A, Palsson BO, 1994. Metabolic flux balancing: Basic concepts, scientific and practical use. Bio/Technology 12, 994–998.

Veiter L, Rajamanickam V, Herwig C, 2018. The filamentous fungal pellet — relationship between morphology and productivity. Appl. Microbiol. Biotechnol. 2997–3006.

Velez-Suberbie ML, Betts JPJ, Walker KL, Robinson C, Zoro B, Keshavarz-Moore E, 2018. High

throughput automated microbial bioreactor system used for clone selection and rapid scale-down process optimization. Biotechnol. Prog. 34, 58–68.

Vemuri GN, Altman E, Sangurdekar DP, Khodursky a B, Eiteman M a, 2006. Overflow Metabolism in *Escherichia coli* during Steady-State Growth : Transcriptional Regulation and Effect of the Redox Ratio. Appl. Environ. Microbiol. 72, 3653–3661.

Villaverde AF, Henriques D, Smallbone K, Bongard S, Schmid J, Cicin-Sain D, Crombach A, Saez-Rodriguez J, Mauch K, Balsa-Canto E, Mendes P, Jaeger J, Banga JR, 2015. BioPreDyn-bench: a suite of benchmark problems for dynamic modelling in systems biology. BMC Syst. Biol. 9, 8.

Visser D, Van Zuylen GA, Van Dam JC, Oudshoorn A, Eman MR, Ras C, Van Gulik WM, Frank J, Van Dedem GWK, Heijnen JJ, 2002. Rapid sampling for analysis of in vivo kinetics using the BioScope: A system for continuous-pulse experiments. Biotechnol. Bioeng. 79, 674–681.

Wang G, Tang W, Xia J, Chu J, Noorman H, van Gulik WM, 2015. Integration of microbial kinetics and fluid dynamics toward model-driven scale-up of industrial bioprocesses. Eng. Life Sci.

Wang G, Zhao J, Haringa C, Tang W, Xia J, Chu J, Zhuang Y, Zhang S, Deshmukh AT, van Gulik W, Heijnen JJ, Noorman HJ, 2018. Comparative performance of different scale-down simulators of substrate gradients in Penicillium chrysogenum cultures: the need of a biological systems response analysis. Microb. Biotechnol. 11, 486–497.

Wang J, Gilles ED, Lengeler JW, Jahreis K, 2001. Modeling of inducer exclusion and catabolite repression based on a PTS-dependent sucrose and non-PTS-dependent glycerol transport systems in *Escherichia coli* K-12 and its experimental verification. J. Biotechnol. 92, 133–158.

Wang Y, Chu J, Zhuang Y, Wang Y, Xia J, Zhang S, 2009. Industrial bioprocess control and optimization in the context of systems biotechnology. Biotechnol. Adv. 27, 989–95.

Wen D, Vecchi MM, Gu S, Su L, Dolnikova J, Huang YM, Foley SF, Garber E, Pederson N, Meier W, 2009. Discovery and investigation of misincorporation of serine at asparagine positions in recombinant proteins expressed in Chinese hamster ovary cells. J. Biol. Chem. 284, 32686–32694.

Wiegmann V, Martinez CB, Baganz F, 2018. A simple method to determine evaporation and compensate for liquid losses in small-scale cell culture systems. Biotechnol. Lett. 40, 1029–1036.

Wolfe AJ, 2005. The Acetate Switch. Microbiol. Mol. Biol. Rev. 69, 12–50.

Wong HE, Huang CJ, Zhang Z, 2018. Amino acid misincorporation in recombinant proteins. Biotechnol. Adv. 36, 168–181.

Xu B, Jahic M, Blomsten G, Enfors SO, 1999. Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*. Appl. Microbiol. Biotechnol. 51, 564–571.

Xu B, Jahic M, Enfors S-O, 1999. Modeling of Overflow Metabolism in Batch and Fed-Batch Cultures of *Escherichia coli*. Biotechnol. Prog. 15, 81–90.

Yordanov B, Dalchau N, Grant PK, Pedersen M, Emmott S, Haselo J, Phillips A, 2014. A Computational Method for Automated Characterization of Genetic Components.

Zelić B, Bolf N, Vasić-Rački D, 2006. Modeling of the pyruvate production with *Escherichia coli*: Comparison of mechanistic and neural networks-based models. Bioprocess Biosyst. Eng. 29, 39–47.

Zemella A, Thoring L, Hoffmeister C, Šamalíková M, Ehren P, Wüstenhagen DA, Kubick S, 2018. Cell-free protein synthesis as a novel tool for directed glycoengineering of active erythropoietin. Sci. Rep. 8, 1–12.

Publications

I

Biochemical Engineering Journal 125 (2017) 23–30

Contents lists available at ScienceDirect

Biochemical Engineering Journal

journal homepage: www.elsevier.com/locate/bej

Regular article

Modelling overflow metabolism in *Escherichia coli* by acetate cycling

Emmanuel Anane*, Diana C. López C, Peter Neubauer, M. Nicolas Cruz Bournazou

Chair of Bioprocess Engineering, Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany

ARTICLE INFO

Article history:
Received 10 March 2017
Received in revised form 10 May 2017
Accepted 15 May 2017
Available online 25 May 2017

Keywords:
Escherichia coli
Overflow metabolism
Acetate
Modelling
Fed-batch

ABSTRACT

A new set of mathematical equations describing overflow metabolism and acetate accumulation in *E. coli* cultivation is presented. The model is a significant improvement of already existing models in the literature, with modifications based on the more recent concept of acetate cycling in *E. coli*, as revealed by proteomic studies of overflow routes. This concept opens up new questions regarding the speed of response of the acetate production and its consumption mechanisms in *E. coli.* The model is formulated as a set of continuous differentiable equations, which significantly improves model tractability and facilitates the computation of dynamic sensitivities in all relevant stages of fermentation (batch, fed-batch, starvation). The model is fitted to data from a simple 2 L fed-batch cultivation of *E. coli* W3110 M, where twelve (12) out of the sixteen (16) parameters were exclusively identified with relative standard deviation less than 10%. The framework presented gives valuable insight into the acetate dilemma in industrial fermentation processes, and serves as a tool for the development, optimization and control of *E. coli* fermentation processes.

1. Introduction

One of the most embattled physiological phenomena in industrial scale *E. coli* cultivation is overflow metabolism and the associated excretion of acetate into the broth. Apart from the fact that extracellular acetate inhibits the growth of *E. coli* [1], redirection of the carbon source to acetate through overflow metabolism is wasteful in recombinant protein production processes [2]. The concept of overflow metabolism in *E. coli* is not new and a number of models in the literature have attempted to describe acetate production mechanistically [3–7]. In all these models, the acetate profile is presented in two distinct phases: an initial batch (overflow) phase during which acetate is produced, followed by a substrate limited phase during which acetate is consumed. Most of these models assume saturation of TCA cycle enzymes and are mainly built on discrete *conditional* statements in the metabolic routes [4,8]. Although these models suffice in describing acetate profiles during *E. coli* cultivation, a major limitation is the discontinuous nature of the functions (e.g, if $q_O < q_{Omax}$, then $qA = 0$ [8]), which makes further mathematical development and the use of sophisticated simulation and optimization programmes difficult. Furthermore, new evidence suggests that acetate conversion is indeed a continuous process. Peebo et al. [9] and Valgepea et al. [10] used advanced proteomic analysis [9,11] and systems biology approaches [10] to show that intracellular production and re-assimilation of acetate (acetate cycling) is a continuous process in *E. coli* metabolism, even under non-overflow conditions. Two intermediates of this cycling process (Acetyl-AMP and Acetyl-P) were shown to play vital roles in *E. coli* motility and osmoregulation [1,12]. Thus, acetate excretion into the extracellular medium only results from an offset of the equilibrium between its production and re-assimilation, which is triggered by carbon catabolite repression, either at higher specific substrate uptake rates [11,13] or under anoxic conditions [11]. Therefore, a new mathematical representation is needed for model-based process development, which i) copes better with the real acetate conversion process in *E. coli* and ii) fulfils the requirements for gradient based algorithms and solution of partial differential equation systems.

1.1. Glucose partitioning and acetate cycling in E. coli

The mechanistic models of Xu [4], Lin [5] and Neubauer [8] are among the most widely used macro-kinetic models for description of *E. coli* fermentations. These were built on the concept of glucose partitioning which was initially developed for *Saccharomyces cerevisiae* [14], and was shown to be equally applicable to the *E. coli* system [15]. The structured model covers the intracellular partitioning of glucose for various physiological demands of the cell

* Corresponding author at: Chair of Bioprocess Engineering, Institute of Biotechnology, Technische Universität Berlin, Ackerstraße 76, ACK24, 13355 Berlin, Germany.
E-mail address: anane@tu-berlin.de (E. Anane).

http://dx.doi.org/10.1016/j.bej.2017.05.013
1369-703X/

Nomenclature

C	Carbon content of (s) substrate, (x) biomass
DOT	Dissolved oxygen tension (%). DOT* represents saturating value of DOT in the broth at the given operating conditions
F	Flow rate (L/h)
K_{ap}	Monod-type saturation constant, intracellular acetate production
K_{sa}	Affinity constant, acetate consumption (g/L)
K_o	Affinity constant, oxygen consumption (g/L)
K_s	Affinity constant, substrate consumption (g/L)
K_{ia}	Inhibition constant, inhibition of glucose uptake by acetate (g/L)
K_{is}	Inhibition constant, inhibition of acetate uptake by glucose (g/L)
p_A	Specific acetate production rate (g/(g.h))
q_A	Specific acetate consumption rate (g/(g.h))
q_m	Specific maintenance coefficient (g/(g.h))
q_S	Max spec glucose uptake rate (g/(g.h))
Y	Yield coefficient (g/g)
Y_{xsof}	Yield of biomass on substrate from auxiliary overflow routes, such as the mixed acid and pentose-phosphate pathways
μ	Specific growth rate (h^{-1})
τ	Dissolved oxygen probe response time (h)

Subscripts

A	Acetate
an	Anabolic
c	Consumption
en	Energetic
glu	Glucose
i	Inlet concentration
m	Maintenance
max	Maximum
O	Oxygen
of	Overflow
ox	Oxidative
S	Substrate (glucose)
X	Biomass
em	Excluding maintenance

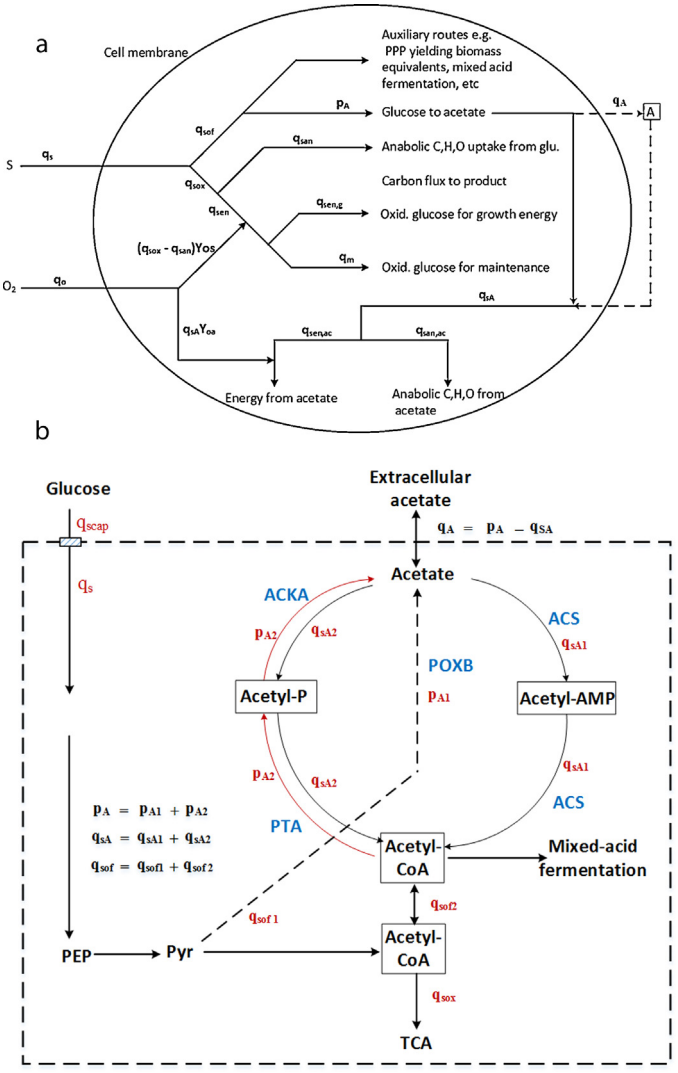


Fig. 1. (A) Glucose partitioning and oxygen usage in E. coli. Scheme for flux distribution used to model glucose (S) and oxygen (O_2) usage in E. coli [4,5] with modifications showing acetate (A) excretion and re-assimilation across the cell membrane. (B) Enzyme-mediated acetate cycling in E. coli, with summaries of the glucose uptake system and upper glycolytic routes, adapted from Valgepea et al. [10], Li et al. [40] and Enjalbert et al. [32]. POXB—pyruvate oxidase, ACS—acetyl-CoA synthetase, PTA—phosphotransacetylase, ACKA—acetate kinase.

and the associated oxygen consumption (Fig. 1A). Additionally, the subsequent conversion of glucose to acetate through the overflow route is based on the ethanol process in *S. cerevisiae*. The large set of metabolic routes are approximated by lumped kinetic parameters, the values of which are estimated from measurable extracellular species in typical cultivation set-ups. Nevertheless, Valgepea et al. [10], Peebo et al. [9], and Martinez-Gomez et al. [16] showed that *E. coli* uses a continuous acetate cycling system for acetate conversion (Fig. 1B). Under *typical* cultivation conditions and at low glucose uptake rates, acetate production is in equilibrium with its re-assimilation ($p_A = q_{sA}$), hence there is neither net accumulation nor release of acetate into the extracellular medium. As the inflow of glucose into the cell increases, acetate consumption through the ACS pathway (acetate consumption) becomes insufficient, hence there is an offset in the cycling equilibrium ($p_A > q_{sA}$), which results in intracellular accumulation of acetate followed by its excretion into the medium. This cycling system can be represented by a set of continuous equations that describe overflow as an off-set of the equilibrium between p_A and q_{sA}.

In the current paper, the concept of acetate cycling is used to derive a set of tractable and continuously differentiable equations at macro-kinetic scale to describe acetate production in *E. coli*. This new set of equations allow a cheap computation of both first and second order sensitivity functions, which are needed in gradient-based methods for optimal experimental design [17,18] and process optimization [19]. Finally, the model is validated with data from fed-batch cultivation of *E. coli* W3110M to show its applicability in real processes.

2. Materials and methods

2.1. Strain and cultivation conditions

The data for model calibration was obtained from a fed-batch culture of non-recombinant *E. coli* W3110M. The cultivation was conducted at a temperature of 37 °C and at pH of 7.0 in a 3.7 L BioEngineering® bench top bioreactor fitted with a polarographic dissolved oxygen probe. The medium consisted of mineral salts and

trace elements, as described by Glazyrina et al. [20], with 5 g/L glucose in the batch phase. To start the batch phase, 2 L of medium was inoculated wit *E. coli* W3110 to OD_{600} of 0.1. The fed-batch phase was initiated with a 300 g/L glucose feed at the exhaustion of the batch phase glucose, signalled by a sudden rise in dissolved oxygen tension. Two fed-batch regimes were implemented: an exponential feed followed by a constant feed. The exponential feed was implemented to maintain a set-point specific growth rate (m_{set}) of 0.22 h^{-1}. Using the biomass concentration (X, g/L) and broth volume (V, L) at the end of the batch phase, the exponential feed rate (F, L/h) was calculated as

$$\boldsymbol{F(t)} = \frac{\boldsymbol{\mu_{set}}}{\boldsymbol{Y_{x/s}S_i}}(\boldsymbol{XV})\boldsymbol{e^{sett}} \quad (1)$$

where S_i represents the glucose concentration in the feed solution (300 g/L) and t is the feed time. After 3 h, the feed was switched to a constant feed, beginning at a specific growth rate of 0.11 h^{-1} for a period of 17 h. To test the robustness of the acetate equations, intermittent glucose pulses were given in the constant feeding phase, and the corresponding response in all profiles were modelled. Cellular growth was monitored by measuring the optical density of samples at 600 nm in a UV–vis spectrophotometer (Novaspec III, Amersham Biosciences, Amersham, UK). Conversion factors that were developed with the same spectrophotometer and the same *E. coli* strain and given in Glazyrina et al. [20] were used to convert OD_{600} values to cell dry weight.

2.2. Analyses

To determine the concentration of residual glucose and acetate, hourly samples were analysed on an Agilent 1200 HPLC system, equipped with a HyperRez™ XP Carbohydrate H+ column (Fisher Scientific, Schwerte, Germany) and a refractive index detector. As eluent, five mM H_2SO_4 was used at a flow rate of 0.5 mL min^{-1}. In total, $N_m = 19$ data points were collected and analysed for each observable variable i.e., biomass (X), glucose (S) and acetate (A), whereas $N_m = 11{,}336$ data points were logged online for measured dissolved oxygen (DOT).

2.3. Parameter estimation

The E. coli model in Section 3 was solved using Matlab R2015a® with the CVODE integrator from SUNDIALS TB [21]. The parameter estimates (PE) were computed using lsqnonlin with the trust region reflective algorithm. The initial parameter values for PE were based on literature [4,5,17]. The initial values for the search in the parameter space were generated using the Minimum bias Latin hypercube design (MBLHD) [22]. The dynamic model can be written in its general form as follows,

$$\dot{x}(t) = f\left(x(t), z(t), u(t); \theta\right) \quad (2)$$

$$0 = g\left(x(t), z(t), u(t); \theta\right) \quad (3)$$

$$y_{Of}(t) = Ax(t) \quad (4)$$

$$y_{On}(t) = Bx(t) \quad (5)$$

$$x(t_0) = x_0, \quad (6)$$

where the set of differential equations f corresponds to Eqs. (12), (14), (18), (22) and (24) whereas the algebraic equation set g refers to Eqs. (13), (15)–(17), (19)–(21) and (23), $t \in [t_o, t_{end}] \subseteq \mathbb{R}$ is the independent time variable, $x(t) \in \mathbb{R}^{N_x}$ and $z(t) \in \mathbb{R}^{N_z}$ are the differential and algebraic state variables, respectively; $u(t) \in \mathbb{R}^{N_u}$ are the time-varying inputs or experimental design variables and $\theta \in \mathbb{R}^{N_p}$ is the unknown parameter vector. The vector $y_{Of}(t) \in \mathbb{R}^{Ny_{Of}}$ are the predicted offline response variables (variables corresponding to sampled measurements) whose elements are defined by the selection matrix $A \in \mathbb{R}^{Ny_{Of} \times N_x}$. The vector $y_{On}(t) \in \mathbb{R}^{Ny_{On}}$ are the predicted response variables measured online (i.e., oxygen) whose elements are defined by the selection matrix $B \in \mathbb{R}^{Ny_{On} \times N_x}$. Note that not all states were measured, therefore $Ny_{Of} + Ny_{On} < N_x$.

The model parameters were estimated by solving the optimization problem

$$\hat{\theta} := arg\min_{\theta} \Phi\left(U, \theta\right) \quad (7)$$

where the cost function $\Phi\left(U, \theta\right)$, which is the weighted nonlinear least-squares criterion between the model predictions $Y\left(U, \theta\right)$ and the experimental data Y^m was calculated as

$$\Phi\left(U, \theta\right) := \frac{1}{2}\left(Y\left(U, \theta\right) - Y^m\right)^T \left(C_y\right)^{-1} \left(Y\left(U, \theta\right) - Y^m\right) \quad (8)$$

All the measured data were collected in the vector $Y\left(U, \theta\right)$,

$$\begin{aligned} Y(U,\theta) := & ((y_{Of_1}(t_1, U, \theta), \cdots, y_{Of_1}(t_{Nm_{Of}}, U; \theta))^T, \ldots, (y_{Of_{Ny_{Of}}}(t_1, U, \theta), \cdots, y_{Of_{Ny_{Of}}}(t_{Nm_{Of}}, U; \theta))^T, \\ & (y_{On_1}(t_1, U, \theta), \cdots, y_{On_1}(t_{Nm_{On}}, U; \theta))^T, \ldots, (y_{On_{Ny_{On}}}(t_1, U, \theta), \cdots, y_{On_{Ny_{On}}}(t_{Nm_{On}}, U; \theta))^T) \in \mathbb{R}^{Ny_{Of} \cdot Nm_{Of} + Ny_{On} \cdot Nm_{On}} \end{aligned} \quad (9)$$

The weighting matrix, $C_y \in \mathbb{R}^{Ny_{Of} \cdot Nm_{Of} + Ny_{On} \cdot Nm_{On} \times Ny_{Of} \cdot Nm_{Of} + Ny_{On} \cdot Nm_{On}}$ was obtained from the measurement errors, which were assumed to be unbiased, independent and normally distributed. Therefore C_y is diagonal matrix with entries given by the variance $\sigma^2_{y,i}$ of each measurement i. Consequently, the observed measured responses in Y^m are normally-distributed random variables, i.e., $Y^m \sim \mathcal{N}(E(Y^m), Var(Y^m))$, hence the expectation $E(Y^m)$ is equal to the model output at the unknown true parameters values θ^*, i.e., $E(Y^m) = Y\left(U, \theta^*\right)$. The previous assumption ensures that the solution $\hat{\theta}$ of the optimization (Eq. (8)) is equivalent to the maximum likelihood estimation solution [23].

2.4. Parameter uncertainty quantification

To quantify the uncertainty associated with the PE, the precision of parameter values was assessed with the variance, relative standard deviation and the confidence interval calculated using a Monte Carlo (MC) method. A total of L–replications of the experimental data $Y^m_j, j = 1, \ldots, L$ were generated, drawing L– random MC samples from the normal distribution $\mathcal{N}(Y^m, C^m)$. For each data set Y^m_j, the PE was repeated to obtain the point estimates $\hat{\theta}_1, \ldots, \hat{\theta}_L$. Then the parameter covariance matrix was computed as [23,24]

$$Cov\left(\hat{\theta}\right) = \frac{1}{(L-1)} \sum_{j=1}^{L} \left(\hat{\theta}_j - E\left(\hat{\Theta}\right)\right)\left(\hat{\theta}_j - E\left(\hat{\Theta}\right)\right)^T \quad (10)$$

where $E\left(\hat{\theta}\right) \approx \frac{1}{L}\sum_{j=1}^{L} \hat{\theta}_j$ represents the mean of the parameter distribution. The variances of the parameters, $\sigma^2_{\theta_i}$ were equal to the diagonal entries of the covariance matrix $Cov\left(\hat{\theta}\right)$ from which the standard deviations σ_{θ_i} were also calculated. The 95% confidence intervals were estimated using σ_{θ_i} and a Student t-distribution.

3. Macro-kinetic model formulation

A significant improvement to the referenced models is the conversion of the inherently discontinuous systems into a continuous one that is more mathematically stable. The model comprises a set of ordinary differential equations (ODEs) describing six state variables, namely biomass X and extracellular concentrations of substrate S, acetate A, dissolved oxygen DOTa and DOT, and the feed F (which can be constant or a function of time). A set of auxiliary algebraic equations describing intracellular interactions relating substrate (glucose), oxygen and acetate consumption as well as biomass formation are coupled to the ODEs to form the kinetic model.

The state variables X, S, A and F are modelled as in a conventional fed-batch fermentation process. The dissolved oxygen, on the other hand, is modelled with two ordinary differential equations (Eqs. (22) and (24)). The difference between the two dissolved oxygen profiles DOTa and DOT is due to the response lag of the sensor, which is approximated by first order response kinetics for the DO-probe [25].

The general form of the governing mass balance is expressed as follows,

$$\frac{dx}{dt} = \frac{F}{V}(x_i - x) + rX \tag{11}$$

where $x \in \{X, S, A\}$ represents the state variable in [g/l] and the subscript i represents the inlet concentration, F the feed, V the volume and r is the corresponding specific rate. Considering that inlet concentration of biomass is zero (sterile feed), we obtain the following expression for biomass balance in the fed-batch case:

$$\frac{dX}{dt} = \frac{F}{V}(0 - X) + \mu X \tag{12}$$

In Eq. (12), X represents the concentration (cell dry weight) of cells and μ (h^{-1}) is the non-inhibited Monod-type specific growth rate, given as

$$\mu = (q_{sox} - q_m)Y_{em} + q_{sof}Y_{xsof} + q_{sA}Y_{xa} \tag{13}$$

where qsox, qsof, qsA, represent the uptake rates of substrate for oxidation, substrate metabolized through the overflow route and acetate respectively, the constants Y. . . define the respective yield coefficients, whereas qm represents the glucose expended for cell maintenance. Thus, according to Eq. (13), the overall growth of the culture results from usage of glucose through the oxidative (q_{sox}), acetate uptake (q_{SA}) and re-use of other products from the overflow route (q_{sof}), which all contribute energy equivalents for cell growth [14]. The mass balance for substrate (glucose) in the fed-batch process is given as

$$\frac{dS}{dt} = \frac{F}{V}(S_i - S) - q_s X \tag{14}$$

The substrate concentration S is modelled taking acetate inhibition into account, as reviewed by Shiloach and Fass [26]. The specific substrate uptake rate is therefore modelled with Mono-type kinetics with non-competitive inhibition:

$$q_s = \frac{q_{smax}}{1 + \frac{A}{K_{ia}}} \cdot \frac{S}{S + K_s} \tag{15}$$

where Kia and Ks are the acetate inhibition and the substrate affinity constants respectively. Not all the substrate consumed is metabolized in the TCA cycle (qsox), but a portion goes to the overflow path qsof.

$$q_{sox} = \left(q_s - q_{sof}\right) \cdot \frac{DOT}{DOT + K_o} \tag{16}$$

$$q_{sof} = \frac{P_{Amax} q_s}{q_s + K_{ap}} \tag{17}$$

where Ko is a dimensionless constant set to 0.1 to increase the stability of the numeric simulation, PAmax and Kap are the maximum acetate production and the production affinity constants. Acetate production/consumption is a cyclic process and considering no addition of acetate in the feed, the mass balance yields

$$\frac{dA}{dt} = \frac{F}{V}(0 - A) + q_{sA}X \tag{18}$$

The equilibrium qsA = 0 is reached when the acetate produced through the overflow route pA is equal to the acetate consumed qsA.

$$q_A = p_A - q_{sA} \tag{19}$$

with

$$p_A = q_{sof}Y_{as} \tag{20}$$

where Yas is the gram of acetate per gram of substrate consumed through the overflow route. The specific acetate consumption rate is modelled as

$$q_{sA} = \frac{q_{Amax}}{1 + \frac{q_s}{K_{is}}} \cdot \frac{A}{A + K_{sa}} \tag{21}$$

were qAmax, Kis, and Ksa are constant parameters representing the maximum acetate uptake rate, the acetate uptake inhibition, and acetate affinity constant. The uptake of acetate, however, is inhibited in a non-competitive way by the presence of glucose in the medium due to *E. coli*'s higher preference for glucose over acetate [27]. Thus, glucose and acetate exhibit a counter inhibition effect on their respective uptake capacities, as presented in Eqs. (15) and (21).

Finally, the actual dissolved oxygen (DOT_a) is calculated in% of saturation with the assumption that the feed solution in the fed-batch phase is fully saturated with dissolved oxygen. The oxygen profile is described with the standard equation

$$\frac{dDOT_a}{dt} = K_{La}(DOT^* - DOT_a) - q_o XH \tag{22}$$

with DOT* being the saturation value of dissolved oxygen in the medium, KLa the volumetric mass transfer coefficient, H the Henry equilibrium constant, and qo the oxygen uptake rate described by

$$q_o = (q_{sox} - q_m)Y_{os} + q_{sA}Y_{oa} \tag{23}$$

where Yos, Yoa are the yield coefficients for substrate and acetate to oxygen consumption respectively. With the probe response, the measured DOT is modelled as:

$$\frac{dDOT}{dt} = K_p(DOT_a - DOT) \tag{24}$$

were $K_p = \frac{1}{\tau}$ is the static gain of the sensor given as the inverse of the probe response time, t. This is important when dealing with pulses in the system and situations where fast changes in the system are expected. In an experimental set-up, t is measured as the time required to reach 63.2% of the final response after a step change in the DOT profile [28]. At the calculated q_O values for the current *E. coli* strain, when the probe response time is greater than 8 s, it is important to include it in the dynamics of the oxygen profile.

4. Results and discussion

Acetate accumulation remains one of the biggest challenges in recombinant protein production using the *E. coli* expression system. In the past, the inclusion of acetate profiles in mechanistic models of *E. coli* meant that the system became inherently discontinuous because acetate was thought to be produced only during

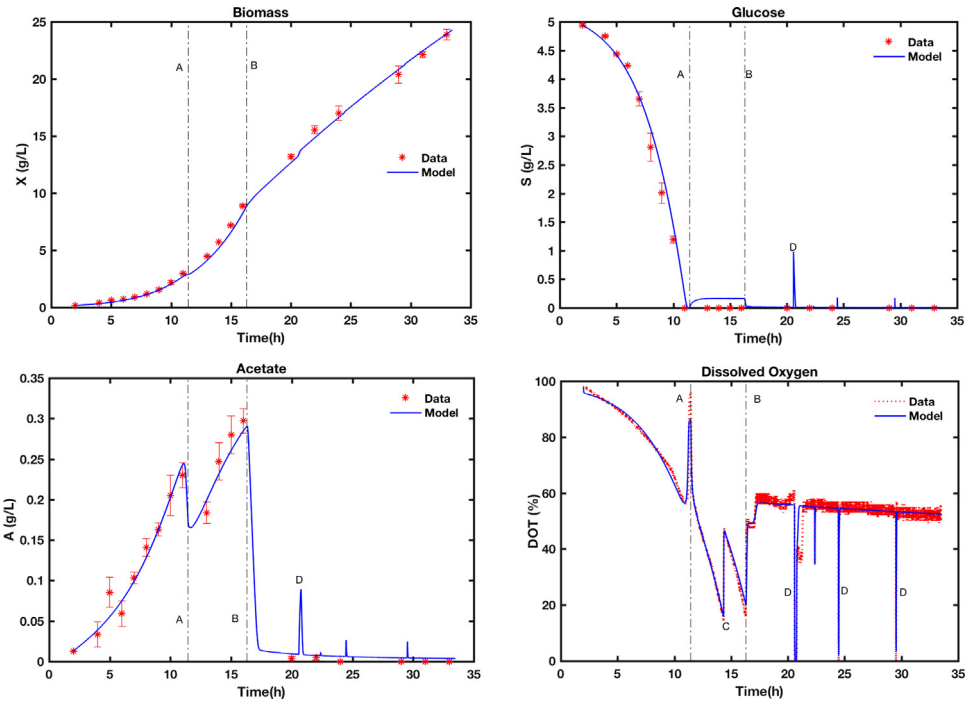


Fig. 2. Comparison of E. coli model with experimental data from fed-batch cultivation of wild type E. coli W3110. A = start of exponential feed fed-batch phase at m_{set} = 0.25 h^{-1} B = start of constant feed fed-batch phase, C = change in K_La, D = glucose pulses.

Table 1
Summary of results of parameter estimation and quantification of the uncertainty associated with the parameter estimators. LB—lower bound, UB—upper bounds, CI—95% confidence interval.

Parameter [a]	Units	Initial guess (literature)[b]	Estimate	PE uncertainty quantification		
				$\%\sigma_\theta$	LB-CI	UB-CI
K_{ap}	g L^{-1}	0.10	0.5052	15.2	0.3539	0.6565
K_{sa}	g L^{-1}	0.05	0.0134	22.0	0.0076	0.0192
K_o	g L^{-1}	10.0	0.0001	0.0	0.0001	0.0001
K_s	g L^{-1}	0.05	0.0370	8.9	0.0305	0.0435
K_{ia}	g L^{-1}	5.00	1.2399	9.6	1.0062	1.4737
K_{is}	g L^{-1}	10.0	2.1231	27.3	0.9788	3.2673
p_{Amax}	g g^{-1} h^{-1}	0.17	0.2268	6.5	0.1977	0.2558
q_{Amax}	g g^{-1} h^{-1}	0.15	0.1148	6.1	0.1009	0.1287
q_m	g g^{-1} h^{-1}	0.04	0.0129	7.0	0.0111	0.0147
q_{Smax}	g g^{-1} h^{-1}	1.37	0.6356	0.3	0.6320	0.6392
Y_{as}	g g^{-1}	0.80	0.9097	4.5	0.8283	0.9911
Y_{oa}	g g^{-1}	1.06	0.5440	9.5	0.4425	0.6455
Y_{xa}	g g^{-1}	0.70	0.5718	9.9	0.4604	0.6833
Y_{em}	g g^{-1}	0.50	0.5333	2.4	0.5085	0.5580
Y_{os}	g g^{-1}	1.06	1.5620	5.4	1.3941	1.7298
Y_{xsof}	g g^{-1}	0.15	0.2268	12.0	0.1730	0.2807

[a] Parameter descriptions in nomenclature.
[b] References: [4,5,8,16].

certain specific growth regimes in *E. coli* cultivation. This property of the system inhibited further mathematical development, especially in the fields of control theory and model-based optimization. In the current paper, we have explored the possibility of developing continuous differential equations for acetate production based on the concept of acetate cycling. The data to validate the new set of continuous equations describing acetate production in *E. coli* was obtained from a fed-batch cultivation of non-recombinant *E. coli* W3110M. The cultivation was done to cover all possible growth regimes in *E. coli* [29], to study the kinetics of acetate production and its consumption. From an excess substrate (glucose) environment in the batch phase to chronic starvation conditions at the end, the acetate profile evolved according to predictable kinetics (Fig. 2). A maximum specific growth rate (m_{max}) of 0.31 h^{-1} was recorded during the batch phase, in the presence of excess glucose. Growth at m_{max} was associated with a steady increase in extracellular acetate concentration. In the exponential feed fed-batch phase, acetate concentration increased further to a maximum of 0.3 g/L due to the higher m_{set} value (0.22 h^{-1}), which corresponds to a specific substrate uptake rate higher than that which would allow fully oxidative growth. This m_{set} was about 75% of m_{max} and was higher than the threshold for the acetate switch at the given m_{max} [1,30]. During the constant feed phase the specific growth rate decreased to values below 0.05 h^{-1}, when the glucose supply fell to values below 0.1 g_{glu} g_x^{-1} h^{-1} (Fig. 3). In the constant, slow feeding phase, acetate consumption was greater than its production, which represented a shift in the cycling equilibrium that favoured the consumption of extracellular acetate (point B, Fig. 2). The profiles of all measurable state variables (X, S, A, DOT) in the model are plotted in Fig. 2, with the corresponding model predictions after parame-

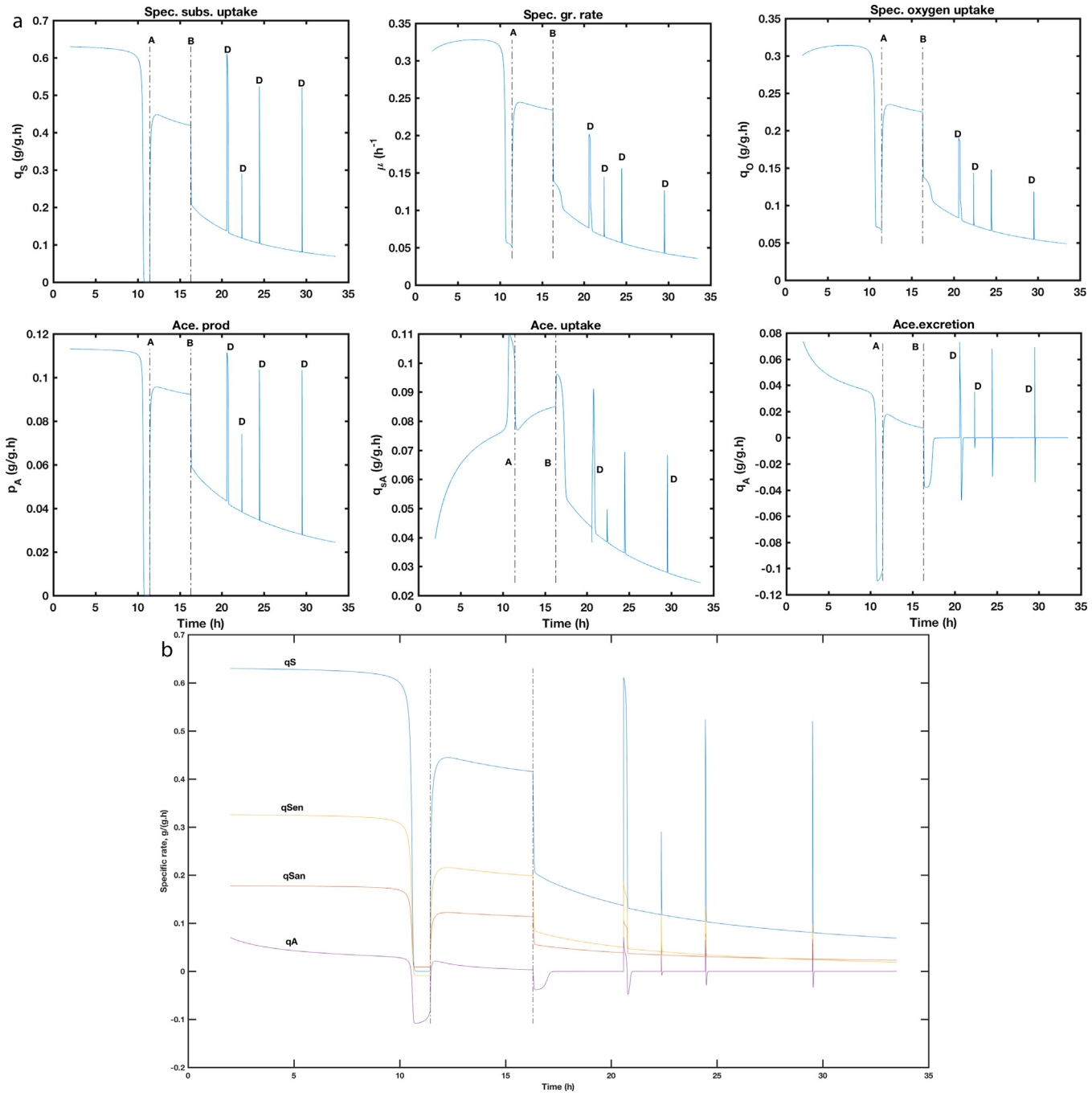


Fig. 3. Simulated specific rates of the major variables of the (A) acetate cycling pathways and (B) overall glucose partitioning during the cultivation of *E. coli*, with dynamic response to intermittent glucose pulses. Negative acetate excretion rates imply that extracellular acetate is taken up by the cell. A = end of batch phase, B = end of exponential feed, start of constant feed, C = change in KLa, D = glucose pulses.

ter estimation. There was an adequate fit between the model and the experimental data. The estimated model parameters based on the experimental data are given in Table 1 together with the uncertainty in the estimated parameter values calculated from the Monte Carlo method. From the measure of the relative standard deviation, it is apparent that the PE gave unique parameter values that describe the *E. coli* system under the given experimental conditions. Twelve (12) out of the sixteen (16) parameters had a σθvalues less than 10%. Although some of the parameters directly related to acetate cycling had somewhat higher σθ values, the acetate profile in the cultivation was sufficiently described by the model.

4.1. Dynamic acetate production and consumption rates

The estimation of model parameters was used to identify parameter values (Table 1) to fully describe the *E. coli* system. Therefore, the partitioning of the carbon source (Fig. 3B), the rate of acetate cycling and other *specific* rates along the different phases of the cultivation could be dynamically simulated from the estimated parameter values (Fig. 3A). The simulated profiles and the model fit (Fig. 2) reveal a rapid response of *E. coli* to acetate production/reconsumption after each glucose pulse. The absence of a diauxic delay during the consumption of acetate after each pulse and after the shift to a lower μ ($0.11\,h^{-1}$) suggests that acetate is contin-

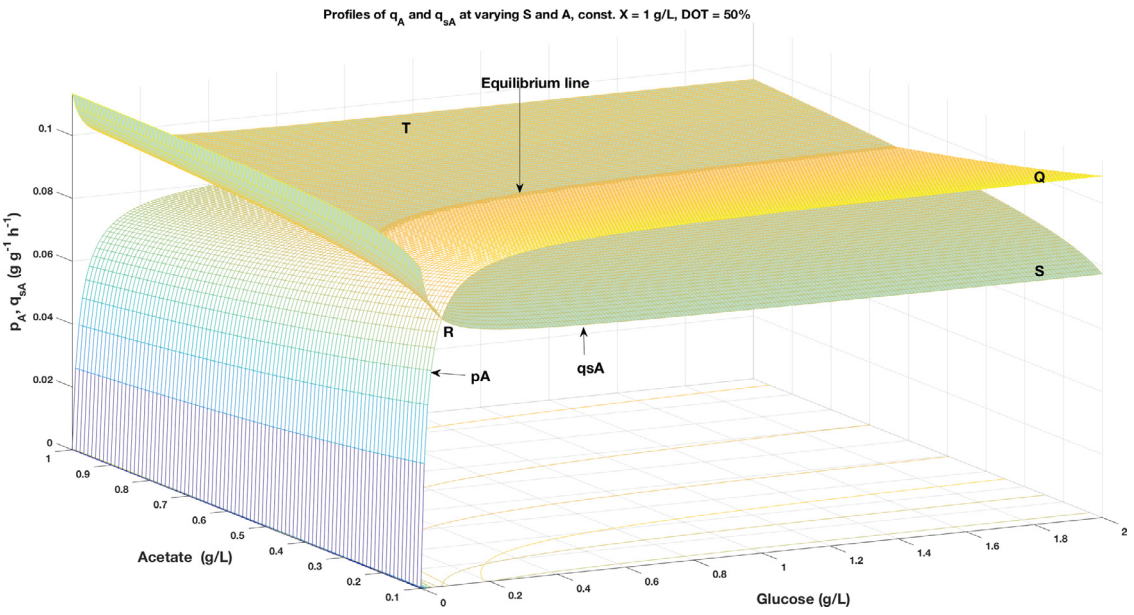


Fig. 4. Dependence of overflow metabolism on residual glucose and acetate concentrations, showing the dynamic equilibrium between intracellular acetate production and re-assimilation at varying glucose concentrations in the medium.

uously being re-assimilated in the background alongside glucose [31–33], which further confirms the acetate cycling concept in *E. coli*. Due to the lack of an active transport mechanism for acetate across the cell membrane in *E. coli* [34,35] (unlike glucose which uses the PTS system), q_A also incorporates the net acetate influx by facilitated transport across the cell membrane by various permease proteins, such as those encoded by the genes *yaaH* [36] and *yjcG* [37]. Re-assimilation of acetate results in lower intracellular acetate concentrations, which alters the membrane gradients and forces extracellular acetate into the cell by facilitated transmembrane balance. In effect, when intracellular acetate production through POXB (Fig. 1) is non-functional (e.g. when $qs = 0$), p_A is derived from extracellular acetate by this principle of transmembrane balance. The inhibition of extracellular acetate uptake by glucose is evidenced by the sharply rising profile of q_{SA} in response to the depletion of glucose during the batch phase (Fig. 3). The delicate equilibrium in the acetate cycling pathways is shown in the 3D plot in Fig. 4, which was generated by solving the set of algebraic equations at constant biomass concentration (X = 1.3 g/L) and DOT (65%) and variable glucose and acetate concentrations. This corresponds to about 9 h in the batch phase. The point labelled T in Fig. 4 represents pure oxidative growth. As the residual glucose concentration increases, the metabolism gradually shifts towards the equilibrium line, until $p_A > q_{sA}$ where acetate excretion begins. At this point, the total overflow flux that leads to extracellular acetate accumulation is proportional to the area enclosed by the envelope QRS and the equilibrium line. Thus, at any given glucose concentration, the overflow flux, and consequently, the excreted acetate concentration can be estimated from the algebraic relations. The Monod-type dependence of the specific acetate uptake rate on the residual acetate concentration as well as the minimal acetate production at low glucose concentrations are also shown in Fig. 4. The formulation is an advanced version of a similar representation of overflow metabolism in *Saccharomyces cerevisiae*, as presented by Pham and co-workers [14]. This formulation of acetate profiles and overflow metabolism is closer to reality since *E. coli* does not have clear cut switching points in its metabolic routes on the onset of overflow, but rather slow or gradual switching systems from oxidative to overflow metabolism [38,39].

5. Conclusions

Model-based process development and optimization is becoming state-of-the-art in the biotechnology industry. Therefore, the mathematical functions used to describe fermentation systems should have the appropriate properties to enable their application in this fast-growing field. We have used the concept of acetate cycling to derive a set of continuously differentiable and tractable equations to describe acetate accumulation in cultivations of *E. coli*. By fitting this model to experimental data, we show that the new set of equations sufficiently describe growth profiles in *E. coli*, as well as the acetate production and re-assimilation rates. Due to its continuous characteristics, this model is suitable for simulations that require higher order gradient calculation such as Computational Fluid Dynamics (CFD) models, as well as in complex optimization problems involving the *E. coli* expression system.

Acknowledgements

The authors are grateful to Prof Emeritus Sven-Olof Enfors (rtd) for his valuable input on the development of the model. EA is grateful for financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie actions grant agreement No. 643056 (Biorapid). DCLC gratefully acknowledges financial support from the Faculty for the Future Fellowship programme (Schlumberger Foundation) for her post-doctoral scholarship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bej.2017.05.013.

The Matlab code and fermentation data can be found at https://gitlab.tubit.tu-berlin.de/nicolas.cruz/E_coli_fed-batch/tree/master/Anane_2017_BEJ.

References

[1] A.J. Wolfe, The acetate switch, Microbiol. Mol. Biol. Rev. 69 (2005) 12–50, http://dx.doi.org/10.1128/MMBR.69.1.12-50.2005.

[2] M. De Mey, S. De Maeseneire, W. Soetaert, E. Vandamme, Minimizing acetate formation in E. coli fermentations, J. Ind. Microbiol. Biotechnol. 34 (2007) 689–700, http://dx.doi.org/10.1007/s10295-007-0244-2.

[3] G. Insel, G. Celikyilmaz, E. Ucisik-Akkaya, K. Yesiladali, Z.P. Cakar, C. Tamerler, D. Orhon, Respirometric evaluation and modeling of glucose utilization by Escherichia coli under aerobic and mesophilic cultivation conditions, Biotechnol. Bioeng. 96 (2007) 94–105, http://dx.doi.org/10.1002/bit.21163.

[4] B. Xu, M. Jahic, S.-O. Enfors, Modeling of overflow metabolism in batch and fed-batch cultures of escherichia coli, Biotechnol. Prog. 15 (1999) 81–90, http://dx.doi.org/10.1021/bp9801087.

[5] 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 (2001) 347–357, http://dx.doi.org/10.1002/bit.1068.

[6] Y.F. Ko, W.E. Bentley, W. a Weigand, A metabolic model of cellular energetics and carbon flux during aerobic Escherichia coli fermentation, Biotechnol. Bioeng. 43 (1994) 847–855, http://dx.doi.org/10.1002/bit.260430903.

[7] A. Cockshott, I. Bogle, Modelling the effects of glucose feeding on a recombinant E-coli fermentation, Bioprocess. Eng. 20 (1999) 83–90, http://dx.doi.org/10.1007/PL00009037.

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

[9] K. Peebo, K. Valgepea, A. Maser, R. Nahku, K. Adamberg, R. Vilu, Proteome reallocation in Escherichia coli with increasing specific growth rate, Mol. Biosyst. 11 (2015) 1184–1193, http://dx.doi.org/10.1039/C4MB00721B.

[10] K. Valgepea, K. Adamberg, R. Nahku, P.-J. Lahtvee, L. Arike, R. Vilu, Systems biology approach reveals that overflow metabolism of acetate in Escherichia coli is triggered by carbon catabolite repression of acetyl-CoA synthetase, BMC Syst. Biol. 4 (2010) 166, http://dx.doi.org/10.1186/1752-0509-4-166.

[11] M. Basan, S. Hui, H. Okano, Z. Zhang, Y. Shen, J.R. Williamson, T. Hwa, Overflow metabolism in Escherichia coli results from efficient proteome allocation, Nature 528 (2015) 99–104, http://dx.doi.org/10.1038/nature15765.

[12] M. Matsubara, T. Mizuno, EnvZ-independent phosphotransfer signaling pathway of the ompR-mediated osmoregulatory expression of OmpC and OmpF in escherichia coli, Biosci. Biotechnol. Biochem. 63 (1999), http://dx.doi.org/10.1271/bbb.63.408.

[13] S. Renilla, V. Bernal, T. Fuhrer, S. Castaño-Cerezo, J.M. Pastor, J.L. Iborra, U. Sauer, M. Cánovas, Acetate scavenging activity in Escherichia coli: interplay of acetyl-CoA synthetase and the PEP-glyoxylate cycle in chemostat cultures, Appl. Microbiol. Biotechnol. 93 (2012) 2109–2124, http://dx.doi.org/10.1007/s00253-011-3536-4.

[14] H.T.B. Pham, G. Larsson, S.O. Enfors, Growth and energy metabolism in aerobic fed-batch cultures of Saccharomyces cerevisiae: simulation and model verification, Biotechnol. Bioeng. 60 (1998) 474–482, http://dx.doi.org/10.1002/(SICI)1097-0290(19981120)60:4<474:AID-BIT9>3.0.CO;2-J.

[15] P. Millard, K. Smallbone, P. Mendes, Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in Escherichia coli, PLoS Comput. Biol. 13 (2017) e1005396, http://dx.doi.org/10.1371/journal.pcbi.1005396.

[16] K. Martínez-Gómez, N. Flores, H.M. Castañeda, G. Martínez-Batallar, G. Hernández-Chávez, O.T. Ramírez, G. Gosset, S. Encarnación, F. Bolivar, New insights into Escherichia coli metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol, Microb. Cell Fact. 11 (2012) 46, http://dx.doi.org/10.1186/1475-2859-11-46.

[17] M.N. Cruz Bournazou, T. Barz, D. Nickel, D. López Cárdenas, F. Glauche, A. Knepper, P. Neubauer, Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities for validation of macro-kinetic growth models using E. coli as an example, Biotechnol. Bioeng. 9999 (2016) 1–10, http://dx.doi.org/10.1002/bit.26192.

[18] D.B. Nickel, N.M. Cruz-Bournazou, T. Wilms, P. Neubauer, A. Knepper, Online bioprocess data generation, analysis and optimization for parallel fed-batch fermentations at mL scale, Eng. Life Sci. (2016) 1–7, http://dx.doi.org/10.1002/elsc.201600035.

[19] L.T. Biegler, Recent advances in chemical process optimization, Chemie Ing. Tech. 86 (2014) 943–952, http://dx.doi.org/10.1002/cite.201400033.

[20] J. Glazyrina, E.-M. Materne, T. Dreher, D. Storm, S. Junne, T. Adams, G. Greller, P. Neubauer, High cell density cultivation and recombinant protein production with Escherichia coli in a rocking-motion-type bioreactor, Microb. Cell Fact. 9 (2010) 42, http://dx.doi.org/10.1186/1475-2859-9-42.

[21] A.C. Hindmarsh, P.N. Brown, K.E. Grant, S.L. Lee, R. Serban, D.E. Shumaker, C.S. Woodward, SUNDIALS: suite of nonlinear and differential/algebraic equation solvers, ACM Trans. Math. Softw. 31 (2005) 363–396, http://dx.doi.org/10.1145/1089014.1089020.

[22] D.C. López C, T. Barz, M. Penula, V. Adriana, S. Ochoa, G. Wozny, Model-based identifiable parameter determination applied to a simultaneous saccharification and fermentation process model for bio-ethanol production, Biotechnol. Prog. 29 (2013) 1064–1082, http://dx.doi.org/10.1002/btpr.1753.

[23] Y. Bard, Non-linear parameter estimation, Academic Press, New York, 1974.

[24] D.C. López, G. Wozny, A. Flores-Tlacuahuac, R. Vasquez-Medrano, V.M. Zavala, A computational framework for identi fi ability and ill-conditioning analysis of lithium-ion battery models, Ind. Eng. Chem. Res. 55 (2016) 3026–3042, http://dx.doi.org/10.1021/acs.iecr.5b03910.

[25] L.A. Tribe, C.L. Briens, A. Margaritis, Determination of the volumetric mass transfer coefficient (k(L)a) using the dynamic gas out-gas in method: analysis of errors caused by dissolved oxygen probes, Biotechnol. Bioeng. 46 (1995) 388–392, http://dx.doi.org/10.1002/bit.260460412.

[26] J. Shiloach, R. Fass, Growing E. coli to high cell density – A historical perspective on method development, Biotechnol. Adv. 23 (2005) 345–357, http://dx.doi.org/10.1016/j.biotechadv.2005.04.004.

[27] H. Ying Lin, P. Neubauer, Influence of controlled glucose oscillations on a fed-batch process of recombinant Escherichia coli, J. Biotechnol. 79 (2000) 27–37, http://dx.doi.org/10.1016/S0168-1656(00)00217-0.

[28] A.C. Badino, M. Cândida, R. Facciotti, W. Schmidell, Improving k(L)a determination in fungal fermentation, taking into account electrode response time, J. Chem. Technol. Biotechnol. 75 (2000) 469–474, http://dx.doi.org/10.1002/1097-4660(200006)75:6<469:AID-JCTB236>3.0.CO;2-4.

[29] W. Chesbro, M. Arbige, R. Eifert, When nutrient limitation places bacteria in the domains of slow growth: metabolic, morphologic and cell cycle behavior, FEMS Microbiol. Ecol. 74 (1990) 103–119.

[30] E. Anane, E. van Rensburg, J.F. Görgens, Optimisation and scale-up of α-glucuronidase production by recombinant Saccharomyces cerevisiae in aerobic fed-batch culture with constant growth rate, Biochem. Eng. J. 81 (2013) 1–7, http://dx.doi.org/10.1016/j.bej.2013.09.012.

[31] V. Bernal, S. Castaño-Cerezo, M. Cánovas, Acetate metabolism regulation in Escherichia coli: carbon overflow, pathogenicity, and beyond, Appl. Microbiol. Biotechnol. 100 (2016) 8985–9001, http://dx.doi.org/10.1007/s00253-016-7832-x.

[32] B. Enjalbert, M. Cocaign-Bousquet, J.C. Portais, F. Letisse, Acetate exposure determines the diauxic behavior of Escherichia coli during the glucose-acetate transition, J. Bacteriol. 197 (2015) 3173–3181, http://dx.doi.org/10.1128/JB.00128-15.

[33] S. Leone, F. Sannino, M.L. Tutino, E. Parrilli, D. Picone, Acetate: friend or foe? Efficient production of a sweet protein in *Escherichia coli* BL21 using acetate as a carbon source, Microb. Cell Fact. 14 (2015) 106, http://dx.doi.org/10.1186/s12934-015-0299-0.

[34] D.D. Axe, J.E. Bailey, Transport of lactate and acetate through the energized cytoplasmic membrane of Escherichia coli, Biotechnol. Bioeng. 47 (1995) 8–19, http://dx.doi.org/10.1002/bit.260470103.

[35] R.R. Wright, J.E. Hobbie, Use of glucose and acetate by bacteria and algae in aquatic ecosystems, Ecology 47 (1966) 447–464, http://dx.doi.org/10.2307/1932984.

[36] J. Sá-Pessoa, S. Paiva, D. Ribas, I.J. Silva, S.C. Viegas, C.M. Arraiano, M. Casal, SATP (YaaH), a succinate-acetate transporter protein in Escherichia coli, Biochem. J. 454 (2013) 585–595, http://dx.doi.org/10.1042/bj20130412.

[37] R. Gimenez, M.F. Nunez, J. Badia, J. Aguilar, L. Baldoma, The gene yjcG, cotranscribed with the gene acs, encodes an acetate permease in escherichia coli, J. Bacteriol. 185 (2003) 6448–6455, http://dx.doi.org/10.1128/JB.185.21.6448-6455.2003.

[38] A. Kayser, J. Weber, V. Hecht, U. Rinas, Metabolic flux analysis of Escherichia coli in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state, Microbiology 151 (2005) 693–706, http://dx.doi.org/10.1099/mic.0.27481-0.

[39] T. Paalme, R. Elken, A. Kahru, K. Vanatalu, R. Vilu, The growth rate control in Escherichia coli at near to maximum growth rates: the A-stat approach, Antonie Van Leeuwenhoek 71 (1997) 217–230, http://dx.doi.org/10.1023/A:1000198404007.

[40] Z. Li, M. Nimtz, U. Rinas, The metabolic potential of Escherichia coli BL21 in defined and rich medium, Microb. Cell Fact. 13 (2014) 45, http://dx.doi.org/10.1186/1475-2859-13-45.

II

Output uncertainty of dynamic growth models: effect of uncertain parameter estimates on model reliability

Emmanuel Anane[1], Diana C. López C.[1], Tilman Barz[2], Gurkan Sin[3], Krist V. Gernaey[3], Peter Neubauer[1], Mariano Nicolas Cruz Bournazou*[1]

[1]Chair of Bioprocess Engineering, Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany

[2]Department of Energy, Austrian Institute of Technology GmbH, Vienna, Austria

[3]Process and Systems Engineering Centre (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Lyngby, Denmark

*Corresponding Author:

Dr.-Ing. Mariano Nicolas Cruz Bournazou, Chair of Bioprocess Engineering, Institute of Biotechnology, Technische Universität Berlin, Ackerstraße 76, ACK24, 13355 Berlin, Germany.

Phone: +49 30 314 72626

Email: nicolas.cruz@mailbox.tu-berlin.de

Submitted to Biochem Eng. Journal (8th February 2019)

Abstract

Mechanistic models are simplifications of bio-physical systems, for which the true values of the parameters of the model are usually unknown. Therefore, before using model based predictions to study or improve a process, it is essential to ensure that the outputs of the model are reliable. Hence, it is imperative that the existing mathematical tools for analysis of the certainty of model outputs are also available for and usable by non-experts in the field. The underlying problem is to quantify the uncertainty generated by inappropriate model structures combined with experimental data that does not contain sufficient information about the system being described.

This paper covers the development and application of a compact framework for practical identifiability and uncertainty analyses of dynamic growth models for bioprocesses. By exploring the basic numerical properties of the sensitivity matrix, a simple algorithm to determine the presence of non-identifiable parameters in models with high output uncertainty is presented. As an example, the framework was used to analyse a macro-kinetic growth model of *Escherichia coli* describing a fed-batch process. The aim of this work is to present a simplified method that will promote a more thorough handling of dynamical models, even for non-experienced modellers, so as to extend the use of mechanistic models in biotechnology.

Keywords

Parameter identifiability; *Escherichia coli*; Fed-batch simulation; Ill-conditioning analysis; Uncertainty analysis

1.0 Introduction

The contribution of mathematical models to our understanding of biological systems from Monod (Monod, 1949) and Andrews (Andrews, 1968) to bioprocess engineering (Batstone et al., 2002), biochemical systems (Savageau, 1970), systems biology (Kitano, 2002), metabolic engineering (Stephanopoulos, Aristidou, & Nielsen, 1998), flux balance analysis (Varma & Palsson, 1994), synthetic biology (Marchisio & Stelling, 2009) and bioinformatics (Saeys, Inza, & Larrañaga, 2007) is undisputed. The reader is referred to Jay Bailey (Bailey, 1998) for a critical overview of the development, application and potential of mathematical models in biotechnology. Unfortunately, the usability of models in biological systems seems to be underestimated, showing only a slow advance in tackling the challenges that were stated two decades ago (Bailey, 1998). There is a general consensus that the current state of mathematical models, and in general model based tools for advanced engineering, "*is not readily applicable to the bio-industry*" (Koutinas, Kiparissides, Pistikopoulos, & Mantalaris, 2012).

One of the major obstacles hampering the use and acceptance of mechanistic models in bioprocesses is the lack of simple reliability tests for the predictions of a model. Especially in bioengineering, first principles models need to describe large, highly nonlinear and state dependent processes using simplified equations and lumped kinetic parameters. Due to the complexity of the underlying metabolic and physiological phenomena, a large number of parameters is typically required to characterise a given biological system. Routinely, a so-called *validation* of the mechanistic model is claimed by simply presenting the outputs of the model (at some estimated parameter values) showing a good agreement with experimental data. But by this the reliability of the parameter estimates is not considered, a typical issue being the existence of infinite combinations of parameter values that can show the same fit (a non-identifiable parameter set). Whereas data-based models are categorically rejected if not presented with its associated confidence regions (e.g. error bars), mechanistic models are rarely shown with similar reliability tests (G Sin, Gernaey, & Lantz, 2009). More importantly, one of the major advantages of mechanistic models is their ability to predict the systems behaviour beyond the region of experimentation (e.g. in scale-up), so that it is crucial to evaluate its predictive power with sound reliability assessment methods. Today,

many mechanistic biological models are published without testing their structural and practical identifiability, the uncertainty of parameter estimates after fitting, and the propagation of this uncertainty on model outputs. The reason for this is not a lack of sufficient literature focused on parameter identifiability of mechanistic models. There are numerous papers and reviews on the topic for both biosystems (Bellman & Astrom, 1970; Cobelli & DiStefano, 1980; Andreas Raue et al., 2009) and general mechanistic models (Bard, 1974; Brun, Martin, Siegrist, Gujer, & Reichert, 2002; Kravaris, Hahn, & Chu, 2013; Vajda, Rabitz, Walter, & Lecourtier, 1989). Additionally, software packages, such as Amigo2 (Balsa-Canto, Henriques, Gábor, & Banga, 2016), Data2Dynamics (A. Raue et al., 2015) and BioPreDyn suite (Villaverde et al., 2015) have been developed as generic plug-and-play platforms for advanced analysis, including identifiability and sensitivity analyses especially suited for systems biology models. Whereas these works and packages are suitable for application by advanced modellers, the complex mathematical concepts presented in them are usually beyond the academic training and mathematical background of biotechnologists and bioprocess engineers (Muñoz-Tamayo et al., 2018; Villaverde, Barreiro, & Papachristodoulou, 2016). Furthermore, there are many methods at different levels of complexity with their respective advantages and disadvantages, so that it is very easy to get lost in the search of a simple, reliable and thorough methodology to test the reliability of parameter estimates.

The aim of this paper is to present a straightforward framework for the analysis of the reliability of the outputs of a given nonlinear dynamical model, specifically directed at non-experts in the field. We deal with this issue from the point of view of an engineer, biotechnologist or a molecular biologist with limited training in these mathematically loaded methods. As an example, we apply this strategy to analyse a macro-kinetic growth model of *Escherichia coli,* and the effect of an ill-conditioned parameter estimation on model predictions assessed by output uncertainty quantification. With this simplified framework, we aim to promote a more thorough handling of relatively complex first principles models in biotechnology so as to increase their reliability and acceptance, especially when these models are used to elucidate intracellular mechanisms or biochemical pathways and when they are applied in biomanufacturing.

2.0 Reliability of model outputs

The outputs y of a dynamical model (Equation 1) depend on the initial values of the state variables x_0, the experimental conditions u, and the parameter vector θ. Assuming the structure of the model is correct with random experimental errors and knowing x_0 and u (already a strong statement), we can confine the parameter values to some region of confidence. Thus, since we do not know the exact values of θ and θ influences y, we know that the outputs of our simulation y must entail some uncertainty.

For the sake of clarity, we consider that we can describe the process of interest by a mathematical model consisting of a non-linear Ordinary Differential Equation (ODE) system with the following characteristics:

$$y(t) = A(x(t))$$

$$\dot{x}(t) = f(x(t), u(t); \theta) \qquad (1)$$

$$x(t_0) = x_0,$$

where t is the independent variable time, $x(t) \in \mathbb{R}^{n_x}$ is the vector containing the dependent state variables. $\theta \in \mathbb{R}^{n_p}$ is the unknown parameter vector (with n_p parameters), $u(t) \subseteq \mathrm{U} \in \mathbb{R}^{\mathrm{n_u}}$ contains the time-varying input signals (i.e. experimental design variables). The initial conditions are given by $x_0 \in X_0$, where $y(t) \in \mathbb{R}^{n_y}$ is a vector of predicted output (response) variables. The measured state variables, y(t) comprising both on-line (y_{on}) and off-line (y_{of}) measurements is a subset of x(t), defined by the constant selection matrix $A \in \mathbb{R}^{n_y \cdot n_x}$.

The uncertainty on $y(t)$ depends on the uncertainty of θ and on how changes in θ are propagated throughout the model. For a given setup, x_0 and $u(t)$, we can compute a point estimate of the model parameters in θ by fitting it to experimental data (performing a Parameter Estimation PE). The important question is: how precisely can we estimate θ at the x_0 and $u(t)$ defined in our experiments, so that $y(t)$ has low uncertainty even at different conditions of x_0 and $u(t)$ defined by e.g. a different reactor or the industrial process at different scale?

We can calculate some approximations to answer this question and for this we need to find out two things i) the identifiability of θ (how large is the confidence interval of the estimates of θ after the PE) and ii) how does this uncertainty on θ affect the reliability of the outputs y at different conditions.

3.0 Some existing methods for parameter identifiability analysis

Parameter identifiability analysis is used to test whether a unique estimate within a confidence region can be assigned to the parameter set of a model after parameter estimation from experimental data. The model is said to be *structurally non-identifiable* (Almquist, Cvijovic, Hatzimanikatis, Nielsen, & Jirstrand, 2014; Saccomani, 2013; Villaverde et al., 2016) when it has inherently redundant parameters that cannot be uniquely estimated independently from the observations available. On the other hand, if some parameters are non-identifiable due to limitations in the quantity or quality of the data (information content of the data), the model is considered to be *practically non-identifiable* (Franceschini & Macchietto, 2008). Whereas structural non-identifiability can be addressed through model reduction (Vora & Daoutidis, 2001), practical non-identifiability can be addressed by improving the data quality (performing more informative experiments (A. Raue, Kreutz, Maiwald, Klingmuller, & Timmer, 2011) or (at the cost of a higher bias) by regularization of the parameter estimation problem (e.g. Tikhonov (Johansen, 1997) or singular value truncation (López C., Barz, Körkel, & Wozny, 2015)).

There are several methods to analyse the identifiability of a model. The methods for structural identifiability analysis include, among others, the Taylor series expansion, Lie derivatives and differential algebra. The reader is referred to the review by Chis and colleges (Chis, Banga, & Balsa-Canto, 2011) for a comprehensive treatment of some of these methods.

Practical identifiability methods include likelihood profiling, eigenvalue decomposition of the Fisher Information Matrix-FIM (Quaiser & Mönnigmann, 2009) and various forms of combinatorial methods that rank parameter subsets according to practical identifiability using the properties of the eigenvalues (Brockmann, Rosenwinkel, & Morgenroth, 2008; Brun et al., 2002; Lencastre Fernandes et al., 2012; Weijers & Vanrolleghem, 1997).

In this work, we focus on the case where i) the model is at hand and has been developed using a priori knowledge of the system (e.g. cell physiology, metabolomics) and ii) the reliability of the existing model at its "true" parameter values is of interest. Hence, the aim is to provide a comprehensive framework for practical (local) identifiability of biochemical models by studying the numerical properties of the model sensitivity matrix.

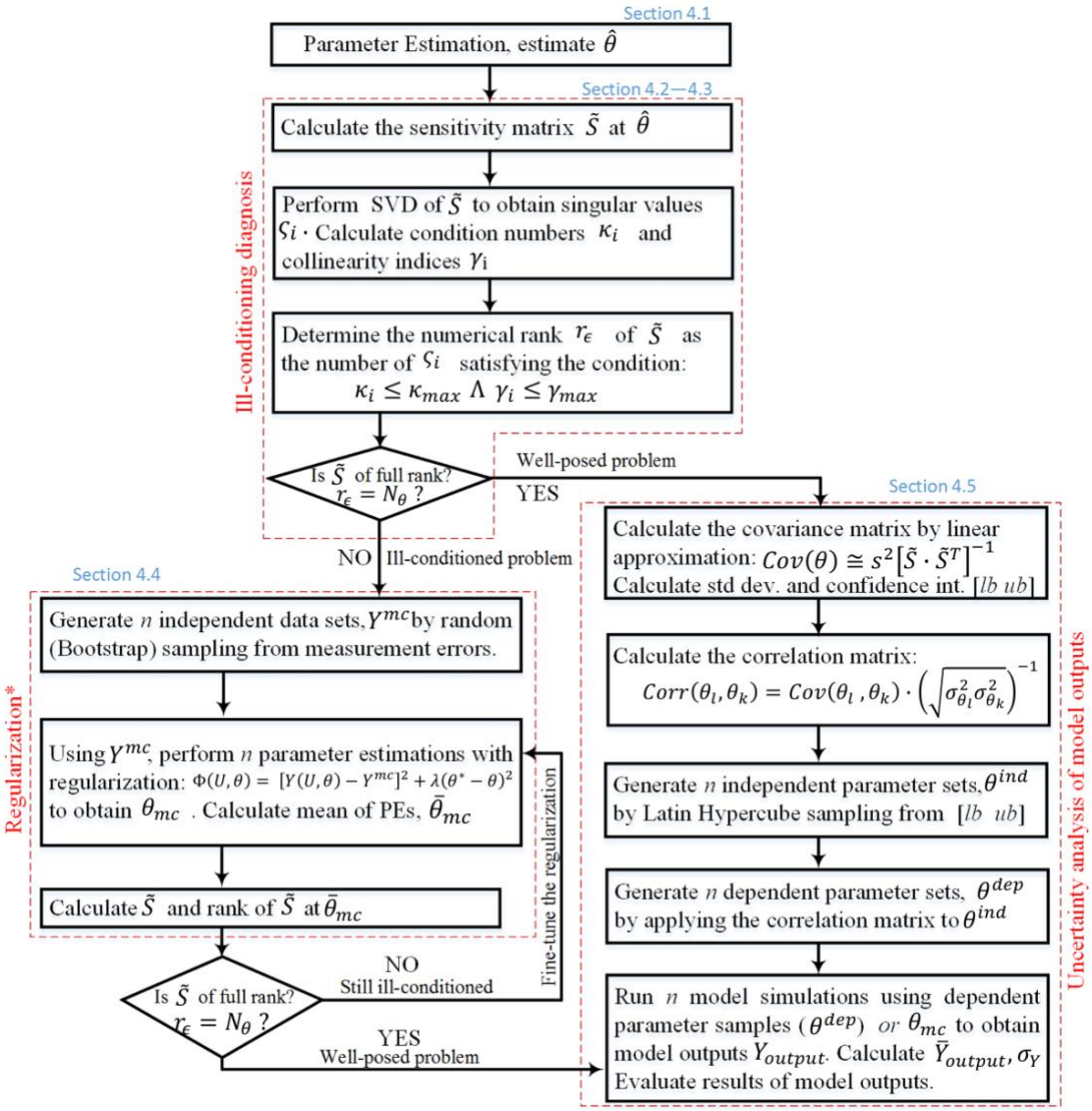


Figure 1 Framework for ill-conditioning diagnosis, regularization and uncertainty analyses of mechanistic growth models. SVD = Singular value decomposition. The thresholds κ_{max} and $\gamma_{\max}$ are empirical coefficients. *An alternative method of regularization, namely Subset Selection (SsS) is presented in the Supplementary material.

4.0 Model analysis framework

The framework comprises three major tasks (Figure 1): (i) ill-conditioning diagnosis, (ii) regularization to handle ill-conditioned problems and (iii) propagation of the parameter uncertainty onto model outputs (output uncertainty analysis). The model analysis technique presented in Figure 1 is similar to the method of eigenvalue decomposition of FIM by Quaiser and Mönnigmann (2009).

However, in order to circumvent problems related to the calculation of an ill-conditioned FIM, it uses the singular value decomposition (SVD) of the sensitivity matrix (Belsley, Kuh, & Welsch, 1980) and further includes a Monte Carlo based method to quantify parameter uncertainty if the problem is ill-conditioned. This is a local analysis of the sensitivity matrix (the results are dependent on the parameter values). All these steps are preceded by data collection from experimental runs and a preliminary parameter estimation. The details of the workflow in the model analysis framework are discussed in the following sections.

4.1 Parameter Estimation (PE)

During parameter estimation, the true parameter values of the model are inferred by minimizing the difference between the observed data and the model predictions. The parameters θ of the dynamical model are typically estimated by solving the optimization problem of the form:

$$\hat{\theta} \coloneqq \arg\min_{\theta} \Phi(U, \theta) \qquad (2),$$

where the objective function $\Phi(U, \theta)$, which in this case is the weighted nonlinear least-squares function between the model predictions $Y(U, \theta)$ and the experimental data Y^m, is calculated using the maximum likelihood estimator:

$$\Phi(U, \theta) \coloneqq \frac{1}{2}(Y(U, \theta) - Y^m)^T \, \mathrm{C_y^{-1}} \, (Y(U, \theta) - Y^m) \qquad (3)$$

In Equation 3, all the measured data collected at the experimental conditions (x_0, u) are compiled in the vector Y^m. To take the measurement errors into account during parameter estimation, the objective function is weighted by the variance-covariance matrix, C_y which in our case is obtained assuming unbiased, independent and normally

identically distributed measurement errors. Therefore C_y is a diagonal matrix with entries given by the variance σ_y^2 of each measurement.

4.2 Sensitivity Matrix

The identifiability of a parameter set (how accurate can we estimate the values of the parameters) is defined by two properties: (i) parameter sensitivity: the effect this parameter has on the objective function; and, (ii) parameter correlation: how correlated this effect is with the effects of the other parameters. Hence, we obtain useful information about how well we can estimate the parameters by analysing the properties of the sensitivity matrix S Equation 4 (Weijers & Vanrolleghem, 1997) computed from the first order partial derivatives of the state variables (y) with respect to the parameters,

$$S = \begin{bmatrix} \frac{\partial y_1}{\partial \theta_1} & \frac{\partial y_1}{\partial \theta_2} & \cdots & \frac{\partial y_1}{\partial \theta_{N_\theta}} \\ \vdots & \cdots & & \vdots \\ \frac{\partial y_{Nv}}{\partial \theta_1} & \frac{\partial y_{Nv}}{\partial \theta_2} & \cdots & \frac{\partial y_{Nv}}{\partial \theta_{N_\theta}} \end{bmatrix} \quad (4)$$

where N_v is the number of state variables and N_θ is the number of parameters. Since the analysis considers dynamical models, S is time-variant. The sensitivity matrix can be calculated in a simple way with the method of finite differences (i.e. perturbation method) (Guisasola et al., 2006). Nevertheless, the computational burden increases exponentially with the number of parameters and has inherent inaccuracies associated with the numerical integrator and the magnitude of parameter perturbations ($\Delta\theta$) (De Pauw & Vanrolleghem, 2006).

In the forward method, the dynamics of the sensitivity functions along the time vector are given by (Franceschini & Macchietto, 2008; Gábor & Banga, 2015)

$$\frac{dS}{dt} = \frac{\partial f}{\partial y} S + \frac{\partial f}{\partial \theta} \quad (5)$$

where $\frac{\partial f}{\partial y}$ represents the Jacobian of the system w.r.t the state variables and $\frac{\partial f}{\partial \theta}$ is the Jacobian w.r.t the parameters. The dynamic sensitivity matrix *S* of the measured state variables *Y* is obtained by integrating Equation 5 together with the model equations, with the initial condition $\mathrm{S}(\mathrm{t}_0) = \frac{\partial f(t_0)}{\partial \theta}$.

$$S = \left[\left.\frac{\partial Y}{\partial \theta_1}\right|_{t1} \quad \left.\frac{\partial Y}{\partial \theta_1}\right|_{t2} \cdots \left.\frac{\partial Y}{\partial \theta_1}\right|_{tn} \right]^T \tag{6}$$

The Symbolic Math Toolbox in Matlab, Maple, or Sympy among others can be used to calculate the partial differentials of Y w.r.t θ, and these can be integrated together with the model equations to obtain the dynamic sensitivity matrix S. The non-dimensional sensitivity matrix $\tilde{S}$ is obtained by normalizing S with the nominal parameter and model output values (Gürkan Sin & Gernaey, 2016). Since the sensitivity matrix is computed at the given parameter values and not at all the possible parameter combinations, the analysis in this framework is local in nature.

4.3 Ill-conditioning Diagnosis

The PE is an optimization problem with no warranty to have a unique solution (a necessary condition for well-posedness). That is, if there exist various parameter combinations that deliver similar values of the objective function Φ, the problem is said to be ill-posed, hence causing ill-conditioned numeric approximation. Practically, this means that very large values are divided by very small ones during the numerical computation of the parameter estimates, causing large inaccuracies and instability in the results. When the PE problem is ill-conditioned, it leads to overestimation of the variance of the parameter estimates. Additionally, the experimental data are inherently subject to errors. Hence, it is important to determine how the conditioning of the parameter estimation problem and measurement errors translate into parameter uncertainty, even for well-posed problems. The analysis presented in the current framework to carry out ill-conditioning diagnosis is based on the orthogonal decomposition of the sensitivity matrix $\tilde{S}$. This decomposition, namely singular value decomposition (SVD) is used to calculate the rank of $\tilde{S}$, used to detect linear dependences that result in a non-invertible Fisher-information and Jacobian matrices (making the estimation ill-conditioned). The N_Y-*by*-N_θ sensitivity matrix $\tilde{S}$ is decomposed according to Equation 7

$$\tilde{S} = U\Sigma V' \tag{7},$$

where U (N_Y-*by*-N_Y) and V (N_θ -*by*-N_θ) are orthogonal unitary matrices and Σ is an N_Y-*by*-N_θ diagonal matrix with ordered non-negative diagonal elements $\varsigma_{11} \geq \varsigma_{22} \geq$

$\varsigma_{33} \ldots \geq \varsigma_{N_\theta}$, which are the singular values of $\tilde{S}$ (Chan & Hansen, 1992; Golub & Van Loan, 1996). These singular values are then used to calculate two metrics (López, Wozny, Flores-Tlacuahuac, Vasquez-Medrano, & Zavala, 2016): the condition number (κ) and collinearity index (γ), as

$$\kappa_i = \frac{\varsigma_{max}}{\varsigma_i} \qquad \text{and} \qquad \gamma_i = \frac{1}{\varsigma_i} \tag{8}$$

The condition number (κ) of parameter *i* is a measure of the sensitivity of the model outputs to perturbations in this parameter, whereas the collinearity index is a measure of the linear independence of the parameter (Brun et al., 2002). The numerical rank r_ϵ is the number of singular values that are greater than ϵ, where ϵ is a threshold defined by the combined empirical conditions

$$\epsilon := \kappa_i \leq \kappa_{max} \ \Lambda \ \gamma_i \leq \gamma_{max} \tag{9}$$

The limits κ_{max} and $\gamma_{\max}$ are empirical thresholds determined as 1000 and 15, respectively, according to the works of Grah (López et al., 2016) and Brun (Brun et al., 2002). The matrix is said to be rank-deficient, indicating a covariance matrix close to singular and an ill posed PE, if $r_\epsilon < N_\theta$, where N_θ is the number of parameters.

4.4 Regularization

If the PE turns out to be ill-posed ($r_\epsilon < N_\theta$), there are several options and methods to regularize the problem (at the cost of adding *a priori* information). If the issue is structural identifiability, methods that range from sophisticated mathematical tools to broad simplifications in model structure exist to reduce the number of states and parameters (Cruz Bournazou, Arellano-Garcia, Wozny, Lyberatos, & Kravaris, 2012; Saccomani, 2013). On the other hand, to deal with practical identifiability the first step is to improve the experiments so as to increase the information content of the data (A. Raue et al., 2011). Still, when technical limitations hamper informative experiments, the PE problem can be regularized to reduce the inflation of the uncertainty in the parameters by adding *a priori* knowledge. This a priori knowledge can be in the form of expert knowledge, experimental data from other sources or literature data. With the emergence of online repositories for quantitative biological data, such as BioNumbers (Milo, Jorgensen, Moran, Weber, & Springer, 2010), BRENDA (Schomburg et al.,

2002) and Elsevier BioBase, it is becoming increasingly easier to find quality biological data on reaction kinetics, yield coefficients, metabolite concentrations, etc. with sound references, which can be used as a priori information for regularization.

(i) *Tikhonov Regularization and Monte Carlo Analysis*: This is a more formal way of getting around ill-conditioned parameter estimation problems. In the commonest form of this method, the ill-conditioned parameter estimation problem is made *regular* by re-defining the objective function (Equation 3) with the addition of a penalty term, $\Gamma(\theta)$ in quadratic form (Johansen, 1997).

$$\Gamma(\theta) = (\theta^* - \theta)^2 \ . \qquad (10).$$

In equation 10, θ^* represents the prior information (literature values of parameters) whereas θ is the vector of current parameter values. The penalization term is weighted by a sufficiently large λ so that the objective function $\Phi(U, \theta)$ can always be inverted.

$$\Phi(U,\theta) \coloneqq \frac{1}{2}(Y(U,\theta) - Y^m)^T \left(C_y\right)^{-1} (Y(U,\theta) - Y^m) \ + \ \lambda\Gamma(\theta) \qquad (11)$$

This penalty term, which is a non-singular symmetric matrix and has a minimum at the initial parameter values, contains prior information (e.g. literature values) of the parameters (Bard, 1974). The contribution of the prior information is inevitably associated with bias so that the influence of the regularization should be kept as small as possible. The weighting factor (λ) in the penalty term determines whether more emphasis is placed on the data or on the prior knowledge of the system during PE. However, the choice of a weighting factor (tuning the regularization) to get the proper balance between the data and the prior information is not a trivial task, and is beyond the scope of the current contribution which focuses the L-curve method (Hansen, 1992). The interested reader is referred to the work of Gábor and Banga (2015) (Gábor & Banga, 2015) for a comprehensive treatment of different iterative methods for tuning the weighting factor.

In the current contribution, the Tikhonov technique is combined with a Monte Carlo (MC) procedure to efficiently regularize ill-conditioned parameter estimation problems. The MC method is a more efficient way to accurately quantify the non-linear correlations inherent a model, which are useful in estimating the output uncertainty of the model. In the method, a total of *n in-silico* datasets are generated by random

sampling with replacement (Bootstrap) from the measurement errors (Gürkan Sin & Gernaey, 2016). The bootstrap sampling technique works well without requiring the assumption of normal distribution of the measurement errors (Campolongo, 1997). Each set of the *in-silico* data is used to carry out *regularized* PE, resulting in a matrix of parameter estimates θ_{mc} (n-*by*-N_θ), from which the mean values of the parameters are calculated. At the mean values ($\bar{\theta}_{mc}$), both the sensitivity matrix and its rank (r_ϵ) are re-calculated. If the problem is still ill-conditioned ($r_\epsilon < N_\theta$), the MC parameter estimation is repeated with a fine-tuned regularization (adjusting λ); otherwise the model calibration is completed and the raw MC results (θ_{mc}) can be used for further output uncertainty analysis.

(ii) *Subset Selection (*SsS*) with Empirical Knowledge:* As opposed to black box models, mechanistic ones entail empirical knowledge, although this knowledge is sometimes difficult to completely define in mathematical equations. Furthermore, the experimental conditions or the observations might be restricted due to practical constraints impeding the estimation of certain parameters that are known to be relevant. Therefore, regularizing the problem by *assigning fixed values to non-identifiable parameters* based on *a priori* information is a very efficient and straight forward method (when properly performed), although the bias caused by *a priori* information and subjective decisions should be kept minimal. In application, the non-identifiable parameters are taken out of the PE problem using subset selection techniques and given fixed values. It should be noted that the subset selection process considers both linear independence of the parameters and the sensitivity of model outputs to each parameter. The parameters in the identifiable set simultaneously satisfy both conditions for sensitivity and linear independence, set by the thresholds κ_{max} and $\gamma_{\max}$. Therefore, it is possible that model output may be highly sensitive to a parameter in the non-identifiable, but this parameter was rejected in the SsS process due to correlation to one or more of the parameters in the identifiable set. Great care must be taken when fixing numerical values for such sensitive yet non-identifiable parameters (Brun et al., 2002) during subsequent parameter estimation as these parameters have a strong influence on the estimated parameter values. A sound knowledge of the process or the biological system being modelled is very important in setting the non-identifiable parameters to fixed values that make sense both physiologically and in process design. This manual regularization technique can be a powerful tool, for example in macro-kinetic models

of biological systems, where there is a huge wealth of knowledge. An example of implementation of this technique is given in the Supplementary material.

4.5 Propagation of Parameter Uncertainty onto Model Outputs

When computing the outputs of the model with the parameter estimates which have some level of uncertainty, we need to consider all possible outputs within this parameter space. That is, we have to consider that the true parameters lie somewhere inside the confidence region of the estimates. The projection of the uncertainties associated with the parameter estimation ($\bar{\theta}_{mc}$ or $\hat{\theta}$) to the model output space and the associated quantification of the output variance should be the determining factor to assess the usability of the model. This output uncertainty can be analysed through a Monte Carlo procedure (MC-2) where various parameter combinations from a defined space are used to run model simulations.

The first step is to define the parameter space. For models with well-posed PE problems ($r_\epsilon = N_\theta$), the parameter input space is defined by the confidence intervals of the parameters. For such problems, the variances of the parameter estimates, $\sigma^2_{\theta_i}$ (and hence the standard deviations, σ_{θ_i}) are given by the diagonal elements of the covariance matrix $Cov(\hat{\theta})$, which is calculated by a first order linear approximation method,

$$Cov(\theta) \approx s^2\left[\tilde{S} \cdot \tilde{S}^T\right]^{-1} \qquad (12)$$

where s^2 is the unbiased variance of parameter estimates (Gürkan Sin & Gernaey, 2016). The correlation coefficient between any two parameters (θ_l and θ_k) can then be calculated from the covariance matrix as

$$Corr(\theta_l, \theta_k) = \frac{Cov(\theta_l, \theta_k)}{\sqrt{\sigma^2_{\theta_l}\sigma^2_{\theta_k}}} \qquad (13)$$

The confidence intervals [*lb ub*] of the estimated parameters are calculated using Student *t*-distribution at 95% confidence level (Gürkan Sin & Gernaey, 2016). In the second step, *n* independent random samples are generated from [*lb ub*] using the Latin hypercube sampling algorithm. The correlation matrix is then used to transform these

samples into quasi-random, dependent samples; to depict the inherent interdependencies among model parameters resulting from fundamental theories (e.g. Monod kinetics, inhibition, Michaelis-Menten) (Iman & Conover, 1987; Mara, Tarantola, & Annoni, 2015). In the third step, each set of quasi-random parameter samples is used to carry out model simulations. The mean and standard deviation of the resulting model outputs are then calculated as a measure of the level of uncertainty associated with model predictions.

For ill-conditioned problems, the parameter space is already defined in the MC parameter estimations. The matrix of parameter estimates θ_{mc} (n-by-N_θ) contains the real, non-linear correlations of the model parameters as well as the real distribution of the values each parameter can take, with the given data. Thus, for uncertainty analysis, each set of parameter estimates from θ_{mc} is used to run model simulations, from which the mean and standard deviations of the outputs are calculated to quantify the uncertainty of model predictions.

5.0 Case Study: Macro-kinetic Growth Model of *E. coli*

To illustrate the use of the proposed framework, we analyse a mechanistic model of *E. coli* cultivated in a fed-batch process. Although such models have been widely used for both bioprocess development and for elucidation of various physiological concepts in *E. coli*, this is the first time a mechanistic model describing *E. coli* physiology is analysed with such methods to ascertain the reliability of the model predictions.

5.1 *E. coli* Model Description

The model describes general growth of *E. coli* and the use of glucose based on the substrate partitioning concept (Neubauer, Lin, & Mathiszik, 2003; Pham, Larsson, & Enfors, 1998). It describes metabolic pathways of glucose uptake and its subsequent conversion to cellular material in anabolic routes, its use for energy generation or its conversion to acetate through overflow metabolism. The interested reader is referred to the referenced works (Anane, López C, Neubauer, & Cruz Bournazou, 2017; Cruz Bournazou et al., 2017; Nickel, Cruz-Bournazou, Wilms, Neubauer, & Knepper, 2017; Xu, Jahic, & Enfors, 1999) for a detailed description of the model, including its physiological basis. The model comprises 4 state variables describing biomass, glucose,

acetate and dissolved oxygen profiles (i.e. x = [X, S, A and DOT]) in *E. coli* cultivations, with a total of 15 unknown parameters ($N_\theta = 15$, Table 1).

5.2 Fermentation Data

The data used for this case study was obtained from a fed-batch cultivation of *E. coli* W3110 in a 3.7 L BioEngineering® bench top bioreactor. The cultivation temperature was 37 °C whilst the pH was controlled at a set point of 7.0. For the batch phase, 2 L of medium was inoculated with an appropriate volume of the pre-culture to OD_{600} of 0.1 and initial glucose concentration of 5 g/L. The exponential fed-batch phase started with 300 g/L glucose solution to maintain a set specific growth rate (μ_{set}) of 0.22 h^{-1}. After 3 hours, the feed was switched to a constant feed, where the last value of the exponential feed profile was maintained for a period of 10 hours. Intermittent glucose pulses were given during the constant feeding phase to explore the dynamics of glucose uptake and acetate re-assimilation in *E. coli*. The measured state variables during the cultivation were dissolved oxygen (with a polarographic DO probe), biomass, acetate and residual glucose. The interested reader is referred to Anane et al (Anane et al., 2017) for further details on the experimental procedures. For model validation, two other independent cultivations were carried out in the same bioreactor using the same protocol, but with significant differences in the cultivation scheme from the parameter estimation dataset. The cultivation for the first validation data was a simple batch followed by exponential fed-batch, without a constant feeding phase. The second cultivation for model validation comprised a batch phase, followed by two exponential fed-batch phases with different slopes (μ_{set} = 0.32 h^{-1} and 0.15 h^{-1}), which are distinct growth phases of overflow and non-overflow conditions, respectively, for *E. coli*. Additionally, two intermittent glucose pulses were given in the second exponential feeding phase to trigger overflow conditions within the slow growth phase. Since the model was calibrated and analysed at a different μ_{set} value, it was of interest to see how it will perform in predicting conditions of distinct overflow phases (e.g. low μ_{set}), which are prevalent and important in *E. coli* process design and operation.

5.3 Parameter Estimation using Fed-batch Data

The computation results of the state variables that describe growth profiles of *E. coli* in the course of the cultivation together with their corresponding measured values are

shown in Figure 2. The PE problem (Equation 2) was solved using the *lsqnonlin* function in Matlab®, and the optimal solution for ($\hat{\theta}$) is given in Table 1. It is noteworthy that, at the optimal parameter values, the model is able to track rapid changes in experimental conditions, such as the intermittent glucose pulses in the constant feed fed-batch phase (Figure 2). Without further analysis, the model seems to properly fit the data at the estimated parameter values. But, as will be seen in the identifiability analysis, this result is inside a large region of possible outputs, and is therefore not unique.

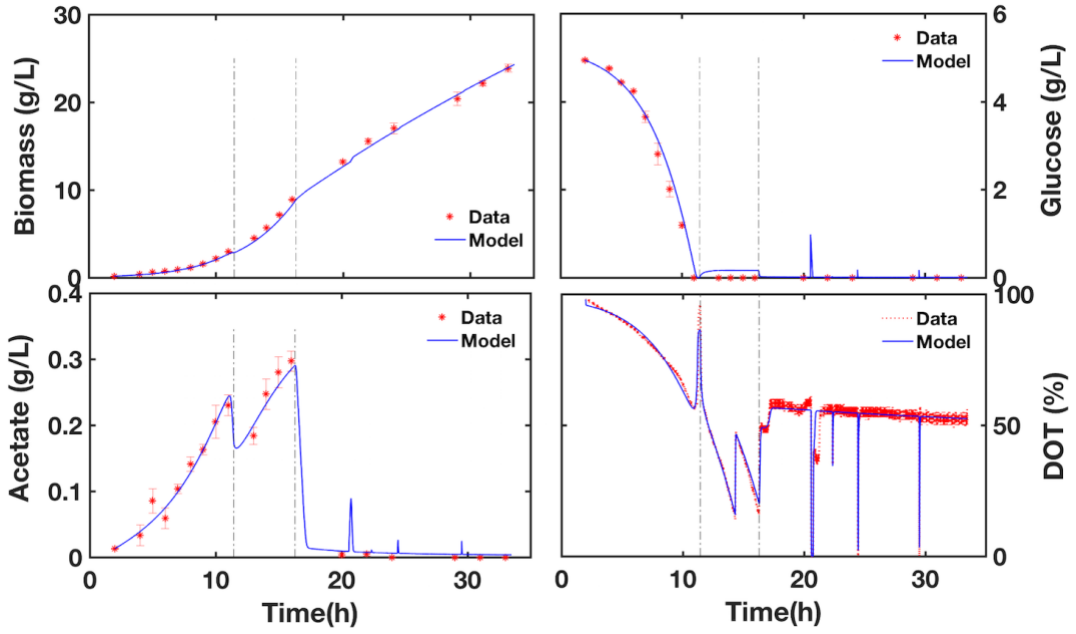


Figure 2 Model fitting of fed-batch fermentation of wild type *E. coli* W3110. Dashed lines signify the beginning and end of exponential feed fed-batch phase.

5.4 Ill-conditioning Diagnosis

A singular value decomposition (Equation 7) was computed at the parameter estimates ($\hat{\theta}$), using the normalized sensitivity matrix of the model outputs with respect to the parameters. The thresholds of 15 and 1000 for the maximum collinearity index ($\gamma_{\max}$) and condition number (κ_{max}), respectively, were selected as defined in Equation 9. A plot of the singular value spectrum, together with the thresholds is shown in Figure 3. Using the thresholds, a numerical rank $r_\epsilon = 10$ was determined for the sensitivity matrix at $\hat{\theta}$, which means only 10 out of the 15 columns of $\tilde{S}$ were linearly independent and sensitive enough to satisfy the conditions of the thresholds.

Table 1 Results of parameter estimation for wild type strain of *E. coli* W3110 (with Tikhonov regularization). Also shown is the quantification of the uncertainty associated with the parameter estimators. LB—lower bound, UB—upper bounds, CI—95% confidence interval.

Parameter [+]	Units	Initial guess (literature)*	Estimate	PE uncertainty quantification		
				% σ_θ	LB-CI	UB-CI
K_{ap}	g g^{-1} h^{-1}	0.10	0.5031	15.2	0.3539	0.6565
K_{sa}	g L^{-1}	0.05	0.0131	22.0	0.0076	0.0192
K_o	g L^{-1}	10.0	0.0001	0.0	0.0001	0.0001
K_s	g L^{-1}	0.05	0.0370	8.9	0.0305	0.0435
K_{ia}	g L^{-1}	5.00	1.2293	9.6	1.0062	1.4737
K_{is}	g L^{-1}	10.0	2.1231	27.3	0.9788	3.2673
p_{Amax}	g g^{-1} h^{-1}	0.17	0.2268	6.5	0.1977	0.2558
q_{Amax}	g g^{-1} h^{-1}	0.15	0.1148	6.1	0.1009	0.1287
q_m	g g^{-1} h^{-1}	0.04	0.0129	7.0	0.0111	0.0147
q_{Smax}	g g^{-1} h^{-1}	1.37	0.6321	0.3	0.6320	0.6392
Y_{as}	g g^{-1}	0.80	0.9097	4.5	0.8283	0.9911
Y_{oa}	g g^{-1}	1.06	0.5440	9.5	0.4425	0.6455
Y_{xa}	g g^{-1}	0.70	0.5718	9.9	0.4604	0.6833
Y_{em}	g g^{-1}	0.50	0.5333	2.4	0.5085	0.5580
Y_{os}	g g^{-1}	1.06	1.5620	5.4	1.3941	1.7298
Y_{xsof}	g g^{-1}	0.15	0.2268	12.0	0.1730	0.2807

[+] Parameter descriptions in nomenclature. * References: (Lin, Mathiszik, Xu, Enfors, & Neubauer, 2001; Martínez-Gómez et al., 2012; Neubauer et al., 2003; Xu et al., 1999)

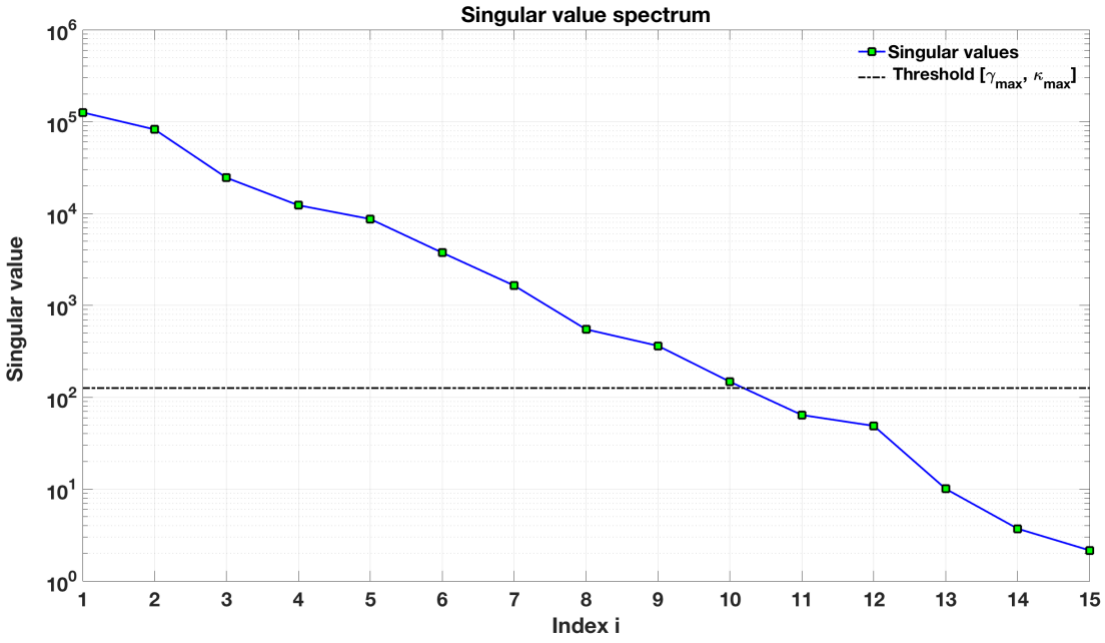


Figure 3 Singular value spectrum of the sensitivity matrix showing the ill-conditioned parameter estimation problem, based on the thresholds for the numerical rank r_ϵ of $\tilde{S}$. The number of singular values above the threshold line is the numerical rank, in this case $r_\epsilon = 10$.

5.5 Regularization and Monte Carlo Analysis

Since the PE problem is ill-posed at $\hat{\theta}$ according to the SVD ($r_\epsilon < N_\theta$), we have to regularize the problem by introducing some *a priori* information on the parameters. Thus, literature values of the parameters (initial guess) were incorporated into the objective function as given in Equation 10, to regularize subsequent parameter estimation runs. A regularization parameter (λ) of 311.6 was determined as the optimal value from a range of 10^6 to 10^{-4}, using the L-curve method (see Appendix for details of tuning the regularization). A total of 500 *in-silico,* Monte Carlo (MC) datasets, Y^{mc} were then randomly generated from the measurement errors (standard deviations of triplicate measurements for biomass, acetate and glucose) using bootstrap sampling with replacement. For high precision measurements such as DOT, the standard deviation was assumed to be 5% of the measurements. The MC datasets were then used to run 500 *regularized* parameter estimations to yield the matrix θ_{mc}. The mean and standard deviations of the MC estimates are given in Table 1, whereas a matrix plot of θ_{mc} is shown in Figure 4, with the distribution of estimates for each parameter on the

diagonal subplots. An extended view of the probability distribution of the parameter estimates is shown in Figure 4B, where the actual distribution of the estimates compared with the assumed normal distribution are also shown.

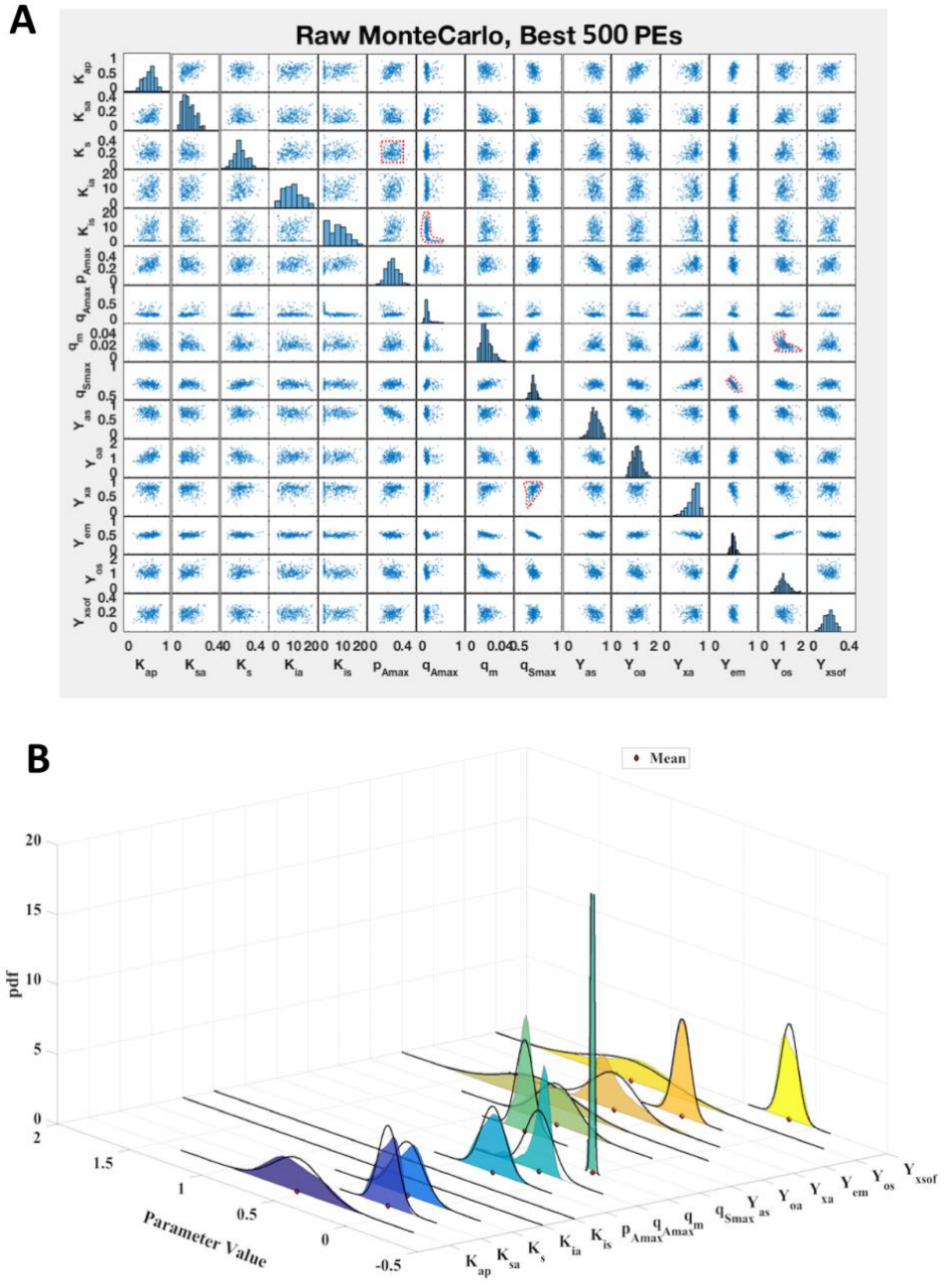


Figure 4 (A) Matrix plot of the non-regularized parameter estimation results showing a few of the inherent non-linearly correlated (q_{Amax} and K_{is}; Y_{os} and q_m), linearly correlated (q_{Smax} and Y_{em}) and non-correlated (p_{Amax} and K_s) parameters of the model. (B) Probability distribution of the Monte Carlo parameter estimates showing the real distributions (coloured inlay) and the assumed normal distribution which was used for subsequent output uncertainty analysis.

The more narrow peaks (q_{Smax}, q_m, Y_{xsof}) indicate parameters for which most of the estimates converged to a single value, implying that these are highly identifiable, whereas parameters with broad peaks show general non-identifiability with the given data. The MC datasets represent different experimental scenarios within the region of the current data and the results show both the inherent non-linear and linear correlations among the parameters. A careful observation of the samples reveals some important physiological perspectives in the parameter correlation patterns. For instance, the maximum specific acetate re-assimilation rate (q_{Amax}) is non-linearly correlated with the acetate inhibition constant K_{is} (circled with red dotted lines in Figure 4) whilst the maximum substrate uptake rate (q_{Smax}) and the biomass yield coefficient (Y_{em}) are negatively correlated in a linear manner. This is because at higher q_{Smax} values, a greater part of the carbon source is directed to overflow metabolism with by-products that eventually inhibit the cell growth (Anane, van Rensburg, & Görgens, 2013), thereby reducing the biomass yield.

5.6 Output Uncertainty Analysis

Each of the parameter estimates in θ_{mc} was used to carry out model simulation, as outlined in Figure 1. The purpose of the regularization is to reduce the variance of the parameter estimates, to improve the accuracy of model predictions for further applications. The major model outputs (biomass, glucose, acetate and dissolved oxygen tension) from the output uncertainty analysis for regularized parameter estimations are plotted in Figure 5B. For comparison, a MC parameter estimations was also carried out (using Y^{mc}) without regularization, and the results were used for output uncertainty analysis (Figure 5A). For the case without regularization, the maximum coefficient of variation ($\sigma_Y/\bar{Y} \times 100\%$) for the predicted model outputs were: 74.4 % for biomass (X), occurring at 6.6 h within the batch phase, 143.5% for glucose (S) occurring at the point of the first glucose pulse and 85.1% for DOT occurring immediately after the first glucose pulse. The prediction error for acetate increased monotonically along the cultivation profile, reaching a maximum of 748.5% at the end of the cultivation. With the regularized PE by the Tikhonov method, the maximum coefficient of variation for the model outputs was: 3.20 % for biomass (X), 1.06 % for glucose (S), 4.75 % for acetate and 4.62 % for DOT, which is a significant improvement in the prediction quality compared to the base case without regularization.

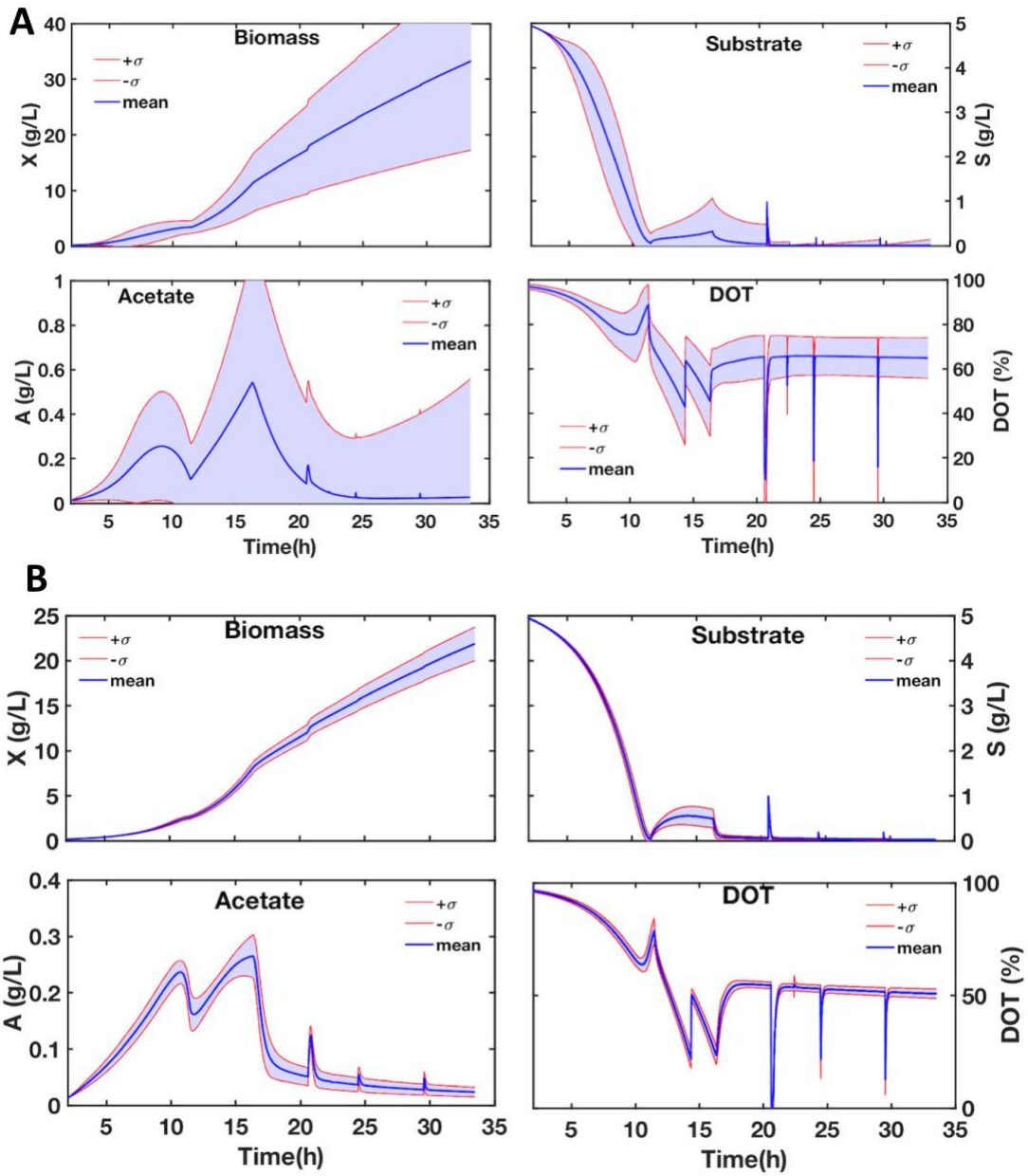


Figure 5 Output uncertainty analysis with (A) non-regularized and (B) regularized Monte Carlo parameter estimates (500 samples). Outer envelops signify one standard deviation from the mean (middle line).

For comparison, the results of the output uncertainty using regularization by subset selection are given in Table A.1 and plotted in Figure A.1 (Supplementary Material). Both regularization methods result in significant improvements in the predictability of model outputs, with low uncertainty, thereby improving the reliability of the model.

5.7 Model Validation Results

Using the mean parameter values of the regularized parameter estimation ($\bar{\theta}_{mc}$), the model was used to predict the outcome of the two validation experiments, using the experimental conditions of each dataset. The predictions were then compared to the actual experimental data, as shown in Figure 6.

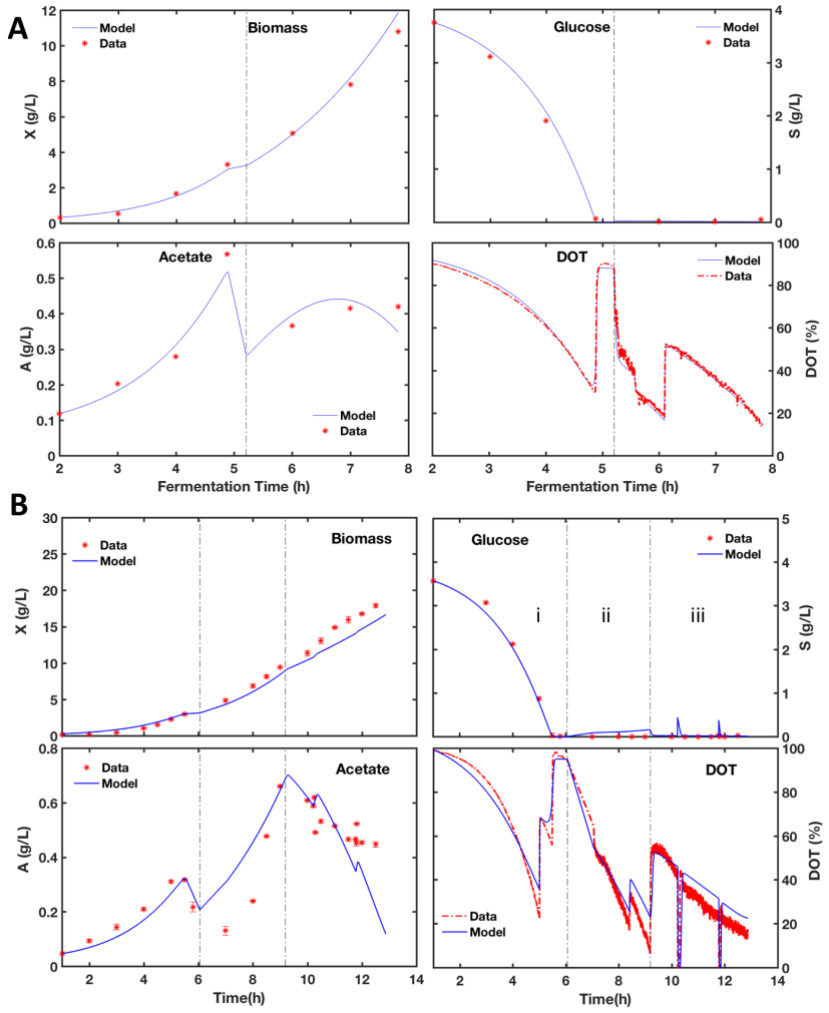


Figure 6 Validation of mechanistic model of *E. coli* after regularized parameter estimation using two independent cultivations, A and B. In B, i—batch phase, ii—exponential feed fed-batch at μ_{set} = 0.32 h^{-1}, iii—exponential feed fed-batch at μ_{set} = 0.15 h^{-1} with intermittent glucose pulses.

The profiles of the relatively simple experiment of the first validation run were accurately predicted by the model (Figure 6A) using $\bar{\theta}_{mc}$ (except for K_La values which were calculated externally from the operating conditions of the bioreactor). As stated earlier, the second validation cultivation was more complex, but the experimental profiles were predicted to acceptable accuracy by the model. In both validation datasets, the least accurately predicted output was the acetate values, where the model seemed

to show higher acetate re-assimilation rates than the actual experimental data towards the end of each cultivation. However, the overall trends in acetate profile were accurately predicted (Figure 6B), showing acetate accumulation at higher specific growth rates (μ_{set} = 0.32 h^{-1}) and its re-assimilation at lower specific growth rates (μ_{set} = 0.15 h^{-1}).

6.0 Discussion

The framework presented in this contribution deals with the identifiability analysis of non-linear growth models in a relatively simple way by calculating the sensitivity matrix and deducing non-identifiability from the numerical properties of this matrix. This is followed by a straightforward, knowledge based regularization to handle the ill-posed problem, which significantly improves the reliability of model predictions as shown in the validation results of the case study. The foregoing analysis is used to improve the prediction quality for experimental scenarios which lie outside the regime of both the calibration and validation data (such as scale-up calculations from pilot scale data). As prediction outside the experimental region is a significant benefit of mechanistic models, it is important that the reliability of such predictions is established, as demonstrated in the current contribution. For instance, considering the subsequent output variance (Figure 5A) associated with the parameter estimation in Figure 2 in the case study, it is apparent that the mere fitting of a model to experimental data (and even validation with similar experimental data) is not enough to guarantee the accuracy and reliability of future predictions. Unfortunately, as pointed out by Sin and co-workers (G Sin et al., 2009) only a few mechanistic models are reported in the literature with reliability analysis. In more intricate applications, such as in using models to elucidate metabolic activity (Chowdhury, Khodayari, & Maranas, 2015), to derive further understanding from molecular level interactions (Stelling, 2004) and in soft sensor design, model reliability is of key importance. It is therefore imperative that the parameters of such models are estimated to very narrow confidence intervals, so as not to produce misleading biological behaviour in their predictions. Therefore, ill-conditioned parameter estimation problems should be avoided in the calibration of such models by thoroughly analysing their identifiability and applying the proper regularization technique.

Finally, we hope that the simplified treatments given here will appeal to biotechnologists and bioprocess engineers and hence promote a more thorough handling of dynamical models, even for non-experienced modellers. This will extend the use of mechanistic models and accelerate the digital revolution in biotechnology.

ACKNOWLEDGEMENTS

We are grateful for financial support from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie actions grant agreement No. 643056 (Biorapid).

Conflict of Interest

The authors declare no financial or commercial conflict of interest.

Supplementary Data

The Matlab® code of the model analysis framework and the fermentation data used in the case study can be found at https://gitlab.tubit.tu-berlin.de/nicolas.cruz/E_coli_fed-batch/tree/master/Anane_2018b_Biotech_Bioeng

References

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., & Jirstrand, M. (2014). Kinetic models in industrial biotechnology – Improving cell factory performance. *Metabolic Engineering*, *24*, 38–60. http://doi.org/10.1016/j.ymben.2014.03.007

Anane, E., López C, D. C., Neubauer, P., & Cruz Bournazou, M. N. (2017). Modelling overflow metabolism in Escherichia coli by acetate cycling. *Biochemical Engineering Journal*, *125*, 23–30. http://doi.org/10.1016/j.bej.2017.05.013

Anane, E., van Rensburg, E., & Görgens, J. F. (2013). Optimisation and scale-up of α-glucuronidase production by recombinant Saccharomyces cerevisiae in aerobic fed-batch culture with constant growth rate. *Biochemical Engineering Journal*, *81*, 1–7. http://doi.org/10.1016/j.bej.2013.09.012

Andrews, J. F. (1968). A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates. *Biotechnology and Bioengineering*, *10*(6), 707–723. http://doi.org/10.1002/bit.260100602

Bailey, J. E. (1998). Mathematical modeling and analysis in biochemical engineering: Past accomplishments and future opportunities. *Biotechnology Progress*, *14*(1), 8–20. http://doi.org/10.1021/bp9701269

Balsa-Canto, E., Henriques, D., Gábor, A., & Banga, J. R. (2016). AMIGO2, a toolbox for dynamic modeling, optimization and control in systems biology. *Bioinformatics*, *32*(21), 3357–3359. http://doi.org/10.1093/bioinformatics/btw411

Bard, Y. (1974). *Non-linear parameter estimation*. New York: Academic Press.

Batstone, D. J., Keller, J., Angelidaki, I., Kalyuzhnyi, S. V., Pavlostathis, S. G., Rozzi, A., … Vavilin, V. A. (2002). The IWA Anaerobic Digestion Model No 1 (ADM1). *Water Science and Technology*, *45*(10), 65–73. http://doi.org/10.2166/wst.2008.678

Bellman, R., & Astrom, K. J. (1970). On structural identifiability. *Mathematical Biosciences*, *7*(3–4), 329–339. http://doi.org/10.1016/0025-5564(70)90132-X

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/0471725153

Brockmann, D., Rosenwinkel, K. H., & Morgenroth, E. (2008). Practical identifiability of biokinetic parameters of a model describing two-step nitrification in biofilms. *Biotechnology and Bioengineering*, *101*(3), 497–514. http://doi.org/10.1002/bit.21932

Brun, R., Martin, K., Siegrist, H., Gujer, W., & Reichert, P. (2002). Practical identifiability of ASM2d parameters — systematic selection and tuning of parameter subsets. *Water Research*, *36*, 4113–4127.

Campolongo, F. (1997). Sensitivity analysis of an environmental model: an application of different analysis methods. *Reliability Engineering & System Safety*, *57*(1), 49–69. http://doi.org/10.1016/S0951-8320(97)00021-5

Chan, T. F., & Hansen, P. C. (1992). Some Applications of the Rank Revealing QR Factorization. *SIAM Journal on Scientific and Statistical Computing*, *13*(3), 727–741. http://doi.org/10.1137/0913043

Chis, O. T., Banga, J. R., & Balsa-Canto, E. (2011). Structural identifiability of systems biology models: A critical comparison of methods. *PLoS ONE*, *6*(11). http://doi.org/10.1371/journal.pone.0027755

Chowdhury, A., Khodayari, A., & Maranas, C. D. (2015). Improving prediction fidelity of cellular metabolism with kinetic descriptions. *Current Opinion in Biotechnology*, *36*, 57–64. http://doi.org/10.1016/j.copbio.2015.08.011

Cobelli, C., & DiStefano, J. J. (1980). Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *The American Journal of Physiology*, *239*(1), R7–R24.

Cruz Bournazou, M. N., Arellano-Garcia, H., Wozny, G., Lyberatos, G., & Kravaris, C. (2012). ASM3 extended for two-step nitrification-denitrification: a model reduction for sequencing batch reactors. *Journal of Chemical Technology & Biotechnology*, *87*(7), 887–896. http://doi.org/10.1002/jctb.3694

Cruz Bournazou, M. N., Barz, T., Nickel, D. B., Lopez Cárdenas, D. C., Glauche, F., Knepper, A., & Neubauer, P. (2017). Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities. *Biotechnology and Bioengineering*, *114*(3), 610–619. http://doi.org/10.1002/bit.26192

De Pauw, D. J. W., & Vanrolleghem, P. a. (2006). Practical aspects of sensitivity function approximation for dynamic models. *Mathematical and Computer Modelling of Dynamical Systems*, *12*(5), 395–414. http://doi.org/10.1080/13873950600723301

Franceschini, G., & Macchietto, S. (2008). Model-based design of experiments for parameter precision : State of the art. *Chemical Engineering Science*, *63*, 4846–4872. http://doi.org/10.1016/j.ces.2007.11.034

Gábor, A., & Banga, J. R. (2015). Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Systems Biology*, *9*(74). http://doi.org/10.1186/s12918-015-0219-2

Golub, G. H., & Van Loan, C. F. (1996). *Matrix Computations* (3rd ed.). Baltimore: The Johns Hopkins University Press. http://doi.org/10.1063/1.3060478

Guisasola, a, Baeza, J. a, Carrera, J., Sin, G., Vanrolleghem, P. a, Lafuente, J., & Matrix, F. I. (2006). The influence of experimental data quality and quantity on parameter estimation accuracy. Andrews Inhibition Model as a Case Study. *Eduacation for Chemical Engineers*, *1*, 139–145. http://doi.org/10.1205/ece06016

Hansen, P. C. (1992). Analysis of Discrete Ill-Posed Problems by Means of the L-

Curve. *SIAM Review*, *34*(4), 561–580. http://doi.org/10.1137/1034115

Iman, R. L., & Conover, W. J. (1987). A Measure of Top – Down Correlation. *Technometrics*, *29*(3), 351–357. http://doi.org/10.1080/00401706.1987.10488244

Johansen, T. A. (1997). On Tikhonov regularization, bias and variance in nonlinear system identification. *Automatica*, *33*(3), 441–446. http://doi.org/10.1016/S0005-1098(96)00168-9

Kitano, H. (2002). Computational systems biology. *Nature*, *420*(6912), 206–210. http://doi.org/10.1038/nature01254

Koutinas, M., Kiparissides, A., Pistikopoulos, E. N., & Mantalaris, A. (2012). Bioprocess Systems Engineering: Transferring Traditional Process Engineering Principles To Industrial Biotechnology. *Computational and Structural Biotechnology Journal*, *3*(4), 1–9. http://doi.org/10.5936/csbj.201210022

Kravaris, C., Hahn, J., & Chu, Y. (2013). Advances and selected recent developments in state and parameter estimation. *Computers and Chemical Engineering*, *51*, 111–123. http://doi.org/10.1016/j.compchemeng.2012.06.001

Lencastre Fernandes, R., Bodla, V. K., Carlquist, M., Heins, A.-L., Eliasson Lantz, A., Sin, G., & Gernaey, K. V. (2012). Applying Mechanistic Models in Bioprocess Development. In *Advances in biochemical engineering/biotechnology* (Vol. 132, pp. 137–166). http://doi.org/10.1007/10_2012_166

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. *Biotechnology and Bioengineering*, *73*(5), 347–357. http://doi.org/10.1002/bit.1068

López C., D. C., Barz, T., Körkel, S., & Wozny, G. (2015). Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. *Computers and Chemical Engineering*, *77*, 24–42. http://doi.org/10.1016/j.compchemeng.2015.03.002

López, D. C., Wozny, G., Flores-Tlacuahuac, A., Vasquez-Medrano, R., & Zavala, V. M. (2016). A Computational Framework for Identifiability and Ill-Conditioning Analysis of Lithium-Ion Battery Models. *Industrial & Engineering Chemistry Research*, *55*(11), 3026–3042. http://doi.org/10.1021/acs.iecr.5b03910

Mara, T. A., Tarantola, S., & Annoni, P. (2015). Non-parametric methods for global sensitivity analysis of model output with dependent inputs. *Environmental Modelling and Software*, *72*, 173–183. http://doi.org/10.1016/j.envsoft.2015.07.010

Marchisio, M. A., & Stelling, J. (2009). Computational design tools for synthetic biology. *Current Opinion in Biotechnology*, *20*(4), 479–485. http://doi.org/10.1016/j.copbio.2009.08.007

Martínez-Gómez, K., Flores, N., Castañeda, H. M., Martínez-Batallar, G., Hernández-Chávez, G., Ramírez, O. T., … Bolivar, F. (2012). New insights into Escherichia

coli metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. *Microbial Cell Factories*, *11*, 46. http://doi.org/10.1186/1475-2859-11-46

Milo, R., Jorgensen, P., Moran, U., Weber, G., & Springer, M. (2010). BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic Acids Research*, *38*(Database issue), D750-3. http://doi.org/10.1093/nar/gkp889

Monod, J. (1949). The Growth of Bacterial Cultures. *Annual Reviews in Microbiology*, *3*(Xl), 371–394.

Muñoz-Tamayo, R., Puillet, L., Daniel, J. B., Sauvant, D., Martin, O., Taghipoor, M., & Blavy, P. (2018). Review: To be or not to be an identifiable model. Is this a relevant question in animal science modelling? *Animal*, *12*(04), 701–712. http://doi.org/10.1017/S1751731117002774

Neubauer, P., Lin, H. Y., & Mathiszik, B. (2003). Metabolic load of recombinant protein production: Inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in Escherichia coli. *Biotechnology and Bioengineering*, *83*(1), 53–64. http://doi.org/10.1002/bit.10645

Nickel, D. B., Cruz-Bournazou, M. N., Wilms, T., Neubauer, P., & Knepper, A. (2017). Online bioprocess data generation, analysis, and optimization for parallel fed-batch fermentations in milliliter scale. *Engineering in Life Sciences*, *17*(11), 1195–1201. http://doi.org/10.1002/elsc.201600035

Pham, H. T. B., Larsson, G., & Enfors, S. O. (1998). Growth and energy metabolism in aerobic fed-batch cultures of Saccharomyces cerevisiae: Simulation and model verification. *Biotechnology and Bioengineering*, *60*(4), 474–482. http://doi.org/10.1002/(SICI)1097-0290(19981120)60:4<474::AID-BIT9>3.0.CO;2-J

Quaiser, T., & Mönnigmann, M. (2009). Systematic identifiability testing for unambiguous mechanistic modeling--application to JAK-STAT, MAP kinase, and NF-kappaB signaling pathway models. *BMC Systems Biology*, *3*(1), 50. http://doi.org/10.1186/1752-0509-3-50

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, *25*(15), 1923–1929. http://doi.org/10.1093/bioinformatics/btp358

Raue, A., Kreutz, C., Maiwald, T., Klingmuller, U., & Timmer, J. (2011). Addressing parameter identifiability by model-based experimentation. *IET Systems Biology*, *5*(2), 120–130. http://doi.org/10.1049/iet-syb.2010.0061

Raue, A., Steiert, B., Schelker, M., Kreutz, C., Maiwald, T., Hass, H., … Timmer, J. (2015). Data2Dynamics: A modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics*, *31*(21), 3558–3560. http://doi.org/10.1093/bioinformatics/btv405

Saccomani, M. P. (2013). Structural vs Practical Identifiability in System Biology. *IWBBIO 2013 Proceedings*, 305–313.

Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*(19), 2507–2517. http://doi.org/10.1093/bioinformatics/btm344

Savageau, M. A. (1970). Biochemical systems analysis. III. Dynamic solutions using a power-law approximation. *Journal of Theoretical Biology*, *26*(2), 215–226. http://doi.org/10.1016/S0022-5193(70)80013-3

Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F., & Schomburg, D. (2002). BRENDA: A resource for enzyme data and metabolic information. *Trends in Biochemical Sciences*, *27*(1), 54–56. http://doi.org/10.1016/S0968-0004(01)02027-8

Sin, G., & Gernaey, K. V. (2016). Data Handling and Parameter Estimation. In M. C. . van Loosdrecht, P. H. Nielsen, C. M. Lopez-Vazquez, & D. Brdjanovic (Eds.), *Experimental Methods in Wastewater Treatment* (1st ed., pp. 201–234). London: IWA Publishing.

Sin, G., Gernaey, K. V, & Lantz, A. E. (2009). Good modelling practice (GMoP) for PAT applications: Propagation of input uncertainty and sensitivity analysis. *Biotechnology Progress*, *25*, 1043–1053. http://doi.org/10.1021/bp.166

Stelling, J. (2004). Mathematical models in microbial systems biology. *Current Opinion in Microbiology*, *7*(5), 513–518. http://doi.org/10.1016/j.mib.2004.08.004

Stephanopoulos, G., Aristidou, A., & Nielsen, J. (1998). *Metabolic Engineering: Principles and Methodologies*. San Diego: Academic Press. http://doi.org/10.1016/B978-012666260-3/50003-0

Vajda, S. V. A., Rabitz, H., Walter, E., & Lecourtier, Y. (1989). Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. *Chemical Engineering Communications*, *83*(1), 191–219. http://doi.org/10.1080/00986448908940662

Varma, A., & Palsson, B. O. (1994). Metabolic flux balancing: Basic concepts, scientific and practical use. *Bio/Technology*, *12*(10), 994–998. http://doi.org/10.1038/nbt1094-994

Villaverde, A. F., Barreiro, A., & Papachristodoulou, A. (2016). Structural Identifiability of Dynamic Systems Biology Models. *PLoS Computational Biology*, *12*(10), 1–22. http://doi.org/10.1371/journal.pcbi.1005153

Villaverde, A. F., Henriques, D., Smallbone, K., Bongard, S., Schmid, J., Cicin-Sain, D., … Banga, J. R. (2015). BioPreDyn-bench: a suite of benchmark problems for dynamic modelling in systems biology. *BMC Systems Biology*, *9*(1), 8. http://doi.org/10.1186/s12918-015-0144-4

Vora, N., & Daoutidis, P. (2001). Nonlinear model reduction of chemical reaction

systems. *AIChE Journal*, *47*(10), 2320–2332. http://doi.org/10.1002/aic.690471016

Weijers, S. R., & Vanrolleghem, P. A. (1997). A procedure for selecting best identifiable parameters in calibrating Activated Sludge Model No. 1 to full-scale plant data. *Water Science and Technology*, *36*(5), 69–79. http://doi.org/10.1016/S0273-1223(97)00463-0

Xu, B., Jahic, M., & Enfors, S.-O. (1999). Modeling of Overflow Metabolism in Batch and Fed-Batch Cultures of Escherichia coli. *Biotechnology Progress*, *15*(1), 81–90. http://doi.org/10.1021/bp9801087

III

Research Article

Received: 9 June 2018 Revised: 9 August 2018 Accepted article published: 13 August 2018 Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI 10.1002/jctb.5798

Modelling concentration gradients in fed-batch cultivations of *E. coli* – towards the flexible design of scale-down experiments

Emmanuel Anane, Annina Sawatzki, Peter Neubauer and Mariano Nicolas Cruz-Bournazou*

Abstract

BACKGROUND: The impact of concentration gradients in large industrial-scale bioreactors on microbial physiology can be studied in scale-down bioreactors. However, scale-down systems pose several challenges in construction, operation and footprint. Therefore, it is challenging to implement them in emerging technologies for bioprocess development, such as in high throughput cultivation platforms. In this study, a mechanistic model of a two-compartment scale-down bioreactor is developed. Simulations from this model are then used as bases for a pulse-based scale-down bioreactor suitable for application in parallel cultivation systems.

RESULTS: As an application, the pulse-based system model was used to study the misincorporation of non-canonical branched-chain amino acids into recombinant pre-proinsulin expressed in *Escherichia coli*, as a response to oscillations in glucose and dissolved oxygen concentrations. The results show significant accumulation of overflow metabolites, up to 18.3% loss in product yield and up to 10-fold accumulation of the non-canonical amino acids norvaline and norleucine in the product in the pulse-based cultivation, compared with a reference cultivation.

CONCLUSIONS: Results indicate that the combination of a pulse-based scale-down approach with mechanistic models is a very suitable method to test strain robustness and physiological constraints at the early stages of bioprocess development.

Supporting information may be found in the online version of this article.

Keywords: *E. coli*; scale-down; glucose gradients; oxygen gradients; modelling; inclusion bodies

NOTATION

K_{ap} = Monod-type saturation constant, intracellular acetate prod. (g g^{-1} h^{-1}).

K_{sa} = Affinity constant, acetate consumption (g L^{-1}).

K_s = Affinity constant, glucose consumption (g L^{-1}).

K_{ia} = Inhibition constant, inhibition of cellular growth by extracellular acetate(g L^{-1}).

K_{is} = Inhibition constant, inhibition of acetate uptake by glucose (g L^{-1}).

p_{Amax} = Maximum specific acetate production rate (g g^{-1} h^{-1}).

q_{Amax} = Maximum specific acetate consumption rate (g g^{-1} h^{-1}).

q_m = Specific maintenance coefficient (g g^{-1} h^{-1}).

q_{Smax} = Maximum specific glucose uptake rate (g g^{-1} h^{-1}).

Y_{as} = Yield of acetate on substrate (g g^{-1}).

Y_{oa} = Specific oxygen used per gram of acetate metabolized (g g^{-1}).

Y_{xa} = Yield of biomass on acetate (g g^{-1}).

Y_{em} = Yield of biomass on glucose, excluding maintenance (g g^{-1}).

Y_{os} = Oxygen used per gram of glucose metabolized per gram biomass (g g^{-1}).

Y_{xsof} = Yield of biomass other products of overflow routes, excluding acetate (g g^{-1}).

C = Carbon content of (s) substrate, (x) biomass.

DOT = dissolved oxygen tension (%). DOT* represents saturating value of DOT in the broth at the given operating conditions.

F = Flow rate (L h^{-1})

μ = specific growth rate (h^{-1})

INTRODUCTION

Inadequate mixing and the associated concentration gradients in large-scale microbial bioprocesses have significant impacts on both cell physiology and recombinant protein quality. In these processes, cells are constantly exposed to oscillating concentrations of substrate, metabolites, dissolved oxygen and carbon dioxide. Hence, the study of performance under conditions similar to

* *Correspondence to: MN Cruz-Bournazou, Chair of Bioprocess Engineering, Institute of Biotechnology, Technische Universität Berlin, Ackerstraße 76, ACK24, 13355 Berlin, Germany. E-mail: mariano.n.cruzbournazou@tu-berlin.de*

Chair of Bioprocess Engineering, Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany

industrial scale process is essential to increase scale-up reliability and speed up process development.[1,2] The effects of oscillating cultivation conditions on microbial physiology and product yields have been studied in the laboratory by applying scale-down techniques, either in the form of scale-down bioreactors[3–6] or as pulse-based methods.[7–10]

A scale-down system is a laboratory-scale bioreactor designed to mimic the environmental conditions in large-scale bioreactors. In multi-compartment scale-down bioreactors, a perfectly mixed stirred tank reactor (STR) is connected to one or more STRs[11] or to one or more plug flow reactors (PFR),[11,12] through which the culture is circulated at a rate equivalent to a specified residence time (Fig. 1(a)). A stress inducing agent (e.g. highly concentrated substrate, base or acid) is injected into one of the sections, which is eventually mixed with the bulk of the culture in the other sections.[13] In a pulsing system, the stress inducer is injected into the bioreactor intermittently, at specified intervals[7] or randomly.[14] These operation mechanisms produce zones similar to feeding and starvation zones in large-scale bioreactors and result in periodic exposure of the culture to varying stresses.[15] Scale-down techniques have been applied for the successful study of the impact of large-scale gradients for most industrially relevant organisms, with significant differences in process behaviour compared with standard small-scale cultivations.[1,12,16,17]

The most recent advances in the development of scale-down concepts include the coupling of computational fluid dynamics (CFD) models of bioreactors with cellular growth kinetics (cellular reaction dynamics, CRD)[18–21] and the mechanistic description of population groups in heterogeneous environments.[22,23] The CFD-CRD models have been used to define specific stress exposure times that are assumed to occur at the larger scale, based on mixing characteristics (CFD simulations) and the dynamics of cellular responses.[18] However, the evaluation of the detailed physiological adaptation to oscillations and their incorporation into CRD models can be an enormous amount of work, as is obvious from the works of Vanrolleghem and Canelas.[24,25] In our opinion, the development of such models and especially their parametrization could benefit from the parallelization of scale-down systems. Although some authors have applied pulse-based feeding profiles in parallel mini-bioreactor systems,[26] mostly due to the difficulty that continuous feeding was technically not possible, real scale-down approaches have not been published in parallel systems, to our knowledge.

The objective of this work was to develop a mechanistic model of a typical pulse-based scale-down bioreactor, suitable for application in high throughput parallel cultivation platforms. The mechanistic model of the pulse-based system was developed from simulations of a two-compartment scale-down bioreactor (2CR), which had been used in many studies before. Thus, the principles of two different scale-down concepts (multi-compartment and pulse-based systems) were combined in a mechanistic model to flexibly design the exposure time of the culture to either high or low glucose and oxygen concentrations. The pulse-based system was used to study the influence of model-derived glucose and dissolved oxygen perturbations on the misincorporation of non-canonical amino acids into pre-proinsulin expressed in *E. coli*. The mechanistic modelling concept has the potential to facilitate the incorporation of scale-down studies into experimental set-ups that would already consider scale-up effects at the early stages of bioprocess development. The big benefit is that cellular reaction models which consider the response to oscillations can be developed and parametrized with much lower effort. Additionally, the run of such experiments in efficient parallel robotic experimental facilities would allow for rapid phenotyping of a large number of candidates under process relevant conditions in short times.[26–29]

MATERIALS AND METHODS

Mechanistic model of two-compartment scale-down bioreactor

The 2CR system that is modelled in this study has been thoroughly described by Junne *et al.*[30] It consists of a 12 L (working volume) stirred tank bioreactor connected to a 1.2 L plug flow reactor. Using the ratio of the volumes of the PFR:STR and the feed and recycling rates, it is estimated that a cell, on average, spends about 5 min in the STR before going through one cycle in the PFR, if the residence time in the PFR is set to 30 s.[30] The mathematical model used to describe the macro-kinetic dynamics of the culture has been presented elsewhere,[31] but a summarized version is presented in the Appendix. This model is a nonlinear ordinary differential equation (ODE) system given as:

$$\dot{x}(t) = f(x(t), u(t); \theta)\, x(t_0) = x_0$$

$$y(t) = \mathrm{A}\, x(t)$$

where $t \in [t_o, t_{end}] \subseteq \mathbb{R}$ is the time, $x(t) \in \mathbb{R}^{n_x}$ are dependent state variables, $u(t) \in \mathbb{R}^{n_u}$ are the time-varying inputs or experimental design variables, $\theta \in \mathbb{R}^{n_p}$ the unknown parameter vector, and initial conditions are given by x_0. The vector $y(t) \in \mathbb{R}^{n_y}$ are the predicted response variables whose elements are defined by the selection matrix $A \in \mathbb{R}^{n_y \times n_x}$. The states considered are the biomass concentration (X, g L^{-1}), glucose as substrate concentration (S, g L^{-1}), extracellular acetate concentration (A, g L^{-1}), and dissolved oxygen tension (DOT in % of saturation).

Some important assumptions have to be made to achieve the required level of simplification in the model of the two-compartment reactor, which have to be considered.

- There is no cell history (the cultivation can be described by a time invariant equation system).
- The metabolic activity under anaerobic conditions can be tracked by changes in parameter values of the model.
- The STR is ideally mixed so that there is no distribution in the residence time of the microorganisms.
- The PFR has no concentration gradients in the radial direction, and convection and diffusion effects can be neglected.

The 2CR model is complicated by the fact that the plug flow reactor never reaches steady state because of the continuous recycling of broth from the STR and the exponential feed injected into the PFR section (Fig. 1(a)). To formulate the mechanistic model of the 2CR system, the transient solution of the partial differential equation (PDE) system was solved by finite differences with discretization in space (dividing the 3.6 m long PFR into 100 elements), and solving the biomass, glucose, acetate and dissolved oxygen (DOT) concentration profiles of the *E. coli* model[31] over each finite element. The cumulative time of the transient solution is equal to the desired residence time in the PFR. By varying the flow rate of the recycling stream in the simulations, different residence times, and therefore different biomass and glucose profiles can be achieved in the PFR (Fig. 2). This can then be used to judge how long the culture has been exposed to the stressing agent (e.g. glucose), as demonstrated in the simulations of a single cell

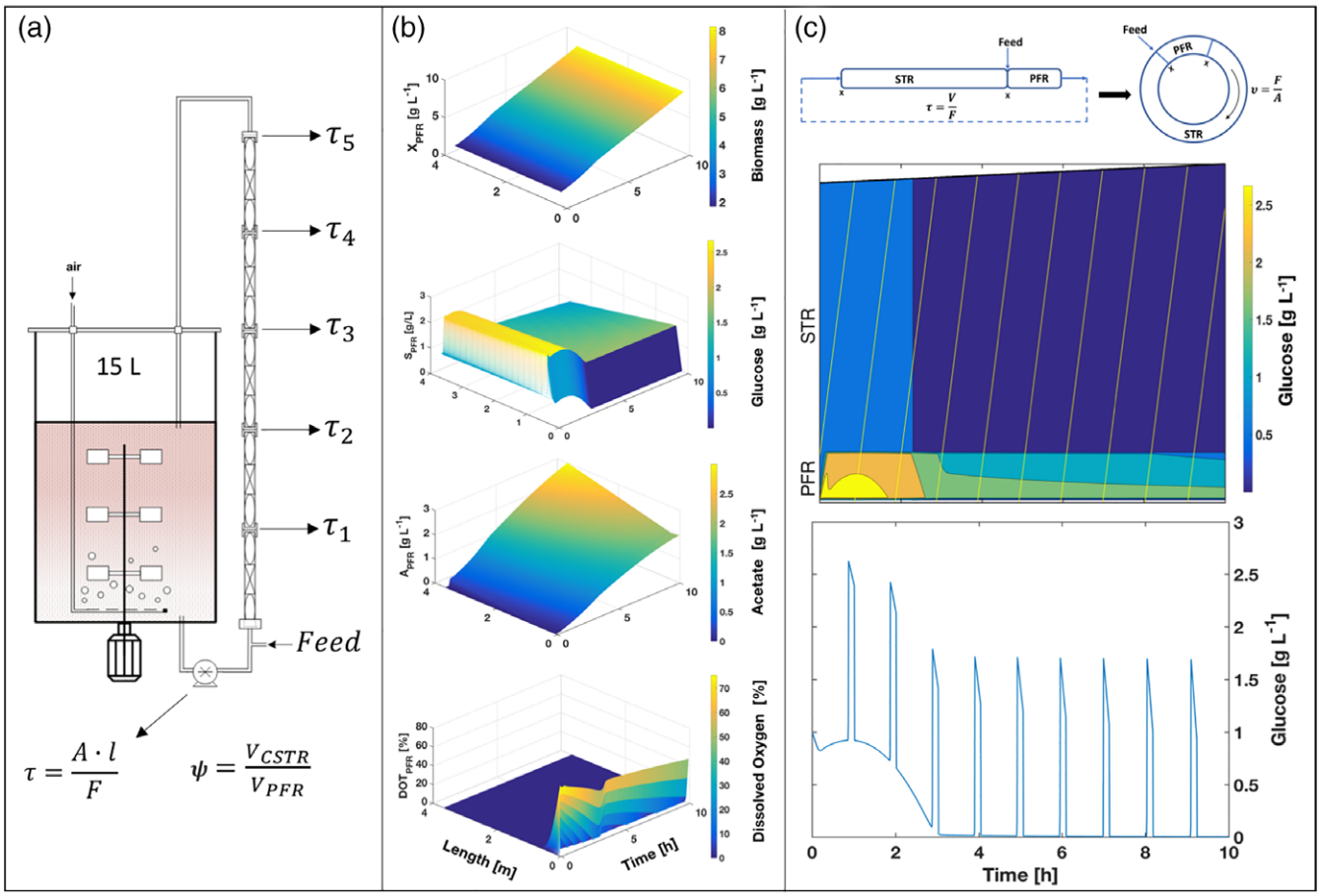


Figure 1. (a) Schematic diagram of two-compartment bioreactor consisting of a stirred tank reactor and a plug-flow reactor. (b) Profiles of dissolved oxygen, biomass, glucose and acetate after running the 2CR for 10 h with feed injection into the PFR. (c) *Top*: cascade configuration of 2CR and toroidal conversion for calculation of cell trajectories; *middle*: gradient profiles of glucose showing the trajectory of a single cell and the glucose concentrations it encounters as it circulates between the STR and PFR; *bottom*: pulse representation of glucose gradients experienced by one cell (glucose in feed = 250 g L^{-1}, biomass in STR = 1.85 g L^{-1} dry weight – Matlab code and full details of simulation are available in the Supplementary material).

trajectory (Fig. 1(c)) for a cell that moves through both the stirred tank and plug flow reactors. The trajectory of an average cell is calculated as

$$\nu(t) = F(t)/A \quad (1)$$

where $\nu(t)$ represents the velocity of the cell (m h^{-1}), $F(t)$ represents the feed flow rate (m^3 h^{-1}), and A the cross-sectional area of the PFR. In order to facilitate the computation of the trajectory including the volume change in the STR due to feeding, the geometry of the 2CR is transformed into a toroid with an ideally mixed flow (Fig. 1(c), top). The volume ratios of STR:PFR and feeding points in the actual 2CR and the toroidal shape are equal. The varying glucose concentrations a traversing cell is exposed to between the STR and PFR are equivalent to intermittent glucose pulses (Fig. 1(c), bottom). Therefore, to transfer these characteristics to a pure glucose pulsing scheme, simulations were done to determine the τ, μ_{set} and feed concentrations at which there would be a glucose carry over from the PFR into the STR. This would ensure the maximum exposure time to the stress in the pulse-based scale-down system.

Recycling rate in the PFR

The recycling rate is determined by the flow rate set on the recycling pump (Fig. 1(a)). The recycling rate of the broth determines three important factors of the 2CR: (i) the duration of the stress, i.e. time in which the culture is exposed to excess glucose environments; (ii) the magnitude of the stress, i.e. the strength of concentration gradients; and (iii) the number of times a particular cell goes through a cycle in the PFR, which is the frequency of exposure to the stress. When the exposure time is of the same or a few orders of magnitude as the characteristic growth time (1/μ), there is a marked influence of the stress on both physiological and metabolic, as well as transcriptional processes in the cell.[8,17,32] Therefore, it is important to determine, in the simulations of the 2CR, how the recycling rate affects both the frequency and magnitude of glucose gradients that the cells are exposed to. These profiles are shown in Fig. 2. At low recycling rates, the cells spend a longer time in the PFR and are therefore exposed to a stronger pulse (higher magnitude, low frequency). At higher recycling rate, the pulse size is smaller, but the cell passes through the vicinity of the stress many times within a given period.

Strain and fermentation conditions

All experiments were performed with *E. coli* W3110 M (*lacI*^{q}) pSW3 (*amp*^{r})[33] expressing a recombinant mini-proinsulin under a *tac*-promoter (inducible with IPTG) (kindly provided by Sanofi-Aventis Deutschland GmbH) in a 3.7 L bioreactor (KLF 2000, Bioengineering AG, Wald, Switzerland). As pre-culture,

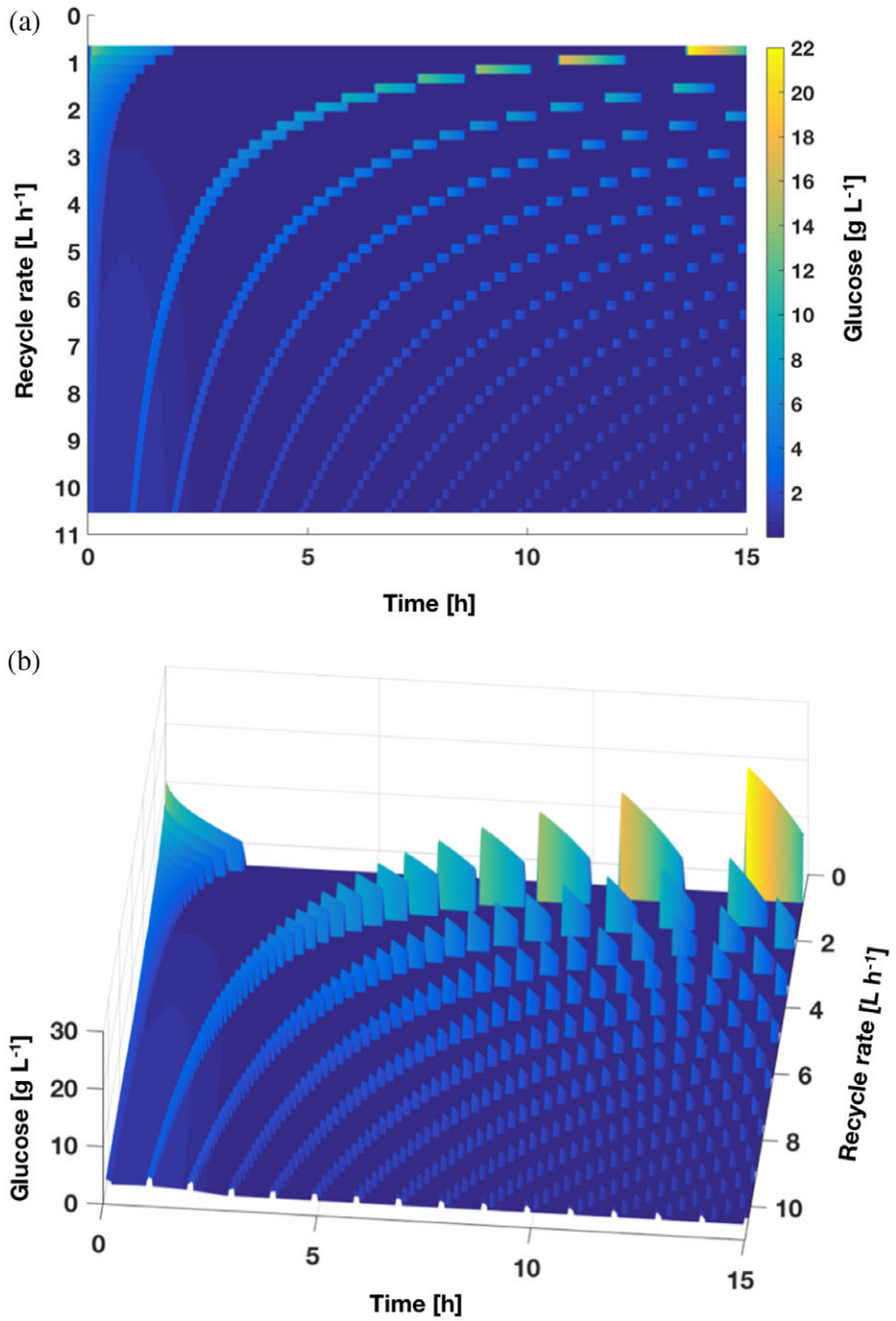


Figure 2. Effect of recycling rate on (a) pulse frequency and (b) pulse size. High recycling rates result in high frequency of exposure, but at lower gradient sizes, and vice versa.

25 mL of bioreactor medium was inoculated with stock cultures of this strain in a 125 mL Erlenmeyer flask and incubated at 37°C, 250 rpm in an orbital shaker (Adolf Kühner AG, Birsfelden, Switzerland). After 12 h, appropriate volumes of the pre-culture were used to inoculate the bioreactor to an OD_{600} of 0.01. The bioreactor medium consisted of mineral salt medium, containing (per L): 2 g Na_2SO_4, 2.468 g $(NH_4)_2SO_4$, 0.5 g NH_4Cl, 14.6 g K_2HPO_4, 3.6 g $NaH_2PO_4 \times 2H_2O$, 1 g $(NH4)_2$-H-citrate and 1 mL antifoam (Antifoam 204, Sigma). Before inoculation, the medium was supplemented with 2 mL L^{-1} trace elements solution, 2 mL L^{-1} $MgSO_4$ solution (1.0 mol L^{-1}) and 1 mL L^{-1} ampicillin (100 mg L^{-1}). The trace element solution comprised (per L): 0.5 g $CaCl_2 \times 2H_2O$, 0.18 g $ZnSO_4 \times 7H_2O$, 0.1 g $MnSO_4 \times H_2O$, 20.1 g Na-EDTA, 16.7 g $FeCl_3 \times 6H_2O$, 0.16 g $CuSO_4 \times 5H_2O$, 0.18 g $CoCl_2 \times 6H_2O$. In all bioreactor cultivations, the initial glucose concentration for the batch phase was 5 g L^{-1}. The temperature was maintained at 37°C and the pH was controlled at 7 by automatic titration with 25% NH_4OH solution.

Reference cultivation

At the end of the batch phase, an exponential feeding mechanism was implemented in the reference cultivation according to Eqn (1),

$$F(t) = \frac{\mu_{set}}{Y_{x/s}S_i}(XV)\,e^{\mu_{set}(t-t_b)} \quad (2)$$

where F represents the feed flow rate ($L\,h^{-1}$), S_i ($g\,L^{-1}$) the concentration of glucose in the feed, $Y_{x/s}$ the biomass yield coefficient ($g\,g^{-1}$), μ_{set} the set-point of the specific growth rate (h^{-1}) and t_b the time at which the batch phase ended. After 3 h of exponential feeding, recombinant protein production was induced by a pulse addition of IPTG to a final concentration of 1 mmol L^{-1}. The feed was then switched to a constant feed, where the constant flow rate was equal to the last flow rate reached in the exponential feeding phase. The feed solution contained 8 mL L^{-1} trace elements solution, mineral salts (same concentration as in batch phase) and 400 g L^{-1} glucose.

A mass balance on gases across the entry and exit points on the STR was used to calculate the specific oxygen consumption rate (q_{O2}) and specific carbon dioxide evolution rate (q_{CO2}) as

$$q_{O_2} = V_G\left(O_2in - \left\{O_2out \times \left(N_2in/N_2out\right)\right\}\right)/X \tag{3}$$

$$q_{CO_2} = \frac{V_G\left(\left\{CO_2out \times \left(\frac{N_2in}{N_2out}\right)\right\} - CO_2in\right)}{X} \tag{4}$$

where V_G represents the gassing rate ($mol\,L^{-1}\,h^{-1}$) and X the biomass concentration ($g\,L^{-1}$). The concentration of all gases in (Eqns (2) and (3)) are in volumetric fractions (%v v^{-1}) and were determined by measuring the composition of the exhaust gas using a BlueSens off-gas analyser (BlueSens Gas Sensor GmbH, Herten, Germany). The biomass concentration for each off-gas measuring point was determined by fitting a biomass profile to values of hourly samples, and interpolating at off-gas measuring points. The respiratory quotient was calculated as $RQ = q_{CO_2}/q_{O_2}$.

Pulse-based cultivation

In order to mimic the oscillations in DOT and glucose concentrations of large industrial-scale bioreactors in the pulse-based system, the calculated exponential feed (Eqn(2)) at the end of the batch phase was divided into discrete pulses of 1 min feeding followed by 9 min of glucose limitation as determined from the 2CR simulations. The amount of glucose fed within the 1 min was equal to that which would have been fed in a continuous exponential feed for 10 min (1 + 9), to maintain the set point of the specific growth rate at 0.25 h^{-1}. During the 1 min feeding period, the agitation rate was manually decreased to 400 rpm (from 800 rpm in the batch phase). In the glucose-limiting phase, it was increased to 1000 rpm (Fig. 3). These cyclic shifts resulted in recurring oxygen-rich and oxygen-deficient conditions, which, together with the glucose pulses, resulted in approximate oscillations in concentration as observed in the 2CR system and at large scale. The feed switching was done manually by turning the feed pump on and off, as required. The kinetics of glucose consumption and acetate production during glucose pulsing were followed by rapidly sampling through a 0.22 μm filter at the point of glucose addition (0 s), 1, 3 and 6 min after each pulse. A total of three pulses were sampled: two in the exponential feeding phase and one after protein induction. The effect of environmental oscillations on overall growth kinetics, cell mass and the accumulation of metabolites and non-canonical branched chain amino acids in the recombinant protein product were assessed by analysing hourly samples. The respiratory quotient (RQ) was calculated for the pulsed-based cultivation as in the reference cultivation. Recombinant protein production was induced by adding IPTG to the same concentration as in the reference cultivation.

Analyses

Cell growth was monitored in both the reference and pulse-based cultivations by measuring the optical density of samples in a UV–vis spectrophotometer (Novaspec III, Amersham Biosciences, Amersham, UK) in triplicate, at a wavelength of 600 nm (OD_{600}). The dry weight was calculated from the OD_{600} values, using a conversion factor of 0.37 g L^{-1} dry weight per OD_{600}, which was previously established for this strain using the same spectrophotometer. Supernatant samples for analysis of residual glucose and acetate were taken from the bioreactor through a 0.22 μm membrane filter at the sampling port and stored at −20°C for further analysis. The concentration of organic acids and glucose were measured with an Agilent 1200 HPLC system (Waldbronn, Germany), equipped with a HyperRezTM XP Carbohydrate H+ column (300 × 7.7 mm, 8 μm) (Fisher Scientific, Schwerte, Germany) and a refractive index detector, using 5 mmol L^{-1} H_2SO_4 as the eluent at a flow rate of 0.5 mL min^{-1}. The column temperature was set to 65°C.

For analysis of the recombinant protein, samples were taken from the bioreactor every hour after protein induction. These samples were normalized to OD_{600} = 15 (in 1 mL) and centrifuged at 16000 × g for 5 min in pre-weighed Eppendorf tubes. The cell pellet was weighed and stored at 4°C for inclusion body (IB) purification. The inclusion body separation was carried out with Bugbuster® protein extraction kit (Novagen, Darmstadt, Germany) as follows. The cell pellets from the normalized samples were resuspended in Bugbuster® reagent and incubated for 20 min at room temperature, taking 5 mL of reagent per gram of wet cell paste. To reduce the viscosity and improve inclusion body extraction, 25 units of benzonase and 1000 units of rLysozyme™ (all from Merck KGaA, Darmstadt, Germany) were added, for each millilitre of Bugbuster reagent used. After the incubation period, the samples were centrifuged and the pellets washed three times in 10× diluted Bugbuster reagent to obtain the purified inclusion bodies. The mass of inclusion bodies per mass of biomass was used as a quantitative measure of the inclusion body fraction. All the purified inclusion bodies obtained from Bugbuster method and 125 μL of internal standard (0.225 mmol L^{-1} α-aminobutyric acid) were hydrolysed in 1 mL of 6 mol L^{-1} HCl at 80°C for 24 h. The hydrolysed samples were then dried in a speed vacuum concentrator (Bachhofer, Reutlingen, Germany), followed by derivatization of the solid residue at 60°C for 60 min. The derivatization reagents were 50 μL N-(tert-butyldimethylsilyl)-N-methyl-trifluoroacetamide (MTBSTFA), 50 μL acetonitrile and 5 μL anhydrous 1-butanol (all from Merck KGaA, Darmstadt, Germany). The derivatized samples were analysed on an Agilent 5975C GC–MS system equipped with a DB-5MS column (30 m × 250 μm, 0.25 μm) and a quadrupole detector, using helium as the carrier gas. All amino acid concentrations were normalized to the mass of inclusion bodies that was hydrolysed.

RESULTS

Model fitting of the pulse-based system

The dynamic model (see Appendix) which was used to derive the glucose pulses was fitted to the pulse feeding experiment to allow a better understanding of growth kinetics under oscillating environmental conditions. The model parameters were estimated by solving the optimization problem

$$\hat{\theta} := \arg\min_{\theta} \mathrm{fl}\,(U, \theta) \tag{5}$$

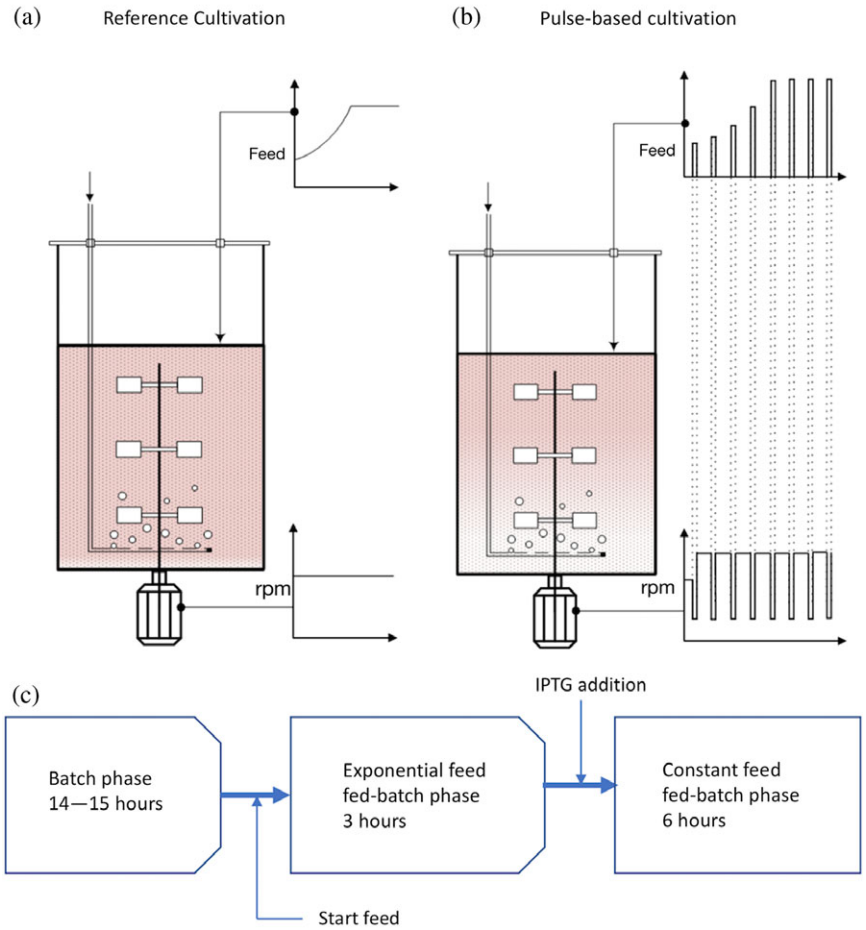


Figure 3. Schematic diagram of (a) reference cultivation showing the constant agitation rate (lower curve) and the smooth glucose feed in two phases: exponential feed followed by constant feeding (upper curve); (b) pulse-based cultivation showing the intermittent supply of glucose feed as pulses in both the exponential and constant feeding phases. The glucose pulses coincide with rpm shifts to achieve the desired gradients; (c) overall cultivation scheme.

where the nonlinear least-square objective function $\Phi(U,\theta)$ was calculated as

$$\Phi(U,\theta) := \frac{1}{2}\left(Y(U,\theta) - Y^m\right)^T \left(Y(U,\theta) - Y^m\right) \quad (6)$$

In Eqn (6), $Y(U,\theta)$ represents the model predictions whereas Y^m represents the experimental data. Both the exponential feeding and constant feeding phases of the pulse-based cultivation and the corresponding model fitting are shown in Fig. 4, with dynamic response of the state variables to oscillating glucose input. The recently described mechanistic model[31] was solved in a pulsing manner, with the addition of residence times of the PFR to dictate the frequency of the pulses. The model fitting resulted in a set of parameter values that sufficiently describe growth under glucose and dissolved oxygen stresses. The model solution fits the experimental data to acceptable accuracy, as given by the relative standard deviations (<20%) of the estimated parameters (Table 1). In further applications, the estimated model parameters can be used to calculate various pulse sizes at varying frequencies, as a means of studying gradient effects under different process conditions.

Growth and metabolic behaviour in pulse-based and reference cultivations

Although the pulse-based experiment and the reference cultivation had similar biomass concentrations at the end of their respective batch phases, they showed different growth patterns during the exponential feeding phase, with the reference cultivation reaching 15% less biomass than the pulse cultivation at the end of this phase. The specific growth rate μ achieved in the exponential feeding phase in the pulse-based cultivation was slightly lower (average of 0.23 h^{-1}) than in the reference cultivation (0.24 h^{-1}), compared with the set point of $\mu_{set} = 0.25\ h^{-1}$. Upon changing from exponential feed to constant feed with protein induction, the specific growth rates declined (Fig. 5(c)) with the decreasing supply of glucose per gram of biomass. However, the pulse-based culture sustained some level of growth (biomass profile, Fig. 5(a)), leading to a more gentle, exponential decline in μ, compared with a sharper decline in μ for the reference cultivation.

In the pulse-based cultivation, all the glucose in a pulse was consumed before the end of the 10 min interval, indicated by a sharp increase in DOT, 60 to 80 s before the next glucose pulse (Fig. 4) during the constant feeding phase. Considering that this pulse contained the same amount of glucose as in 10 min of continuous feed in the reference cultivation, the pulse-based cultivation seemed to have a higher specific glucose uptake rate than in the reference cultivation. As the major product of overflow metabolism in *E. coli*, acetate accumulated to higher levels in the pulse-based system than in the reference cultivation.

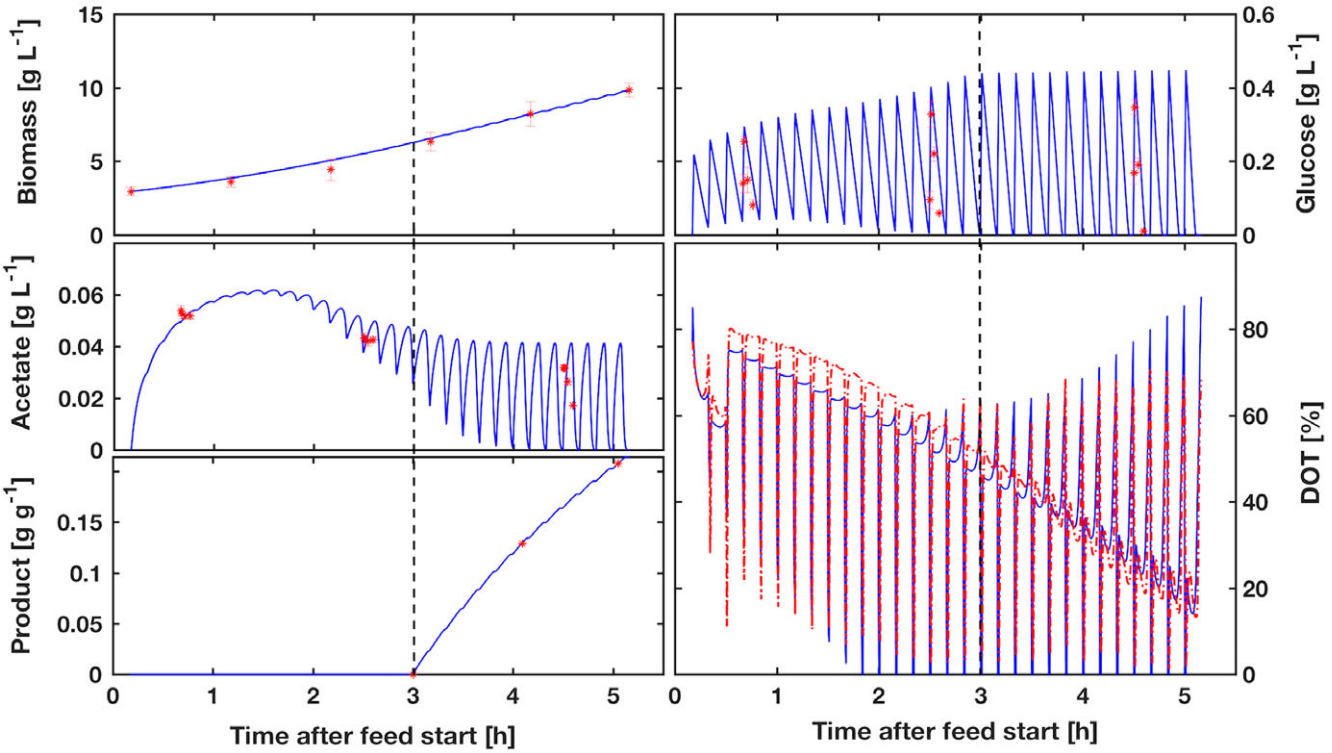


Figure 4. Model fitting of fed-batch phase of the pulse-based scale down cultivation (experimental data *, model ____). The dashed line represents the time of induction (3 h). The Matlab code for the model fitting is available in Supplementary material.

Table 1. Results of model fitting for pulse-based scale-down experiments

			Estimate	Standard Dev		95% CI	
Par	Unit	Initial[a] guess	$\hat{\theta}_1$	σ	%σ	LB	UB
K_{ap}	g g^{-1} h^{-1}	0.438	0.349	0.029	8.30	0.340	0.3570
K_{sa}	g L^{-1}	0.016	0.020	0.003	15.0	0.019	0.0210
K_s	g L^{-1}	0.035	0.019	0.003	15.7	0.0182	0.0198
K_{is}	g L^{-1}	1.111	14.26	1.338	9.40	13.890	14.629
K_{ip}	g g^{-1}	1.562	0.870	0.108	12.4	0.835	0.9051
p_{Amax}	g g^{-1} h^{-1}	0.203	0.326	0.041	12.6	0.3150	0.3274
q_{Amax}	g g^{-1} h^{-1}	0.106	0.250	0.026	10.1	0.2428	0.2572
q_m	g g^{-1} h^{-1}	0.013	0.040	0.001	2.40	0.0397	0.0403
q_{Smax}	g g^{-1} h^{-1}	0.635	0.519	0.025	4.80	0.5121	0.5259
Y_{as}	g g^{-1}	0.827	0.970	0.203	20.9	0.9039	0.9999
Y_{oa}	g g^{-1}	1.100	1.199	0.101	8.42	1.1710	1.2270
Y_{xa}	g g^{-1}	0.611	0.501	0.091	18.2	0.4760	0.5260
Y_{em}	g g^{-1}	0.546	0.481	0.134	20.8	0.4440	0.5187
Y_{os}	g g^{-1}	1.100	1.079	0.073	6.80	1.0530	1.0930
Y_{xsof}	g g^{-1}	0.206	0.351	0.041	11.7	0.3399	0.3620
Y_{px}	g g^{-1}	0.250	0.532	0.039	7.40	0.5211	0.5434

[a] Initial parameter guesses taken from literature [26,31].

After the change to constant feeding, acetate was immediately re-assimilated in the reference cultivation, but acetate re-assimilation in the pulse-based system was delayed up to 1 h after protein induction (Fig. 5(b)).

The RQ remained constant at around 1.2 for the reference cultivation as shown in Fig. 4(d). The RQ for the pulse cultivation shows the periodic availability and depletion of glucose in response to the pulses, with an average value (middle line, Fig. 5(d)) that declined continuously over the course of the cultivation from 1.48 to 1.05 at the end. In the analysis of *E. coli* cultures growing on glucose, the RQ value is unaffected by overflow metabolism due to the fact that the substrate (glucose), the major overflow product (acetate) and the biomass all have the same degree of reduction of approximately 4.[14,34] Therefore, the RQ value is not as informative

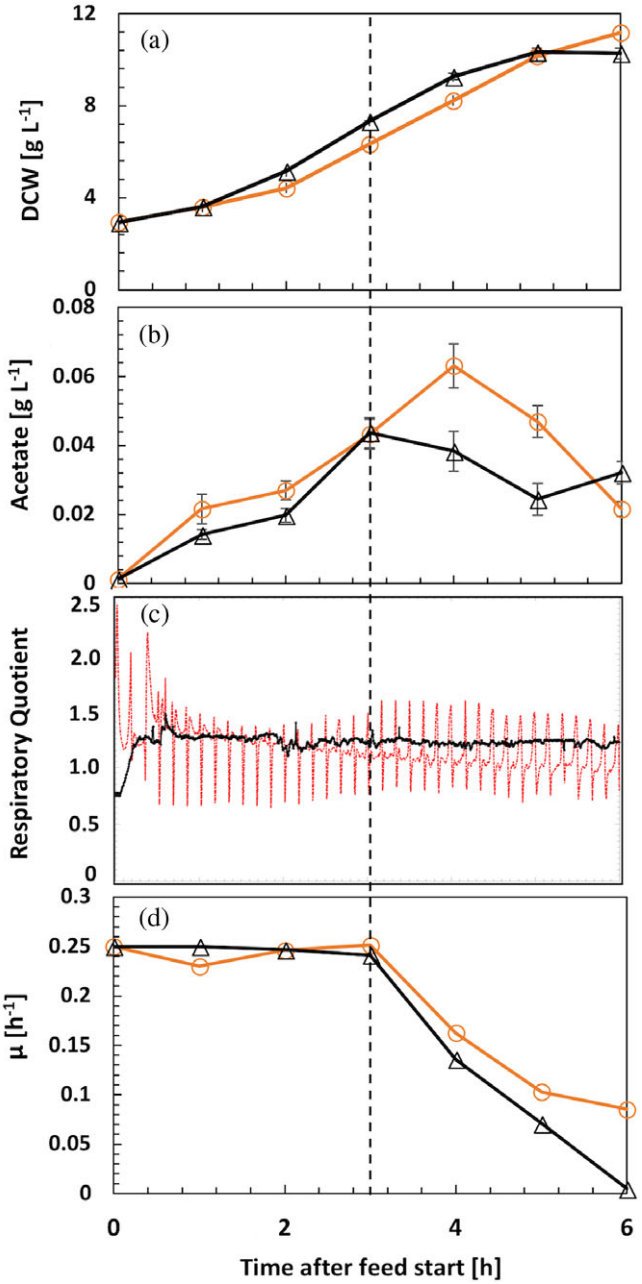


Figure 5. Growth profiles and metabolic activity of *E. coli* during the fed-batch phase of cultivations under pulse-based (O) and reference cultivation (Δ) methods. (a) Biomass concentration, (b) extracellular acetate concentration, (c) respiratory quotient and (d) Specific growth rate during the exponential and constant feed fed-batch phases. The dashed line indicates the point of induction by IPTG. Error bars show the standard deviation.

in *E. coli* as it is in cultivations of *Saccharomyces cerevisiae* where it can indicate important overflow metabolic states due to differences between the degree of reduction of glucose and that of ethanol. However, in the pulse-based cultivation, the intermittent exposure of the culture to anaerobic conditions can lead to the formation of formate and lactate which should have an influence on the cumulative degree of reduction of the metabolites. Therefore, the metabolic behaviour showed a slightly higher average RQ value for the reference cultivation than in the pulse-based cultivation during recombinant protein production.

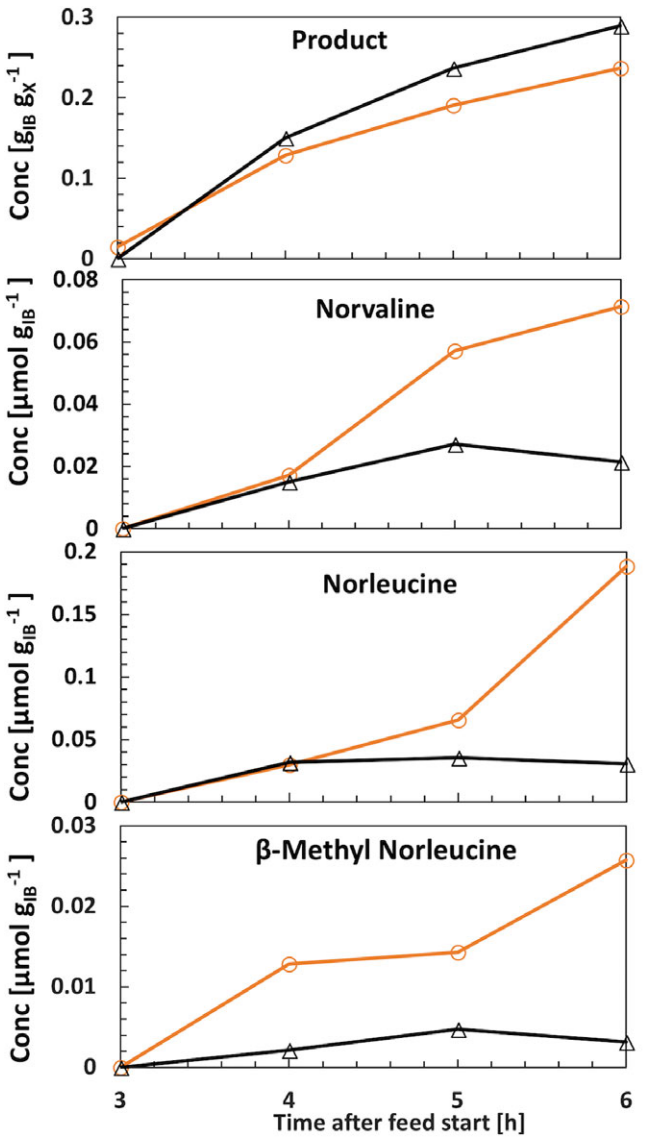


Figure 6. Product formation profile and the concentration of non-canonical amino acids in the purified inclusion bodies from the pulse-based cultivation (O) and the reference cultivation (Δ).

Inclusion body formation and recombinant product quality

Prior to protein induction by IPTG in the pulse-based cultivation, there was already some accumulation of inclusion bodies in the cells (Fig. 6(a)). This seemingly leaky expression, which was not observed in the reference cultivation, can be attributed to the partial de-repression of the *tac* promoter by the intermittent glucose limitation[35] between pulses in the exponential feeding phase. As shown in Fig. 6(a), the pulse-based scale-down cultivation condition led to a lower amount of recombinant pre-proinsulin per gram of biomass than in the reference cultivation. The final product yield in the pulse-based cultivation was 18% lower than in the reference cultivation.

The quality of the recombinant product as used in this text refers to the concentration of non-conventional amino acids (norvaline, norleucine and beta-methyl-norleucine) present in the purified product. These amino acids replace their corresponding canonical forms (leucine, methionine and isoleucine, respectively) in the product under stressful cultivation conditions[36] and

thus reduce the efficacy of therapeutic proteins produced recombinantly in *E. coli*. The results of the scale-down studies show a steady increase in the concentration of non-canonical amino acids in the product for the pulse-based cultivation, which is about 10-fold higher than their concentrations in the reference cultivation.

DISCUSSION

Physiological response of *E. coli* to glucose and oxygen pulses

Scale-down bioreactors provide an effective method to investigate the challenges of bioprocess scale-up. Their ability to mimic the environmental heterogeneity in large-scale bioreactors provides the possibility to look into the dynamics of industrial cultivations, and to study the effects of such dynamics on microbial physiology and process efficiency. The model-based pulsing scheme implemented in the current study as a scale-down methodology resulted in significant differences in both the metabolic behaviour and recombinant protein quality, compared with the reference cultivation. Earlier studies report that *E. coli* cells in oscillating environmental conditions show a higher cell viability in comparison with laboratory-scale cultivations.[5,15] This observation was also confirmed in the current study as seen in the more gentle decline of specific growth rate for the scale-down cultivation (Fig. 5(c)). This higher viability can be attributed to the higher specific uptake capacities under oscillating conditions, as observed in the glucose uptake rates in the current study and also discussed by Lin *et al.*[37] The loss of product yield in heterogeneous conditions is a common occurrence during bioprocess scale-up.[38,39] The results of the scale-down cultivation indicate a significant loss in product yield in the presence of heterogeneous environmental conditions. This lower product accumulation rate (Fig. 5(b)) may be due to the loss of valuable carbon source through overflow metabolism and the associated low ATP generation under stressful cultivation conditions.[10] These factors lead to a lower biomass accumulation rate and consequently, a lower specific product formation rate in large-scale bioreactors, as demonstrated in the scale-down cultivation in the present study.

There was also a significant accumulation of non-canonical amino acids in the purified inclusion bodies in the pulse-based cultivation. According to recent reviews,[36,40] these wrong amino acids can arise from metabolic by-products originating from environmental stresses in *E. coli* cultivation. The results obtained by Soini *et al.* also directly link heterogeneous fermentation conditions to norvaline accumulation in *E. coli* W3110.[41] Thus, the higher accumulation of these non-canonical amino acids in the pulse-based cultivation shows the ability of this cultivation set-up to reproduce environmental stresses that trigger unfavourable responses in *E. coli*, under scale-down conditions.

Model application in scale-down bioreactor systems

A major outcome of the current study is the dynamic description of concentration gradients in a mechanistic framework, that allows the estimation of physiological parameters under oscillating environmental conditions. These parameters, when estimated in high throughput platforms under scale-down conditions can be used for CFD-CRD applications,[21,42] as well as for screening large libraries under real cultivation conditions. The application of glucose pulses for physiological studies in fermentation is not a new concept.[7,14] However, in most of the previous studies the pulses applied were randomly determined and not based on any physiological basis. Here, an exponential feed profile to maintain a certain specific growth rate in the fed-batch phase was calculated, then a mechanistic model (see pseudocode in Algorithm S1) was used to discretize this feed into pulses according to gradient profiles simulated in the 2CR. The duration, frequency and magnitude of concentration gradients in an actual large-scale bioreactor are all dynamic parameters.[15,43] In effect, the exposure time in the scale-down bioreactor should be a flexible parameter that can be changed easily, to suit a specific large-scale bioreactor.[44] In the current contribution, this flexibility of scale-down design is offered by the modelling framework of the pulse-based system. Thus, data for the inclusion of variable exposure times to study response kinetics of specific zones (organism lifelines) in CFD-CRD models can be generated easily in such an experimental set-up. Although the pulse-based feed leads to a synchronized response of the culture to the stress, advanced process analytical technology (PAT) tools such as online *in situ* microscopy[45] could be used to monitor the population heterogeneity.

CONCLUSIONS

We have demonstrated in this study that physiological behaviour of cells under environmental oscillations can be described fully with mechanistic models. The model can then be used to design simple scale-down experimental set-ups, which is a step in simplifying scale-down bioreactor systems for application in parallelization. Such experimental set-ups could be used to study the effects of scale-up stresses on the efficiency of bioprocesses at the early stages of process development. As demonstrated by Cruz *et al.*,[26] model-based automation can be used to achieve faster bioprocess characterization. Therefore, incorporating scale-up effects into such platforms through modelling provides further opportunities to facilitate bioprocess development with scale-up in mind.

Supplementary Data

The Matlab® codes and fermentation data can be found at https://gitlab.tubit.tu-berlin.de/nicolas.cruz/E_coli_fed-batch/tree/master/Anane_2018_JCTB

ACKNOWLEDGEMENTS

We are grateful for financial support from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie actions grant agreement No. 643056 (Biorapid).

CONFLICT OF INTEREST

The authors declare no financial or commercial conflict of interest.

A. Appendix

MACRO-KINETIC MODEL OF *ESCHERICHIA COLI*

The mechanistic model of *E. coli* used in this publication is based on the physiological use of glucose, in a glucose partitioning framework as well as the overflow of glucose to acetate through the acetate cycling concept. The two physiological concepts, as

given by Ying Lin and Neubauer[9] and Lin *et al.*[37] and basic growth concepts such as Monod kinetics and acetate inhibition are used to derive simple algebraic equations that describe intracellular pathways of glucose and oxygen usage. Details of the derivation of the model and its subsequent usage in *E. coli* processes can be found in t he literature.[26,31,37] The algebraic equations that describe these intracellular activities are as follows:

$$q_s = \frac{q_{smax} S}{S + K_s} \cdot e^{-P * K_{ip}} \tag{A.1}$$

$$q_{sox} = \left(q_s - \frac{P_{Amax} q_s}{q_s + K_{ap}} \right) \cdot \frac{DOT}{DOT + K_o} \tag{A.2}$$

$$q_{sof} = q_s - q_{sox} \tag{A.3}$$

$$p_A = q_{sof} Y_{as} \tag{A.4}$$

$$q_{sA} = \frac{q_{Amax}}{1 + \frac{q_s}{K_{is}}} \dot{c} \frac{A}{A + K_{sa}} \tag{A.5}$$

$$q_A = p_A - q_{sA} \tag{A.6}$$

$$\mu = \left(q_{sox} - q_m \right) Y_{em} + q_{sA} Y_{xa} + \left(q_{sof} - p_A \right) Y_{xsof} \tag{A.7}$$

$$q_o = \left(q_{sox} - q_m \right) Y_{os} + q_{sA} Y_{oa} \tag{A.8}$$

$$q_p = \mu Y_{px} \tag{A.9}$$

The algebraic equations are coupled with mass balances for a fed-batch process to yield the full ODE system. The ODE system for the *E. coli* mechanistic model is derived from mass balances on biomass (X), glucose (substrate, S), acetate (A) and dissolved oxygen measured as the percentage saturation at the operating conditions in the bioreactor (DOT).

$$\frac{dX}{dt} = \frac{F}{V} (0 - X) + \mu X \tag{A.10}$$

$$\frac{dS}{dt} = \frac{F}{V} \left(S_i - S \right) - q_s X \tag{A.11}$$

$$\frac{dA}{dt} = \frac{F}{V} (0 - A) + q_{sA} X \tag{A.12}$$

$$\frac{dDOT}{dt} = K_{La} \left(DOT^* - DOT \right) - q_o X H \tag{A.13}$$

$$\frac{dP}{dt} = q_p - \mu P \tag{A.14}$$

The model (Eqns (A.1)–(A.14)) was compiled as a single mathematical function (e_colimodel) and implemented in Matlab® R2016a. The model was integrated with ode15s solver and parameter estimation was done with the *fmincon* optimization routine in Matlab.

CODE FOR FITTING THE PULSE-BASED CULTIVATION

The mechanistic model of *E. coli* was solved in a pulsing manner to fit the data of the pulse-based scale-down cultivation. The pulsing scheme was determined by two time spans: t_{sp} and t_{swof}, which are respectively the times of glucose feed and glucose limitation. For each pulse cycle ($t_{se} = t_{sp} + t_{swof}$), the exponential feed profile was integrated to find the volume of feed added within the time t_{sp}. The details of the calculations are given in the pseudocode in Algorithm A.1.

Supporting Information

Supporting information may be found in the online version of this article.

REFERENCES

1 Neubauer P and Junne S, Scale up and scale down methodologies for bioreactors, in *Bioreactors: Design, Operation and Novel Applications*, ed. by Mandenius C-F. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp. 323–354 (2016).

2 Neubauer P, Cruz N, Glauche F, Junne S, Knepper A and Raven M, Consistent development of bioprocesses from microliter cultures to the industrial scale. *Eng Life Sci* **13**:224–238 (2013).

3 Neubauer P, Häggström L and Enfors SO, Influence of substrate oscillations on acetate formation and growth yield in Escherichia coli glucose limited fed-batch cultivations. *Biotechnol Bioeng* **47**:139–146 (1995).

4 Bylund F, Castan A, Mikkola R, Veide A and Larsson G, Influence of scale-up on the quality of recombinant human growth hormone. *Biotechnol Bioeng* **69**:119–128 (2000).

5 Hewitt CJ, Nebe-von-Caron G, Axelsson B, McFarlane CM and Nienow AW, Studies related to the scale up of high cell density E. coli fed batch fermentations using multiparameter flow cytometry: effect of a changing microenvironment with respect to glucose and dissolved oxygen concentration. *Biotechnol Bioeng* **70**:381–390 (2000).

6 Lara AR, Leal L, Flores N, Gosset G, Bolívar F and Ramírez OT, Transcriptional and metabolic response of recombinant Escherichia coli to spatial dissolved oxygen tension gradients simulated in a scale-down system. *Biotechnol Bioeng* **93**:372–385 (2006).

7 Neubauer P, Åhman M, Törnkvist M, Larsson G and Enfors SO, Response of guanosine tetraphosphate to glucose fluctuations in fed-batch cultivations of Escherichia coli. *J Biotechnol* **43**:195–204 (1995).

8 Lara AR, Taymaz-Nikerel H, Mashego MR, van Gulik WM, Heijnen JJ, Ramírez OT *et al.*, Fast dynamic response of the fermentative metabolism of Escherichia coli to aerobic and anaerobic glucose pulses. *Biotechnol Bioeng* **104**:1153–1161 (2009).

9 Ying Lin H and Neubauer P, Influence of controlled glucose oscillations on a fed-batch process of recombinant *Escherichia coli*. *J Biotechnol* **79**:27–37 (2000).

10 Jaén KE, Sigala JC, Olivares-Hernández R, Niehaus K and Lara AR, Heterogeneous oxygen availability affects the titer and topology but not the fidelity of plasmid DNA produced by *Escherichia coli*. *BMC Biotechnol* **17**:1–12 (2017).

11 Limberg MH, Pooth V, Wiechert W and Oldiges M, Plug flow versus stirred tank reactor flow characteristics in two-compartment scale-down bioreactor: setup-specific influence on the metabolic phenotype and bioprocess performance of *Corynebacterium glutamicum*. *Eng Life Sci* **16**:610–619 (2016).

12 Li J, Jaitzig J, Lu P, Süssmuth RD and Neubauer P, Scale-up bioprocess development for production of the antibiotic valinomycin in *Escherichia coli* based on consistent fed-batch cultivations. *Microb Cell Fact* **14**:1–13 (2015).

13 Lemoine A, Limberg MH, Kästner S, Oldiges M, Neubauer P and Junne S, Performance loss of *Corynebacterium glutamicum* cultivations under scale-down conditions using complex media. *Eng Life Sci* **16**:1–13 (2016).

14 Sunya S, Bideaux C, Molina-Jouve C and Gorret N, Short-term dynamic behavior of *Escherichia coli* in response to successive glucose pulses on glucose-limited chemostat cultures. *J Biotechnol* **164**:531–542 (2013).

15 Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen B *et al.*, Physiological responses to mixing in large scale bioreactors. *J Biotechnol* **85**:175–185 (2001).
16 Brand E, Junne S, Anane E, Cruz-Bournazou MN and Neubauer P, Importance of the cultivation history for the response of *Escherichia coli* to oscillations in scale-down experiments. *Bioprocess Biosyst Eng* **28**:1–9 (2018).
17 Delvigne F, Takors R, Mudde R, van Gulik W and Noorman H, Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond. *J Microbial Biotechnol* **10**:1267–1274 (2017).
18 Haringa C, Tang W, Deshmukh AT, Xia J, Reuss M, Heijnen JJ *et al.*, Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: an analysis of organism lifelines. *Eng Life Sci* **16**:652–663 (2016).
19 Lapin A, Müller D and Reuss M, Dynamic behavior of microbial populations in stirred bioreactors simulated with Euler–Lagrange methods: traveling along the lifelines of single cells. *Ind Eng Chem Res* **43**:4647–4656 (2004).
20 Wang G, Tang W, Xia J, Chu J, Noorman H and van Gulik WM, Integration of microbial kinetics and fluid dynamics toward model-driven scale-up of industrial bioprocesses. *Eng Life Sci* **15**:20–29 (2015).
21 Haringa C, Deshmukh AT, Mudde RF and Noorman HJ, Euler-Lagrange analysis towards representative down-scaling of a 22 m3 aerobic S. cerevisiae fermentation. *Chem Eng Sci* **170**:653–669 (2017).
22 Delvigne F, Lejeune A, Destain J and Thonart P, Modelling of the substrate heterogeneities experienced by a limited microbial population in scale-down and in large-scale bioreactors. *Chem Eng J* **120**:157–167 (2006).
23 Baert J, Delepierre A, Telek S, Fickers P, Toye D, Delamotte A *et al.*, Microbial population heterogeneity versus bioreactor heterogeneity: evaluation of redox sensor green as an exogenous metabolic biosensor. *Eng Life Sci* **16**:643–651 (2016).
24 Canelas AB, Ras C, ten Pierick A, van Gulik WM and Heijnen JJ, An in vivo data-driven framework for classification and quantification of enzyme kinetics and determination of apparent thermodynamic data. *Metab Eng* **13**:294–306 (2011).
25 Vanrolleghem PA, De Jong-Gubbels P, Van Gulik WM, Pronk JT, Van Dijken JP and Heijnen S, Validation of a metabolic network for *Saccharomyces cerevisiae* using mixed substrate studies. *Biotechnol Prog* **12**:434–448 (1996).
26 Cruz Bournazou MN, Barz T, Nickel DB, Lopez Cárdenas DC, Glauche F, Knepper A *et al.*, Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities. *Biotechnol Bioeng* **114**:610–619 (2017).
27 Nickel DB, Cruz-Bournazou NM, Wilms T, Neubauer P and Knepper A, Online bioprocess data generation, analysis and optimization for parallel fed-batch fermentations at mL scale. *Eng Life Sci* **17**:1–7 (2016).
28 Hemmerich J, Noack S, Wiechert W and Oldiges M, Microbioreactor systems for accelerated bioprocess development. *Biotechnol J* **13**:e1700141 (2018).
29 Barz T, Sommer A, Wilms T, Neubauer P and Cruz Bournazou MN, Adaptive optimal operation of a parallel robotic liquid handling station. *IFAC-PapersOnLine* **51**:765–770 (2018).
30 Junne S, Klingner A, Kabisch J, Schweder T and Neubauer P, A two-compartment bioreactor system made of commercial parts for bioprocess scale-down studies: impact of oscillations on *Bacillus subtilis* fed-batch cultivations. *Biotechnol J* **6**:1009–1017 (2011).
31 Anane E, López Cardenas DC, Neubauer P and Cruz Bournazou MN, Modelling overflow metabolism in *Escherichia coli* by acetate cycling. *Biochem Eng J* **125**:23–30 (2017).
32 Gao X, Kong B and Vigil RD, Characteristic time scales of mixing, mass transfer and biomass growth in a Taylor vortex algal photobioreactor. *Bioresour Technol* **198**:283–291 (2015).
33 Dörschug M, Habermann P, Seipke G and Uhlmann E, Mini-proinsulin, its production and use. Germany; EP0347781B1 (1988), pp. 1–10.
34 Jobé AM, Herwig C, Surzyn M, Walker B, Marison I and von Stockar U, Generally applicable fed-batch culture concept based on the detection of metabolic state by on-line balancing. *Biotechnol Bioeng* **82**:627–639 (2003).
35 Wong P, Gladney S and Keasling JD, Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol Prog* **13**:132–143 (1997).
36 Reitz C, Fan Q and Neubauer P, Synthesis of non-canonical branched-chain amino acids in *Escherichia coli* and approaches to avoid their incorporation into recombinant proteins. *Curr Opin Biotechnol* **53**:248–253 (2018).
37 Lin HY, Mathiszik B, Xu B, Enfors SO and Neubauer P, Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of *Escherichia coli*. *Biotechnol Bioeng* **73**:347–357 (2001).
38 Xu B, Jahic M, Blomsten G and Enfors SO, Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*. *Appl Microbiol Biotechnol* **51**:564–571 (1999).
39 Junker BH, Scale-up methodologies for *Escherichia coli* and yeast fermentation processes. *J Biosci Bioeng* **97**:347–364 (2004).
40 Bullwinkle T, Lazazzera B and Ibba M, Quality control and infiltration of translation by amino acids outside of the genetic code. *Annu Rev Genet* **48**:149–166 (2014).
41 Soini J, Falschlehner C, Liedert C, Bernhardt J, Vuoristo J and Neubauer P, Norvaline is accumulated after a down-shift of oxygen in *Escherichia coli* W3110. *Microb Cell Fact* **7**:30 (2008).
42 Morchain J, Gabelle J-C and Cockx A, A coupled population balance model and CFD approach for the simulation of mixing issues in lab-scale and industrial bioreactors. *AIChE J* **60**:27–40 (2014).
43 Larsson G, Törnkvist M, Ståhl Wernersson E, Trägårdh C, Noorman H and Enfors SO, Substrate gradients in bioreactors: origin and consequences. *Bioprocess Eng* **14**:281–289 (1996).
44 Simen JD, Löffler M, Jäger G, Schäferhoff K, Freund A, Matthes J *et al.*, Transcriptional response of Escherichia coli to ammonia and glucose fluctuations. *J Microbial Biotechnol* **10**:858–872 (2017).
45 Marbà-Ardébol A-M, Emmerich J, Muthig M, Neubauer P and Junne S, Real-time monitoring of the budding index in *Saccharomyces cerevisiae* batch cultivations with in situ microscopy. *Microb Cell Fact* **17**:73 (2018).

IV

Scaling down further: model-based study of scale-up effects in mini-bioreactors for accelerated phenotyping

Emmanuel Anane[1], Ángel Corcoles Garcia[2], Benjamin Haby[1], Sebastian Hans[1], Niels Krausch[1], Florian Glauche[1], Sebastian L. Riedel[1], Peter Hauptmann[2], Peter Neubauer[1], M. Nicolas Cruz-Bournazou[#1, 3, 4]

[1] Bioprocess Engineering, Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany

[2]C & BD Biotechnologie, Sanofi-Aventis Deutschland GmbH, Industriepark Höchst, Frankfurt (aM), Germany

[3] Dept. of Chemistry and Applied Biosciences, ETH Zurich, Institute of Chemical and Bioengineering, Zurich, Switzerland

[4] DataHow AG, c/o ETH Zürich, HCl, F137, Vladimir-Prelog-Weg 1, CH-8093 Zurich, Switzerland

[#]Corresponding author: Dr.-Ing. Mariano Nicolas Cruz Bournazou, Bioprocess Engineering, Institute of Biotechnology, Technische Universität Berlin, Ackerstraße 76, ACK24, 13355 Berlin, Germany.

E-mail: nicolas.cruz@ethz.ch

(Manuscript ready for Submission)

Abstract

Concentration gradients that occur in industrial large-scale bioreactors due to mass transfer limitations have significant effects on process efficiency. Hence, it is desirable to investigate the response of strains to such bioreactor heterogeneities so as to reduce risk of failure during process scale-up. Although there are various scale-down experimental techniques to study these effects, scale-down strategies are rarely applied in the early developmental phases of a bioprocess as they have not yet been implemented on small scale parallel cultivation devices.

In this work, we combine mechanistic growth models with a parallel mini-bioreactor system to create a high-throughput platform for studying the response of *Escherichia coli* strains to heterogeneities in the glucose and dissolved oxygen concentration as they appear in industrial large-scale glucose-limited fed-batch processes. Here, as a scale-down approach, a model-based glucose pulse feeding scheme is applied and compared to a continuous feed profile, to study the influence of glucose and dissolved oxygen gradients on both cell physiology and the mis-incorporation of non-canonical amino acids into recombinant proinsulin. The results show a significant increase in the mis-incorporation of the non-canonical amino acid norvaline in the soluble intracellular extract and in the recombinant product in cultures with glucose/oxygen oscillations. Interestingly, the amount of norvaline depended on the pulse frequency and was negligible with continuous feeding, confirming observations from large-scale cultivations. Most importantly, the results also show that a larger number of the model parameters are significantly affected by the feeding model, i.e. continuous or discontinuous and also by the cycle interval. This indicates that the model based approach which has been applied here, is an important basis for implementing physiological aspects of the cell in large-scale bioreactor models.

In this example, it was possible to describe the effects of oscillations in a single parallel experiment. The platform offers the opportunity to combine strain screening with scale-down studies to select the most robust strains for bioprocess scale-up.

1. Introduction

Scale-down bioreactors have been applied in bioprocess development to study the response of microbial cell factories and other expression systems to various stress-inducing concentration gradients that occur in large-scale bioreactors [1,2]. In the past three decades, both multi-compartment and pulse-based scale-down bioreactor systems were developed to mimic gradients which are expected in large-scale bioreactors, in an attempt to better understand and characterise microbial behaviour at industrial production scale. For instance, a 2-compartment scale-down bioreactor with a total volume of 7 L was applied to study the loss in product yield and accumulation of undesirable metabolites in a 12 000 L industrial-scale cultivation of *E. coli* producing human growth hormone [3]. The successful operation of a scale-down bioreactor to study a particular process depends on how accurate the scale-down bioreactor mimics the environmental heterogeneity existing in the large-scale bioreactor. In the past, 2- or 3-compartment bioreactors have often been used as scale-down systems [1,4]. However, although they can provide comprehensive information on cellular responses, they are complex to operate and not easy to parallelize for early screening approaches.

The widespread use of parallel mini-bioreactor systems for strain screening has been adopted in many bioprocess development settings to help reduce product life cycles and accelerate R&D [5]. The parallelization and automation of such cultivation platforms enable screening of large libraries within shorter times [6]. An initial challenge of these systems was the implementation of controlled feeding technologies, as they are standard in laboratory bioreactors. However, various solutions have been developed recently to enable fed-batch-like conditions, such as the gradual supply of glucose to the culture through enzyme-based glucose release systems [7], the application of micro-pumps for continuous feed supply [8] and model-based intermittent feeding [9]. These improvements bring the screening system closer to actual cultivation conditions, such that a better performing strain can be selected for scale-up. However, it is important to note that under normal fed-batch-like screening conditions, the cells are still growing in a homogeneous, perfectly mixed environment without perturbations, which may be quite different from actual industrial-scale cultivation conditions. Therefore the strains selected in these high-throughput platforms for scale-up may not be robust enough for scale-up conditions. High-throughput experimental facilities that perform fed-batch experiments with frequent glucose perturbations have been reported [10–12]. Nevertheless, the objective of glucose perturbations was to increase process dynamics information in Optimal Experimental Design and no reference to bioreactor heterogeneities or scale-down efforts was done.

The aim of this work was to apply a high-throughput parallel mini-bioreactor system as a scale-down bioreactor together with a mechanistic, model-based approach to test and evaluate the performance of a possible production strain under oscillating conditions relevant on an industrial large-scale already in the screening phase. Therefore a platform of 21 parallel mini-bioreactors was used, that was coupled to a robotic liquid handling station (LHS) for automated operation of the cultivations. The operation of the LHS was based on mechanistic model outputs which describe both the dynamic physiological behaviour of the strain [13] and gradient profiles of scale-down bioreactors determined from typical mixing times of large-scale bioreactors [14]. As a demonstration, the platform was used to study the effects of the oscillating glucose/dissolved oxygen concentrations on the amount and quality of a recombinant pre-proinsulin expressed in *E. coli* as inclusion bodies. The quality of the protein was assessed by the amount of mis-incorporation of non-canonical branched chain amino acids (ncBCAA) into the recombinant product. The intracellular synthesis and incorporation of the non-canonical branched-chain amino acids norvaline, norleucine and ß-methyl-norleucine has been earlier described to occur as a result of conditions where oxygen depletion is connected with high glucose concentrations, which can appear in feeding zones of large-scale industrial bioreactors (for a recent review see [15]).

Such a high-throughput scale-down system is suitable for the generation of large physiological data under real oscillating conditions/gradients, that can be used for the validation of metabolic network models which may be applied in scale-up studies and models of large scale bioreactor scenarios, as in e.g. Haringa et al. [16]. Additionally, the platform also offers the possibility to compare the response of different candidate strains to concentration gradients during screening, to select the best phenotype for scale-up.

A few studies that consider the operating conditions of larger-scale bioreactors in high-throughput screening systems are reported in the literature. For instance, Janakiraman and co-workers matched the volumetric aeration rates (vvm) between parallel *ambr15*^{TM} cultivations of CHO cells and a 15 000 L production scale bioreactor [17], whereas Velez-Suberbie *et al*. used the power per unit volume as a scale-down criterion to compare *ambr15f* cultivations of *E. coli* with 20 L bioreactor cultivations [18]. Other works involving high-throughput cultivations in complex integrated facilities have also been reported [see recent review here [19]]. However, although these works point to the right direction in terms of high-throughput bioprocess development, matching only one engineering criterion between scales is not necessarily the same as replicating specific environmental heterogeneities in the smaller bioreactors for scale-down studies. The engineering criteria are global parameters for

the whole system, whereas scale-up effects arise from zone-specific heterogeneities within industrial bioreactors [20]. These heterogeneous environments have specific space-time dynamics in relation to the physiology of the organism. These dynamics must be reproduced in the smaller scale for proper scale-down studies. The smaller volumes of mini-bioreactor systems pose a difficulty in running such representative scale-down cultivations in high-throughput systems. Therefore, in the current work, we explore the applicability of mechanistic models combined with programmable liquid handling stations to reproduce heterogeneous conditions in mini-bioreactors. The methodological approach described here for calibration of model parameters under perturbations is an important basis for the generation of large-scale models which integrate fluid dynamics with cellular reactions, as pioneered by [21–23].

2. Materials and Methods

2.1 Strain, Cultivation Conditions and Mini-bioreactor Configuration

The scale-down experiments in mini-bioreactors were carried out with *E. coli* BW25113 (*lacI*$^{+}$) pSW3 (*amp*^{r}), expressing a recombinant proinsulin under the IPTG-inducible *tac*-promoter (provided by Sanofi-Aventis Deutschland GmbH). To make the starting culture, 20 ml of mineral salt medium, supplemented with trace elements [14], 5 g L^{-1} glucose and 100 mg L^{-1} ampicillin was inoculated with 100 µL of cryo stock culture in a 125 mL Erlenmeyer flask and incubated at 37°C, 200 rpm in an orbital shaker (Adolf Kühner AG, Birsfelden, Switzerland). The single-use baffled flask was equipped with optical sensor spots for online monitoring of pH and dissolved oxygen tension on the Presens platform (PreSens-Precision Sensing GmbH, Regensburg, Germany). After 10 hours of cultivation, whilst in exponential growth phase, an appropriate volume of the preculture was used to inoculate the bioreactor medium to a final optical density (OD_{600}) of 0.1. 10 ml of the inoculated medium was then aliquoted into each mini-bioreactor, under aseptic conditions to start the batch phase of the cultivation. The bioreactor medium consisted of mineral salt medium and trace elements (same as in the preculture), 5 g L^{-1} glucose and 30 g/L of EnPump200® glucose polymer (Enpresso GmbH, Berlin, Germany). During cultivation, the pH was maintained at a set point of 7.0 by by a PI controller with addition of 6% NH_4OH, the temperature was maintained at 37 °C. The cultivations were done in 15 ml disposable mini-bioreactor vessels (bioREACTOR® 48 system, 2mag AG, München, Germany) equipped with a magnetically driven impeller and photo-optical sensor spots for pH and DOT measurement. The bioreactor is connected to a Tecan®

pipetting robot (Tecan Group Ltd, Männedorf, Switzerland) and connected to a central database where all online and at-line data were directly saved. The aeration of each mini-bioreactor was achieved by an induced draft mechanism, where air was drawn into the broth through a hollow shaft on which the impeller was mounted [24]. By this mechanism, the whole bioreactor block was aerated to 10 L_{air} min^{-1} during cultivation. The detailed configuration of hardware-software protocols, sampling algorithms, pH control routines, pulse feed scheduling and off-line, at-line and on-line analytics on the robot station during parallel cultivations were presented previously [9,11,25].

2.2 Calculation of Gradient Profiles

In aerobic fed-batch cultivations in large industrial-scale bioreactors, zones of higher glucose concentrations (feed zones) are invariably associated with depletion of dissolved oxygen, whereas zones with low glucose concentrations have higher dissolved oxygen concentrations. Thus, both glucose and dissolved oxygen gradients occur simultaneously, although inversely, in larger bioreactors. To mimic such effects in the mini-bioreactors, a mechanistic solution of the pulse-based scale-down concept was developed [14]. The pulse frequencies and amplitudes (set point of growth rate μ) were calculated based on previous simulation of probable gradient profiles in large-scale bioreactors [14]. At the end of the batch phase, both online and at-line data were read directly from the central database of the cultivation platform into a programming environment to ascertain the current state of the cultures. Based on the measured data, model-based predictions of biomass and the glucose requirements were calculated; these were then used to define the feeding scheme of the respective cultivation. The mechanistic model of *E. coli* was solved in discrete intervals to obtain glucose concentrations that are equivalent to concentrations experienced by a cell in a large-scale bioreactor at the given mixing times. The pulsing scheme was determined by two time spans: t_{sp} and t_{swof}, which are respectively the times of glucose excess and glucose limitation, the sum of which reflects the pulse cycle. The 5 min pulse cycle results in approximate gradient profiles equivalent to those in larger bioreactors with about 30 seconds mixing time, whereas the 10 min pulse cycle is equivalent to 90 seconds mixing time in actual large-scale bioreactors [14]. For each pulse cycle ($t_{se} = t_{sp} + t_{swof}$), the exponential feed profile was integrated to find the volume of feed to be added to obtain a glucose excess (ca. 0.4—0.8 g L^{-1}) within the period t_{sp}. This was followed by the resting (starvation phase) when glucose was depleted from the medium. The details of the calculations and the actual code are given in a downloadable Matlab file in the Supplementary Material.

2.3 Scale-down Cultivations in Mini-bioreactors

A total of seven (7) cultivation conditions were implemented in triplicates (3 parallel mini-bioreactors for each condition) as follows (Table 1):

i. B1—B3: these were the reference or control cultivations. With 30 g L^{-1} of EnPump® polymer (EnPresso GmbH, Berlin, Germany) in the medium already, these mini-bioreactors were fed with glucoamylase (also from EnPresso GmbH), in distinct additions every 5 min to a final concentration of 13 U/L at the end of the exponential feeding phase. The linear enzyme feed resulted in an approximately exponential glucose release rate, which then switched to a near constant release rate when enzyme feeding was stopped [26].

ii. C1—C3: these were fed with a glucose pulse every 5 min during the fed-batch phase to mimic concentration gradients experienced in a larger bioreactor with a mixing time of approximately 40 sec [14]. In the first 3 hours of feeding, the pulse size was increased exponentially to maintain a specific growth rate of 0.35 h^{-1} in this phase, after which the pulse size was kept constant, upon IPTG induction. They were induced to a final concentration of 0.5 mM IPTG.

iii. D1—D3: These had the same feeding condition as C1—C3, but were induced to a final concentration of 1.0 mM IPTG.

iv. E1—E3: glucose pulse feed and induction conditions in these bioreactors were the same as those in C1—C3. However, a single dose of amylase (3 U/L) was added to this set of bioreactors to ensure a continuous supply of limiting glucose in the background, in addition to the 5-min glucose pulses. The residual glucose supply was calculated to be less than 10% of the glucose requirements to maintain the specific growth rate at the set value. This was to ensure that the culture in these bioreactors does not fall into acute starvation conditions between the pulses.

v. The last set of 9 mini-bioreactors were similar to the previous set of nine, with the only exception that the glucose pulses were administered every 10 min, producing approximate glucose and dissolved oxygen gradients experienced in bioreactors with mixing times of about 90 sec [14]. Thus, C1—C3 ≈ F1—F3, D1—D3 ≈ G1—G3 and E1—E3 ≈ H1—H3, with the exception of a glucose pulse frequency of 10 min in the latter sets, as shown in Table 1.

Table 1. Cultivation conditions in 21 parallel mini-bioreactors

Bioreactors			Pulse freq. (min)	IPTG mM	Enzyme-based feeding
B1	B2	B3	-	0.5	linearly to 13 U/L
C1	C2	C3	5	0.5	-
D1	D2	D3	5	1.0	-
E1	E2	E3	5	0.5	once, 3 U/L
F1	F2	F3	10	0.5	-
G1	G2	G3	10	1.0	-
H1	H2	H3	10	0.5	once, 3 U/L

2.4 Analyses

For analyses, 300 µL sample of culture broth was taken from each mini-bioreactor at the end of the batch phase, and thereafter on hourly basis in the fed-batch phase. Cell growth was monitored by optical density (OD) at 600 nm. Therefore, 10 µL of each sample was diluted by a factor of 20 with saline water (0.9 % NaCl) in a microwell plate, and the absorbance was measured in a plate reader (Synergy™ MX microwell plate reader, BioTek Instruments GmbH, Bad, Germany). The OD_{600} values were corrected against a blank (0.9 % NaCl), and divided by a factor of 0.5 for path length correction to a 1 cm cuvette, as described previously [11]. The OD_{600} values were converted to dry cell weight using a conversion factor of 0.37 g/L dry weight per 1 OD_{600}, which was previously determined for this strain using the same equipment. The rest of the samples were centrifuged at 15 000 ×g for 5 min and the supernatant and pellets were separated and stored at -20 °C for further analysis.

The concentrations of glucose, acetate and formate were measured enzymatically on the Cedex Bio HT Analyzer (Roche Diagnostics Int. Ltd.) with automatic dilution for samples where the calculated analyte concentration was outside the range of concentrations within the calibration curve. A maximum of 10 µL of thawed supernatant from the samples was required for the enzymatic analysis of the metabolites.

Protein quantification was done by analysing the cultivation samples and protein standards on a Labchip® GX II Touch (PerkinElmer, Waltham, US). The cell pellets were re-suspended in 500 µL PBS and normalized to OD_{600} = 5 in 1 ml. After centrifugation, the pellet from the normalized samples was lysed using Bugbuster® reagent, according to a previously described procedure [14]. Further inclusion body purification was carried out using the Bugbuster protocol, as also described previously [14]. From this procedure, three sample fractions were isolated for further processing: the raw cell lysate, intracellular soluble fraction and purified

inclusion bodies. For protein quantification, 2 µL of the raw cell lysate fraction was added to 7 µL of denaturing solution (700 µL PBS, 24.5 µL of 1M DTT) in PCR plates. The plate was incubated in a thermal cycler at 100 °C for 5 min. After cooling for 5 min at room temperature, the plate was loaded into the Labchip device to run the microchip capillary electrophoresis. Protein quantification was done by running the cultivation samples together with protein standards of similar size (10 kDa) as the recombinant proinsulin (11 kDa). Data analysis (calibration with standards, imaging and quantification) were done using the Labchip Reviewer® software (PerkinElmer). To quantify the concentration of non-canonical amino acids in each sample, both the purified inclusion body and the intracellular soluble fractions were hydrolysed in concentrated (6 M) HCl, followed by derivatization and analysis of derivatized amino acids on an Agilent GC-MS system (Agilent Technologies, Waldbronn, Germany) according to the procedure described previously [14]. To measure the level of mis-incorporation of ncBCAAs for each cultivation condition, the ratio of the concentrations of the non-canonical form to the respective canonical form of the amino acid in the product was calculated (e.g. moles of norvaline per mole of leucine), given that all samples were normalized to the same biomass concentration before analysis.

3. Results

3.1 Growth Profiles of *E. coli* in Responses to Glucose and Oxygen Pulses

To investigate the influence of scale-up effects on the performance of fed-batch processes in a high-throughput platform, *E. coli* BW25113 pSW3 was cultivated under seven different growth conditions. The conditions include glucose and dissolved oxygen gradients, different induction strengths, exposure to intermittent acute starvation zones and repetitive exposure to overflow conditions. The average of triplicate runs (biological replicates) for each condition is reported in the following sections, except for the concentration of non-canonical amino acids where the average of triplicate measurements (technical replicates) is reported. As there were no differences in set-up among the parallel runs in the batch phase of the cultivation, the discussion is limited to the fed-batch phase of the cultivation. With the same concentration of glucose in the batch phase, all 21 parallel cultivations reached the same average biomass of 2.3 ± 0.08 g L^{-1} at the end of this phase.

3.1.1 Effect of Pulse Frequency in the Fed-batch Phase

The pulse frequencies of 5 and 10 min led to lower biomass concentrations in the scale-down cultivations in comparison to the continuous reference cultivation, as shown in Figure 2A. The mini-bioreactors that were subjected to 5 min and 10 min glucose pulses respectively reached 17 % and 11 % less final biomass concentrations than the reference cultivation. Although the same mass of glucose was fed in both of the pulse-based cultivations, the 10 min pulse frequencies led to about 7 % higher biomass than the 5 min frequency pulses (Figure 2A).

Despite the relatively lower biomass, the highest concentration of overflow metabolites were measured in the pulse-based cultivations. The final acetate concentration in the 5 min pulse cultivation was about 5-fold higher than in the reference cultivation and 3-fold higher than that in the 10 min pulse cultivations (Figure 2B). In the 10 min pulse cultivations, a higher concentration of formate was recorded than in the 5 min pulses, but the formate was rapidly consumed when the feed was switched to constant feed. The frequency of the pulses also influenced the accumulation of inclusion bodies (product) during the cultivations. The 5 min pulse cultivations had 16 % and 22 % lower final concentration of inclusion bodies per gram of biomass (dry weight basis) than the 10 min pulse and reference cultivations, respectively (Figure 2D).

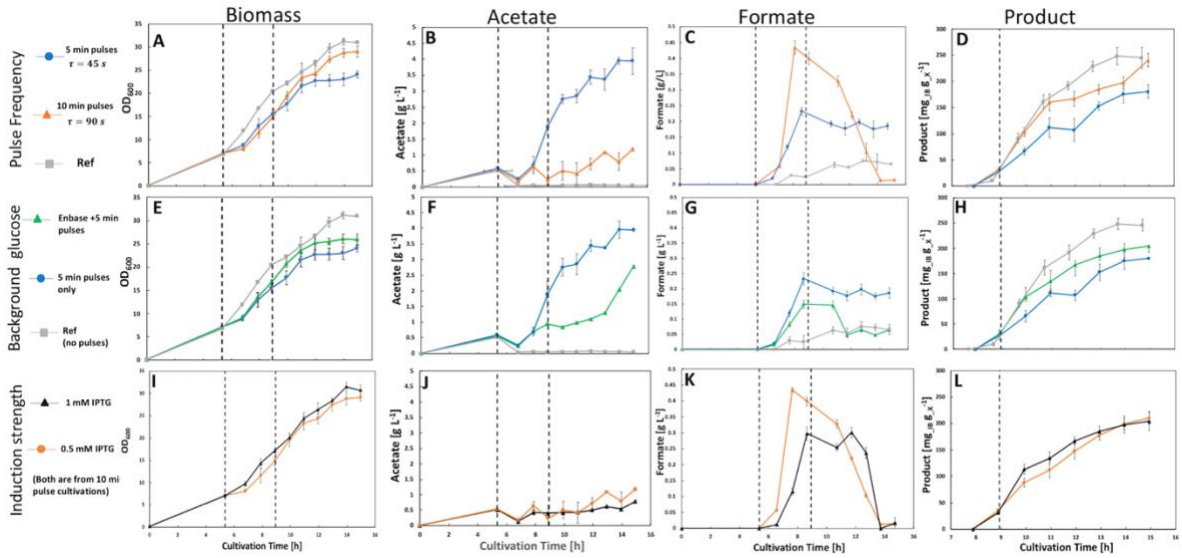


Figure 1. Growth profiles of fed-batch cultures of *E. coli* BW25113 pSW3 and metabolite concentrations from 21 parallel cultivations. The detailed conditions are shown in Table 1 and in the Materials and Methods section. The cultivations have been grouped according to variation in pulse frequency (A-D), presence or absence of an enzymatic glucose supply in the background (E-H) and the strength of IPTG induction (I-L). Profiles of 5 min pulse frequencies in A-D are repeated in E-H. Error bars represent standard deviation of triplicate cultivations.

3.1.2 Effect of Background/Residual Glucose Supply

The supply of background glucose through the enzymatic glucose release system led to higher biomass than the corresponding cultivations without background glucose. The 5 min pulse frequency with background glucose led to about 7 % higher biomass than the same cultivation without glucose release (Figure 1E), whereas the 10 min frequency only recorded 4 % higher biomass with the background glucose release. However, this marginal increase in biomass may be due to the 3% higher total glucose fed in these cultivations, i.e. the contribution of the background glucose from the enzymatic system.

However, contrary to expectations, the residual glucose supply led to a lower extracellular metabolite concentration: 34 % less acetate in 5 min pulse frequency (Figure 1F) and 10 % less acetate in the 10 min pulse frequency, compared to the reference cultivation. The same trends were also recorded for formate profiles (Figure 1G), where the 5 min and 10 min frequencies accumulated lower concentrations of formate compared to the corresponding cultivations without background glucose.

The accumulation of proinsulin was also slightly influenced by the background glucose supply, with the 5 min frequency (Figure 1H). The final amount of inclusion bodies in the cultivations with background feeding was approx. 6 % higher compared to the corresponding cultivations without background glucose feeding. The recombinant product in the reference cultivation was still higher than that in the pulse-based cultivations with the background enzymatic glucose release system.

3.1.3 Effect of Induction Strength

As observed in earlier studies [14], the intermittent starvation of *E. coli* cells due to the pulse feeding mechanism partially de-represses the *tac* promoter under which the recombinant protein is produced. Therefore, before induction, some amount of inclusion bodies already accumulated in both the 5 min and 10 min frequency cultivations (Figure 1 D, H and L). This pre-induction product formation was not observed in the reference cultivation (sample taken immediately before induction) where glucose supply was smooth and continuous. A slightly higher induction strength (1 mM IPTG) did not have any observable influence on cell growth and extracellular metabolite accumulation for the cultivations with 5 and 10 min pulses respectively (Figure 1 I—K). However, the rate of product accumulation was predictably influenced by the IPTG concentration. The higher IPTG concentration (1 mM) led to a more rapid accumulation of inclusion bodies, but this also resulted in a slightly faster saturation of the recombinant protein production rate compared to the cultures which were induced by 0.5

mM IPTG. Apart from the product formation kinetics which were different between the two IPTG concentrations, the final inclusion body concentration was the same for both (Figure 1 L).

3.2 Model Fitting of Scale-down Cultivations

To further investigate the influence of glucose and dissolved oxygen pulses (gradients) on the metabolic behaviour of *E. coli* BW25113, datasets from the scale-down cultivations were fitted to the mechanistic model (see Appendix) to estimate the values of parameters that were achieved during the cultivations. The criterion of the model fitting was the minimization of the sum of squares of residuals between the measured data and model predictions, using the interior-point algorithm, which was executed within the global optimisation framework *Globalsearch* in Matlab®. As described previously [13], the use of structured mechanistic models to describe the physiological states of the culture also allows us to estimate some metabolic states of the culture that are otherwise difficult to measure. The results of the model fitting are shown in Figure 2, whereas parameters that describe maximum specific uptake rates, yield coefficients and maintenance energy levels are compared in Table 2 for the 5 min pulse, 10 min pulse and reference cultivations. Also the original parameter values that were used to calculate the feed profiles in the pulse-based fed-batch culture are provided in Table 2.

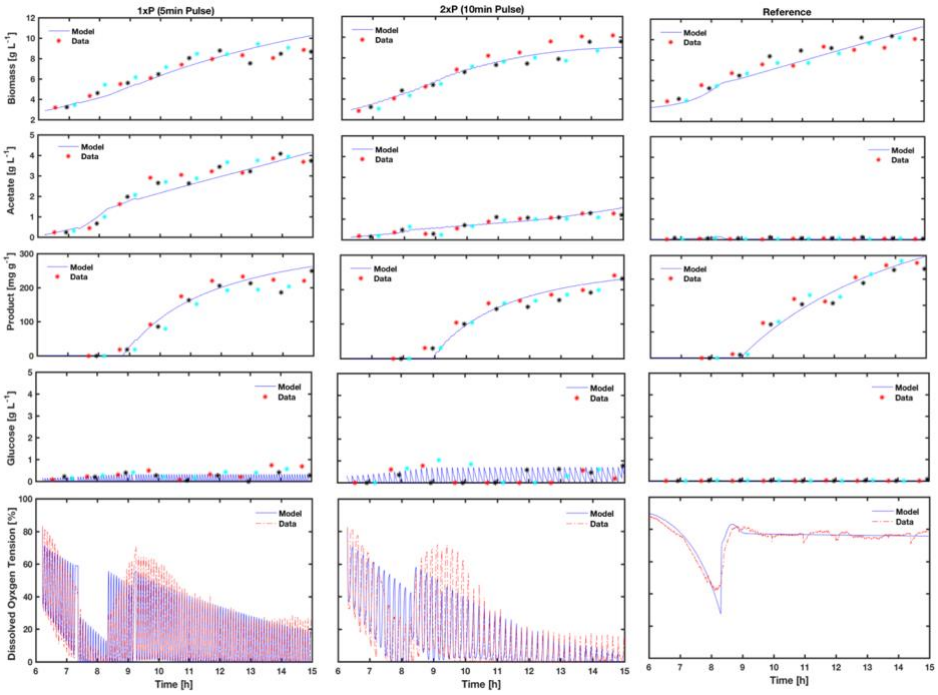
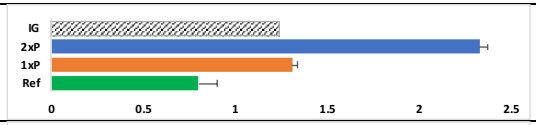
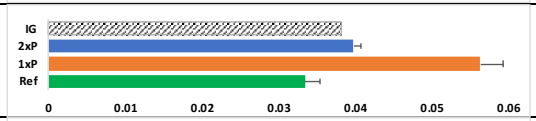
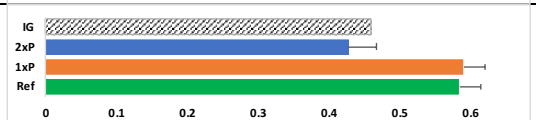
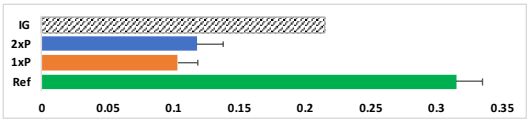
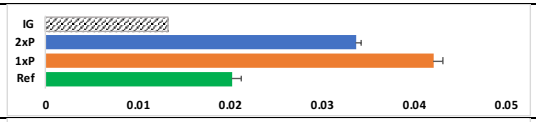
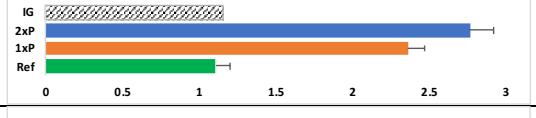
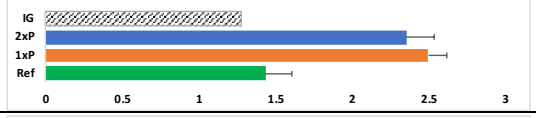
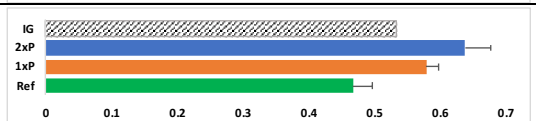
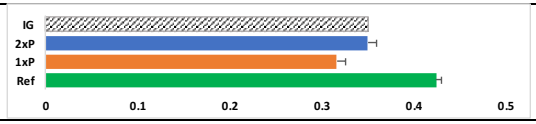


Figure 2 Fitting of mechanistic model to data of pulse-based scale-down cultivations in the mini-bioreactors. All the data points of the triplicate runs for each condition were pooled together for the parameter estimation (except for DOT). Symbols: * , *, * are triplicate runs, — model. Cultivations are shown from the start of the feeding phase at 6 hours. Induction with IPTG (0.5 or 1 mM respectively) was performed at 9 h.

Table 2. Results of parameter estimation for scale-down fed-batch cultivation in miniaturized bioreactors, showing the maximum specific rates and yield coefficients of *E. coli* in response to the induced glucose pulses. IG—initial guesses, which were also used to calculate the feed profiles; 2xP—10 min glucose pulse cultivation; 1xP—5 min glucose pulse cultivation; Ref—reference cultivation. Error bars represent the standard deviation for each parameter estimate, which was calculated using the method of *covariance matrix based on sensitivity matrix*, as described by López et al. 2016 [27]

Parameter	Estimated Values	Notes
q_{Smax}		In agreement with the results of [28] and [29], the *E. coli* strain developed a higher capacity for glucose uptake in the presence of glucose pulses, where the adaptation of the maximum uptake capacity seems to be proportional to the pulse size.
K_s		Both pulse-based cultivations resulted in higher K_s values than the reference cultivation, which implies that the *E. coli* strain developed a lower affinity for glucose under the induced concentration gradients.
p_{Amax}		Interesting to note that the specific maximum acetate production rate is the same in both the 5 min pulse-based and reference cultivations. The difference in acetate profiles must therefore be accounted for by other parameters such as differences in its re-assimilation rates
q_{Amax}		Under homogeneous conditions, there is a higher acetate re-assimilation rate, which results in lower extracellular acetate than under pulse-based conditions. Within the pulses, the re-assimilation rate of acetate in the 5 min pulses is slightly lower than in 10 min pulses which is reflected in the acetate profiles.
q_m		Homogeneous cultivation conditions led to lower maintenance energy. Higher frequency glucose pulses divert more energy towards maintenance than low frequency pulses.
Y_{oa}		Both pulse based cultivations used up more oxygen for oxidation of acetate than in the reference. This yield coefficient is directly linked to the specific oxygen uptake rate (qO) in the model.
Y_{os}		Oxygen requirements for glucose oxidation in the pulse-based cultivation was higher than in reference. This may be linked to the higher maintenance energy requirements under heterogeneous cultivation conditions.
Y_{xs}		Without taking maintenance energy into consideration and biomass gained from acetate consumption, there seemed to be a higher yield of biomass on glucose in the pulse-based cultivation than in the reference cultivations.
Y_{px}		In terms of recombinant product yield, the pulses had a negative effect on product formation. The reference cultivation had the highest mass of inclusion bodies of proinsulin per gram of biomass.

3.3 Metabolic Responses of *E. coli* to Glucose Pulses based on Model Fitting

The parameters in Table 2 were estimated for the overall duration of the cultivation by collating data points from the triplicate runs of each condition. However, to track the time-course evolution of the metabolic fluxes, the estimated parameters and experimental data were used to calculate the specific rates during the cultivation using equations A.1—A.9 given in the Appendix. The results are shown in Figure 3. Whereas Table 2 depicts the metabolic potential or the capacity of the strain under each cultivation condition, Figure 3 depicts the actual metabolic states achieved during cultivation under each of the conditions.

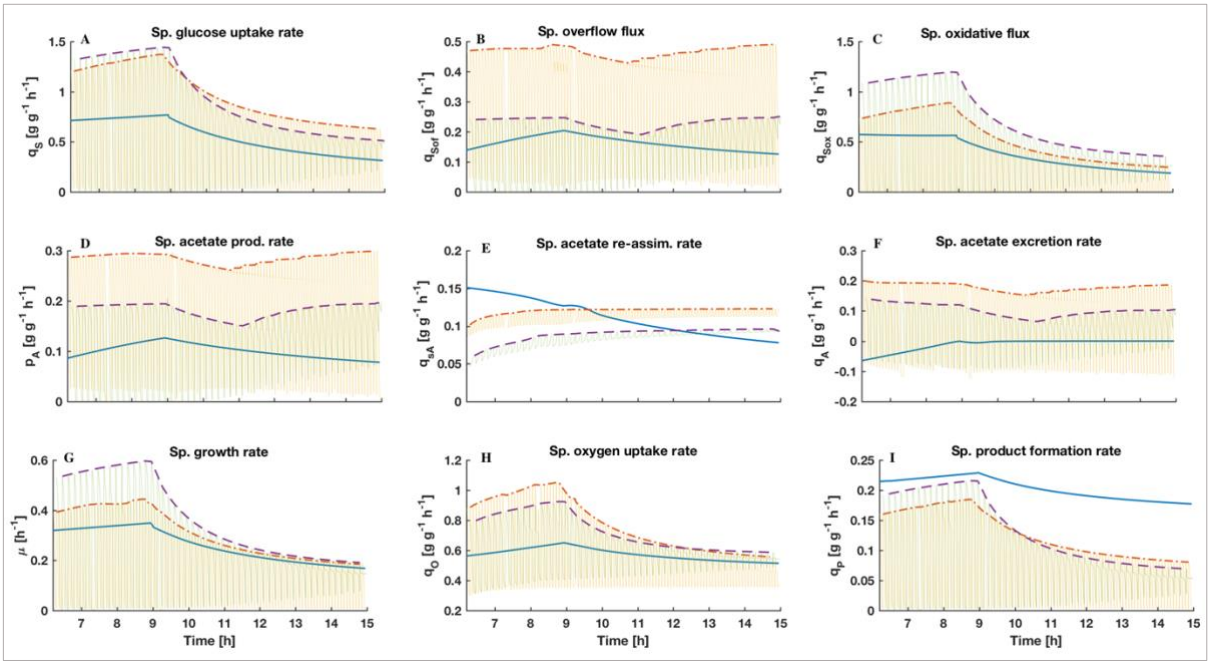


Figure 3. Profiles of metabolic fluxes of *E. coli* BW25113 pSW3 during the different cultivation conditions: ref—reference cultivation with smooth glucose supply; the other two are fed-batch cultivation with pulse-based feed of glucose every 5 min (1xP) and every 10 min (2xP); sp.—specific. The broken lines depict the overall trend of the fluxes, whereas the actual data of the pulse-based cultivations are shown as the oscillations in the background. Note that negative secretion rates are shown when acetate is consumed. Cultivations are shown from the start of the feeding phase at 6 hours. Induction with IPTG (0.5 or 1 mM) was performed at 9 h with the switch to constant feeding.

The metabolic profiles show lower glucose uptake rates in the reference cultivation, with a lower overflow flux than in the pulse-based cultivations (Figure 3 A-C). Noting that the sum of the overflow and oxidative fluxes for each condition is equal to the total glucose uptake rate for that condition, it is apparent that more carbon source was wasted through overflow metabolites in the pulse-based cultivations. This is evident from the ratios of the average

overflow flux to the average glucose uptake rate ($^{\bar{q}_{sof}}/_{\bar{q}_s}$) during the exponential fed-batch phase for each cultivation, which are 0.37, 0.22 and 0.16 for the 5 min pulse, 10 min pulse and the reference cultivation, respectively.

The acetate cycling fluxes (Figure 3 D, E and F) show a higher specific acetate re-assimilation rate for the reference cultivation. Therefore, although there was some amount of acetate produced in these cultivations, the higher re-assimilation rate did not allow acetate accumulation in the medium. According to the acetate cycling concept [13], the difference in acetate re-assimilation rates, relative to its production rates, is what actually leads to different acetate concentrations in the extracellular medium in the cultivations.

Additionally, the specific growth rates that were achieved in the exponential feed fed-batch phase were higher than the set point ($\mu_{set} = 0.35\ h^{-1}$) in the pulse based cultivations (13 % higher for 5 min, 33 % higher for 10 min). The deviation from the set point of the specific growth rate seemed to be directly proportional to the magnitude of the glucose pulses (Figure 3 G). Probably due to a higher maintenance energy (Table 2) in the pulse-based cultivations, there was a higher oxygen consumption rate in these bioreactors (Figure 3H) than in the reference cultivation, whereas the specific product formation rate was higher in the reference cultivation than in the pulse-based cultivations.

3.4 Accumulation and Mis-incorporation of the Non-canonical Amino Acid Norvaline into Proinsulin

Generally, there was a higher concentration of norvaline in the intracellular soluble fraction than in the purified inclusion bodies for all cultivation conditions. For instance, in the 5 min and 10 min glucose pulse cultivations (without background glucose feeding), there were respectively 5 and 3 times higher norvaline in the intracellular soluble fraction than in the purified inclusion bodies (Figure 4A). Since the recombinant protein used for this study is very rich in leucine with only a few methionine and isoleucine residues [15], much of the discussion is based on the mis-incorporation of norvaline for leucine in the product.

The ratio norvaline to leucine in the inclusion bodies from the 5 min pulse cultivations was about 50 times higher than in the reference cultivation, whereas the 10 min pulse cultivations had about 30 times higher norvaline mis-incorporation levels (norvaline-leucine ratio) compared to the reference cultivation (Figure 5B).

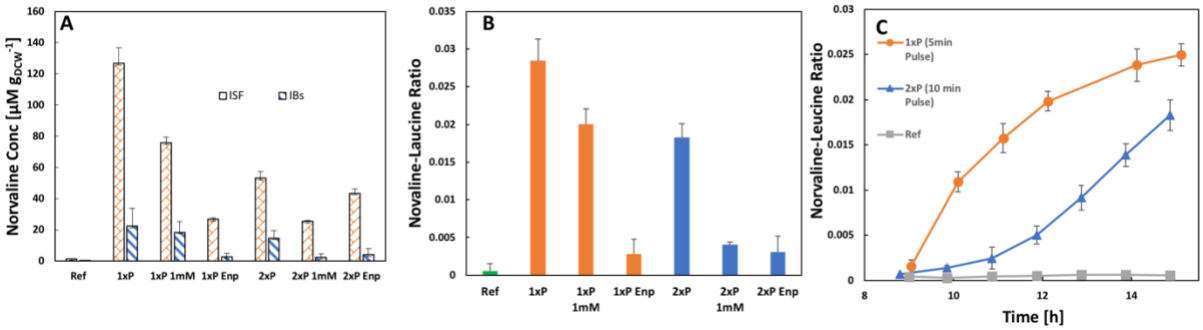


Figure 4. Influence of glucose pulses on the concentration of the non-canonical amino acid norvaline: **A**: intracellular soluble fraction (ISF) and purified inclusion bodies (IBs); **B:** mis-incorporation levels in the purified inclusion bodies; **C**: Dynamic profile of mis-incorporation (1xP and 2xP, both 0.5 mM IPTG) as a function of glucose pulse frequency. Ref: reference cultivation; 2xP—10 min glucose pulse cultivation; 1xP—5 min glucose pulse cultivation; Enp—cultivation with enzymatic glucose release.

The frequency of the pulses also affected the kinetics of norvaline mis-incorporation (Figure 4C). At higher pulse frequencies, there was a more rapid mis-incorporation rate, which seemed to reach saturation levels about 4 hours after protein induction. Conversely, the low frequency pulses led to a more gradual mis-incorporation rate at the beginning, but this seemed not to reach any saturation levels for all the 6 hours that were investigated after induction

The presence of glucose in the background of the pulses lowered the concentration of norvaline in both the recombinant product and cellular material (Figure 4B). For the same cultivation conditions (e.g. 5 min pulse, 0.5 mM IPTG), the runs that had background enzymatic glucose release had a 10 times lower norvaline mis-incorporation than the cultivations without the background glucose feed. Thus, in situations where the cells were prevented from experiencing acute glucose starvation, the level of mis-incorporation was less than when complete glucose depletion occurred between pulses.

4. Discussion

We report here the development and application of a high-throughput mini-bioreactor scale-down platform for investigating the effects of heterogenous cultivation conditions, particularly glucose and dissolved oxygen gradients on *E. coli* fed-batch cultivation, at the early bioprocess development phase. As opposed to strain screening and fermentation development under

homogeneous cultivation conditions, the results of the scale-down cultivations in the current study show significant variations in strain behaviour, in response to different cultivation conditions. In particular, the cultivations that were subjected to pure glucose pulses with accompanying limitation in dissolved oxygen supply (Figure 2) showed higher extracellular metabolite concentrations, low biomass, low product yields and overall poor performance compared to the reference cultivations. The higher concentration of acetate in the high frequency glucose pulses may be due to the inability of *E. coli* to re-assimilate this metabolite in the presence of rapidly changing glucose concentrations, as seen in the parameters of the specific acetate re-assimilation rates of the strain in the three conditions (Figure 3E). However, with long-term stress exposure, the cells seem to build the whole framework of stress response mechanisms, such as production of sigma factors, membrane modifications and other adaptations as observed in the 10 min pulses. The metabolic adaptation of the strain to the various conditions over the period of the cultivation could be deduced from the mechanistic model fitting, as given by the maximum achievable specific rates under each cultivation condition in Table 2. The *E. coli* BW25113 psW3 strain showed a significant shift in these rates from efficient utilisation of glucose (low ratios of $\bar{q}_{sof}$ to $\bar{q}_s$) under homogeneous cultivation conditions to carbon wasting through overflow routes (high ratios of $\bar{q}_{sof}$ to $\bar{q}_s$) under the pulse-based cultivations. This shift in metabolic states is further seen in the higher values of the parameters q_{Amax} and p_{Amax} in the pulse-based cultivations compared to the reference cultivation (Table 2). The model fitting further showed strong parameter dependencies on specific cultivation conditions. For instance, variations in the values of the affinity constant K_s in relation to the residual glucose under each cultivation condition determines the sensitivity of the strain to the heterogeneous environment, as previously reported by Haringa and co-workers [23]. The model fitting revealed an increasing K_s value with increasing pulse frequency, which means under the higher pulse frequency, the cells were adapting to a different metabolic regime (lifeline) that would be less sensitive to the heterogeneous conditions.

Additionally, the glucose pulses led to a significant deviation of the specific growth rates from the set-points in the exponential feed fed-batch phase. In the presence of low frequency gradients, which are usually present in large industrial bioreactors [30], the cells are occasionally exposed to substrate concentration pockets that may trigger growth at μ_{max}, although they are still in fed-batch phase where growth is supposed to be controlled well below μ_{max}. This loss of physiological control may contribute to the lack of process reproducibility in

industrial scale cultivations. Another important aspect of the metabolic response to the heterogenous cultivation conditions is that *E. coli* tends to maintain higher metabolic states in the presence of gradients (Figure 3), which in some cases can lead to higher viability in heterogenous environments than in homogenous environments [31]. However, as shown in the product profiles (Figure 2), this higher viability does not necessarily translate into higher productivity in recombinant protein processes. Rather, the higher metabolic states may channel energy towards maintenance and adaption of the strain to the heterogenous environment.

The scale-down condition with the background enzymatic glucose feed represents the situation in the large-scale bioreactor where the cells are occasionally exposed to excess glucose conditions, but not acute starvation zones. This *zone* may be in the vicinity of recirculation loops in broths that are mixed by Rushton turbine impellers in the large scale. This scale-down condition led to relatively lower extracellular metabolite concentrations, slightly higher biomass concentrations and lower concentration of norvaline. The continuous supply of background glucose by the enzymatic feeding system may maintain substrate concentrations close to the K_s value of the organism. This ensures that a steady pool of intracellular metabolites is maintained in the cell, which prevents the starvation responses that were apparent in the pure pulse feed conditions.

As observed for the concentration of extracellular metabolites, the higher frequency pulses have clear influence on the product quality (mis-incorporation of norvaline) more adversely than the low frequency gradients (Figure 4B). We observed a close relationship between the profiles of extracellular metabolites and the accumulation of norvaline in the cells, possibly due to the ability of the leucine synthesis genes to use the accumulated pyruvate as a substrate to synthesise α-keto-butyrate and further α-keto-valerate [32], which are precursors for norvaline production. The higher concentration of norvaline in the intracellular soluble fraction than in the inclusion bodies (Figure 4A) confirms the hypothesis that it first accumulates in the cellular material under stressful cultivation conditions. Then, as its concentration increases, there is an increase in the probability of mis-acylation of tRNA molecules, due to the similar structure to its canonical isoform leucine [15,33]. The cultures that were induced to a higher inducer concentration had a lower level of mis-incorporation of norvaline than those with lower IPTG concentration (Figure 4B). This observation is interesting, since it shows that the induction strength can influence the quality of the product.

5. Conclusions

Bioprocess development is very time consuming and labour intensive, requiring several experimental runs to characterise a specific product and its production process. The development of high-throughput cultivation platforms is a promising technology which facilitates both strain screening and fermentation process development. As pointed out by Tannaz *et al*. 2019, the configuration of miniaturized bioreactors to achieve real gradient profiles for scale-down studies is challenging. In this study, we leverage on the capability of mechanistic modelling and LHS to achieve approximate gradients that give physiological responses similar to those previously observed in other scale-down studies. The use of the mechanistic model (which is derived from the physiology of the strain) to define the gradient profiles for the various scale-down conditions is key to achieving real stress responses that are also observed in large-scale bioreactor cultivations of *E. coli*. We prove that the specific perturbation conditions have a direct impact on many of the model parameters which describe the cell physiology. This indicates that the use of an appropriate scale down model is important for the development of large-scale models which include the cell physiology. Furthermore, the model based identification of the cellular parameters provides an important tool for the selection of strains already in the screening phase which provide an improved performance or higher robustness under large-scale conditions. This would help to select the most robust strains for further development and subsequent bioprocess scale-up.

Acknowledgements

We are grateful for financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie actions grant agreement No. 643056 (Biorapid).

Conflict of interest

The authors declare no financial or commercial conflict of interest.

Supplementary Data

The Matlab® codes and fermentation data can be found at https://gitlab.tubit.tu-berlin.de/nicolas.cruz/E_coli_fed-batch/tree/master/Anane_2018_JCTB

References

[1] P. Neubauer, S. Junne, Scale-Up and Scale-Down Methodologies for Bioreactors, in: C.-F. Mandenius (Ed.), Bioreact. Des. Oper. Nov. Appl., Wiley-VCH Verlag GmbH, Weinheim, Germany, 2016: pp. 323–354. doi:10.1002/9783527683369.ch11.

[2] A.W. Nienow, W.H. Scott, C.J. Hewitt, C.R. Thomas, G. Lewis, A. Amanullah, R. Kiss, S.J. Meier, Scale-down studies for assessing the impact of different stress parameters on growth and product quality during animal cell culture, Chem. Eng. Res. Des. 91 (2013) 2265–2274. doi:10.1016/j.cherd.2013.04.002.

[3] F. Bylund, A. Castan, R. Mikkola, A. Veide, G. Larsson, Influence of scale-up on the quality of recombinant human growth hormone, Biotechnol. Bioeng. 69 (2000) 119–128. doi:10.1002/(SICI)1097-0290(20000720)69:2<119::AID-BIT1>3.0.CO;2-9.

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

[5] T. Tajsoleiman, L. Mears, U. Krühne, K. V. Gernaey, S. Cornelissen, An Industrial Perspective on Scale-Down Challenges Using Miniaturized Bioreactors, Trends Biotechnol. 0 (2019) 1–10. doi:10.1016/j.tibtech.2019.01.002.

[6] A. Back, T. Rossignol, F. Krier, J.-M. Nicaud, P. Dhulster, High-throughput fermentation screening for the yeast Yarrowia lipolytica with real-time monitoring of biomass and lipid production, Microb. Cell Fact. 15 (2016) 147. doi:10.1186/s12934-016-0546-z.

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

[8] G. Gebhardt, R. Hortsch, K. Kaufmann, M. Arnold, D. Weuster-Botz, A new microfluidic concept for parallel operated milliliter-scale stirred tank bioreactors, Biotechnol. Prog. 27 (2011) 684–690. doi:10.1002/btpr.570.

[9] A. Sawatzki, S. Hans, H. Narayanan, B. Haby, N. Krausch, M. Sokolov, F. Glauche, S. Riedel, P. Neubauer, M. Cruz Bournazou, Accelerated Bioprocess Development of Endopolygalacturonase-Production with Saccharomyces cerevisiae Using Multivariate Prediction in a 48 Mini-Bioreactor Automated Platform, Bioengineering. 5 (2018) 101. doi:10.3390/bioengineering5040101.

[10] T. Barz, A. Sommer, T. Wilms, P. Neubauer, M.N. Cruz Bournazou, Adaptive optimal operation of a parallel robotic liquid handling station, IFAC-PapersOnLine. 51 (2018) 765–770. doi:10.1016/j.ifacol.2018.04.006.

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

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

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

[14] E. Anane, A. Sawatzki, P. Neubauer, M.N. Cruz-Bournazou, Modelling concentration gradients in fed-batch cultivations of E. coli - towards the flexible design of scale-down experiments, J. Chem. Technol. Biotechnol. 94 (2019) 516–526. doi:10.1002/jctb.5798.

[15] C. Reitz, Q. Fan, P. Neubauer, Synthesis of non-canonical branched-chain amino acids in Escherichia coli and approaches to avoid their incorporation into recombinant proteins, Curr. Opin. Biotechnol. 53 (2018) 248–253. doi:10.1016/j.copbio.2018.05.003.
[16] C. Haringa, W. Tang, G. Wang, A.T. Deshmukh, W.A. van Winden, J. Chu, W.M. van Gulik, J.J. Heijnen, R.F. Mudde, H.J. Noorman, Computational fluid dynamics simulation of an industrial P. chrysogenum fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization, Chem. Eng. Sci. 175 (2018) 12–24. doi:10.1016/j.ces.2017.09.020.
[17] V. Janakiraman, C. Kwiatkowski, R. Kshirsagar, T. Ryll, Y.-M. Huang, Application of high-throughput mini-bioreactor system for systematic scale-down modeling, process characterization, and control strategy development, Biotechnol. Prog. 31 (2015) 1623–1632. doi:10.1002/btpr.2162.
[18] M.L. Velez-Suberbie, J.P.J. Betts, K.L. Walker, C. Robinson, B. Zoro, E. Keshavarz-Moore, High throughput automated microbial bioreactor system used for clone selection and rapid scale-down process optimization, Biotechnol. Prog. 34 (2018) 58–68. doi:10.1002/btpr.2534.
[19] J. Hemmerich, S. Noack, W. Wiechert, M. Oldiges, Microbioreactor Systems for Accelerated Bioprocess Development, Biotechnol. J. 13 (2018) 1–9. doi:10.1002/biot.201700141.
[20] S.O. Enfors, M. Jahic, A. Rozkov, B. Xu, M. Hecker, B. Jürgen, E. Krüger, T. Schweder, G. Hamer, D. O'Beirne, N. Noisommit-Rizzi, M. Reuss, L. Boone, C. Hewitt, C. McFarlane, A. Nienow, T. Kovacs, C. Trägårdh, L. Fuchs, J. Revstedt, P.C. Friberg, B. Hjertager, G. Blomsten, H. Skogman, S. Hjort, F. Hoeks, H.Y. Lin, P. Neubauer, R. van der Lans, K. Luyben, P. Vrabel, A. Manelius, Physiological responses to mixing in large scale bioreactors., J. Biotechnol. 85 (2001) 175–85.
[21] C. Haringa, W. Tang, A.T. Deshmukh, J. Xia, M. Reuss, J.J. Heijnen, R.F. Mudde, H.J. Noorman, Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines, Eng. Life Sci. 16 (2016) 652–663. doi:10.1002/elsc.201600061.
[22] A. Lapin, J. Schmid, M. Reuss, Modeling the dynamics of E . coli populations in the three-dimensional turbulent field of a stirred-tank bioreactor — A structured – segregated approach, 61 (2006) 4783–4797. doi:10.1016/j.ces.2006.03.003.
[23] C. Haringa, A.T. Deshmukh, R.F. Mudde, H.J. Noorman, Euler-Lagrange analysis towards representative down-scaling of a 22 m3 aerobic S. cerevisiae fermentation, Chem. Eng. Sci. 170 (2017) 653–669. doi:10.1016/j.ces.2017.01.014.
[24] R. Puskeiler, K. Kaufmann, D. Weuster-Botz, Development, parallelization, and automation of a gas-inducing milliliter-scale bioreactor for high-throughput bioprocess design (HTBD), Biotechnol. Bioeng. 89 (2005) 512–523. doi:10.1002/bit.20352.
[25] D.B. Nickel, M.N. Cruz-Bournazou, T. Wilms, P. Neubauer, A. Knepper, Online bioprocess data generation, analysis, and optimization for parallel fed-batch fermentations in milliliter scale, Eng. Life Sci. 17 (2017) 1195–1201. doi:10.1002/elsc.201600035.
[26] J. Panula-perälä, J. Šiurkus, A. Vasala, R. Wilmanowski, M.G. Casteleijn, P. Neubauer, Enzyme controlled glucose auto-delivery for high cell density cultivations in microplates and shake flasks, 12 (2008) 1–12. doi:10.1186/1475-2859-7-31.
[27] D.C. López, G. Wozny, A. Flores-Tlacuahuac, R. Vasquez-Medrano, V.M. Zavala, A Computational Framework for Identifiability and Ill-Conditioning Analysis of Lithium-Ion Battery Models, Ind. Eng. Chem. Res. 55 (2016) 3026–3042. doi:10.1021/acs.iecr.5b03910.

[28] H.Y. Lin, B. Mathiszik, B. Xu, S.-O.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 (2001) 347–357. doi:10.1002/bit.1068.

[29] E. Brand, S. Junne, E. Anane, M.N. Cruz-Bournazou, P. Neubauer, Importance of the cultivation history for the response of Escherichia coli to oscillations in scale-down experiments, Bioprocess Biosyst. Eng. 41 (2018) 1305–1313. doi:10.1007/s00449-018-1958-4.

[30] A.R. Lara, E. Galindo, O.T. Ramírez, L.A. Palomares, Living With Heterogeneities in Bioreactors: Understanding the Effects of Environmental Gradients on Cells, Mol. Biotechnol. 34 (2006) 355–382. doi:10.1385/MB:34:3:355.

[31] S. Sunya, C. Bideaux, C. Molina-Jouve, N. Gorret, Short-term dynamic behavior of Escherichia coli in response to successive glucose pulses on glucose-limited chemostat cultures, J. Biotechnol. 164 (2013) 531–542. doi:10.1016/j.jbiotec.2013.01.014.

[32] J. Soini, K. Ukkonen, P. Neubauer, Accumulation of amino acids deriving from pyruvate in Escherichia coli W3110 during fed-batch cultivation in a two-compartment scale-down bioreactor, Adv. Biosci. Biotechnol. 02 (2011) 336–339. doi:10.4236/abb.2011.25049.

[33] T.E. Jones, R.W. Alexander, T. Pan, Misacylation of specific nonmethionyl tRNAs by a bacterial methionyl-tRNA synthetase, Proc. Natl. Acad. Sci. 108 (2011) 6933–6938. doi:10.1073/pnas.1019033108.

[34] H. Ying Lin, P. Neubauer, Influence of controlled glucose oscillations on a fed-batch process of recombinant Escherichia coli, J. Biotechnol. 79 (2000) 27–37. doi:10.1016/S0168-1656(00)00217-0.

[35] 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 (2001) 347–357. doi:10.1002/bit.1068.

[36] A.C. Badino, M. Cândida, R. Facciotti, W. Schmidell, Improving k(L)a determination in fungal fermentation, taking into account electrode response time, J. Chem. Technol. Biotechnol. 75 (2000) 469–474. doi:10.1002/1097-4660(200006)75:6<469::AID-JCTB236>3.0.CO;2-4.

Appendix

A.1 Macro-kinetic model of *Escherichia coli*

The mechanistic model of *E. coli* used in this publication is based on the physiological use of glucose, in a glucose partitioning framework as well as the overflow of glucose to acetate through the acetate cycling concept. The two physiological concepts, as given by Neubauer et al. 2000 [34] and Lin et al., [35] and basic growth concepts such as Monod kinetics and acetate inhibition are used to derive simple algebraic equations that describe intracellular pathways of glucose and oxygen usage. Details on the derivation of the model and its subsequent usage in *E. coli* processes can be found in the literature [11,13,35]. The algebraic equations that describe these intracellular activities are as follows:

$$q_s = \frac{q_{smax} S}{S+K_s} . e^{-P*K_{ip}} \quad \text{(A.1)}$$

$$q_{sox} = \left(q_s - \frac{P_{Amax} q_s}{q_s+K_{ap}}\right) \cdot \frac{DOT}{DOT+K_o} \quad \text{(A.2)}$$

$$q_{sof} = q_s - q_{sox} \quad \text{(A.3)}$$

$$p_A = q_{sof} Y_{as} \quad \text{(A.4)}$$

$$q_{sA} = \frac{q_{Amax}}{1+\frac{q_s}{K_{is}}} \cdot \frac{A}{A+K_{sa}} \quad \text{(A.5)}$$

$$q_A = p_A - q_{sA} \quad \text{(A.6)}$$

$$\mu = (q_{sox} - q_m) Y_{em} + q_{sA} Y_{xa} + (q_{sof} - p_A) Y_{xsof} \quad \text{(A.7)}$$

$$q_o = (q_{sox} - q_m) Y_{os} + q_{sA} Y_{oa} \quad \text{(A.8)}$$

$$q_p = \mu Y_{px} \quad \text{(A.9)}$$

The algebraic equations are coupled with mass balances for a fed-batch process to yield the full ODE system. The ODE system for the *E. coli* model is derived from mass balances on biomass (X), glucose (substrate, S), acetate (A) and dissolved oxygen measured as the percentage saturation at the operating conditions in the bioreactor (DOT).

$$\frac{dX}{dt} = \frac{F}{V}(0 - X) + \mu X \quad \text{(A.10)}$$

$$\frac{dS}{dt} = \frac{F}{V}(S_i - S) - q_s X \quad \text{(A.11)}$$

$$\frac{dA}{dt} = \frac{F}{V}(0 - A) + q_{sA} X \quad \text{(A.12)}$$

$$\frac{dDOT}{dt} = K_{La}(DOT^* - DOT) - q_o X H \quad \text{(A.13a)}$$

$$\frac{dDOT^m}{dt} = K_p(DOT - DOT^m) \quad \text{(A.14)}$$

$$\frac{dP}{dt} = q_p - \mu P \quad \text{(A.15)}$$

There is a significant difference between biological *times* of macroscopic phenomena (growth and cell division, as captured in equation A.7 & A.10) and mass transfer times (Equations A.8 and A.13). The time required for diffusion of oxygen across the gas-liquid interphase into the fermentation broth is several orders of magnitude greater than the cell division time [30]. The coupling of two differential equations describing a slow process and a very fast process results in a stiff ODE system, which is mathematically difficult to handle. Since oxygen transfer was considered to be very fast, the concentration of dissolved oxygen at any time was considered to be at steady state. Therefore, Equation A.13a was reduced to the algebraic steady-state solution (Equation A.13b), which was used in the mechanistic model instead of Equation A.13a.

$$DOT = \frac{K_{La}DOT^* - q_o XH}{K_{La}} \quad \text{(A.13b)}$$

Additionally, in the presence of a fast changing DOT signal due to the induced gradients (glucose pulses), the response time of the dissolved oxygen probe becomes significant in predicting the DOT profile. This is especially important when the response time of the probe is about 5 or less times slower than the inverse of the K_La ($^1/_{5K_{La}} > \tau$) [36]. Since this condition was satisfied for the cultivation system used, a differential equation was added to simulate the measured DOT by the probe (DOT^m) which takes into account the probe response time τ, where $K_p = {}^1/_\tau$. Therefore, the actual dissolved oxygen, which was solved algebraically was equal to the measured DOT, only after the elapse of the response time under equilibrium conditions.

The model (Equations A.1—A.14) was compiled as a single mathematical function (e_colimodel) and implemented in Matlab® R2016a. The model was integrated with ode15s solver and parameter estimation was done with the *fmincon* optimization routine in Matlab, using the interior-point algorithm.

Appendix to Thesis

i. General Methods and Protocols

a. Plaque Test for Bacteriophage Contamination in *E. coli* Cultivations

STEPS

1. Take 1 ml sample from the culture you suspect to have bacteriophages, into a sterile Eppendorf tube.
2. Centrifuge the sample at 15000 rpm for 5min, 4 °C.
3. Transfer the supernatant into another sterile Eppendorf tube under the laminar hood.
4. Filter the supernatant with a sterile 0.22 µm filter under the laminar. The filtrate is the '*culture medium*'.
5. Grow another culture of *E. coli* to OD_{600} = 5 (in principle, any OD should work). This may be wild type or any other strain, *but make sure it is growing very w*ell, without contamination. This will be referred to as '*test culture*' in this protocol.
6. In four sterile Eppi's, labelled A, B, C and D, do the following:
   - in Eppi A, mix 100 µL of *culture medium* with 900 µL of *test culture*. This forms 10x dilution of the phage-laden medium. Vortex gently or pipette up and down to mix.
   - to Eppi B, add 100 µL from Eppi A and 900 µL of *test culture*. This forms 100x dilution of the phage contaminated medium. Vortex gently or pipette up and down to mix.
   - nto Eppi C, add 10 µL from Eppi B and 990 µL of *test culture*. This forms 10000x dilution of the phage contaminated medium. Vortex gently or pipette up and down to mix.
   - to Eppi D, add 10 µL from Eppi C and 990 µL of *test culture*. This is 10^6 x dilution. Vortex gently or pipette up and down to mix.
7. Pipette 100 µL from Eppi A and spread this evenly on an LB agar plate (no antibiotic please!!).
8. Repeat Step 7 for Eppis B, C and D on different LB agar plates
9. Incubate at 37°C and look at the plates every 2 hours, up to 8 hours.

Observations and Conclusions

1. When there is bacteriophage contamination in the culture medium, only the plates with highly-diluted culture medium (10^6 or 10000x) will show growth. The plates that show growth will also show '*needle-tip spots'* in the growth layer, i.e. the culture does not cover the growing area evenly. The needle-tip spots are plaques of bacteriophages.

2. Conversely, when there is growth on the plates, i.e. all dilutions, then there may be no bacteriophage contamination.

What to do if you have phage contamination

1. Change/re-sterilize all solutions
2. Observe the highest standards to maintain sterility

b. Inclusion Body Purification using Bugbuster Reagent

Much of the experimental work in this thesis revolved around the influence of concentration gradients on recombinant proinsulin produced in *E. coli*. Since the product is accumulated intracellularly as inclusion bodies, the protocol below for inclusion body extraction from whole cell pellet was applied in all cultivations involving the recombinant strain, to purify, quantify and further analyse the product. The protocol is the same as the online version from Novagen, except for the enbolded entries where changes were made for this specific work.

1. Harvest cells from liquid culture by centrifugation at **15,000 x g for 5 min** using a pre-weighed centrifuge tube. Decant the supernatant and determine the weight of the wet cell pellet.
2. Resuspend the cell pellet in room temperature Bugbuster, using 5 ml reagent pre gram of wet cell paste.
3. Add **25 U benzonase and 1 kU rLysozyme** for each ml of Bugbuster reagent used to reduce the viscosity of the extract.
4. Incubate the cell suspension on a shaking platform or rotating mixer at slow speed for **20 min** at room temperature.
5. Remove the soluble fraction by centrifugation at 15,000 x g for 10 min at 4 degrees C. Save the pellet.
6. Resuspend the pellet in the same volume of Bugbuster reagent that was used to resuspend the original pellet. Complete resuspension by gently pipetting up and down and vortexing is required.
7. Addition **of 1 kU/ml of Bugbuster reaget is** recommended, but was not added in this work.
8. Incubate at room temperature for 5 min, centrifuge at 15, 000 x g for 5 min and save the pellet.
9. Resuspend the inclusion bodies in 1:10 diluted Bugbuster reagent (10 ml per gram of original cell paste), mix by vortexing and centrifuge as in Step 8.
10. Repeat Step 9 twice, for a total of 3 washes with the 1:10 diluted Bugbuster solution.
11. After the final wash, remove all supernatant by carefully pipetting all liquid, and store the inclusion bodies at -80 °C for further analysis.

ii. Cultivations in Minibioreactors: Extra Data and Figures

a. Structure of the minibioreactor cultivation platform

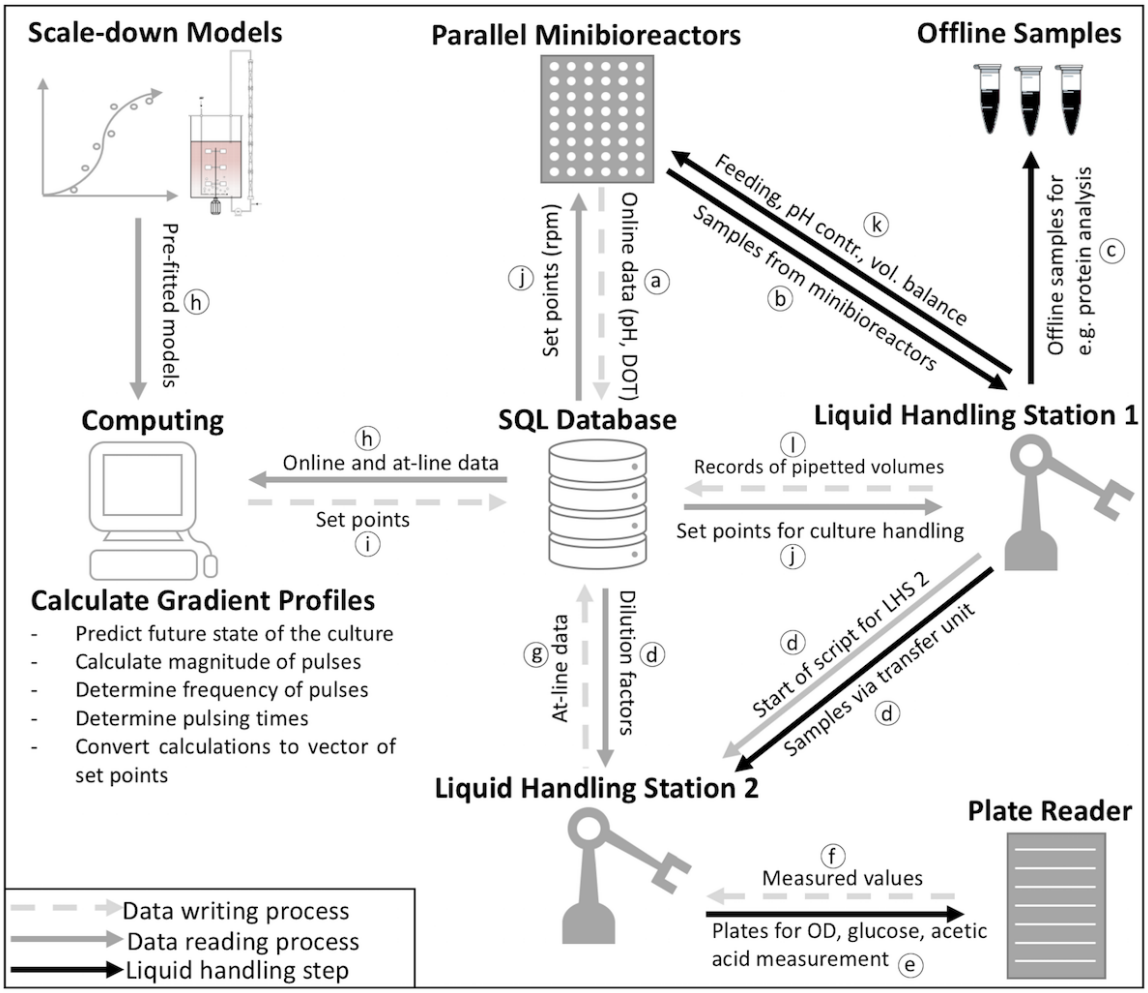


Figure A.1 Schematic representation of the cultivation platform of the minibioreactor scale-down system showing the data exchange routes via the centralized database—DB (iLab database) with end-to-end communication over the whole robotic platform. The circled numbers show the sequence of events that are required to calculate the concentration gradients during a scale-down cultivation. The system is set-up to use pre-validated models of the strain and the scale-down system, thereby enabling the accurate description of the proposed gradients in the large-scale bioreactor.

b. Pulse feed profiles

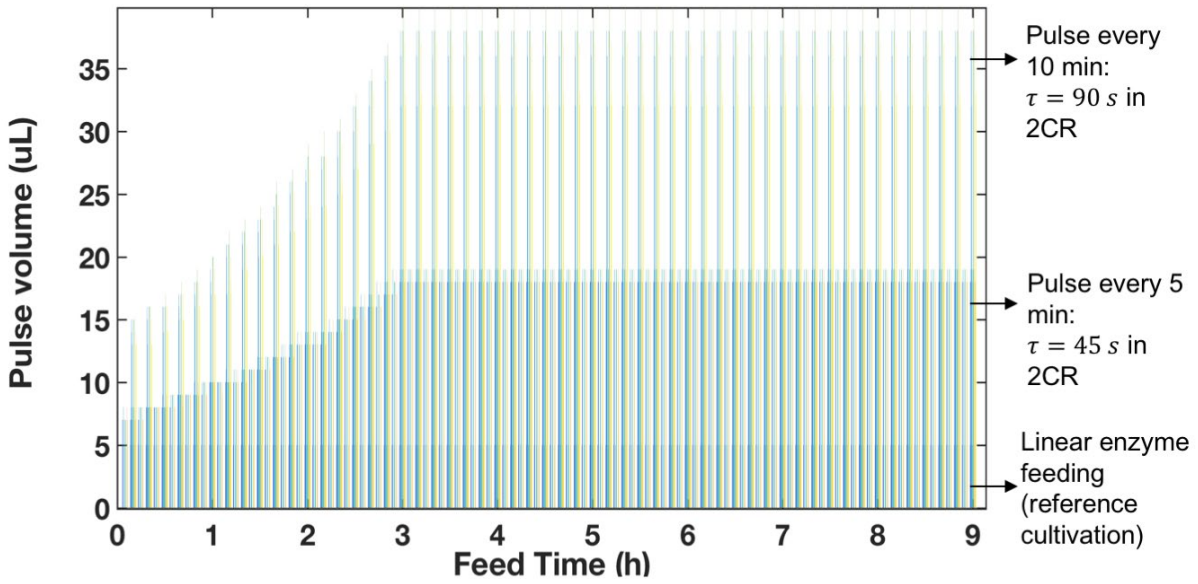


Figure A. 2 Model-derived feed profiles for minibioreactors at the end of batch phase in minibioreactor cultivations. Shown here are the feed profiles for 10 minute pulses, 5 minute pulses and enzymatic glucose addition. However, the enzymatic addition to the reference cultivations was stopped at the end of the exponential fed-batch phase (see Figure A.3 below).

c. Glucose release profile of Enbase system in reference cultivations

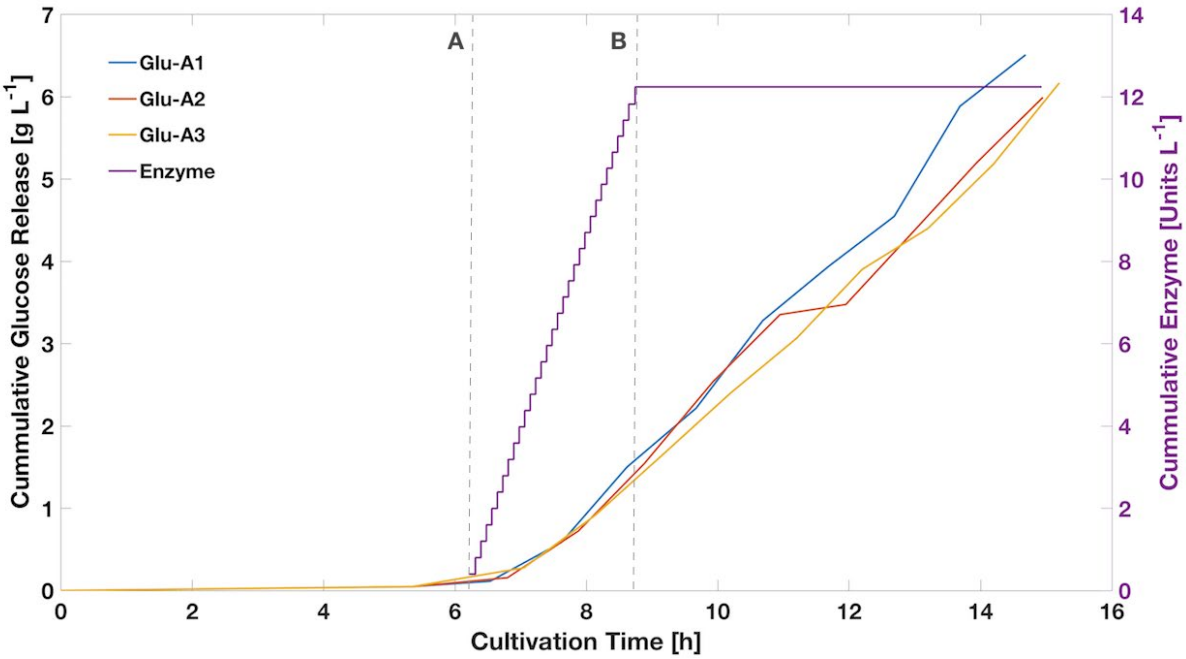


Figure A.3 Glucose release profile of reference cultivation in minibioreactors. The three bioreactors (A1—A3) were treated the same as the reference cultivation, except that these bioreactors were not inoculated. Tehrefore, the glucose released by the intermittent addition of enzyme was the same as the glucose released in minibioreactors B1—B3. The enzyme addition was done every 5 minutes,adding 5 µL of 800 U/L stock solution of amylase every 5 minutes until the end of the

exponential feeding phase. Dashed line A = start of fed-batch phase, Dashed line B: end of exponential feed fed-batch phase and beginning of protein induction. However, comparing the total amount of biomass obtained in the fed-batch phase of the reference cultivation to the glucose released in the non-inoculated bioreactors, it seems the glucose release rate was lower in the non-inoculated bioreactors. This characteristic of the Enbase glucose release system may be explained by the inhibition of the enzyme by the released glucose, which stays in the medium in the case of the non-inoculated bioreactors. Since the glucose in the inoculated bioreactors is readily consumed upon release, this inhibition is minimal, possibly leading to a higher overall glucose release in these bioreactors.

d. Quantification of recombinant protein (proinsulin)

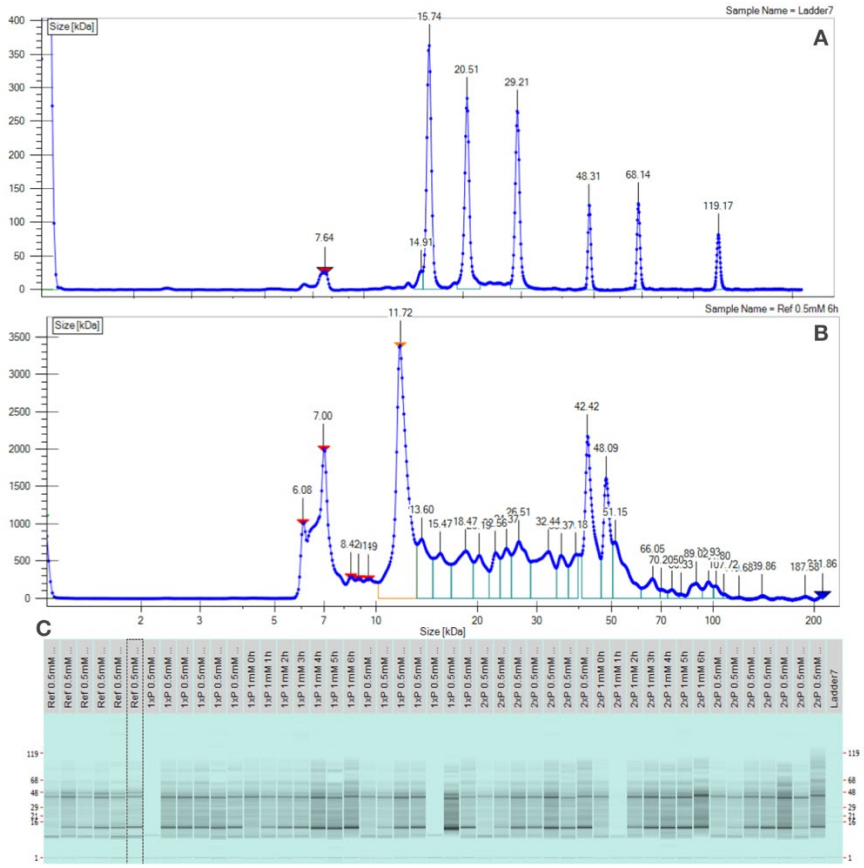


Figure A.4 Results of high throughput protein analysis with the technique of capillary gel electrophoresis using the Labchip GX II Touch (Perkin Elmer). *Top*: Electropherogram of a sample from *E. coli* cultivation, 6 hours after induction, showing the peak of the recombinant proinsulin (orange arrow) at 11.72kDa. The area under each peak is integrated and compared to peak area of protein standards, by which the quantity of each protein band can be determined. *Bottom*: virtual gel of 48 samples that were quantified in parallel, showing the band of the proinsulin (red arrows). Sample names are as follows: Ref—samples from reference cultivation, without glucose pulses; 1xP—samples from 5 min glucose pulse cultivations; 2xP—samples from 10 min glucose pulse cultivations; 0.5mM and 1mM refer to the IPTG concentration used for protein induction in the respective cultivation; Enp—referes to cultivations with background glucose release.

e. Visualisation of inclusion body formation during cultivations in minibioreactors

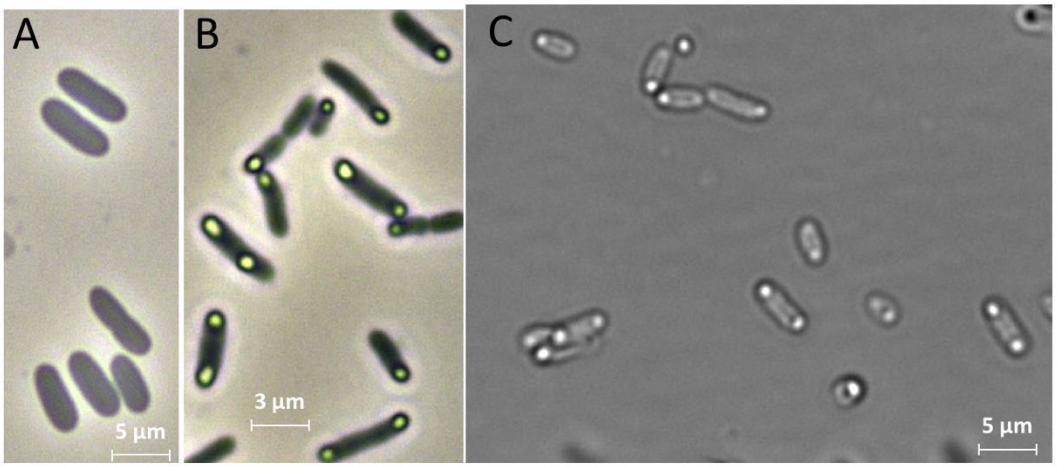


Figure A.5 Microscopic pictures of *E. coli* during cultivations in the minibioreactors, showing the accumulation of inclusion bodies. A—before induction in the reference cultivations; B—3 hours after induction in the reference cultivations; C—5 hours after induction in the reference cultivation. It is apparent from the pictures that the size of the inclusion bodies are not increasing very much with time, but as time goes on, each cell is accumulating inclusion bodies in two loci, compared to the early hours after induction.

f. Misincorporation of non-canonical amino acids into proinsulin

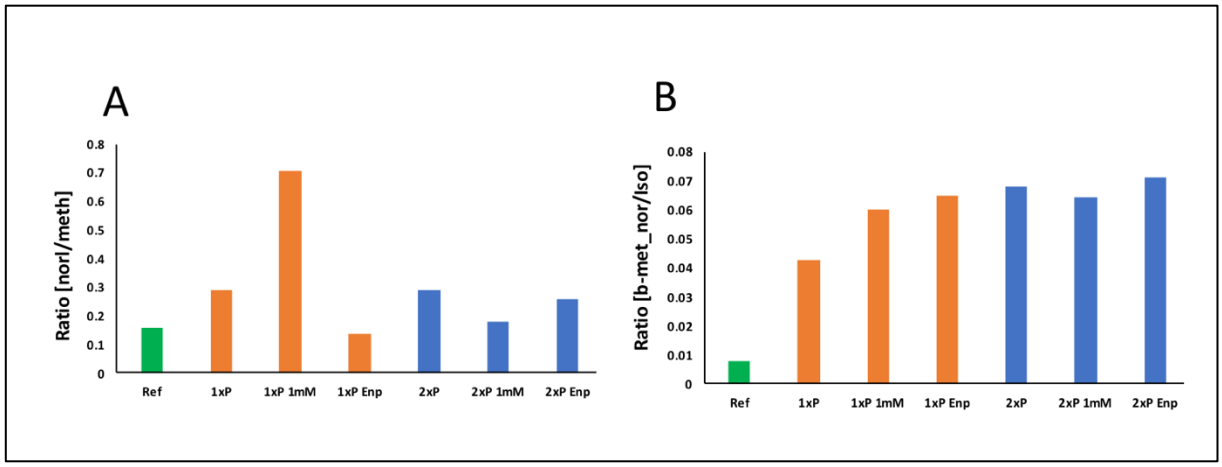


Figure A.6 Ratio of the non-canonical amino acid to their canonical forms in the purified inclusion bodies, for the various scale-down cultivation conditions investigated in the minibioreactors. A—norleucine to methionine ratio, B—beta-methyl norleucine to isoleucine ratio. In all ratios, the reference cultivation has lower ratios (lower misincorporation rates) than the cultivations where the cultures were subjected to glucose gradients. Notable here is low misincorporation of norleucien for methionine in the background glucose feeding cultivations. However, since the number of methionine and isloleucine residues in the proinsulin are very few, only the results of norvaline replacing leucine were used for concrete conclusions. 1xP—samples from 5 min glucose pulse cultivations; 2xP—samples from 10 min glucose pulse cultivations; 0.5mM and 1mM refer to the IPTG concentration used for protein induction in the respective cultivation; Enp—referes to cultivations with background glucose releas