Diplomarbeit

Experimentelle Charakterisierung & Simulation der Produktbildungsdynamik eines induzierten *Escherichia coli* Prozesses

ausgeführt zum Zwecke der Erlangung des akademischen Grades Diplom-Ingenieur (Dipl. Ing oder DI) in

Verfahrenstechnik

eingereicht an der TU Wien, Fakultät für Maschinenwesen und Betriebswissenschaften, von

Philipp Pably, BSc
Mat. Nr.: 01616237

unter der Leitung von
Univ.Prof. Dipl.-Ing. Dr.techn. Christoph Herwig

und der Assistenz von
Projektass. Don Fabian Müller, MSc

Institut für Verfahrenstechnik, Umwelttechnik und technische Biowissenschaften, E166

Wien, März 2023

Unterschrift

Diploma Thesis

Experimental characterisation & simulation of the product formation dynamics of an induced *Escherichia coli* cultivation

carried out for the purpose of obtaining the degree of Master of Science (MSc or Dipl. Ing. or DI) in

Chemical and Process Engineering

submitted at TU Wien, Faculty for Mechanical and Industrial Engineering, by

Philipp Pably, BSc
Mat. Nr.: 01616237

under the supervision of
Univ.Prof. Dipl.-Ing. Dr.techn. Christoph Herwig

and assistance of
Projektass. Don Fabian Müller, MSc

Institute of Chemical, Environmental and Bioscience Engineering, E166

Vienna, March 2023

Signature

Declaration of Authorship

I declare in lieu of oath, that I wrote this thesis and performed the associated research myself, using only literature cited in this volume. If text passages from sources are used literally, they are marked as such.

I confirm that this work is original and has not been submitted elsewhere for any examination, nor is it currently under consideration for a thesis elsewhere.

I acknowledge that the submitted work will be checked electronically-technically using suitable and state-of-the-art means (plagiarism detection software). On the one hand, this ensures that the submitted work was prepared according to the high-quality standards within the applicable rules to ensure good scientific practice "Code of Conduct" at the TU Wien. On the other hand, a comparison with other student theses avoids violations of my personal copyright

City and Date *Signature*

Acknowledgements

This work was conducted within the COMET Centre CHASE, funded within the COMET Competence Centers for Excellent Technologies programme by the BMK, the BMDW and the Federal Provinces of Upper Austria and Vienna. The COMET programme is managed by the Austrian Research Promotion Agency (FFG). The project was done in cooperation with Festo SE & Co KG and I want to thank Daniel Wibbing and Florian Zieker for their support.

I would like to express my sincerest gratitude to my thesis supervisor Prof. Christoph Herwig for his guidance and help throughout this research work. A special thank you is due to Fabian Müller for introducing me to all the experimental work, teaching me the ways of bioprocess modeling and providing valuable feedback and advice in the course of this project and beyond. I hope our findings will help your scientific career and contribute to your successful PhD thesis. Furthermore, I want to express my thankfulness to Julian Kager for his path-setting input and guiding influence, particularly in helping to shape my results into a scientific essay.

I want to further extend my gratitude to all the people associated with the bioprocess engineering research unit for their extended support and the friendships formed, while sharing lab space, equipment and occasional frustrations. Thank you also to my study colleagues and friends, with out you the time at this university would by much harder than it already is.

Thank you Misiu for listening to my hardships and being my personal anchor in the storm, even if it was just one in a teacup.

I want to dedicate this work to my parents, without whos support I would not have been able to carry through with my academic education. To my mother who has been cheering on to my success since I can remember and to my father whos influence and example has lead me onto this path.

Abstract

In the biotech industry, the optimization and the control of fermentation processes are essential in order to repeatedly meet quality standards and to increase competitiveness. Process models are the key to success as they provide a great summary of the available knowledge and many areas in engineering capitalize on them, from process development to monitoring and advanced control strategies. Traditional approaches utilize empirical models by including technical process parameters as the describing variables, but a shift towards mechanistic equations and physiological descriptors comes with several advantages. The cell itself as the catalytic entity in bioprocesses becomes the focal point and is thereby considered more thoroughly throughout up-scaling or process transfer efforts. The goal of this master thesis was to explore the influence of such physiological variables on the product formation dynamics of an *Escherichia coli* process and mould the findings into a novel mechanistic model.

The investigated organism expresses the recombinant enzyme L-Lactatedehydrogenase upon induction and is prone to product aggregation into biochemically inactive inclusion bodies (IBs). The ambition was to comprehensively describe the protein formation, while also considering the viability of the biomass. As a starting point, a previously established state-space model for growth and death of the organism was extended with equations for total product and IB formation. Several experiments were conducted to obtain data, which was used to find suiting kinetic expressions and estimate the model parameters. Following the guidelines for Good Modeling Practices (GMoP), the resulting structures were analysed for their structural and practical identifiability. The final model described the total productivity through a Monod term driven by the specific substrate uptake rate and was combined with a Haldane inhibition term that used the specific metabolised substrate as an indicator for the cell age. The protein localisation in form of the specific IB formation was expressed through a Moser kinetic utilizing the specific total productivity. Although characterized as overall ill-conditioned, the resulting model was identifiable and tracked the production dynamics with high accuracy, achieving $NRMSE$ values of 0.094 for the overall product amount and 0.153 for IBs. It does so by implementing a mechanistic model describing both, the total product formation as well as the accumulation into IBs using physiological descriptors only and considers the influence of the cell viability explicitly. To the knowledge of the author, this holistic approach and the achieved results pose as a novelty within the described field of application.

Furthermore, the developed model identified the specific substrate uptake rate as an influential factor on the efficiency of the product formation and the tendency to form IBs. In a simulation study this was used to optimize the production process for a variety of objective functions. Through numerical determination of the optimal substrate uptake rate trajectory, the product titer was maximized in-silico and the share of IBs was found to be adjustable within 9 to 29 %. After

implementing according economical functions the space-time yield could be improved by 17 % and the substrate-to-product yield by 50 %, compared to the maximum titer case. Apart from the resulting opportunities in process optimization and development, this model holds potential to be utilized as a soft-sensor for the productivity and further in advanced control strategies like Model Predictive Control (MPC). Simple online-measurements are sufficient for these implementations and their capabilities serve as a strong motivator for industrial applications. Finally, they enable knowledge based process monitoring and quality control, which falls in line with recent initiatives of regulatory bodies pushing for more Quality by Design (QbD) and advanced Process Analytical Technologies (PAT).

Kurzfassung

Die Optimierung und Kontrolle von biotechnologischen Prozessen sind zentrale Aspekte für industrielle Anwendungen um dauerhaft Qualitätsstandards einhalten zu können und konkurrenzfähig zu bleiben. Der Schlüssel zum Erfolg sind dabei Prozessmodelle, da sie eine hervorragende Sammlung des vorhandenen Wissens darstellen und viele Vorteile für Prozessentwicklung, -überwachung und -regelung bieten. Traditionelle Ansätze setzen dabei auf empirische Modelle und beschreiben den Einfluss von technischen Prozessparametern. Die Verwendung mechanistischer Gleichungen und physiologischer Prozessvariablen birgt jedoch großes Potential. Dabei rückt die Zelle als zentrale katalytische Einheit des Bioprozesses in den Fokus des Modells und wird bei Scale-up und Prozesstransfer umfassender berücksichtigt. Das Ziel dieser Diplomarbeit war die Untersuchung des Einflusses von physiologischen Variablen auf die Produktbildungsdynamik eines *Escherichia coli* Prozesses und die Entwicklung eines mechanistischen Modells aus den gewonnenen Erkenntnissen.

Der untersuchte Organismus produziert das rekombinante Enzym L-Laktatedehydrogenase nach Induzierung und ist anfällig für Bildung von biochemisch inaktiven Einschlusskörpern (IBs). Zur Identifizierung dieser Proteinbildungsdynamik und des Einflusses auf die Aggregation zu IBs wurden mehrere Experimente mit variierenden Substratfütterungsraten durchgeführt. Als Ausgangspunkt für die Modellbildung wurden die Ergebnisse zusammen mit einem bereits vorhandenen Zustandsraummodell für Zellwachstum und -sterben verwendet. Dabei wurden verschiedene kinetische Gleichungen getestet, um sowohl die gesamte Proteinproduktion als auch die IB-Bildung zu beschreiben. Den Empfehlungen der Good Modeling Practices (GMoP) folgenden wurde zur Auswahl der geeigneten Gleichungen sowohl die Fähigkeit die Messdaten zu replizieren, also auch die strukturelle und praktische Identifizierbarkeit der resultierenden Modellstruktur beurteilt. Das finale Modell beschreibt die spezifische Gesamtproduktivität in dem es die spezifische Substrataufnahmerate als treibende Variable und die spezifische metabolisierte Substratmenge inhibierend als Indikator für das Zellalter enthält. Der Anteil des Produkts, welcher zu IBs aggregiert wird über eine separate spezifische Rate beschrieben, welche die spezifische Gesamtproduktivität als Variable einer Moserkinetik einsetzt. Die entwickelte Modellstruktur wurde insgesamt als schwach konditioniert, aber identifizierbar eingestuft und beschrieb die Messdaten mit hoher Genauigkeit. Dabei wurden $NRMSE$-Werte von 0.094 für die gesamte Proteinmenge und 0.153 für die IB-Menge erzielt. Das entwickelte mechanistische Modell verwendet nur physiologische Variablen und berücksichtigt die Viabilität der Zellmasse explizit. Es beschreibt dabei sowohl die Bildung der gesamten Proteinmenge als auch die Unterscheidung in lösliches und aggregiertes Produkt. Dem Wissen des Autors zur Folge stellt dieser holistische Ansatz und die dabei erzielten Resultate eine Neuheit auf diesem Anwendungsgebiet dar.

Das entwickelte Modell zeigte, dass die spezifische Substrataufnahmerate einen wesentlichen Einfluss auf die Effizienz und die Aggregationsneigung der Proteinbildung darstellt. Innerhalb einer Simulationsstudie wurde daraufhin der Produktionsprozess durch numerische Ermittlung entsprechender Trajektorien für die Substrataufnahmerate in-silico optimiert. Über die Wahl geeigneter Zielfunktionen konnte der Produkttiter maximiert und der IB-Anteil des Produkts innerhalb von 9 bis 29 % variiert werden. Durch das Einführen von ökonomischen Kennzahlen als Optimierungsziele konnte außerdem die Raum-Zeit-Ausbeute um 17 % und die Substrat-Produkt-Ausbeute um 50 % verbessert werden, verglichen mit der Prozessführung für maximalen Produkttiter. Die Ergebnisse zeigen das Potential des Modells für Prozessoptimierung und -entwicklung, aber auch die Möglichkeit für den Einsatz als Softsensor für die Produktivität. Die Implementierung verwendet einfache, bereits vorhandene Online-Messungen und kann in weiterer Folge für fortgeschrittene Regelungsstrategien verwendet werden, wie Model Predictive Control (MPC). Diese Anwendungsgebiete zusammen mit den Möglichkeiten für wissensbasierte Prozessüberwachung und Qualitätskontrolle dienen als starker Motivator für industrielle Umsetzungen. Der Einsatz solcher mechanistischer Modelle deckt sich dabei auch mit den Empfehlungen der zuständigen Aufsichtsbehörden bezüglich Quality by Design (QbD) und modernen prozessanalytischen Technologien (PAT) für Bioprozesse.

Contents

1 Introduction

There is an on-going outcry for a change towards sustainable and more environmental friendly practices over all industrial sectors. The dependency on finite fossil resources, together with the negative impact they have on our planet and society is evident and changes are necessary [113, 79]. Therefore, political figures and many economic sectors push towards greener and more efficient production processes. One of the approaches is the transition from fossil to renewable sources such as sun, wind or biomass [6, 70]. The later is often combined with the utilization of microorganisms that convert readily available natural and renewable resources into a variety of products, ranging from bulk chemicals [38] up to highly complex human proteins [66]. Due to the fast cycle times of the used raw materials they appear to be a valid and steadily growing industry in replacing harsh chemical methods [1, 2].

These bioprocesses are convincing alternatives that operate at rather mild conditions and mostly metabolise simple substrates like glucose and other monomeric sugars [53]. More recent developments even push away from inputs that are direct competitors to the food market towards sources that are often regarded as waste streams such as lignocellulosic biomass fractions that arise from processing biological raw materials [2]. The key of those processes is the metabolism of the organisms that utilize a complex reaction network, where a manifold of enzymes act as bio-catalysts and enable the transition of simple molecules into more complex ones [4]. The manipulation of these metabolic pathways through modern genetic modifications has accelerated the improvement of common host organisms and creation of new strains for challenging tasks. This field of gene engineering has opened the door for bio-based methods into many industries and a vast amount of research is going into the transition of traditional production into bioprocesses [82].

This work contributes to the transformation of established processes towards more sustainability and efficiency by increasing the understanding of the underlying mechanisms of the utilized microorganisms. Ultimately, a tool in form of a productivity model for the examined strain was established and its potential for process development and optimization was shown.

1.1 *Escherichia coli* as cell factory

Bioprocesses use organisms like bacteria, fungi, yeast or mammalian cells for the synthesis of a variety of products. One of the most established representatives in biotechnological applications is the gram-negative *Escherichia coli* (*E. coli*), which was also the investigated microbe of this work. It serves as the cell factory for 24 % of all biopharmaceuticals approved by the FDA and the European Medicines Agency from 2004 to 2013 [69, 10]. While the trend shifted more

towards mammalian cells in recent years, prokaryotic cells still play a major role, especially for the production of biosimilars [110].

E. coli was utilized early on for the expression of therapeutic proteins, being the host of the first licensed drug produced using recombinant DNA technology in 1982. The product was insulin for the treatment of diabetes, which replaced the previous method of extraction from pig pancreases [102, 10]. Its genome was first sequenced already in 1997 and since then poses as one of the most researched and documented microorganisms in biotechnology [15]. This contributed to the development of a variety of efficient and precise genetic tools, which allowed the adaptation of *E. coli* and its metabolic pathways to produce a manifold of different non-native compounds [56]. This versatility stands as the biggest advantage of this microorganism and makes it a metabolic generalist. Combined with its quick growth and inexpensive requirements towards the carbon source and other nutrients, the bacteria drew the interest of many industries as a cost-effective cell factory [107, 102, 69].

Substrates and products of *E. coli* production

E. coli was genetically transformed to grow on a wide palette of substrates. Next to simple mono- and disaccharides like glucose, more complex carbohydrates were made available for industrial processes. This enabled the growth on hemicellulose, which is generally regarded as a waste stream in many industries and poses no competition to the food market [92, 16]. Furthermore, a metabolic pathway utilizing CO_2 was successfully integrated, allowing carbon dioxide fixation through microbial fermentation [34].

The list of compounds produced with *E. coli* after genetic optimization is comprehensive and reaches from small commodity chemicals to secondary metabolites spanning several kDa in size . Organic acids, alcohols and even amino acids can be synthesized through this cell factory with high titers reaching more than $100\,\mathrm{g\,L^{-1}}$ and very favourable substrate-to-product yields [102, 90]. As examples succinic acid, named on of the top value added chemicals from biomass by the US Department of Energy, or Hydroxycarboxylic acids, precursors to lactones which are used in a range of pharmaceutical products, can be mentioned here [117, 29].

A special attention should be brought to the production of recombinant proteins. Metabolic engineering opened up the possibility of producing complex proteins with *E. coli*, that demand a number of mechanisms not native to this microorganism [41, 56]. Nowadays, numerous systems for initiation of transcription and translation for protein expression were implemented in different strains. Often Isopropyl β-D-1-Thiogalactopyranoside (IPTG) inducible promoters are incorporated, but temperature shifts or other mechanisms are also available as less toxic and cheaper induction alternatives [105, 56]. After synthesizing the product it can be located in the periplasm, intracellularly or released to the extracellular media by installing appropriate transport and secretion mechanisms into the cells. The conditions in the periplasmic space are often the most favourable for the proteins, but not always an option. In the periplasm degradation of the protein is reduced and the oxidizing environment helps with the correct folding and formation of disulfide bonds [95, 59].

Challenges with *E. coli* as host cell

Disadvantages of *E. coli* are its sensitivity towards bacteriophage infections which has to be tended to in industrial processes [11]. Further, the endotoxic nature of the lipopolysaccharides in the outer membrane of the cells can be problematic when producing pharamceuticals used for humans or animals [61]. A struggle for the metabolic engineering of *E. coli* is its general inability to provide complex protein glycosylation or other post-translational modifications due to the absence of adequate mechanisms in the cell [55, 81]. Although it was managed to import a N-linked glycosylation machinery from a different organism, more complicated mechanisms are often not possible [54]. For those tasks animal cells are usually used in the pharmaceutical industry, which have a number of downsides themselves. Among those are slower growth, more fragility and higher demands throughout the cultivation process compared to *E. coli* [14].

Another issue special to the production of recombinant proteins in *E. coli* is Inclusion Body (IB) formation, where the expressed poly-peptides tend to form nonfunctional aggregations inside the cell. Several circumstances can lead to the miss-folding of the amino-acid chains and cause their precipitation as IBs [106, 81]. Their characteristics and challenges for the industrial processes are discussed further in the next chapter.

1.2 Inclusion bodies

E. coli and other prokaryotic organisms are prone to the formation of IB, when expressing recombinant proteins. They are amorphous aggregates mostly consisting of the encoded target molecules, which are held together by covalent and non-covalent bonds [55]. IBs are relatively homogeneous in composition, but can differ in their conformations to the native form. The remainder is typically made up of traces of proteolytic fragments of the recombinant protein, membrane proteins, phospholipids and nucleic acids [106, 21, 43]. IBs usually occur in the polar region of the cell and are of porous structure, spherical or rod-shaped and have diameters in the magnitude of a few µm. Most of the time, they are located in the cytoplasm, but if the organism contains a secreting machinery these aggregates were also observed in the periplasm [17, 78, 106]. IBs are generally inactive, but especially when they are formed from recombinant enzymes they are known to show some enzymatic activity. Their porous nature then allows substrates to reach active areas within the aggregate [48, 32].

Kinetics and reason of formation

IB formation in prokaryotic organisms occurs during over-expression of heterologous proteins [55]. Several factors are assumed to play a role in this process. First of all, the protein folding process can be overwhelmed leading to miss-folding or intermediate states of the molecules. Further, the bacteria might not be able to perform the post-translational modifications needed to ensure a correct folding or lacks the appropriate modulators required. As a consequence expressed protein chains that do not not reach their native form or modulator in time, are prone to form IBs [12, 42]. Hydrophobic interactions between exposed polypeptide chains, the reducing environment of the cytoplasm and the improper formation of disulfide bonds are also contributing towards their formation [62, 93, 106]. Small changes in the amino acid sequence can have significant effects on

the solubility and tendency for aggregation. Additionally, existing IBs can work as seeds and even trigger correctly folded proteins to aggregate [93, 19]. The rate of agglomeration is often much higher than for proper folding, especially when the protein reaches high cellular concentrations [12]. But the organism is also able to reverse this aggregations, as it was observed that IBs are in dynamic transition between soluble and insoluble fractions to a certain extent. Further, cells also have the ability to disintegrate some of the aggregates inside the cytoplasm, when they are not occupied with protein synthesis [21, 20].

Apart from the physiological susceptibility to IB formation, process parameters have a strong influence too. Temperature and pH of the culture media and the inside the cells affect the conformation, stability and hydrophobic interactions of the proteins [93, 106]. High metabolic stress during the protein expression phase favors IBs and is influenced by the choice of substrate, the feeding rate and the strength of induction, which is a result of the incorporated mechanism and inducer concentration [88, 48].

Downsides of IBs

Traditionally, IBs were regarded as a drawback of heterologous protein expression in prokaryotes, since they diminish the overall yield of active protein and caused some processes to be economically infeasible [106]. This lead to efforts in genetic engineering and process development to reduce IB formation, where four main approaches can be distinguished [106, 93, 62, 12].

- Adjusting the sequence and post-translational modifications of the heterologous protein
- Improving the availability and efficiency of folding modulators and chaperons
- Improving the translational machinery to result in less miss-folded proteins
- Adjusting the physio-chemical conditions to reduce conformational and metabolic stress

Chaperone proteins can either shield intermediates from aggregation or even disaggregate IBs and release the protein again to fold correctly. Additionally, proteases help with in-situ digestion and reduction of such material [24, 103, 12]. Process parameters mainly influence the last factor listed where they play a role in manipulating the overall metabolic stress on the cells and the physical environment of the expressed protein. Lowering the temperature has shown good results in reducing miss-folding, as it slows down transcription and translation but also reduces the strength of hydrophobic interactions, where as changes in pH have similar effects [12, 47].

Utilization of IBs and their processing

In more recent developments, refolding procedures have been identified that transform IBs into correctly folded native proteins. Under this considerations, the formed aggregates have the benefit of being easily recoverable with high purity, as well as being more thermally and mechanically stable. They also show higher resistance against proteolysis and the high yield and relatively cheap production with bacterial systems in contrast to other organisms can be exploited. Their sturdiness is also beneficial for the necessary cell disruption step during downstream processing

[106, 42, 78, 55]. This leads to some processes being optimized for enhanced IB formation, which can be done by inverting the methods mentioned earlier for reducing aggregation [106, 12, 93, 62].

Processing of IBs as product starts with cell breakage and the separation of other components [51]. The recovered IB fraction then gets solubilized by a suitable dissolving agent, such as urea or guanidine hydrochloride [55, 8]. In the refolding step these straightened amino-acid chains are guided to fold into the native soluble protein and are ultimately concentrated and purified by a chromatography step [55, 51, 74]. The complexity of the process leading to the native form depends strongly on the protein at hand and the requirements of its folding mechanism [55, 106, 8].

This makes the design of the downstream challenging as unit operations become specific to the protein. Furthermore, the industry traditionally relies heavily on empirical methods resulting in laborious process development projects and low space-time-yields [60, 23]. The used unit operations often bring about huge liquid volumes and even cause partial fragmentation and agglomeration of the product, resulting in yield losses [71, 42]. More novel approaches try to act by Quality by Design methods and implement more mechanistic understanding, as well as transforming batch-wise production into a continuous downstream process [42, 67]. The application of process models can further help with improving the problems of refolding procedures or act as digital twins for online monitoring and product quality assurance [68, 31]. Mechanistic modeling can lead to soft-sensors and enable advanced control strategies to optimize the process with the given online measurements [68].

1.3 Role of models in bioprocesses and their control

Modeling plays an essential role in modern engineering and science, especially in the on-going digital transformation and automation of many industrial sectors. Process models are utilized in plenty ways, but the fermentation based and especially biopharmaceutical industry seem to lag behind and still rely strongly on testing and empirical development. Regulatory bodies like the U.S. Food and Drug Administration (FDA) and European Medicine Agency (EMA) are emphasizing advanced technologies and recommend initiatives such as Quality by Design (QbD) and Process Analytical Technology (PAT), which lead to an increased interest in modeling approaches. The intention is to use them not only for economical optimizations but also to increase process understanding, prediction and consistency [65, 85].

Types of bioprocesses models

When developing a new model for a process its goal and application should always be kept in mind. This decides the overall structure and most importantly the level of complexity needed. When focusing on the utilization of living organisms to produce a compound, the modeling of the metabolic activity is usually the center of attention [26]. The developed models are distinguished by different features. Empirical or statistical models fit mathematical equations to the observed data set using regressions and statistical analyses, without further regard to the physical or chemical relationships of the system. Their applicability is constrained by the range of the conducted measurements and they loose validity outside of them. Mechanistic models utilize existing knowledge of underlying mechanisms to create models that are potentially more universal

in their application. For that matter they use balances and natural laws and double as an excellent summary of the existing process knowledge [57, 33].

Bioprocess models are further characterized by whether they are segregated or structured. Segregation describes if the model differentiates between populations of cells with different properties that exist simultaneously during the process. In this way the model can account for several cell populations co-fermenting or with the help of continuous independent variables for the age of the cells. Structured models consider internal structural elements of a cell and balance them and their interaction separately. A model can be a mixture of those types, where the complexity increases along with the numerical effort to solve them. Therefore, it is a virtue to describe the process as simple as possible but with sufficient accuracy for the defined purpose [84, 33, 39].

In industrial applications models are usually built with a mixture of empirical and mechanistic relations, with a shift towards the latter when more process knowledge is available. Those hybrid models reduce the level of abstraction of data-driven approaches, decrease the data requirements and increase the extrapolation properties [57, 65]. They are mostly used in upstream processes for production models where they are kept rather simple and used for process design, control and optimization. Unstructured models are often implemented to describe the growth, substrate consumption and metabolite production using empirical kinetic equations. In the scientific field of bioprocesses, more complex models are used to test hypothesis in order to expand the process knowledge and fundamental understanding of metabolic mechanisms [80, 33, 39].

Challenges in modeling bioprocesses

Biological processes such as the cultivation of microorganisms are characterized by non-linear and highly interrelated dynamics that are under active research and many mechanisms are still not fully explained. Furthermore, cells are known to be very sensitive to their surroundings such as temperature, pH or nutrient availability [9, 108]. With many aspects influencing the course of a fermentation, decisions have to be made during the development about model complexity, which are detrimental to its later performance. Another challenging aspect is the limited amount and availability of online measurements and the time consuming analysis of off- or at-line samples, especially when the goal is the implementation in advanced control algorithms. The identification and parameterization phases of the model need quantified values for model outputs and inputs in a time resolved manner to capture the inherently non-linear dynamics of the cells [33, 9, 65]. This can lead to a cost and time intensive experimental effort, which is critical minding the pressure for short development cycles in the industry [80]. The introduction of a systematic model development methodology and orientation towards Good Modeling Practices (GMoP) should help with creating a more efficient process and assessing the quality and transferability of the results[33, 87].

Application of models

Models of an organism or process are developed for a manifold of reasons. Applications in the biotechnological industry usually try to optimize bioprocesses to enhance cell growth and increase productivity. Another field is within PAT to describe the influence of process variables on critical quality attributes (CQAs), that are difficult to measure directly. Stochastic models of the latter

can be used for QbD and create design spaces for biopharmaceutical processes to be used for real-time batch releases that reduce elaborate and time consuming testing in the laboratory [65, 86].

For process optimization and control, models can be used to consider the underlying dynamics of the organism and predict future changes and necessary actions. Meanwhile, industrial bioreactors still predominantly use PID controllers that ensure simple set-point tracking [86]. Typically used for critical process parameters (CPPs) such as temperature, pH or dissolved oxygen, which are maintained at a constant level or above a certain threshold [36]. For the substrate addition predefined strategies are applied, that are developed empirically or on scale-down screenings. These methods disregard the actual process state and complex information about the fermentation are not used in a feedback loop, which often results in sub-optimal manufacturing conditions or high batch to batch variability [36, 41]. Adaptive and model predictive controllers utilizing hybrid models are capable of enabling the direct control of specific physiological rates, given adequate online measurements. Substrate uptake or cell growth can be held to defined trajectories or optimized for user-defined objective functions, like maximizing the product titer or the space-time-yield [36, 58, 46]. When working with limited available online measurements, models are implemented as a state observer to estimate the missing values. These soft-sensors can also be used together with delayed offline analytics, which then trigger a recalculation that updates the estimation of the states as well as the model parameters [45]. Altogether, these algorithms help with consistent control of product quality and quantity, as well as batch reproducibility [65].

Finally, mechanistic or hybrid models are of great use in process development, as well as scale-up and scale-down. Together with modern Design of Experiment (DoE) approaches, they have the potential to reduce the development time and costs significantly. Instead of trial-and-error experiments and other empirical approaches, models can help finding non-intuitive operation conditions and help with transferability during scale-up for production [65, 76].

Bioprocess models for *E. coli*

Due to the popularity of *E. coli* as a cell factory for industry and academia, there are many attempts at modeling the bacteria with different levels of complexity. The simplest approaches often consider only a single variable like the logistic law with the cell mass or are purely empirical black-box models, disregarding effects of substrate limitation and other influences [84]. One step further are unstructured models using first-order elemental and degree of reduction balances to calculate theoretical yields and rates for steady-state conditions. When combined with empirical kinetics for substrate uptake or inhibition, models with good predictive power are created with only a minor increase in complexity [33, 63]. There are also many scientific contributions towards segregated and structured models, which are very detailed representations of the biochemical reactions inside the cell. Metabolic flux analyses are used as an in-silico method to identify optimal pathways and possible optimizations for the genetic engineering of production strains [111]. For real-time applications such as monitoring and control of industrial production processes, simplified models are better suited as they are easier to develop, parameterize and execute [104].

Among unstructured and unsegregated models a big variety of models have been developed for the palette of available *E. coli* strains [107]. The influence of technological process parameters like temperature, pH or inducer strength on the cell growth and productivity in induced fed-batch

processes has been researched extensively. But in scale-up or process transfer, the equipment and their influences might change while the organism usually stays the same, which is why a focus on physiological descriptors was proposed by Reichelt et al [76]. They developed a model that accounted for the physiological state of the cells by describing a yield decay over the amount of metabolised substrate, instead of assuming it as constant. The model was implemented in a soft-sensor, using the online measurements of the off-gas concentrations and achieved good accuracy [77]. In a different work they connected the critical specific substrate uptake rate $q_{S,crit}$ of a Monod kinetic with time, temperature and the metabolic activity [75].

Kager et al. used a similar approach to model the production of an antibody fragment with a rahmnose-inducible *E. coli* strain. They implemented a yield reduction as a function of the metabolised substrate. The productivity was modeled with a Monod term describing the dependency on the specific substrate uptake rate combined with a Haldane term acting inhibiting as the amount of metabolised substrate increases. A further equation was implemented to model product loss through cell division. The implemented controller utilized the model to directly control the specific productivity through the feeding profile and showed promising results, as well as potential to stabilize and prolong the induction phase of the recombinant protein production [44].

For *E. coli* processes that show IB formation, most scientific results describe the dependency of quality attributes on technical process parameters. Slouka et al. showed the dependency of the IB titer, purity and bead size on temperature and pH, concluding that lowering both values during the induction phase has favourable results. They further empirically investigated the dependence on physiological parameters and correlated them to the substrate uptake but did not put forward a mechanistic model [88]. Kopp et al. experimentally produced a model describing the IB size dependent on the cumulative substrate uptake [48].

Gnoth et al. described a data-driven model in form of two trained Artificial Neural Networks (ANNs) to determine the specific growth rate and IB formation rate of an *E. coli* strain. They used the growth rate, the time after induction and specific product concentration as inputs to estimate the productivity. It was acknowledged that purely data-driven models need a substantial amount of measurements to perform well and the ANN might not be transferable to other products [37].

Wurm et al. on the other hand, investigated an *E. coli* process producing an antibody fragment and established a model with more mechanistic features. The strain was lactose inducible and the process utilized a two feed system with glucose and lactose. By adjusting the specific glucose uptake rate and lactose uptake rate together they were able to show a correlation with the properties of the resulting IBs and the total product [116, 115].

Finally, Hoffmann et al. should be mentioned for establishing a model describing the product in three different conformations. They modeled an initial translational product that was later split into the protein folded into its native form and IBs. The implemented kinetic for the specific production rate of hFGF-2 was designed as exponentially decreasing with the time after induction and further considered the growth rate [40].

1.4 Problem statement and aim of thesis

Overall it can be seen that many researches tried to model the product formation and certain aspects of the recombinant protein aggregation for bacterial processes. The advantages of using physiological descriptors over technical parameters are pointed out by the aforementioned authors, which are better mechanistic understanding and transferability of the process. Nevertheless, the focus was either on overall productivity or describing quality attributes of formed IBs, but not combining the kinetics into a single model. The few approaches in this direction that were identified used empirical equations, which limits their explanatory power. Finally, the examined works that considered the decline in metabolic activity during the process described it through a decrease of the yield coefficient and thereby disregarded the viability of the present cell mass.

Therefore, this work set out to build a novel mechanistic model for the production of Lactate dehydrogenase (LDH) and the accompanying IB formation of a provided *E. coli* strain. The goal was to use physiological descriptors for the production of the recombinant protein and provide a holistic model, readily usable for process development, soft-sensors and potentially process control. Furthermore, the inherited model for biomass growth also included a cell death kinetic, describing the viability of the bacteria in the broth and introducing a segregation-like dynamic. In order to find fitting model equations, experiments were carried out to identify variables that are connected to the product formation. The gathered data was applied to test kinetics and different hypothesis regarding the process dynamics. Using statistical tools and identifiability analyses, the established models were compared and verified before deciding on the final structure. Finally, the resulting model was applied by simulating different operating conditions and optimizing the substrate uptake rate for selected objectives.

To the knowledge of the author this work presents a novelty, as no similar model, combining overall product and IB formation using physiological descriptors and accounting for cell viability, could be identified for the expression of a recombinant protein like LDH in *E. coli* in the literature. The goals and demands towards this thesis are summarized once again below.

- Construct a mechanistic model for the total productivity and IB formation using physiological descriptors
- Include the influence of cell viability into the product formation dynamics
- Conduct cultivation experiments for model identification, parameterization and validation
- Apply statistical tools to verify the identifiability of the model structure
- Explore process optimization opportunities with the developed model

2 | Material & Methods

In the following sections all the applied theoretical and practical procedures are presented, such as details on the lab-scale cultivation experiments and wetlab analytics. The gathered data was then used to create a model with the methods given in the modeling sections of this chapter. Furthermore, all the analysis tools are introduced that were applied to compare, evaluate and explore the implemented kinetics and overall model structure.

2.1 Strain

For the scope of this work the recombinant *Escherichia coli* (*E. coli*) strain BL21(DE3) was used. The gram-negative bacteria is known for high replication rates, low acetate formation and can be grown on relatively cheap media [108, 5].

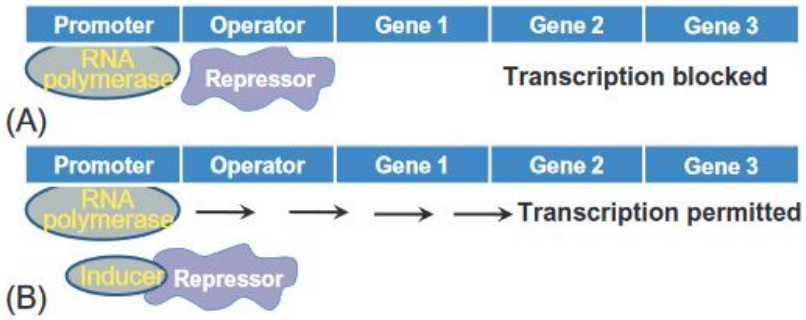

Figure 2.1: Principle of T7-Polymerase induction system (A) repressor blocks operator (lac-operon) (B) inducer Isopropyl β-D-1-Thiogalactopyranoside (IPTG) reduces affinity of repressor for operator gene [53]

The strain contains the T7-polymerase induction system that is activated by binding an inducer to the repressor of the lac operon and thereby changing its affinity. This causes the lactose repressor to be unable to block the operator and the lac-promoter then drives the expression of the T7 RNA polymerases. These bind to a T7-promoter and initiate the transcription of the heterologous genes that encode the production of Lactate dehydrogenase (LDH). An illustration of the principle can be seen in fig. 2.1. For the conducted experiments the induction was done by adding IPTG to the broth and the cells started to produce the enzyme thereafter [94, 108]. The addition of the induction agent ensued as a one time bolus injection to the reactor.

2.2 Cultivation

The five fermentation experiments were all performed with the same physical bioreactor setup and were carried out as three runs with two tanks operated simultaneously. Both were set up identically and the same conditions were applied except for the feeding-profile in the fed-batch phases.

2.2.1 Equipment

The overall setup consisted of two Labfors 5 bioreactors (Infors AG, Switzerland) with a total volume of 3.6 L. Each was equipped with a potentiometric pH sensor (Hamilton, Switzerland) and optical Dissolved Oxygen (DO) probes (Hamilton, Switzerland). Their signals were fed into and used by the local control tower of the reactor system to manipulate the process. The tower used the *INFORS HT Parallel Bioreactor* Software, which controlled the cultures temperature through the water jacket, powered by a Alpha RA8 heat exchanger (Lauda Dr. R. Wobser GmbH & Co. KG, Germany), providing 1.5 kW heating / 0.225 kW cooling. Furthermore, several peristaltic pumps were used for adding or removing liquids. The acid and base was supplied with the pumps attached to the Labfors 5 system, while the substrate solution was fed by a calibrated Preciflow pump (Lambda Instruments GmbH, Switzerland). The samples were drawn through a custom software that accessed a pump and an automatic FC 203B Fraction Collector (Gilson Inc., USA) to distribute the samples into their respective glass tubes. Those were positioned in a cooling rack and constantly cooled to a temperature between 4 and 10 °C.

The DO-level in the fermentation broth was controlled by the bioreactor system through the gas inflow passing a L-sparger and the stirrer speed (Infors AG, Switzerland). The bioreactor system also measured and recorded the pressure inside the reactor. The in-flowing gas source was manually switched upon demand from air provided by the in-house compressed air system to additionally added oxygen from a gas bottle. The reactor, as well as all the added substances were put on scales and their weight was monitored throughout the process. The off-gas of the bioreactors was first cooled by a heat exchanger to condensate possible volatile components and then led to an online gas analyzer measuring the CO_2 and O_2 content of the off-gas (BlueInOne, BlueSens, Germany). The flow velocity was measured for the in-flowing gas of the reactor.

2.2.2 Media & Chemicals

All experiments used the same media compositions and differentiated between the preculture and batch medium, as established by Müller et al. [64]. They are based on the DeLisa minimal medium [28] but use different concentrations for glucose as seen in tab. 2.1. They contained 13.3 g L^{-1} KH_2PO_4, 4.0 g L^{-1} $(NH_4)_2HPO_4$, 1.2 g L^{-1} $MgSO_4$, 1.7 g L^{-1} citric acid, 100 mg L^{-1} Fe(III) citrate, 13 mg L^{-1} $Zn(CH_3COO)_2$, 8.4 mg L^{-1} EDTA. Further it was complemented with 5 mL L^{-1} trace element solution, which consisted of 2.5 mg L^{-1} $CoCl_2 \cdot 6\,H_2O$, 15 mg L $MnCl_2 \cdot 4\,H_2O$, 1.2 mg L^{-1} $CuCl_2 \cdot 2\,H_2O$, 3.0 mg L^{-1} H_3BO_3 and 2.5 mg L^{-1} $Na_2MoO_4 \cdot 2\,H_2O$. Lastly the medium contained two different antibiotics with 4.5 mg L^{-1} Thiamine HCl and 50.0 mg L^{-1} Ampicillin.

The added feed had a slightly different composition than the batch media apart from the higher substrate concentration as indicated in tab. 2.1. It contained 20.0 g L^{-1} $MgSO_4$, 40 mg L^{-1} Fe(III) citrate, 13.0 mg L^{-1} EDTA, 16.0 mg L^{-1} $Zn(CH_3COO)_2$, as well as 8 mL L^{-1} trace element solution. Additionally approximately 5 g L^{-1} Polypropylene glycol (PPG) were added to work as anti-foam during the fermentation.

Table 2.1: Glucose concentration of used media

Media type	Process 1-4	Process 5
Preculture	8 g L^{-1}	8 g L^{-1}
Batch	15 g L^{-1}	10 g L^{-1}
Fed-batch	440 g L^{-1}	650 g L^{-1}

The salt and glucose solutions were prepared and sterilised with an autoclave separately to prohibit unwanted reactions catalysed by the heat. Afterwards the according amounts were mixed and the sterile-filtered trace elements and antibiotics were added in the sterile workbench.For the pH-control 2 M H_3PO_4 was used as an acid and 25 % NH_4OH as a base. For the induction a sterile-filtered stock solution of 100 mM IPTG was produced and kept in the freezer. Furthermore, 0.9 % NaCl solution (saline) was prepared for the live-dead analysis via Flow Cytometry (FCM).

2.2.3 Fermentation strategy

The parallel reactor setup was prepared, calibrated and sterilised in an autoclave (121 °C, 20 min). The media without the glucose and trace elements was added to the reactors before heat-sterilisation, while the missing elements were separately sterilised or sterile-filtered and added afterwards through a septum with a syringe. The cultivations 1 to 4 used a media volume of 1.5 L per reactor for the batch phase, while for experiment 5 only 1 L was filled into the vessel.

Figure 2.2: Shake flasks containing preculture medium with growing *E. coli* strain

For the preculture two autoclaved 1 L shake-flasks were filled with 100 mL of preculture media and a thawed sample of the used *E. coli* strain stored at −80 °C was added to each of them under sterile conditions. Then they were put into the shaker cabinet at 37 °C and 230 rpm for at least 12 h as seen in fig. 2.2. The preculture was left to grow overnight and on the next day the broth of both shake-flasks was combined to obtain a homogeneous culture and then separated into two equal parts of 100 mL, which were added to the bioreactors through syringes.

With the inoculation the batch phase was started, as well as the online monitoring, pH and DO-control of each vessel. The setpoints for the control units were 37 °C for the temperature and a pH-value of 6.75, while the DO-threshold was set to 30 %. At the same time the repetitive sampling routine was started, where approximately 12 mL of broth was drawn every 2 h. After the batch phase was recognized as finished by analyzing the online DO-signal and the offline glucose concentration in the sampled broth, the fed-batch phase started. The feed-forward control of the feed-pump was activated with defined set-points for the specific glucose uptake rate q_S, using eq. 2.1.

After 24 h of cultivation the bioreactors were induced with 10 mL L^{-1} of 1 mM IPTG through the septum of the reactors with a syringe. This started the phase of induced fed-batch and LDH production within the cells. At the point of induction the temperature setpoint was lowered to 32 °C, since in previous studies lowering the temperature in the culture benefited the productivity [48]. The two reactor system during a fermentation can be seen in fig. 2.3.

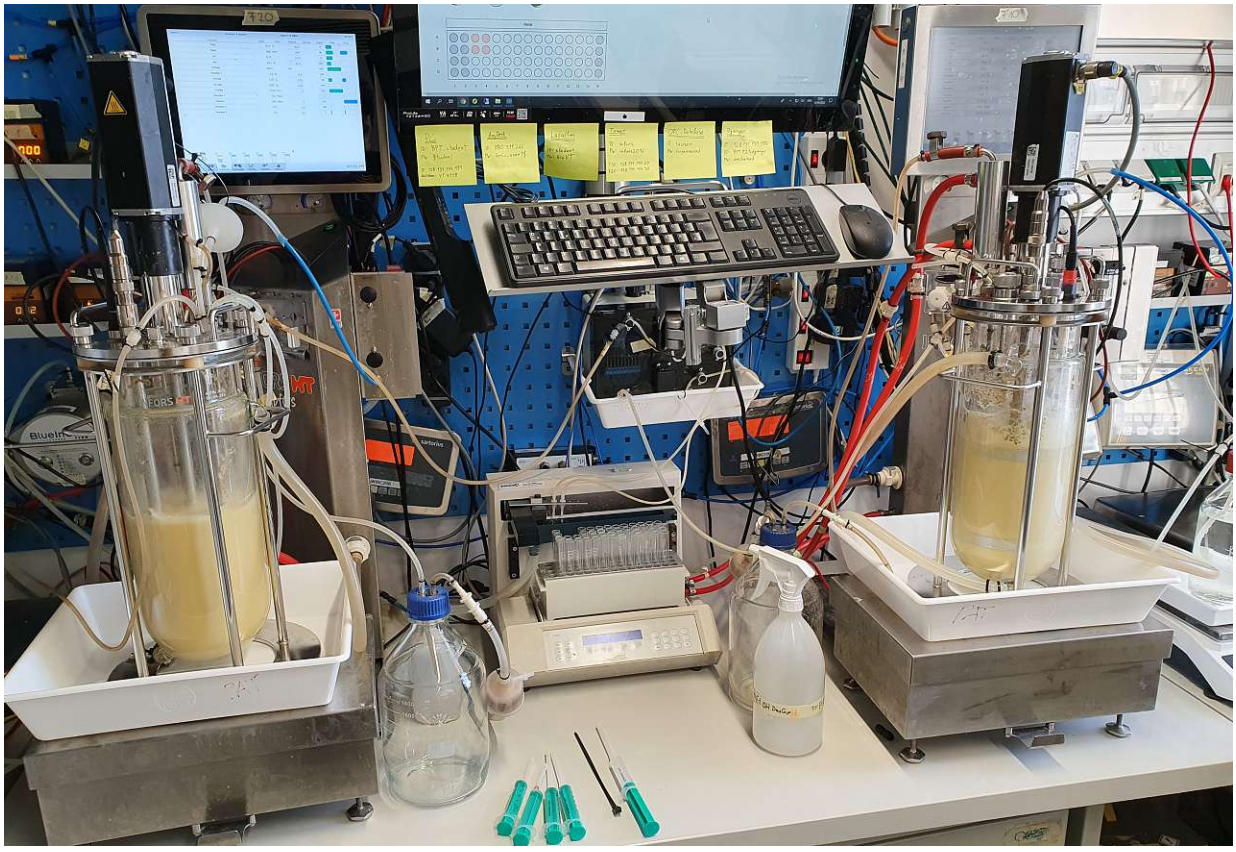

Figure 2.3: Reactor setup during fermentation with automatic sampler in the middle

2.2.4 Feed-control implementation

The real-time implementation of the feed-rate was facilitated with the Lucullus Process Information Management System PIMS (Securecell, Switzerland). The system communicated through an OPC with a computer running a live-script in MATLAB® which is further explained by Müller et al. [64]. A state-estimation algorithm was used to estimate the biomass m_X, based on the initial biomass and the actual feed rate. To convert the feed balance signal into the measured feedrate, the signal was put through a hampel filter [52] (window size = 31, outlier threshold = 3 standard deviations) to remove outliers. Next, its discrete derivative was computed in order to transform the weight signal into a feedrate. In a final step a Savitzky-Golay-Filter [83] (window size = 11, polynomial order= 1) was applied to smooth the data points.

$$F_{RS}(t) = q_S(t) \cdot \frac{m_X(t)}{c_{SR}} \tag{2.1}$$

In order to get an accurate guess for the initial conditions, a sample was drawn right after inoculating the reactors with the preculture and the measurement of the Optical Density at 600 nm (OD_{600}) was used to estimate the biomass at the end of the batch phase. The algorithm calculated the feed-forward trajectory for the feed-rate F_R through eq. 2.1, using the user-defined setpoints for the specific substrate uptake rate q_S, the online calculated biomass and the glucose concentration of the feed c_{SR} from tab. 2.1. The trajectory was then used to calculate the setpoints for the peristaltic pump, which was calibrated beforehand. The values were updated at the control tower in 30 s intervals through the live-script.

2.3 Analytical procedure

The analytical procedure covers on how all the collected data was obtained, which formed the foundation of the modeling attempts. For this work the data was separated into three main categories that differed in their frequency and the time until the result was available.

First was the group of online measurements that had a narrow sampling resolution and could be retrieved at any point during the process. They were partly used during the process for control measures and were exported as a total data set from the control tower after the experiment concluded. Second were the offline measurements that get processed within a few hours after the sample has been drawn from the bioreactor and no prolonged storing was necessary. This included the measurements quantifying the growth of the cell, the remaining substrate concentration and an analysis determining the live-dead relation of the biomass. The samples had to be processed swiftly to minimize the skewedness of the results. The final category included all the product related analyses of the cell pellets. The samples derive from the original broth samples but they were stored in a freezer in order to analyse them at a later point.

2.3.1 Sampling

The broth in the automatically sampled glass tubes was manually transferred into a different container and kept on ice or in the fridge for further processing. The Dry Cell Weight (DCW)

was determined gravimetrically in doublets by centrifuging 1 mL of broth for 10 min at 4 °C and 14 000 rpm. The supernatant was separated and cooled, while placing the remaining wet cell pellet into a drying cabinet at 105 °C. The supernatant samples were either frozen for a later analysis or the glucose concentration was directly measured with a Cedex Bio HT (Roche Diagnostics Deutschland GmbH, Germany) and the Glucose Bio HT photometric assay kit using hexokinase.Furthermore, the OD_{600} of the broth was determined with a photometer and 100 µL of the broth were diluted 1:100 in saline solution for the Flow Cytometry (FCM) measurements. The resulting 10 mL samples were stored in the fridge at 4 °C for a later analysis.

Starting with the time of induction additional product samples were prepared from the deducted broth. Depending on the remaining amount of liquid, 4 to 6 mL were pipetted into two previously dried and weighted 15 mL conical centrifuge tubes. They were then centrifuged for 10 min at 4 °C and 14 000 rpm. The supernatant was discarded and the wet cell pellet was stored in the freezer for the analysis of its LDH content at a later point.

2.3.2 Growth & death analysis

The growth of the cells was measured through the offline analysis of the DCW, by measuring the difference in weight of the dried 1.5 mL reaction tube while empty and filled with the centrifugate. Additionally the OD_{600} measurements were used in combination with a DCW-OD_{600} correlation factor as an additional method to obtain the biomass values.

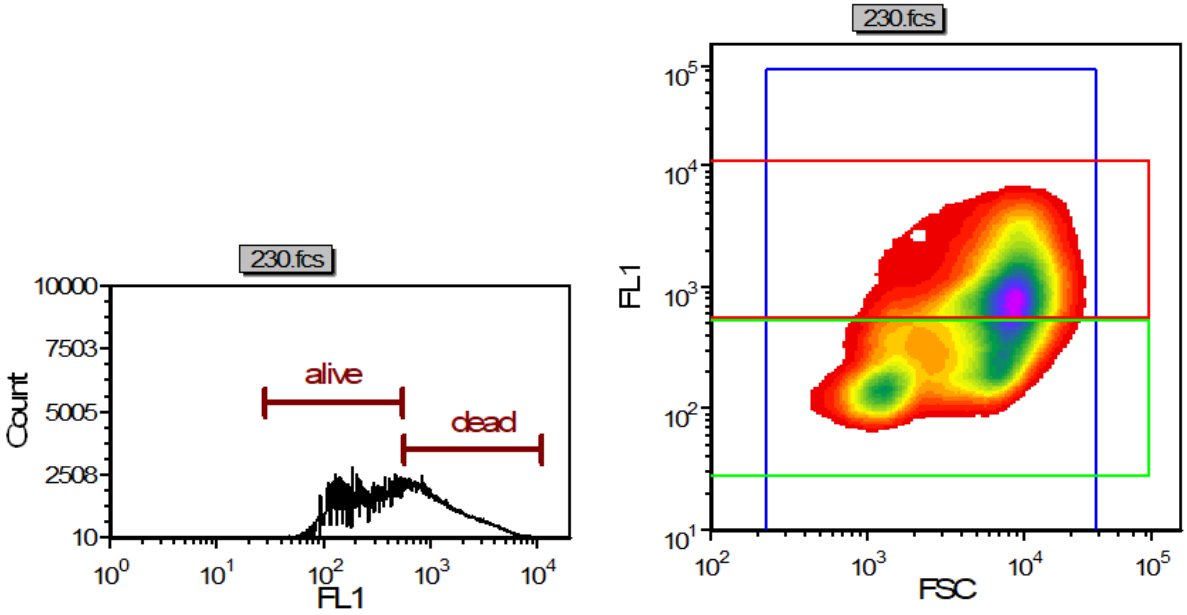

Figure 2.4: FCM analysis for process 2 at 34 h after induction. left: counts of events at the fluorescence port at respective wavelength. right: fluorescence against forward-scatter intensity with boundaries for alive cells (green) and dead cells (red)

The live-dead analysis of the cells was carried out through a FCM for which the staining protocol described by Langemann et al. [50] was followed. First, the saline-buffered cell solution was further

diluted to a final concentration of 1:10 000 of the original broth sample. Then 1 mL of this dilution was mixed with 1.5 µL of the 0.75 µM membrane potential-sensitive dye $DiBAC_4(3)$ and the 3.0 µM fluorescent dye RH414 respectively and incubated for 5 min. The two chemicals bind differently to damaged and undamaged cell-walls. The sample was then analysed with the CyFlow® Cube 6 flow cytometer (Sysmex Partec GmbH, Germany), which recorded the forward-scatter and fluorescence signals. Afterwards a software related the two signals to the percentage of dead and alive cells in the sample [50, 114]. Fig. 2.4 shows a graphical interpretation of the measurement results. The software ultimately delivered the count of events classified as dead and alive, which were put into relation to estimate the relative amount of either cell state in the broth.

2.3.3 Product analysis

Since the product was excreted by the *E. coli* cells, they had to be disrupted and further treated to be able to qualitatively and quantitatively analyse the product amount. The used protocol for product exposure and Inclusion Body (IB) refolding, together with the applied reagents were developed by Igwe et al. (Paper in preparation). In a first step, the used washing and solubilization solutions listed in tab. 2.2 had to be prepared in Deionized water (DI water). Additionally, the needed reagents for the enzymatic and protein assays were produced.

Table 2.2: Recipes for the chemicals used for the product analytics from Igwe et al. (Paper in preparation).

Lysis buffer		
Tris	0.1	M
$Na_2EDTA \cdot 2\,H_2O$	0.01	M
pH (HCl)	7.4	

Solubilization buffer		
$NaH_2PO_4 \cdot H_2O$	0.15	M
GuHCl	4	M
pH (HCl)	6	

Washing buffer A		
Tris	0.05	M
NaCl	0.5	M
Tween 80	0.02	% w/v
pH (HCl)	8	

Refolding buffer		
$NaH_2PO_4 \cdot H_2O$	0.15	M
pH (HCl)	6	

Washing buffer B		
Tris	0.05	M
$Na_2EDTA \cdot 2\,H_2O$	0.005	M
pH (HCl)	8	

Pellet disruption

In order to extract the produced intracellular LDH a cell disruption was performed. First, the frozen samples were thawed and diluted in lysis buffer towards a similar DCW concentration. To solubilize the pellets a T10 basic Ultra-Turrax® disperser (IKA®-Werke GmbH & CO. KG, Germany) was used. Concerning the disruption of the cell-walls an ultrasonic rod was applied

with 75 % Amplitude for 5 min where it performed 20 s of sonication followed by a 40 s pause. The sample was put on ice during the procedure minding the denaturing effect of excessive heat towards the LDH enzymes. The chemical influence of the lysis buffer in combination with the mechanical disruption resulted in a decomposed cell suspension, which was centrifuged for 10 min at 4 °C and 14 000 rpm. A sample of the supernatant was taken to analyse the amount of solubilized LDH later, while the rest of the liquid was discarded.

IB washing

The centrifugate was further processed to analyse the content of IBs in the sample. Therefore, the pellet was purified from some of the cell debris and impurities by adding two different washing solutions, described in tab. 2.2, consecutively. First, 6 mL of the washing buffer A was added and the pellet was again resuspended by the disperser. After incubating it for 15 to 30 min the suspension was centrifuged with the previous settings and the supernatant was discarded. Afterwards the pellet was resuspended again in 6 mL of washing buffer B with the help of the disperser. From the suspension a sample of 1.5 mL was pipetted into a 2 mL reaction tube for the refolding protocol. Another 50 µL were diluted in 0.95 mL of DI water for a later analysis of the protein content with a Bicinchoninic Acid (BCA)-assay. The sample with the remaining 4.5 mL was centrifuged for 10 min at 4 °C and 14 000 rpm, decanted and put into the drying cabinet for several days. Ultimately this was done to determine the dry weight of the disrupted and washed cell pellet. But since no correlation to the IB content could be determined from these dry weight samples, they were not further considered for this work.

IB refolding procedure

The sample for the refolding protocol was also centrifuged with the same settings and the pellet was decanted. Next, 1.5 mL of solubilization solution was added to the pellet to dissolve and straighten the LDH enzymes aggregated into IBs. The pellet was slightly resuspended by using a VORTEX 2 laboratory shaker (IKA®-Werke GmbH & CO. KG, Germany) and then incubated for 2 h on a rocking plate at room temperature. After the solubilization period a sample of the solution was diluted by 1:40 in refolding buffer to a total volume of 1.5 mL and placed on a rocking plate inside a fridge at 4 °C for approximately 16 h. After the incubation was finished the enzymatic activity and protein content were analysed with the respective assays. It has to be noted, that the IB refolding procedure is accompanied by product losses.

Enzyme assay for LDH activity

The concentration of active enzymes in the supernatant after cell disruption and after refolding of the IBs were determined using an adaptation of the enzyme assay procedure described by Ponsoda et al. [72]. There the catalytic nature of LDH for the reduction of pyruvate to lactate in the presence of Nicotinamide-Adenine-Dinucleotide-Hydrate (NADH) was utilized. The change in absorption at a wavelength of 340 nm was recorded with the automatic Spark® multimode microplate reader (Tecan Trading AG, Switzerland). The time-resolved data was analysed with a MATLAB® script (MathWorks, USA), which performed a linear regression on the decline in absorption and used eq. 2.2 to calculate the activity.

$$Act = \frac{k_{abs} \cdot V_t}{\varepsilon_{340} \cdot l \cdot V_s} \cdot f_d \tag{2.2}$$

Act	Enzymatic activity $[\frac{U}{L} = \frac{\mu mol}{min \cdot L}]$	k_{abs}	Slope of the absorbance regression $[\frac{Abs}{min}]$
V_t	Total volume per well [µL]	V_s	Sample volume per well [µL]
f_d	Sample dilution factor	l	Path length through well [cm]
ε_{340}	Molar extinction factor of NADH at 340 nm $[\frac{L}{mmol \cdot cm}]$		

The wells were filled with 66 µL of sample and mixed with the reaction solution to 200 µL in total. The path length in the machine was 0.624 cm and the molar extinction factor of NADH was implemented as 6.22 L mmol^{-1} cm^{-1} [13]. In order to obtain a reasonably long linear region the dilution factor had to be adjusted accordingly. All samples were measured in triplets and corrected against a blank to compute the weighted average of enzymatic activity at the sampling point and their standard deviation.

BCA-assay for protein content

To evaluate the total protein concentration of the product samples a BCA-assay was performed. Bovine serum albumin (BSA) was used in different dilutions to calculate a standard regression as reference for the measurements. For this assay BCA was mixed with $CuSO_4$ and together they reacted with the proteins in the sample and changed their photometric properties. The well-plate with the sample-reagent-mixture was analysed with a Spark® multimode microplate reader (Tecan Trading AG, Switzerland). The machine incubated the samples at 37 °C for 30 min before their absorbance at a wavelength of 562 nm was measured once. The samples were analysed in triplets and corrected against a blank. The protein concentration was determined for the supernatant after cell disruption, the IBs resuspended in washing buffer B, the solubilised IBs and the refolded IBs. The results were evaluated by a MATLAB® (MathWorks, USA) script, where the reference samples were used to calculate a standard regression and the samples were compared against it.

Further analytics

Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC) measurements were performed with an UltiMate 3000 HPLC-system (Thermo Fisher Scientific Inc., USA) for some of the samples. Small amounts before and after the refolding of the solubilized IBs were analysed to gain an additional comparison to the assay results. The used method was developed by Kopp et al. [49] for the analysis of mAb and large proteins. A BioResolve RP mAb Polyphenyl column (100x3 mm, particle size 2.7 µm, Waters Corporation, USA) was used to quantify the LDH content of the samples. The mobile phase consisted of mixtures of type-1 purified water and Acetonitrile (AcN), both of them containing 0.1 % Trifluoroacetic acid (TFA). After loading the column with the sample the concentration of AcN was continuously increased as seen in fig. 2.5. At the end of the column the absorbance of the passing proteins was measured at 280 nm and the resulting curve was integrated to evaluate the protein amount. BSA at different concentrations was used as a reference standard for quantification.

Figure 2.5: Mobile phase concentration profile over time with A - type-1 purified water and B - AcN, both containing 0.1 % TFA

Finally a Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) was carried out for selected samples as a qualitative information on the molecular-weight distribution of proteins and particles at certain points of the analytical procedure. The pre-cast gel plates (10% Criterion^{TM} TGX Stain-Free^{TM} Protein Gel, Bio-Rad Laboratories Inc., California, USA), were prepared with the different samples and a protein ladder (Precision Plus Protein^{TM} Unstained Protein Standards, Strep-tagged recombinant, Bio-Rad Laboratories Inc., California, USA) was added as a reference before the current was applied. After the electrophoresis procedure finished the gels were stained with coommassie blue (G-250), destained with diluted methanol and desalted water. A microscope was finally used to take pictures of the result.

2.4 Modeling methods

Overall the modeling workflow described by Daume et al. [73] was followed in regards to adhering to GMoP and applying a systematic approach. The six central steps of this iterative process can be seen in fig. 2.6. In the first step the main focus of this work was defined. A model should be developed, which is able to explain the product formation and localisation in the organism with appropriate kinetics, using physiological descriptors. The obtained experimental data should be sufficiently explained by it, while also developing a universally applicable, overall model structure.

For the formulation of the overall reaction scheme in step 2 a state-space model was used as a starting point, which was previously developed by Müller et al. [64]. It described the growth and death of the examined *E. coli* strain and was extended in this work to include the product formation. A more detailed description follows in the next chapters.

2.4.1 Bioprocess growth model

The general structure of the implemented model was of state-space kind, where the state-vector x contains the process variables of interest. The mathematical structure can be seen in eq. 2.3, which describes the change over time of the state-vector through the inner and outer influences and dynamics of the modeled process. The vectors u and f represent the controlled inputs and the uncontrolled disturbances respectively. The matrices A, B and E contain the influences of the states, inputs and outputs on the system. Furthermore, the observable output y can be expressed through eq. 2.4, where the inner states get translated by the matrix C [30, 118].

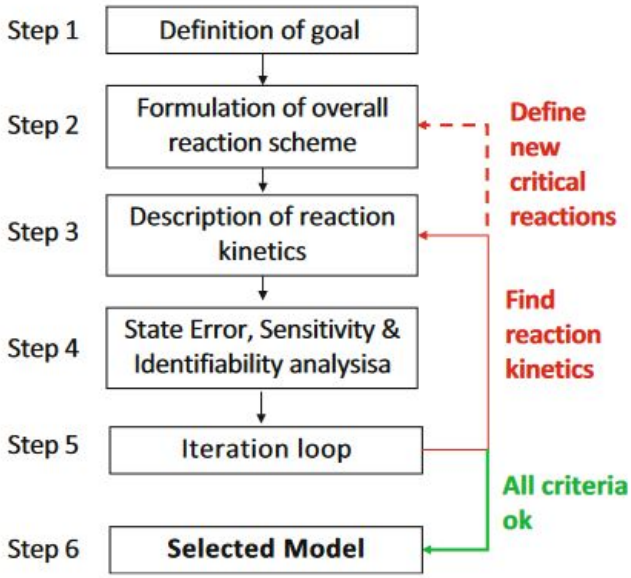

Figure 2.6: Iterative modeling workflow following GMoP from Daume et al. [73]

$$\frac{d\mathbf{x}}{dt} = A \cdot \mathbf{x} + B \cdot \mathbf{u} + E \cdot \mathbf{f} \tag{2.3}$$

$$\mathbf{y} = C \cdot \mathbf{x} \tag{2.4}$$

In order to describe a bioreactor system the state vector was assembled with several variables (states) describing the growth and death of cells, as well as process dynamics, as can be seen in eq. 2.5 [64].

$$x = \begin{bmatrix} V \\ X_V \\ X_d \\ S_1 \\ S_2 \\ P \\ I \\ CO2 \\ O2 \end{bmatrix} \quad \begin{matrix} \\ \\ \\ \\ \dots \\ \\ \\ \\ \end{matrix} \quad \begin{matrix} \text{Volume of broth } [L] \\ \text{Viable cell mass } [g] \\ \text{Dead cell mass } [g] \\ \text{Mass of substrate 1 } [g] \\ \text{Mass of substrate 2 } [g] \\ \text{Mass of product } [g] \\ \text{Mass of induction agent } [g] \\ \text{Cummulated } CO_2 \text{ produced } [g] \\ \text{Cummulated } O_2 \text{ produced } [g] \end{matrix} \tag{2.5}$$

As mentioned in step 2 of the modeling workflow the overall reaction scheme seen in eq. 2.6 to 2.7 had been implemented in the inherited model.

$$S_{1,2} \xrightarrow{q_{S1,2}} X_V \tag{2.6}$$

$$X_V \xrightarrow{q_D} X_D \tag{2.7}$$

Eq. 2.6 symbolizes a simplification of the cell metabolism when no product is formed. A more comprehensive version can be seen in 2.8.

$$X_V + S_{1,2} + O_2 \rightarrow X_V + CO_2 + P \tag{2.8}$$

In order to explain the dynamics of the *E. coli* strain, balances for each compound were established, expressing the change of a variable through the in- and outgoing flows, as well as sources and sinks in the form of the specific rates q_i. The resulting equations can be seen in eq. 2.9 to 2.14. Except for the volume, all of the equations embody mass balances of the respective compounds. Since the productivity was not yet modeled the product formation in eq. 2.13 was set to zero.

$$\frac{dV}{dt} = F_{in} - F_{out} \tag{2.9}$$

$$\frac{dX_V}{dt} = Y_{X/S_{1,2}} \cdot q_{S_{1,2}} \cdot X_V - q_D \cdot X_V - F_{out} \cdot X_V \tag{2.10}$$

$$\frac{dX_D}{dt} = q_D \cdot X_V - F_{out} \cdot X_D \tag{2.11}$$

$$\frac{dS_{1,2}}{dt} = F_{RS} \cdot c_{SR} - q_{S1,2} \cdot X_V - F_{out} \cdot S_{1,2} \tag{2.12}$$

$$\frac{dP}{dt} = q_P \cdot X_V = 0 \tag{2.13}$$

$$\frac{dI}{dt} = F_{RI} \cdot c_{I_R} - F_{out} \cdot I \tag{2.14}$$

The change of the variable for the cummulated CO_2 and O_2 are the equivalents of the Carbon Evolution Rate (CER) and Oxygen Uptake Rate (OUR), as expressed with eq. 2.15 and 2.16. Their governing equations are elaborated at a later point.

$$\frac{dCO2}{dt} = q_{CO2} \cdot X_V \mathrel{\widehat{=}} CER \tag{2.15}$$

$$\frac{dO2}{dt} = q_{O2} \cdot X_V \mathrel{\widehat{=}} OUR \tag{2.16}$$

This set of equations was lumped together in a general state-space structure as seen in eq. 2.18. The change of of the state-vector x is expressed as a function of the added substances F_{in} translated to the respective states with the input concentration matrix C_{in}, the outgoing stream F_{out} multiplied with the current concentration of the state variables c and the rates of the states r. The flow entering the reactor F_{in} is assembled by the separate input flows of the substrate F_{RS} and the induction agent F_{RI} by eq. 2.17.

$$F_{in} = [F_{RS} \quad F_{RI}] \tag{2.17}$$

$$\frac{dx}{dt} = C_{in} \cdot F_{in} + \sum_{j=1}^{m} F_{out,j} \cdot c + r \tag{2.18}$$

The independent rates of the states r describe the changes due to reactions or the cell metabolism and is expressed with the more detailed formula of eq. 2.19. The growth and death kinetics as well as other dynamics are used to compute it. The specific reaction rates q_i contain the uptakes for two different substrates q_{S1}, q_{S2} and the cell death rate q_D. In bioprocesses the specific rates q_i are expressions relative to the biomass X to make their dynamics more comparable over different setups and scales [84].

$$r = Y_X \cdot q_i \cdot X_V = Y_X \cdot \begin{bmatrix} q_{S1} \\ q_{S2} \\ q_D \end{bmatrix} \cdot X_V \tag{2.19}$$

Several mechanistic relations were established to describe the kinetics of the specific rates in step 3 of the workflow in fig. 2.6. In the implemented model a Monod kinetic was used for glucose uptake in q_{S1}, where the substrate concentration was obtained by dividing the state for substrate mass by the broth volume. The cell death was modeled with the constant factor k_D and both equations can be seen in eq. 2.20. A second substrate was not implemented setting its specific rate q_{S2} to zero throughout the process. From this point forward the general notation for substrate S describes the effects concerning glucose and substitutes the explicit symbol S_1 in this work.

$$q_{S1} = q_S = q_{S,max} \cdot \frac{c_S}{c_S + K_S} \quad \text{and} \quad q_D = k_D \tag{2.20}$$

The yield matrix Y_X contains all the yield coefficients that translate the substrate uptake q_S to the specific rate of the other states, as seen for the specific growth rate μ with the biomass-per-substrate yield $Y_{X/S}$ in eq. 2.21 [84].

$$\mu = q_X = q_S \cdot Y_{X/S} \tag{2.21}$$

Implementing eq. 2.19 into eq. 2.18 leads to the more detailed notation of eq. 2.22 that implicitly includes the kinetics. These mathematical expressions were implemented in a MATLAB® environment using scripts and SIMULINK®-models to simulate the fermentation. The state-update was calculated discretely through eq. 2.23 using the mass balance of eq. 2.22. The inputs were the feeding profile F_{in} and the disturbance was the sampling, which was detected through weight changes of the bioreactor. The algorithm recalculated the specific rates q_i at every step with the current value of the state vector x. The yield matrix Y_X was implemented as a constant and obtained through a calibration using the obtained data. In a later approach of this work certain values were also expressed through kinetics, which made the matrix a time-dependent variable itself.

$$\frac{dx}{dt} = C_{in} \cdot F_{in} + \sum_{j=1}^{m} F_{out,j} \cdot c + Y_X \cdot q_i \cdot X_V \tag{2.22}$$

$$x_{i+1} = x + \frac{dx}{dt} \cdot \Delta t \tag{2.23}$$

The inherited model was adapted and extended according to steps 2 and 3 of the workflow, to represent the product formation and its dynamics for the *E. coli* strain at hand. Therefore, a number of different kinetics were implemented for the specific total productivity $q_{P,total}$ and the specific inclusion body productivity $q_{P,IB}$ to match the experimental data. The approaches and results are documented in chapter 3 Results & Discussion.

2.4.2 Model parameterization and goodness of fit

The different candidates for the adapted model structures and kinetics were evaluated for their performance to represent the given process. The evaluations of step 4 of the modeling workflow in fig. 2.6 were done in three parts. Usually a structural identifiability analysis is done first, followed by the model parameterization with experimental data in combination with an evaluation of the goodness of fit and finally a practical identifiability analysis [73]. To minimize the effort for the identifiability analysis, the parameterization and goodness of fit analysis were prioritised in this work.

A numerical procedure was used to globally estimate the parameters of the kinetic equations over the four training data sets, by minimizing a cost-function consisting of the Sum of Squares Error (SSE) between the simulated data $\hat{x}_i$ and the measurements x_i as seen in eq. 2.24 [22]. Since an unconstrained optimization algorithm was used, some additional penalty terms were necessary that increased the cost-term significantly in case the algorithm pushed the parameters towards implausible values, such as negative numbers for the Monod parameters.

$$SSE = \sum_{i=0}^{n} (x_i - \hat{x}_i)^2 \tag{2.24}$$

$$SSR = \sum_{i=0}^{n} (\hat{x}_i - \bar{x})^2 \ , \qquad SST = \sum_{i=0}^{n} (x_i - \bar{x})^2 \tag{2.25}$$

After the parameterization algorithm found the local optimum, the model with the parameter set was evaluated for its goodness of fit by calculating the coefficient of determination (R^2) with eq. 2.26 for the validation data set. The Sum of Squares Regression (SSR) and Sum of Squares Total (SST) of eq. 2.25 were used to evaluate if the model explained the measured data better than the simple mean $\bar{x}$ [22]. In the case of polynomial terms like taylor series but also for more complex algebraic structures an increase of parameters will lead to a higher R^2, but not necessarily to a better model structure. Therefore, the indicator was corrected for its complexity by eq. 2.27 to yield the adjusted coefficient of determination ($R^2_{adjusted}$), where n is the amount of measurements and p the number of parameters considered. With this adjustment the risk of over-fitting the data was decreased and a better decision basis for choosing a model structure was created [22].

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad \text{with} \qquad SST = SSR + SSE \tag{2.26}$$

$$R^2_{adjusted} = 1 - \frac{(1 - R^2) \cdot (n - 1)}{n - p - 1} \tag{2.27}$$

Furthermore, the Normalized Root-Mean-Square Error (NRMSE) value was calculated for each model through eq. 2.28 as an additional comparison for the performance of different kinetics. It computed the remaining average error between the model $\hat{x}_i$ and the measured data x_i and was normalized by dividing through the range of the data set [73].

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=0}^{n}(x_i - \hat{x}_i)^2}}{x_{max} - x_{min}} \tag{2.28}$$

2.4.3 Model sensitivity and identifiability

After choosing the model structure based on the values of $R^2_{adjusted}$ the system was evaluated for its sensitivity and identifiability, as indicated under step 4 of fig. 2.6. The results of these analyses gave an insight if the parameters can be quantified given the overall choice of equations or if an inappropriate model structure has been implemented [73]. The structural and practical identifiability analysis are based on the local sensitivity **S**, which evaluates the effects of the parameters on the outputs of the model. Less sensitive parameters have little influence on the model output and are therefore difficult to estimate with measured data. Moving forward, such parameters should be neglected or the kinetic equations changed in an iteration loop (step 5 in fig. 2.6) or model reduction step [73].

The sensitivity matrix **S** of the different models was obtained by numeric approximation. The parameter set was diverted by 1-2 % and the resulting change in each state variable ∂c_i was divided by the parameter variation $\partial\theta_p$, as seen for state i and parameter p in eq. 2.29 [27]. This was done for every $state \times time \times parameter$ combination of the chosen model together with the obtained data sets. A 3-dimensional matrix was built ($n \times d \times r = rows \times depth \times columns$), which was transformed into a 2D matrix by stacking the n state matrices on top of each other. Every state was represented by a $d \times r$ matrix in which it was derived by the parameter at each time incident [73, 25, 18].

$$S_{i,p} = \frac{\partial c_i}{\partial \theta_p} \tag{2.29}$$

The result was also used for the importance factor $\delta_{i,p}$, which ranks the parameters by their overall influence on the model output through the mean square root of the sensitivity values. This allowed a first insight on the model dynamics and was calculated with eq. 2.30, where high values indicate a high influence on the selected outputs [73, 25].

$$\delta_p^{msqr} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\bar{S}_{i,p}^2} \quad \text{with} \quad \bar{S}_{i,p} = S_{i,p} \cdot \theta_p \tag{2.30}$$

Structural Identifiability

The structural identifiability is a property of the model and evaluates if its parameters can theoretically be determined when the exact values for the state variables are known. It is therefore a necessity for parameter estimation and the algorithm was described in detail by Daume et al. and Brun et al.[73, 18]. The outcome of the analysis categorizes them into three different groups depending on the amount of solutions that exist for a parameter of the model.

- **structural unidentifiable:** An infinite number of solutions exists
- **non-uniquely structurally identifiable:** More than one countable number of solutions exist
- **structural identifiable:** A unique solution exists

The analysis was done by firstly calculating the discrete-time sensitivity matrix **S**. It was built by stacking the local sensitivity matrices as described previously with eq. 2.29. The resulting matrix was checked for rank deficiency by eq. 2.31. If the number of parameters is equal to the column-rank r the model has full rank and is structurally identifiable.

$$\text{rank}(\mathbf{S}) = r \tag{2.31}$$

In a last step the conditional number of the system was evaluated by dividing the largest by the smallest absolute singular value of the discrete-time sensitivity matrix **S** as seen in eq. 2.32. Ideally conditioned systems have a condition number of 1, well-conditioned below 10^3 and systems with values above are called ill-conditioned. The last category may lead to unstable solutions and result in even more problems for iterative solvers when the condition number exceeds 10^{30}, which describes very ill-conditioned systems [73].

$$\text{cond}(\mathbf{S}) = \frac{\lambda_{largest}}{\lambda_{smallest}} \tag{2.32}$$

Practical Identifiability

For the practical identifiability the experimental measurements were used together with the accompanying level of noise to determine if the parameters are uniquely identifiable. This final step evaluated how reliable the estimations for the parameters were, given the amount and accuracy of the available measurements. The used algorithm is described in detail by Daume et al. [73], where the central step is to compute the Fisher Information Matrix (**FIM**) with eq. 2.33.

$$\mathbf{FIM} = \mathbf{S}_{sum}^T \mathbf{W} \mathbf{S}_{sum} \quad \text{with} \quad \mathbf{S}_{sum} = \sum_{l=1}^{d} \mathbf{S}_{t_l} \ , \quad \mathbf{W} = \text{diag}(\frac{1}{\sigma_{i,l}^2}) \tag{2.33}$$

$\mathbf{S}_{sum}$ is the sum of the sensitivity matrices of the measured states c_i at the sample time t_l. The covariance matrix of the state variables **W** is build with the variance values of the state variables $\sigma_{i,l}$ given by their measurement techniques. Furthermore the Minimum parameter covariance

matrix ($\text{Cov}_{min}(\hat{\theta})$) was calculated from the **FIM** with eq. 2.34 and used to obtain the Parameter corrrelation matrix ($\text{Cor}_{min}(\hat{\theta})$).

$$\mathbf{Cov}_{min}(\hat{\theta}) = \mathbf{FIM}^{-1} \tag{2.34}$$

The covariance matrix is a $r \times r$ - matrix, where r is the number of parameters and the entries are the minimum covariances between the respective parameters. They were subsequently normalized by using eq. 2.35 to obtain the correlation between the pth and qth parameter and build the $\text{Cor}_{min}(\hat{\theta})$.

$$\mathbf{Cor}_{min}(\hat{\theta}) = \begin{bmatrix} \rho_{11} & \dots & \rho_{1r} \\ \dots & \rho_{pq} & .... \\ \rho_{r1} & \dots & \rho_{rr} \end{bmatrix} \quad \text{with} \quad \rho_{pq} = \frac{\sigma_{pq}}{\sqrt{\sigma_p} \cdot \sqrt{\sigma_q}} \tag{2.35}$$

Values close to one are indications that the correlated parameters influence each other strongly and it becomes harder to determine them globally [73].

Iteration loop

The results of all analytical tools applied in step 4 of the workflow in fig. 2.6 were used to select a model structure. When certain kinetics did not fit the data well enough or failed the identifiability analysis, a step was taken back to look for and implement alternatives. This started an iterative process until a final model structure was found that satisfied the original modeling goals best.

2.4.4 Carbon Evolution Rate (CER) and Oxygen Uptake Rate (OUR)

One of the central pillars of modeling in process engineering are elemental balances and two of the fundamental constituents in biology are carbon and oxygen. Microorganisms take up carbon in form of substrate and process it into biomass, metabolites and most of the time CO_2. The conversion was already described symbolically by eq. 2.8 and can be translated into a elemental balance expressed by eq. 2.36. The equation reflects the distribution of the carbon introduced to the metabolism with the substrate and incorporated into the biomass and product or excreted as CO_2. This balance reflects a stochiometric calculation, therefore the rates had to be transformed into molar rates through the Molar mass (C-mol based) (MW) of the compounds. Through simplifications it could be shown that the modeled CER is a function of the substrate uptake and the biomass, when assuming constant yields.

$$\text{CER}_{model} = q_{CO2} \cdot X_V = \left(\frac{1}{M_S} - \frac{Y_{X/S}}{M_X} - \frac{Y_{P/S}}{M_P} \right) \cdot q_S \cdot X_V \cdot M_{CO2} \tag{2.36}$$

On the other hand, the formed CO_2 can be calculated through a balance around the reactor, using the online off-gas measurements together with the known input consisting of the volume flow of ambient air F_{AIR} and pure oxygen F_{O2}, if applied. The measured CER was obtained through eq. 2.37, where the molar volume of an ideal gas V_M was used to convert the result into a molar rate and the MW of CO_2 transformed it into a mass flow. Changes in dissolved CO_2 were neglected

in the calculation. In order to account for the varying humidity of the gaseous phase the inert gas fraction R_a with eq. 2.38 was used. The resulting rate presents an observed CER and can be used as an indicator for the metabolic activity in the reactor. Combined with the feedrate and assumed yields it serves as an estimate of how much carbon has accumulated in the broth either as biomass, product or non-metabolised substrate [7, 112].

$$\text{CER}_{meas} = \frac{F_{AIR} + F_{O2}}{V_M} \cdot (x_{CO2,out} \cdot R_{a,inert} - x_{CO2,in}) \cdot M_{CO2} \tag{2.37}$$

$$\text{with } R_{a,inert} = \frac{1 - x_{O2,in} - x_{C02,in}}{1 - x_{O2,out} - x_{CO2,out} - \left(1 - \frac{y_{wet}}{x_{O2,in}}\right)} \tag{2.38}$$

A second elemental balance can be used to monitor the process with the given online off-gas measurements. Aerobe organisms need oxygen in form of O_2 for their catabolism, which can be used to conduct a Degree of Reduction (DoR) balance with the metabolic activity described in eq. 2.8. When creating the elemental balance for oxygen, the respective amount in the composition of the molecule has to be considered. For the DoR the charges and reductive potential of all compounds are expressed through the degree of reduction of every molecule γ_i. The advantage of this consideration is that γ_{CO_2} is zero and the equation simplifies into eq. 2.39, where the final division with γ_{O2} and multiplication with MW_{O2} results in the modeled oxygen mass consumed by the metabolic process [3, 112].

$$\text{OUR}_{model} = q_{O2} \cdot X_V = \left(\frac{\gamma_S}{M_S} - \frac{Y_{X/S} \cdot \gamma_X}{M_X} - \frac{Y_{P/S} \cdot \gamma_P}{M_P}\right) \cdot q_S \cdot X_V \cdot \frac{M_{O2}}{\gamma_{O2}} \tag{2.39}$$

The OUR was measured as well with the same approach as for the CER, seen in eq. 2.40. The difference in oxygen put into the vessel through the aeration and the concentration measured in the off-gas was used to build a balance and calculate the consumed O_2. The change in dissolved oxygen in the broth was neglected due to its overall small solubility in water.

$$\text{OUR}_{meas} = \frac{F_{AIR} + F_{O2}}{V_M} \cdot (x_{O2,out} \cdot R_{a,inert} - x_{O2,in}) \cdot M_{O2} \tag{2.40}$$

The DoR values of the different compounds were calculated from their elemental composition, with $\gamma_C = 4$, $\gamma_H = 1$, $\gamma_O = -2$, $\gamma_N = -3$ and $\gamma_S = 6$ [108]. The results for the considered molecules and the respective MWs are listed in tab. 2.3 and calculated for the product below.

Table 2.3: MW and DoR of molecules considered in CER and OUR

Compound	MW	γ
Glucose	30 g C-mol^{-1}	4
Biomass [64]	26.5 g C-mol^{-1}	4.113
CO_2	44.01 g C-mol^{-1}	0
O_2	32 g mol^{-1}	-4

MW and DoR of LDH

The values for the produced protein LDH were obtained with the help of the UniProt Knowledgebase [100, 101], where the protein mass and the amino acid sequence was given. A MATLAB® script was then used to derive the MW and the sequence was used together with a amino acid table to calculate the elemental composition and DoR for the protein. The respective results were normalized to a C-mol basis for the C-balance and the DoR balance. The collection of the outcome can be seen in tab. 2.4.

The modeling and online measuring of the CER and OUR hold the potential to be used as directly observable states in order to update state estimators and enable control of the process. They could be used together to calculate the specific substrate uptake rate q_S and the viable biomass X_V, assuming constant and correct yields. But, the CER and OUR only represent a measurement of the overall metabolic activity and cannot differentiate between biomass and product formation.

Table 2.4: Protein specific values for LDH calculated from UniProt [100, 101]

Physical quantity	Value
formula (C-based)	$CH_{1,599}N_{0,266}O_{0,316}S_{0,0053}$
m_{LDH} (database)	34 206 Da
MW_{LDH}	22.58 g C-mol^{-1}
γ_{LDH}	4.202

3 | Results & Discussion

The goal of this work was the establishment of a mechanistic model describing the dynamics of the product formation and its aggregation into IBs, utilizing physiological descriptors. Therefore, data was collected to investigate and quantify the productivity of the organism during an induced fed-batch phase. The main focus was on the influence of the feeding-profile and the resulting substrate uptake rate. The existing model had to be adjusted in order to accommodate the cell behaviour of an induction triggered enzyme expression system. The developed models were further analysed for their identifiability to ensure that meaningful parameter sets were obtainable with the given measurement techniques and their uncertainties. Finally, the resulting model was explored for its characteristics and the possible capabilities in process optimization through simulation experiments.

In total 5 different cultivation experiments were carried out with different trajectories for the feed-rate. The collected data sets were then split into 4 training runs used to obtain the parameter estimations for the chosen model structure and 1 validation run. The setpoints for the substrate uptake rates q_S after induction for the different processes can be seen in tab. 3.1.

Table 3.1: Setpoints for the substrate uptake rate q_S in $\mathrm{g\,g^{-1}\,h^{-1}}$

Name	Purpose	q_S setpoint	End of process
Process 1	Model training	0.25	over-foaming
Process 2	Model training	0.13	manual
Process 3	Model validation	0.33	over-foaming
Process 4	Model training	0.22	over-foaming
Process 5	Model training	0.16	over-foaming

These setpoints were used in eq. 2.1 to calculate the feedrate, during the induction phases. As explained in section 2.2.4, the biomass content was estimated through a simple mass balance approach. Therefore, the actual values for q_S differed from the setpoints throughout the process. A comparison between the intended uptake rates and the rates calculated with the offline DCW values is shown in fig. 3.1.

Some processes had to be shut down prematurely due to an overflowing of the foam built up during the fermentation, as indicated in the last column of tab. 3.1. This triggered a substantial amount of the fluid to be sucked out of the reactor through the off-gas line and into the foam trap located before the gas analyzers, as can be seen in fig 3.2.

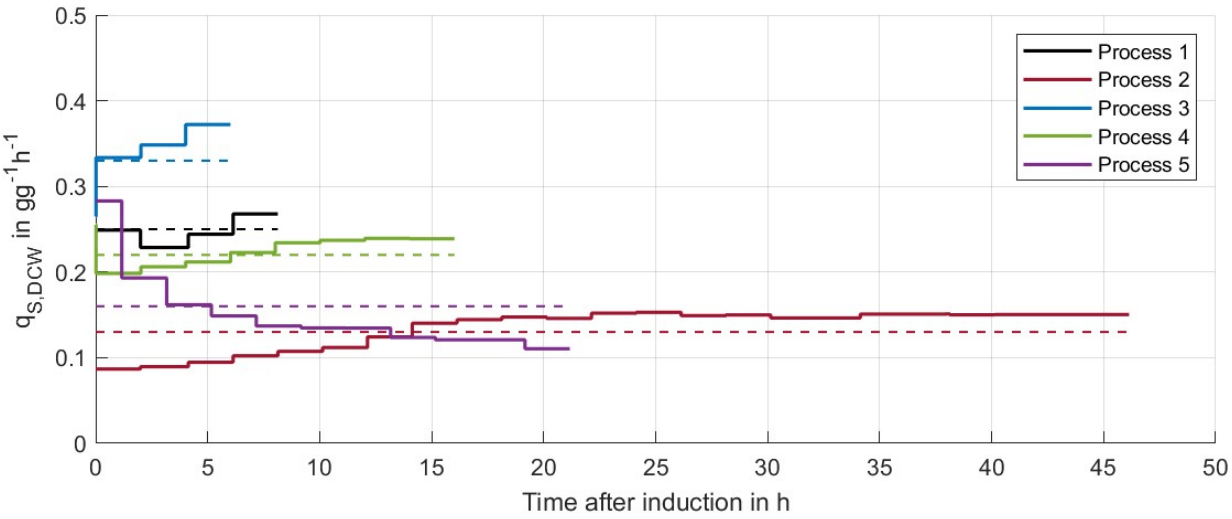

Figure 3.1: Substrate uptake rate q_S after induction of the 5 cultivation experiments. Dashed lines represent the intended q_S setpoints and the solid lines the calculated rates using the DCW measurements for the biomass.

Figure 3.2: End of process through over-foaming of reactor and consequently loss of broth

In order to prohibit this effect substantial amounts of PPG were added as anti-foam agent before and during the process. Nevertheless, some of the experiments suffered from over-foaming, usually occurring over night when the process was not actively observed. The accompanying loss of fermentation broth resulted in the termination of the process, but the collected data up until the extensive foaming was used for the model development.

3.1 Data exploration

The analysis methods of the experimental processes were divided into 3 main groups, where the first one consists of the online and the other two of the offline measurements, as explained in section 2.3. The offline data was further distinguished into data that was related to the biomass growth and data describing the product amount and localisation. Since the growth and death of the biomass directly influenced the protein production and were the main focus prior to the induction, those measurements were analysed first.

3.1.1 Growth related data

The central variable in fermentation processes is the total amount or the concentration of biomass in the broth. For many processes maximizing the growth rate and a high output of biomass is the primary objective, but even if other metabolites are of interest the amount of cells play a key role. They pose as the micro-factories of the target compounds and a higher concentration in the broth translates to a higher production potential [4, 56].

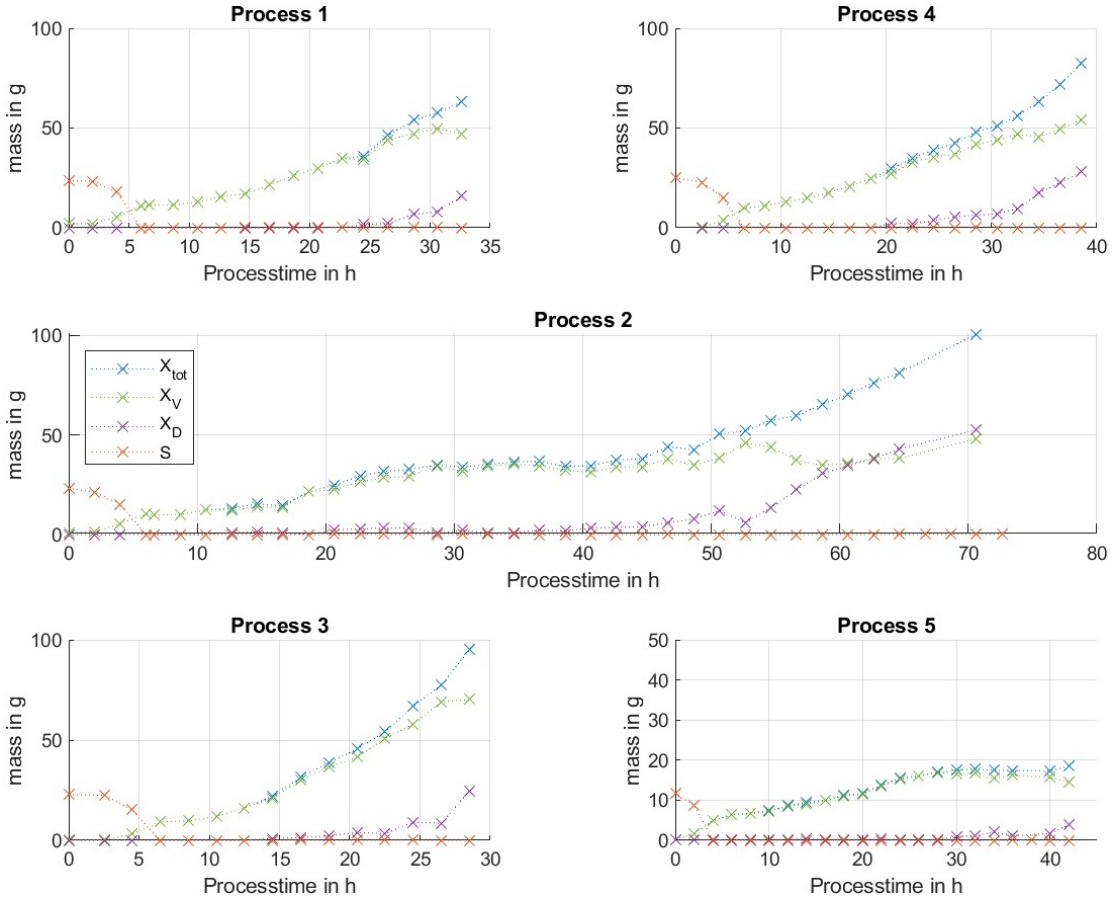

Figure 3.3: Results for the total X_{tot} (blue), alive X_V (green) dead X_D (purple) biomass and glucose in the broth S (orange) using DCW, FCM and Cedex measurements. Batch phase from 0 to 5 h, followed by uninduced fed-batch up until induction at ~24 h.

For the conducted experiments the offline biomass measurements were split into alive and dead cells by considering the FCM-analysis in fig. 3.3. In addition the substrate content is shown, indicating possible overfeeding. The time-resolved data gives insight on the processes taking place during the cultivations. In the graphs the switches between the three stages of these cultivations are observable. First, the end of the batch-phase was signaled by the depletion of the substrate in the broth. The following fed-batch was carried out with substrate limitation to control the growth rate which was characterized by seemingly no present substrate but significant increase in biomass. Finally, the induced fed-batch resulted in lower growth rates and a visible increase in cell death.

3.1.2 Product measurements

Three different methods to measure the LDH content of the samples were carried out and compared against each other. The two described protein assays were the preferred analyses due to a smaller experimental effort and faster output. But their results had to be compared and correlated with the more reliable RP-HPLC measurements to confirm their validity. For that matter a selection of samples were analyzed with all three methods. First, the RP-HPLC chromatograms for the soluble and refolded content of the samples were assessed. As seen in fig. 3.4, the soluble fraction was accompanied by other proteins that eluted close to the LDH peak, contrary to the refolded sample, where the response was cleaner.

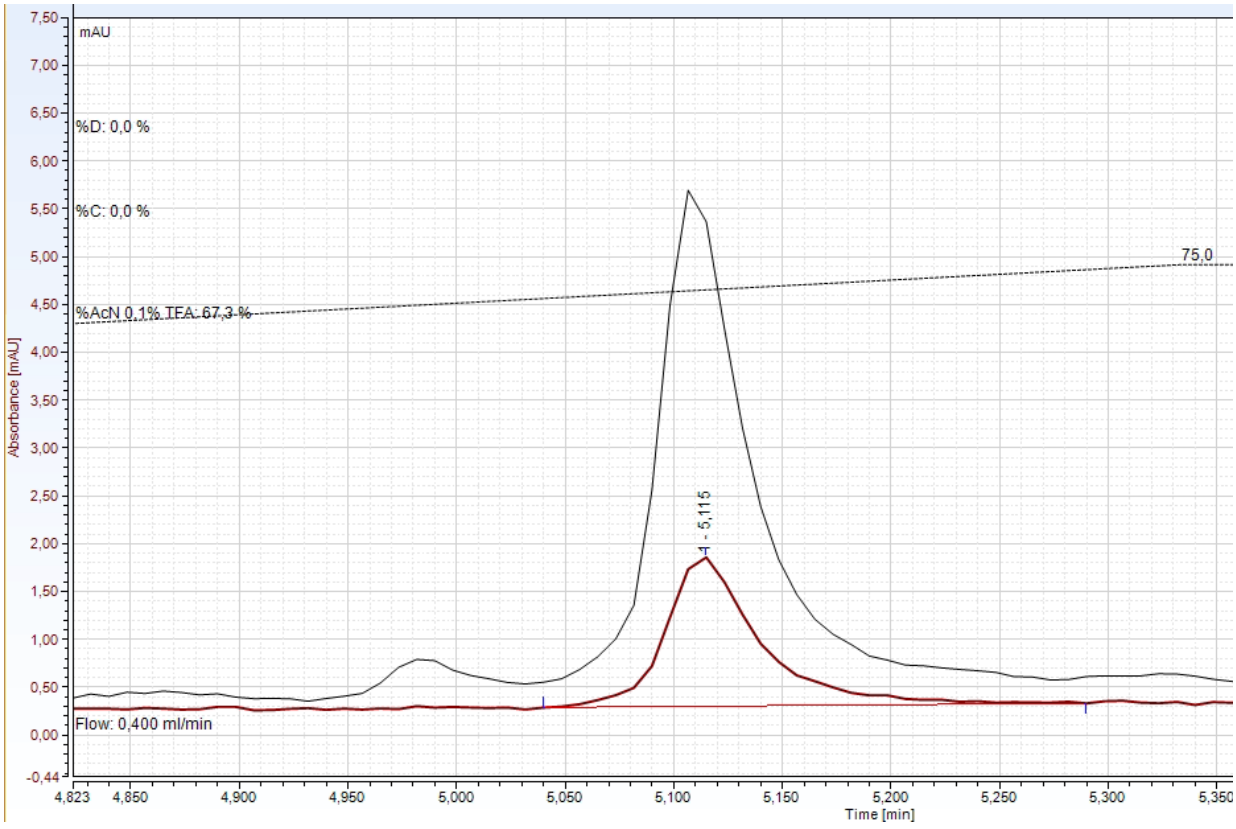

Figure 3.4: Chromatogram of the soluble fraction (black) and after refolding (red) at 8 h after induction

The SDS-PAGE of fig. 3.5 clearly showed the difference in the samples purity towards the targeted enzyme and the bandwidth of accompanying particles. Due to that matter only the RP-HPLC results of the refolded enzyme were used to compare the methods.

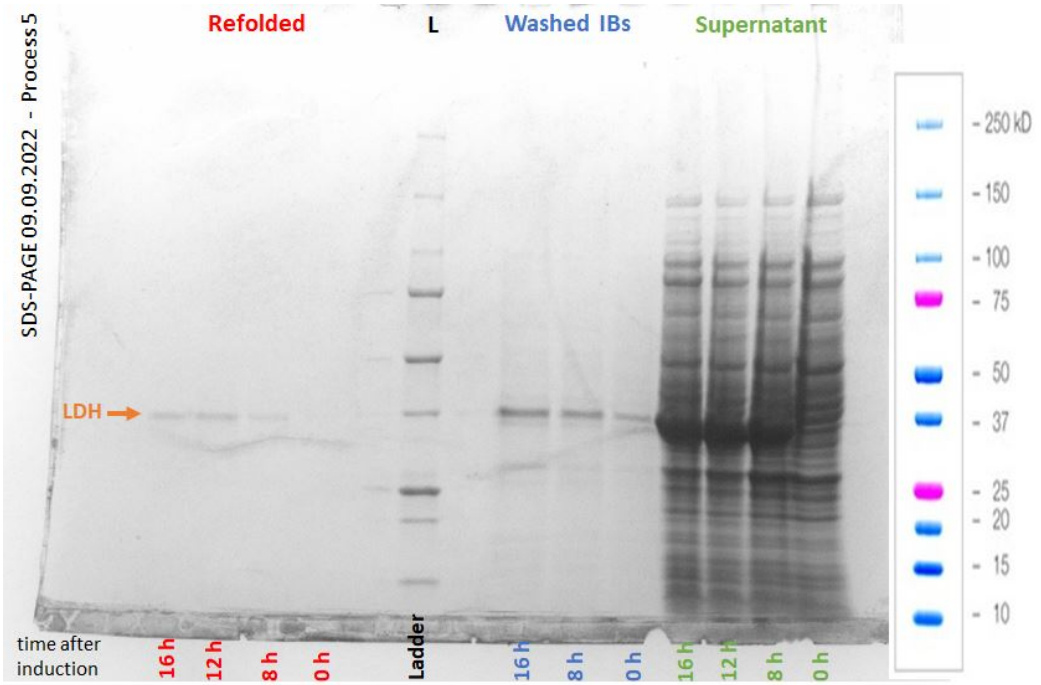

Figure 3.5: SDS-PAGE for selected samples of process 5. Soluble (green), washed IBs (blue) and refolded IBs (red). The mass of LDH was marked (orange), see sec. 2.4.4.

The evaluation of the measurement methods was done by plotting them against each other and looking for linear trends. In fig. 3.6 it can be seen that the higher enzymatic activity translated to higher concentrations in the BCA assay. But at the lower concentrations the test got unreliable, which was evident by the wide range of concentration values for activities close to zero and a poor R^2 of 0.39 for the linear regression. Since the BCA assay is unspecific to the type of proteins present an offset was expected but from the different processes no correlation was found to its magnitude. Further, the RP-HPLC plot showed no correlation with the BCA results, which made this assay impractical for the LDH determination.

On the other hand, a clear connection between the enzymatic activity and the RP-HPLC measurements was identified. Subsequently, a linear regression was done with a R^2 value of 0.77 and the slope was used as a conversion factor of 53 400 U g^{-1} to translate the enzymatic assay results into LDH concentration values. These values were ultimately used for the modeling of the productivity.

The product mass in the fermentation broth was obtained by multiplying the measured and converted activity results with the volume of the broth. Fig. 3.7 shows the resulting data for all 5 processes separated into soluble LDH and IBs, as well as their total sum. The behaviour exhibited over all processes was first a strong rise in product followed by a flattening of the curve. Processes carried out for a longer period like process 2 even showed a decline in product mass detected by the measurement methods.

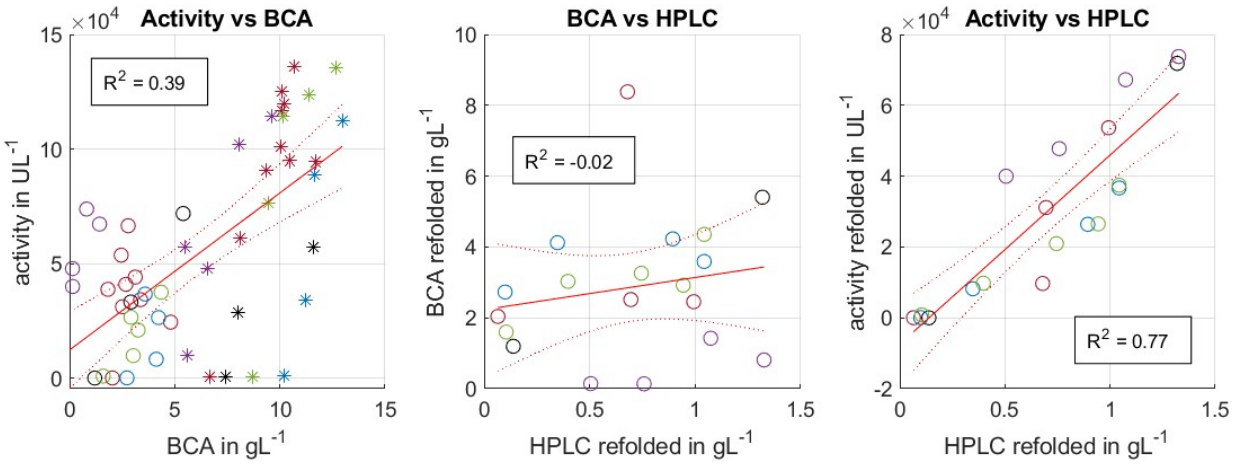

Figure 3.6: Plots of the measurements of the 3 product evaluation methods against each other. ∘ refolded, ∗ soluble, Process 1-5: black, red, blue, green, purple

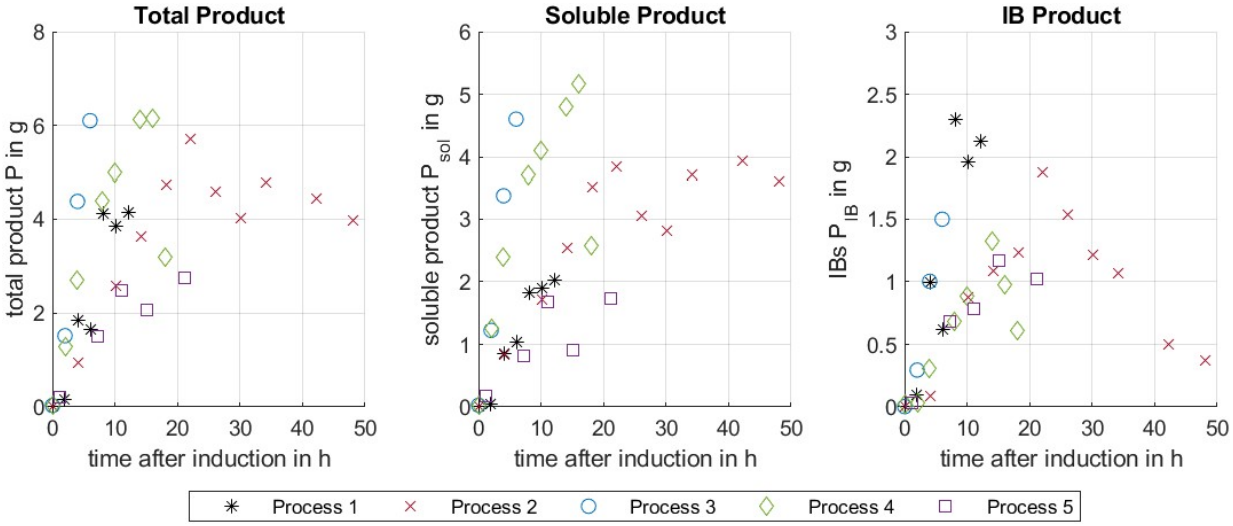

Figure 3.7: Amounts of total (left), soluble (middle) and IB (right) product measured for each process.

As mentioned in chapter 2, the values for IBs were obtained through a refolding procedure, which is accompanied by losses. For this work these losses were assumed to be a constant factor describing the difference between true and measured values. This would cause only a scaling effect on the dynamics of the process and was further neglected for modeling. The modeled IB formation and output therefore represents the mass obtainable through a downstream processing similar to the presented.

3.1.3 Correlations of productivity and physiological variables

In order to create a model for the productivity, the correlations between various physiological variables and the product formation were explored. First, the total specific product formation rate $q_{P,total}$ was investigated before turning towards the inclusion body formation. Inspired by Kager et

al. [44] a connection between the specific substrate uptake rate q_S and the enzyme expression rate was hypothesised. Similar to Reichelt et al. [77] they also included the amount of metabolised substrate S_{met} into their models, which acted as a time-independent measure for the metabolic load and age of the cells. To make the variable more comparable between different cultivations and more representative of the metabolic load of the cells, it was divided by the cell mass consuming the specific metabolised substrate s_{met}, as seen in eq 3.1.

$$s_{met}(t) = \frac{S_{met}(t)}{X_{tot}(t)} \tag{3.1}$$

Subsequently, the two variables were plotted against the total productivity in fig. 3.8, where two general trends were observed. The total productivity $q_{P,total}$ was overall declining with the progressing specific metabolised substrate s_{met}. Secondly, a slight enhancing influence of q_S could be observed, although the measurement points stayed rather grouped by their respective processes and a linear regression only scored a poor R^2 value of 0.1.

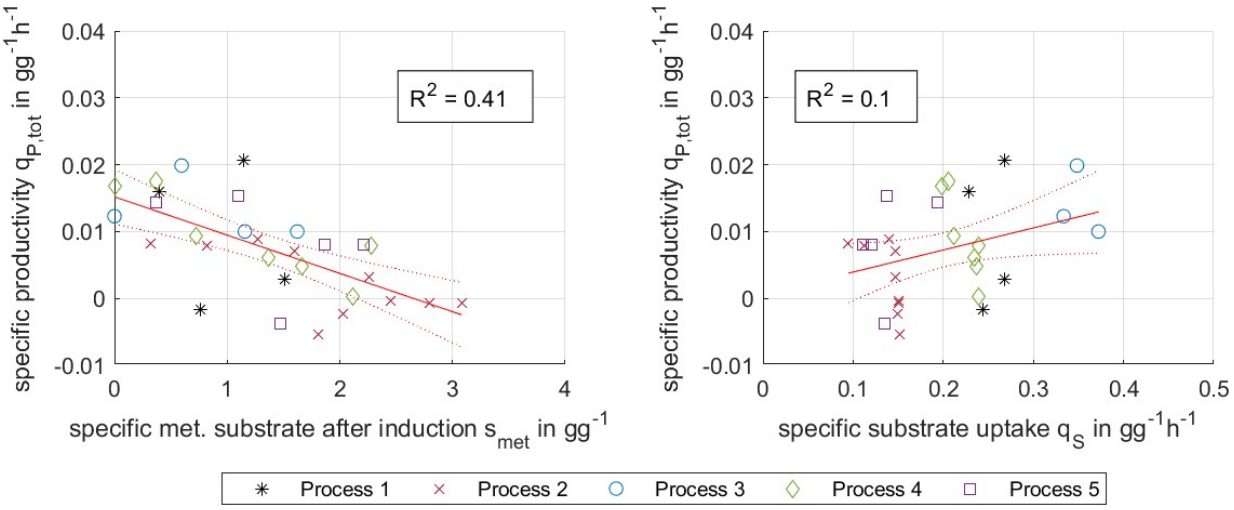

Figure 3.8: Correlation of the total specific productivity $q_{P,tot}$ with the specific metabolised substrate s_{met} (left) and the specific substrate uptake rate q_S (right)

Since the latter connection was weak, the maximum values of the specific product formation rate $q_{P,tot,max}$ of each process were examined additionally. When plotted against the same variables as before, a stronger linear connection between the substrate uptake rate and the maximal productivity was evident in fig. 3.9. Meanwhile no significant correlation with s_{met} was identified other than the maximum being generally located at the beginning of the process and the ones occurring later in the process tend to be lower, as seen for process 2 and 5.

These findings in the data were the motivation to use the specific substrate uptake rate q_S and the metabolised substrate per biomass s_{met} as the two variables describing the model. Their combined connection with the specific productivity $q_{P,tot}$ and the product titer specific to the cell mass $P_{tot,specific}$ is shown in fig. 3.10.

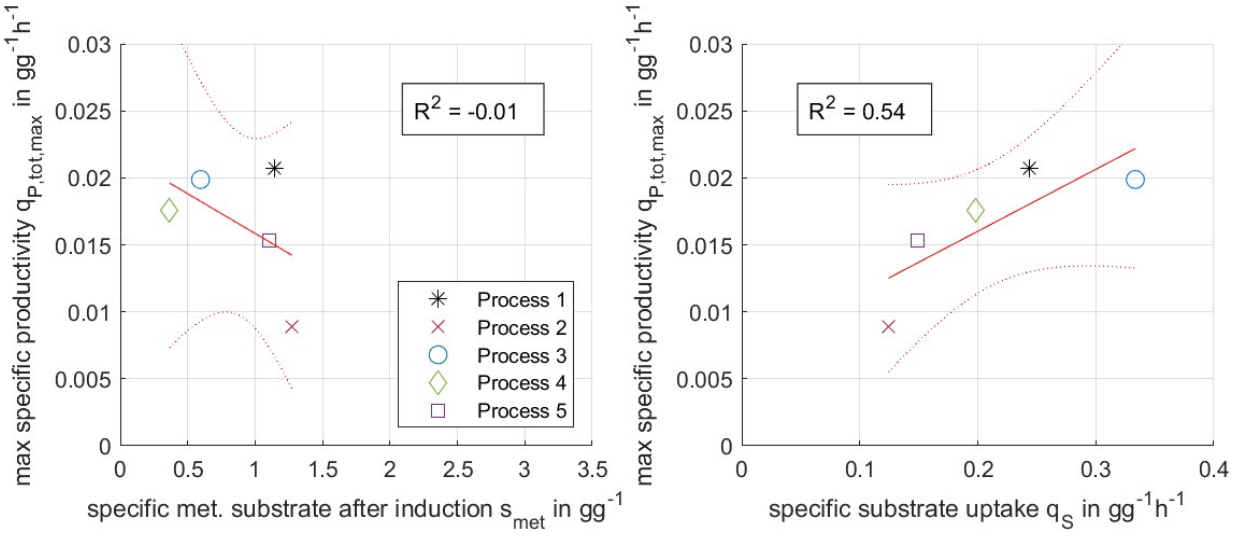

Figure 3.9: Correlation of the maximum total specific productivity $q_{P,tot,max}$ with the specific metabolised substrate s_{met} (left) and the specific substrate uptake rate q_S (right)

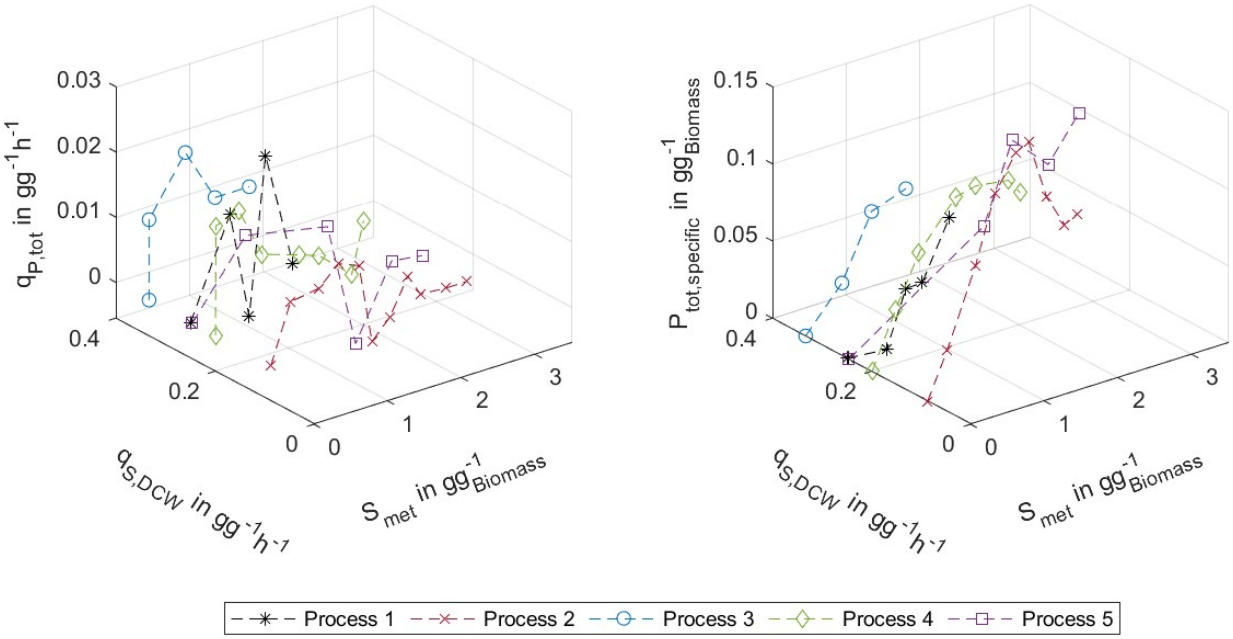

Figure 3.10: 3D plots of the measured total specific productivity $q_{P,tot}$ (left) and specific product titer $P_{tot,specific}$ (right) with the specific substrate uptake rate q_S and the specific metabolised substrate s_{met} for each process

After identifying the key contributors to the total productivity, connections between variables and the inclusion body formation were explored. As seen in fig. 3.11 both fractions of the product were linearly connected to the total production rate $q_{P,total}$. It was hypothesised that the expressed enzymes aggregate into IBs stronger when the expression system is working harder or when bigger amounts of product are already present. Therefore it was decided to test the total productivity as the single variable for the IB formation $q_{P,IB}$.

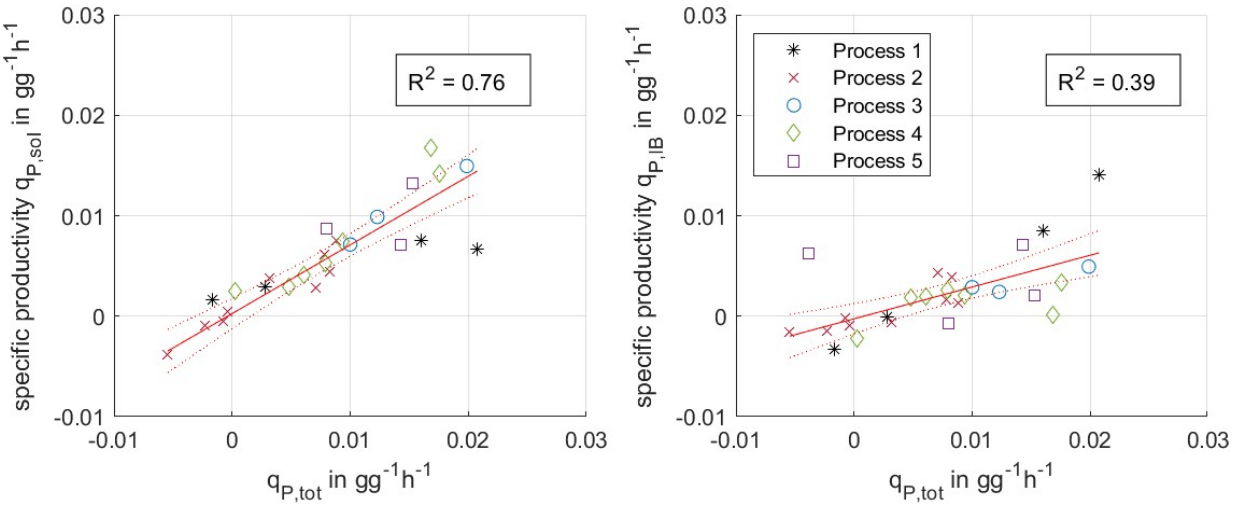

Figure 3.11: Correlation of the specific soluble productivity $q_{P,sol}$ (left) and the specific IB formation $q_{P,IB}$ (right) with the total specific productivity $q_{P,tot}$

3.2 General model adaptations

Before starting with investigating the kinetic equations to model the product and IB formation, the original bioprocess model had to be adapted to suit the characteristics of an induced production process. Upon analysing the existing model structure and the nature of the examined fermentations, the decision was to divide the process into two phases, before and after induction with IPTG.

3.2.1 Model structure

The model was partitioned into a pre- and post-induction structure, where the amount of induction agent represented by the state-variable I acted as a trigger. As soon as IPTG is added to the system ($I > 0$) the kinetics within the model change. Since the induction agent I was not measured, it was introduced into the model through F_{in}. There it acted as a second input source with a predefined trajectory in form of a bolus shot. In order to implement this phase-switch the original state-update equation in eq. 2.22 had to be adapted. An additional product yield matrix Y_P was introduced to account for the direct effect of the product mass on the states and the entries of the existing yield matrix Y_X had to be updated. The general form of the calculating equation is shown in eq. 3.2.

$$\frac{dx}{dt} = C_{in} \cdot F_{in} + \sum_{j=1}^{m} F_{out,j} \cdot c + Y_X \cdot q_i \cdot X + Y_P \cdot q_i \cdot P \tag{3.2}$$

The data exploration yielded the hypothesis that the productivity is connected to the amount of metabolised substrate S_{met} and the specific substrate uptake rate q_S. Furthermore the model should describe the kinetics of the product aggregating into IBs and how much stays soluble. Consequently the state vector x had to be extended, as well as the sizes of the other matrices, which is depicted in eq. 3.3.

$$x = \begin{bmatrix} V \\ X_V \\ X_D \\ S_1 \\ S_2 \\ P \\ I \\ CO2 \\ O2 \end{bmatrix} \quad \rightarrow \quad x = \begin{bmatrix} V \\ X_V \\ X_D \\ S_1 \\ S_2 \\ P \\ I \\ CO2 \\ O2 \\ S_{met} \\ P_{sol} \\ P_{IB} \end{bmatrix} \quad \dots \quad \begin{array}{l} \text{Volume of broth } [L] \\ \text{Cell mass viable } [g] \\ \text{Cell mass dead } [g] \\ \text{Mass of substrate 1 } [g] \\ \text{Mass of substrate 2 } [g] \\ \text{Mass of product } [g] \\ \text{Mass of induction agent } [g] \\ \text{Cummulated } CO_2 \text{ produced } [g] \\ \text{Cummulated } O_2 \text{ produced } [g] \\ \text{Mass of metabolised substrate } [g] \\ \text{Mass of product soluble } [g] \\ \text{Mass of product in IBs } [g] \end{array} \tag{3.3}$$

In order to account for the accumulation of product inside the cells, the specific productivity of the different conformations had to be incorporated into the model equations. Therefore, the specific rates were assigned to their respective positions within the biomass yield matrix Y_X, since the production of proteins is catalysed by the biomass. This caused the matrix to transform from being filled with constants into a state-dependent variable itself, similar to the product yield matrix Y_P. The change comes into effect as soon as the induction is triggered, which is schematically depicted by eq. 3.4.

$$Y_X = \text{const.} \;\&\; Y_P = 0 \quad \xrightarrow{\text{Induction}} \quad Y_X = f(x) \;\&\; Y_P = g(x) \tag{3.4}$$

This work followed the approach to first model the total specific productivity $q_{P,tot}$ and afterwards the specific IB formation rate $q_{P,IB}$ to account for the localisation of the expressed enzyme. The specific soluble product formation rate $q_{P,sol}$ results from a mass balance for the protein. The procedure of finding the optimal kinetics is described in section 3.3. The productivity rates were implemented in the biomass yield matrix Y_X as seen in eq. 3.5. In order to avoid computational issues the according entries for Y_X were set to zero in cases where q_S was zero.

$$Y_X(q_P/q_S) = \frac{q_P}{q_S} \;, \quad Y_X(q_{P,sol}/q_S) = \frac{q_P - q_{P,IB}}{q_S} \;, \quad Y_X(q_{P,IB}/q_S) = \frac{q_{P,IB}}{q_S} \tag{3.5}$$

Over the course of the fermentation the data showed not only a decrease in the specific productivity but also a decline of specific product in the broth towards the end of the process. The hypothesis was that cell death with the consequent lysis and connected degradation process at the end of the cells life caused a loss of product. A further theory was, that after the cell death the cell membrane became leaky and the product was released into the broth. There, either the present environment or the conditions of the sample processing resulted in a denaturation of the product, or the soluble part was simply not recovered after the centrifuging step. All these theories were combined in the approach to connect the product loss to the cell death, which was implemented through the product yield matrix Y_P as seen in eq. 3.6.

$$Y_P(q_P/q_D) = -1 \;, \quad Y_P(q_{P,sol}/q_D) = -\frac{P_{sol}}{P} \;, \quad Y_P(q_{P,IB}/q_D) = -\frac{P_{IB}}{P} \tag{3.6}$$

For the new state variable S_{met} the respective Y_X entry in eq. 3.7 was changed from zero to one after the induction. Thereby it represented the accumulation of the substrate taken up by the organism and used for the production of the enzyme.

$$Y_X(q_{S_{met}}/q_S) = 1 \tag{3.7}$$

The resulting yield matrix related to the biomass Y_X and to the product Y_P with their respective entries is summerized in eq. 3.8.

$$Y_X = \begin{array}{c} \begin{matrix} q_{S1} & q_{S2} & q_D \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 \\ Y_{X/S} & 0 & -1 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{q_P}{q_{S1}} & 0 & 0 \\ 0 & 0 & 0 \\ q_{CO2} & 0 & 0 \\ q_{O2} & 0 & 0 \\ 1 & 0 & 0 \\ \frac{q_P - q_{P,IB}}{q_{S1}} & 0 & 0 \\ \frac{q_{P,IB}}{q_{S1}} & 0 & 0 \end{bmatrix} \end{array} \begin{array}{l} V \\ X_V \\ X_D \\ S_1 \\ S_2 \\ P \\ I \\ CO2 \\ O2 \\ S_{met} \\ P_{sol} \\ P_{IB} \end{array} \qquad Y_P = \begin{array}{c} \begin{matrix} q_{S1} & q_{S2} & q_D \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\frac{P_{sol}}{P} \\ 0 & 0 & -\frac{P_{IB}}{P} \end{bmatrix} \end{array} \begin{array}{l} V \\ X_V \\ X_D \\ S_1 \\ S_2 \\ P \\ I \\ CO2 \\ O2 \\ S_{met} \\ P_{sol} \\ P_{IB} \end{array} \tag{3.8}$$

3.2.2 Changes in the growth metabolism

Since the production of LDH consumes energy and resources from the metabolic cycle, the biomass to substrate yield $Y_X(q_X/q_S)$ was updated at the point of induction. The metabolised substrate is divided between cell growth and the expression of the product enzyme, therefore the yield from substrate to biomass is smaller than before the addition of IPTG. Furthermore the cell death factor k_D had to be adapted for after the induction, since an increase was observed through the FCM measurements. This was believed to be connected to the increased metabolic stress on the cells during the production phase [91, 84]. It was assumed that the parameters for the Monod kinetic of the specific substrate uptake q_S stay unchanged throughout the whole process.

The growth and death parameters were fitted twice with all five processes, once for the pre-induction phase and then for the production phase. Only after reaching a globally satisfying parameter set describing the biomass, the modeling was continued for the productivity of the *E. coli* strain. The final parameter set for the Monod kinetic of the substrate uptake, the biomass-to-substrate yield and the cell death rate are shown in tab. 3.2.

The results for each process with the fitted growth and death model are shown in fig. 3.12. The simulation showed a good tracking of the total biomass and substrate measurements, but some variations were evident for the differentiation between viable and dead cells. The offset started to be evident post-induction, with the biggest deviations for process 5. The observed dying rate suggests that the dynamic is possibly of higher order than a constant. This difference in predictive performance was shown through the NRMSE values for the states, expressing the remaining error

Table 3.2: Parameter values for growth and death related kinetics after optimization

Symbol	Name	Parameter values		Units
		pre-induction	post-induction	
$Y_X(q_X/q_S)$	biomass-to-substrate yield	0.3903	0.2331	$\mathrm{g\,g^{-1}}$
k_D	cell death rate	$0.86 \cdot 10^{-2}$	$2.73 \cdot 10^{-2}$	$\mathrm{g\,g^{-1}\,h^{-1}}$
$q_{S,max}$	max substrate uptake rate	1.7913		$\mathrm{g\,g^{-1}\,h^{-1}}$
K_S	Monod constant	$7.02 \cdot 10^{-2}$		$\mathrm{g\,g^{-1}\,h^{-1}}$

between prediction and measurement. The total biomass X_{tot} was described the best followed closely by the viable biomass with 0.0503 and 0.0538 NRMSE respectively. The dead cell mass had a higher prediction error at a NRMSE value of 0.0991. Nevertheless, the results were deemed as satisfying and the given model and parameter fit were used as basis for the modeling efforts.

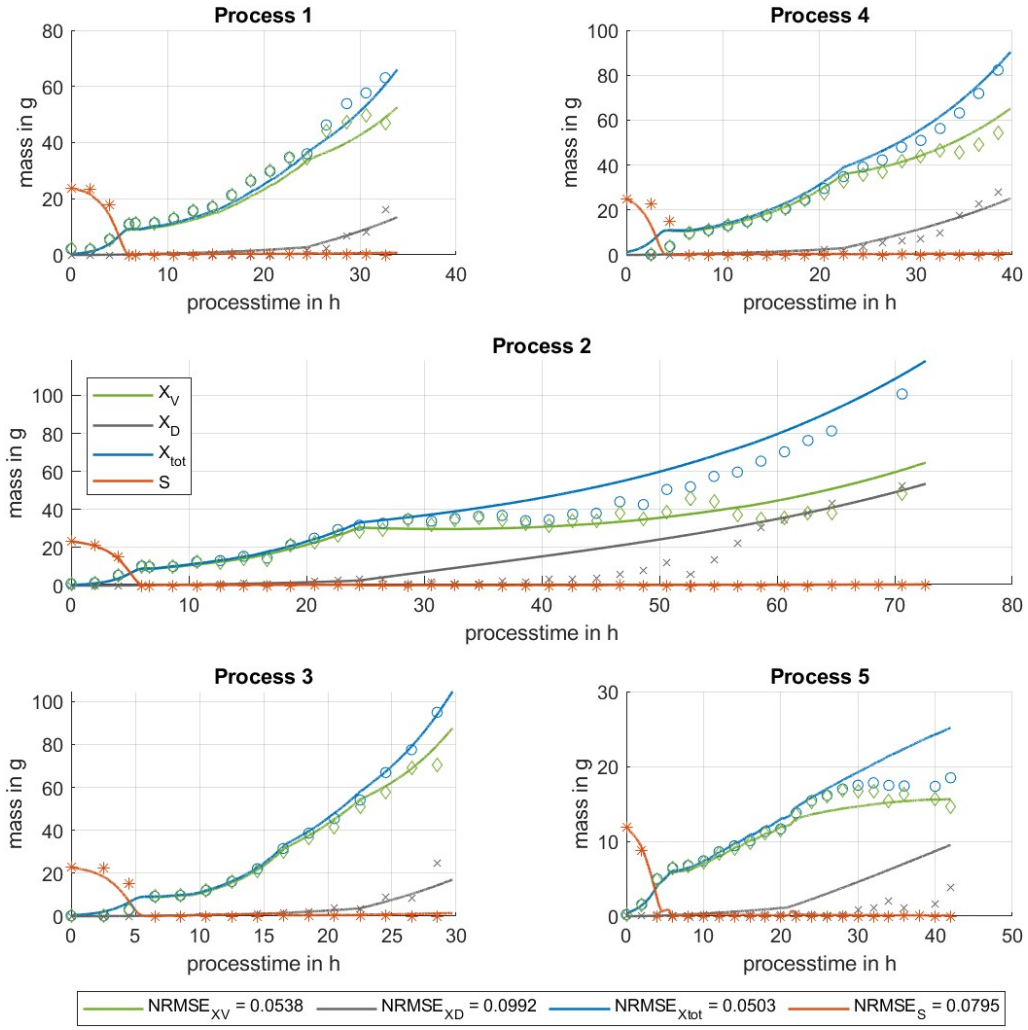

Figure 3.12: Simulation and measurements for each process with the fitted growth model: total biomass (blue,○), biomass alive (green,◇) & dead (grey,×), substrate (orange,∗).

3.2.3 Data set separation for training and validation

In data science, measurements are separated into training and validation sets in order to have a quantified statement about the models behaviour on data that has not directly been used to fit the parameters. Usually a distribution of the data with 80 % for parameterization and 20 % for validation is done. When a big number of measurements is available and their sampling is independent of each other, this allocation is done randomly and repeated several times for cross validation. In bioprocesses often only small amounts of samples can be obtained due to the experimental effort. This work further kept samples from one fermentation together and did not break them up between training and validation data. This resulted in limited possibilities for data combinations by assigning whole fermentations as training and validation bioprocesses. Furthermore, the individual data sets differed in amount of sample points due to the premature terminations explained in section 3.1. The respective shares in the collected data total used for the parameterization of the productivity is shown in fig. 3.13.

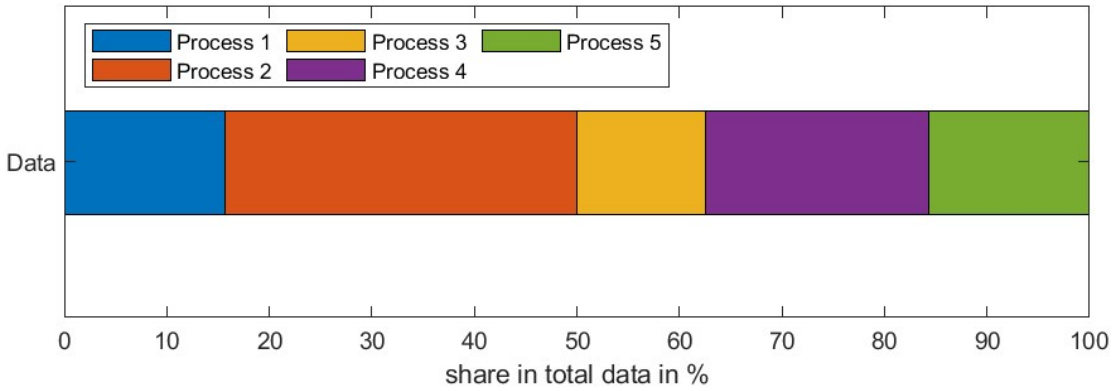

Figure 3.13: Share of processes on the total data used for modeling the productivity

Due to the small amount of possible combinations for the validation sets, all of them were investigated to find the best performing allocation. Since process 2 accounts for over 30 % of the data it was ruled out as a validation set, since it resulted in big variances in the obtained values or caused a model to be unidentifiable altogether. The others were considered separately to reach a similar division between training and validation data, which was closer to a 85-15 % ratio, due to the data set lengths. The parameterization later yielded process 3 as the optimal validation set.

3.2.4 Kinetic terms to model specific rates

In order to model the behaviour of specific rates, kinetic equations are implemented. An often used example is the Monod kinetic, which was already introduced in chapter 2.4.1 for the substrate uptake. These kinetic terms were utilized to express the dynamics of the specific rates with correlating process variables. Tab. 3.3 lists kinetic equations that typically characterize the growth on a substrate and have a promoting effect on the specific rate. The higher the values of the driving variable x_1, the closer q_i is to its maximum value. Tab. 3.4 on the other hand, shows several inhibition terms that dampen the specific rate with increasing values in x_2 [84]. The symbols x_1 and x_2 represent an arbitrary bioprocess variable.

Table 3.3: Kinetic terms promoting the productivity, described by Schügerl et al. [84]

Name	normalized kinetic equation $\tilde{f}_{promote}(x_1,(x_2))$
Monod	$\frac{x_1}{x_1 + K_{s,x_1}}$
Moser	$\frac{x_1^N}{x_1^N + K_{s,x_1}^N}$
Competitive[a]	$\frac{x_1}{x_1 + K_{s,x_1} \cdot (1 + \frac{x_2}{K_{x_2}})}$
Uncompetitive[a]	$\frac{x_1}{x_1 \cdot (1 + \frac{x_2}{K_{x_2}}) + K_{s,x_1}}$
Han[a]	$\frac{x_1}{x_1 + K_{s,x_1} \cdot (1 - \frac{x_2}{K_{x_2}})^M}$

[a] Inhibition term included in promoting kinetic

Table 3.4: Additional kinetic terms for inhibition, described by Schügerl et al. [84]

Name	normalized kinetic equation $\tilde{f}_{inhibit}(x_2)$
Ierusalimsky (non-competitive)	$\frac{1}{1 + \frac{x_2}{K_{x_2}}}$
Yano and Koya	$\frac{1}{1 + (\frac{x_2}{K_{x_2}})^N}$
Levenspiel	$(1 - \frac{x_2}{K_{x_2}})^N$
Luong	$(1 - (\frac{x_2}{K_{x_2}})^N)$
Haldane	$\frac{x_2}{\frac{x_2^k}{K_{I,x_2}} + x_2 + K_{S,x_2}}$
Webb	$\frac{x_2 \cdot (1 + \frac{x_2}{K_1})}{K_{S,x_2} + x_2 \cdot (1 + \frac{x_2}{K_2})}$

These normalized kinetic expressions were combined to described the LDH production and IB formation and had to be multiplied with a maximum rate $q_{i,max}$, as indicated in eq. 3.9. The resulting modeling equations were then tested on the data to find a combination that described the measurements sufficiently well.

$$q_i = q_{i,max} \cdot \tilde{f}_{promote}(x_1) \cdot \tilde{f}_{inhibit}(x_2) \tag{3.9}$$

3.3 Modeling the total specific productivity q_P

As mentioned before, the total specific productivity q_P was modeled first disregarding the physical state of the expressed LDH enzymes. This was done due to the understanding that the cells gene expression mechanisms produce the enzymes RNA and consequently the protein in its basic form. During the folding process of the protein chains into the native LDH enzyme some of them fold correctly, while others miss-fold or are faced by other struggles and aggregate into IBs [12, 78]. The data exploration hinted at a connection between the total specific productivity q_P and the specific substrate uptake rate q_S. Furthermore, a decline of the productivity over the progress of the specific metabolised substrate s_{met} was identified. In order to include this specific variable eq. 3.1 was used, where the total biomass X_{tot} was build by adding up the viable cell mass X_V and dead cell mass X_D in the model. The observed correlations in chapter 3.1 lead to the decision to investigate kinetics equations with the overall structure of eq. 3.10.

$$q_P = q_{P,max} \cdot \tilde{f}_{promote}(q_S) \cdot \tilde{f}_{inhibit}(s_{met}) \tag{3.10}$$

3.3.1 Kinetic formulation and parameterization of q_P

In order to model the measured production of LDH several of the discussed combinations of productivity and inhibition terms were implemented. The performance after optimization of the respective parameter sets were evaluated with the methods described in chapter 2.4. The analysed combinations of kinetic terms are listed in tab. 3.5.

After the parameterization of the different kinetic combinations the goodness of fit quantified how suitable they were to explain the given measurements. Since different combinations for training and validation data sets yield different performance values for the same kinetic, the best separation for the measured data sets was determined through a cross validation. In a first step, process 5 was used for validation and the kinetic with the best score on the validation experiment was selected. With this chosen model all possible data set combinations were explored and their $R^2_{adjusted}$ and NRMSE values were compared. Then the best combined score of training and validation data was used to run another parametrization experiment with this data allocation for all kinetics. This either verified the chosen kinetic as best option or the procedure was repeated with a new kinetic until a satisfying solution was found.

Table 3.5: Examined combination of kinetic terms for productivity q_P

ID	$\tilde{f}_{promote}(q_S)$	$\tilde{f}_{inhibit}(s_{met})$
Monod	Monod	-
HalSolo	-	Haldane
Competitive	Competitive	(included)
Uncompetitive	Uncompetitive	(included)
Ieru	Monod	Ierusalimsky
Yano	Monod	Yano and Koya
Levenspiel	Monod	Levenspiel
Luong	Monod	Luong
Haldane	Monod	Haldane
Webb	Monod	Webb
Moser	Moser	Haldane
HL	Han	Levenspiel

The achieved values for $R^2_{adjusted}$ and the NRMSE with process 3 as the optimal validation set can be seen in fig. 3.14 and 3.15. Scores that had negative values were set to zero, which was the case when the model described the data set worse than the simple average over all measurements. Since the validation sets often consisted of fewer samples, the procedure of adjusting the R^2 values had to be skipped in order to obtain a result.

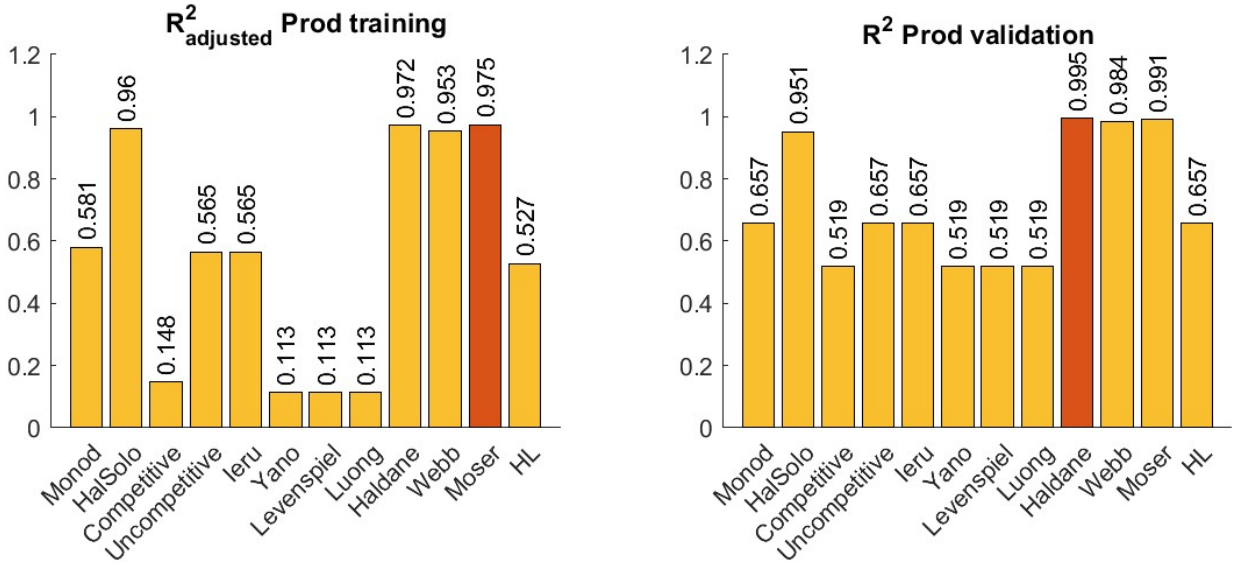

Figure 3.14: R^2_{adj} score for the training data and R^2 results for the validation data (process 3) for the total productivity q_P. Highest score was marked orange.

The results showed that the highest values for $R^2_{adjusted}$ as well as the lowest for NRMSE were scored for a Monod kinetic incorporating the specific substrate uptake rate combined with a Haldane term describing the inhibition through the increasing s_{met} over the course of the process.

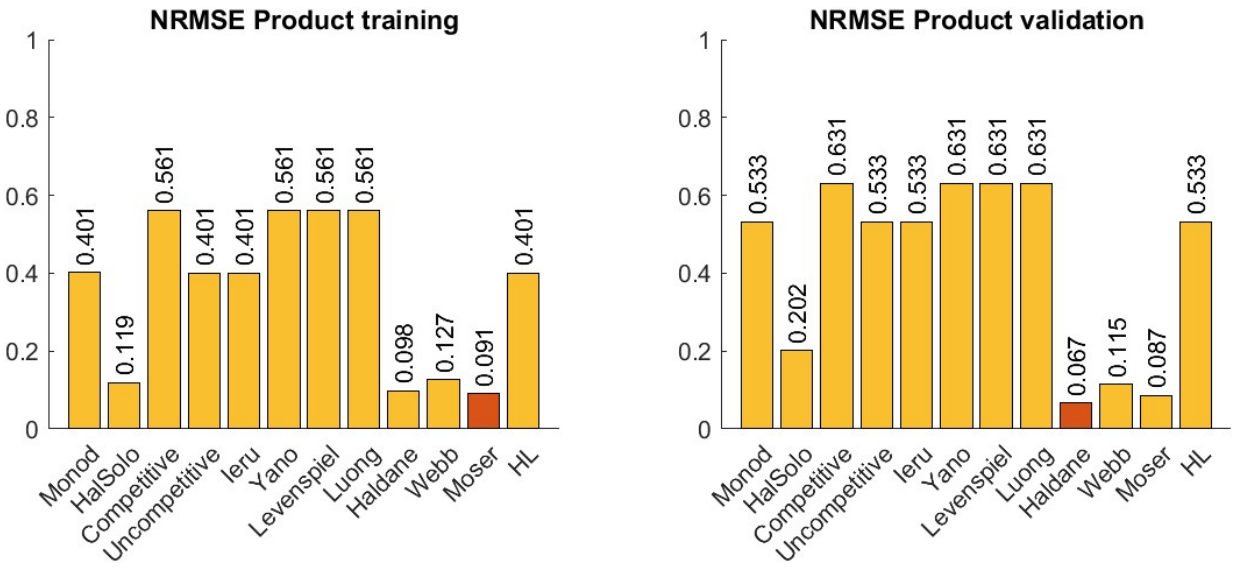

Figure 3.15: NRMSE scores for the total productivity q_P (orange = lowest score).

The fitted parameter set described the training and the validation data better than other options, where an emphasize was put on the low NRMSE values. The Moser kinetic was disregarded since it would introduce a sixth parameter, complicating the model further without providing significant improvements. Furthermore, the outcome for the cross validation can be seen in fig. 3.16, where the lowest combined NRMSE score for the training and validation set was marked green. The best overall fit for the chosen kinetic was achieved when processes 3 was used as validation.

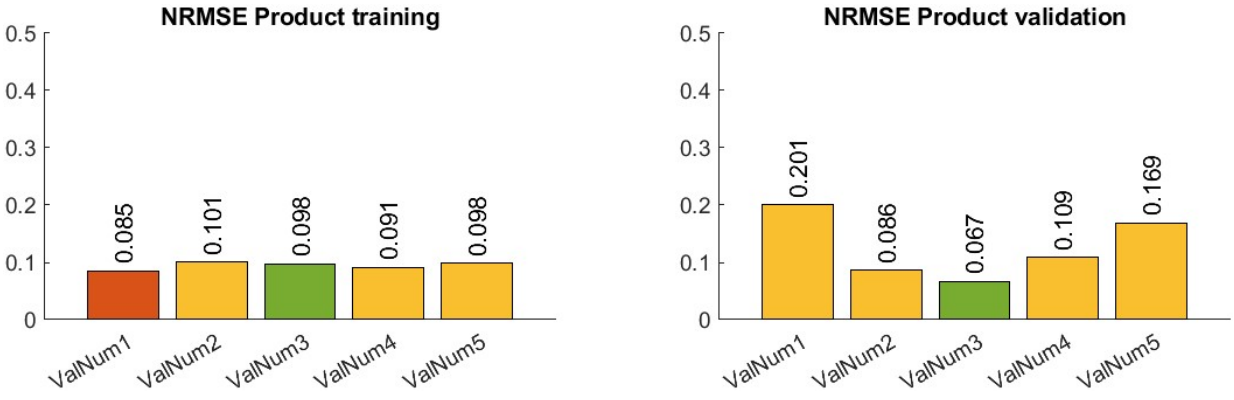

Figure 3.16: NRMSE scores for the total specific productivity q_P of different data set combinations. Best individual (orange) and combined score (green) for training and validation were marked.

3.3.2 Identifiability analysis - q_P

As mentioned in section 2.4 usually a structural identifiability analysis is done by evaluating the model symbolically before the parameterization, which was attempted with the STRIKE-GOLDD toolbox developed by Villaverde et al. [109]. But the computational effort of analysing a simplified model describing the total product was already too much and the software repeatedly crashed

during the construction of the nonlinear observability matrix. Therefore, structural and practical identifiability were determined numerically, which required to insert values for the respective set of model parameters and hence the parameterization was done beforehand. Additionally, the practical identifiability required the standard deviation of the state variables, listed in tab. 3.6. The standard deviation σ for the online measurements was taken from the respective sensor specifications and were calculated from the doublets and triplets of the measurements for the offline analyses.

Table 3.6: Variance levels for the state variables

Symbol	Name	σ	Unit
L^a	broth volume	0.1	mL
${X_V}^b$	cell mass alive	0.43 - 2.57	g
${X_D}^b$	cell mass dead	0.28 - 2.66	g
$S^{c,d}$	substrate	5	%
P^b	product mass	0.001 - 0.07	g
${CO_2}^a$	cummulated CO_2 produced	3	%
${O_2}^a$	cummulated O_2 produced	3	%
${S_{met}}^d$	metabolised substrate	5	%
${P_{sol}}^b$	soluble product mass	0.001 - 0.07	g
${P_{IB}}^b$	IB mass	0.001 - 0.06	g

[a] variance from measurement device specification, [b] calculated variance from repetitive measurements
[c] variance at least 0.02 g, since the detection limit of the CEDEX was 0.111 mmol L^{-1}
[d] standard deviation assumed higher than CEDEX specification of 0.7 %

The identifiability analysis was carried out for the four highest scoring kinetics after the parameter optimization, since they scored similarly well. The models were considered as structurally identifiable when their time-resolved sensitivity matrix S had full rank. This matrix was also used to calculate the conditional number of the system to assess and compare its quality. The same was done for the practical identifiability by constructing the **FIM** with eq. 2.33. The results are shown in fig. 3.17, where the conditional number of unidentifiable model structures was set to zero.

The conditional numbers for the practical and the structural identifiability assessments characterised all analysed models as ill-conditioned, since their scores exceeded 10^3. Naturally, the conditional number worsened when taking the uncertainty at the time points of the measurements into account to construct the **FIM** for the practical identifiability. The Moser kinetic performed best among the analysed models, followed by Haldane and HalSolo achieving similar results. Nevertheless, the Haldane kinetic was chosen as the model to describe the total product formation, since it performed more favourable describing the data. It was preferred over the HalSolo kinetic because it included the specific substrate uptake rate q_S, which was deemed as an essential influence on the productivity due to the findings of the data exploration in section 3.1. Furthermore, the Haldane kinetic implements a Monod term for the promoting effect of q_S, which poses as a simpler variant than the Moser kinetic.

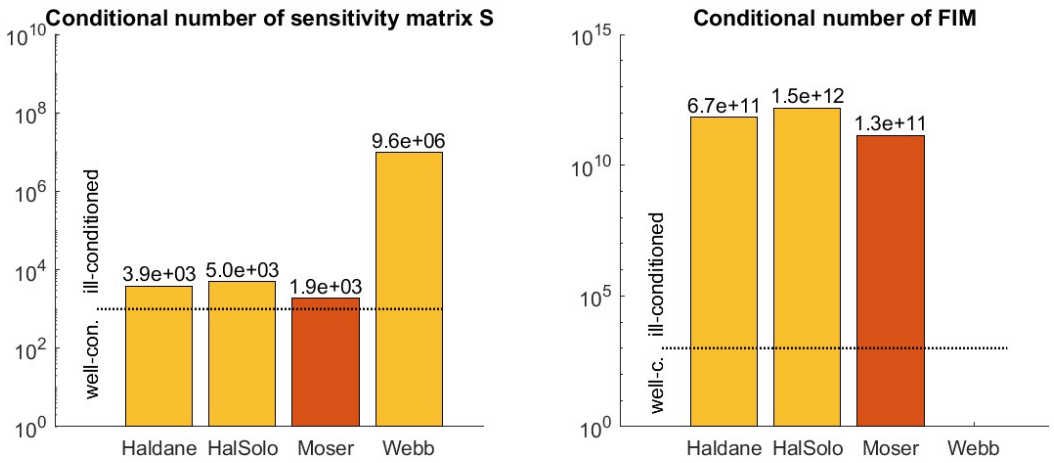

Figure 3.17: Conditional numbers for the structural (left) and practical (right) identifiability analyses of the model incorporating the productivity, including the border between well and ill-conditioned models. Lowest score emphasized in orange.

3.3.3 Parameter importance of chosen q_P kinetic

The results of the identifiability analysis were further used to calculate the significance of the model parameters δ_p^{msqr} on the model output in form of the produced product. Additionally, the general variance of the parameter values was obtained with the methods presented in section 2.4.3 from the main diagonal of the $\mathrm{Cov}_{min}(\hat{\theta})$. As seen in fig. 3.18, the parameters connected to the substrate conversion ($Y_{X/S}$, $q_{S,max}$ and K_S) had the biggest influence, since they were directly connected to the biomass formation which is the essential driver for the production of LDH. When taking a closer look at the productivity parameters it can be seen that all of them were of the same magnitude. Therefore, it was followed that all of the parameters of the chosen model kinetic had a significant impact on the product output and none of them were negligible.

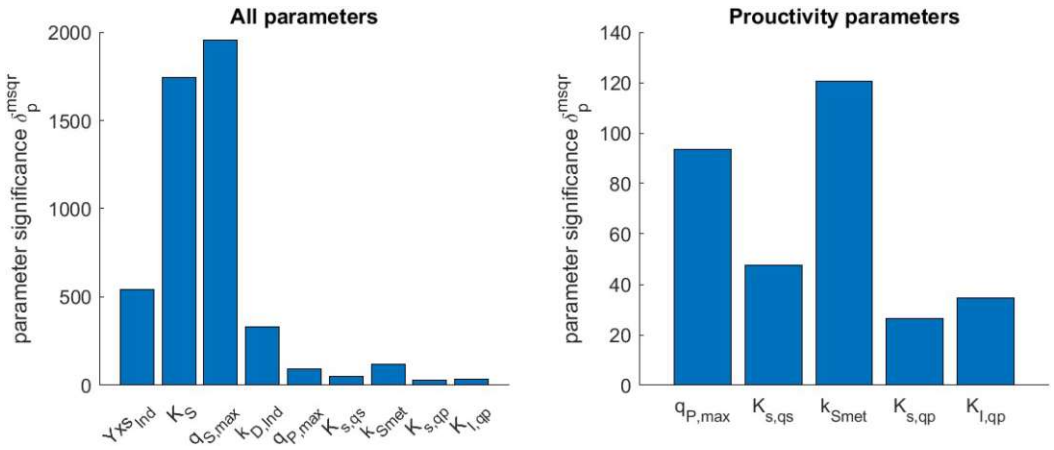

Figure 3.18: Parameter significance for all parameters associated with product formation (left) and just for the total productivity model (right)

The classification as an ill-conditioned model structure can lead to unstable solutions during the parameterization since noise in the data or numerical errors result in bigger variances and therefore poorer estimates of some parameters. This is due to the structure of the model which causes two columns of the sensitivity matrix **S**, and further of the **FIM**, to be almost parallel to each other. These columns represent otherwise structurally identifiable parameters, which means that they are difficult to be estimated together, which is connected to the non-linearity of the problem to a certain extent. The ill-conditioned nature can be partially circumvented by providing meaningful initial conditions and allowing only plausible intervals for the solution. It can further be improved by increased sampling in order to catch the full dynamic, which was a problem for the prematurely terminated experiments [30].

3.4 Modeling the IB formation $q_{P,IB}$

In order to create a model expressing the formation of IBs inside the cells, the same procedure as for the total specific productivity q_P was used. The kinetic terms introduced in section 3.3 were adapted and combined as done before. The previous training data allocation was used, but an additional cross validation procedure was carried out to compare its performance for the IB formation.

3.4.1 Kinetic formulation and parameterization of $q_{P,IB}$

It was hypothesised that the IB formation is connected to the intensity of the product expression and the data exploration showed a connection to the measured $q_{P,tot}$, which supports this theory. Therefore the specific substrate uptake rate q_S of the related terms in tab. 3.3 was substituted with the total specific productivity q_P. These adjusted kinetic terms were then combined with the different types of inhibition according to tab. 3.7 and tested for their suitability to match the measured IB content.

Table 3.7: Examined combination of kinetic terms for IB formation rate $q_{P,IB}$

ID	q_P related term $\tilde{f}_{promote}(q_P)$	Inhibition term $\tilde{f}_{inhibit}(s_{met}, x_{P,sol})$
Monod	Monod	-
Moser	Moser	-
HalSolo	-	Haldane[a]
Competitive	Competitive	(included)
CompBoost	Monod	Levenspiel[a,b]
Uncompetitive	Uncompetitive	(included)
Ieru	Monod	Ierusalimsky
Webb	Monod	Webb
Haldane	Monod	Haldane

[a] The specific soluble product concentration $x_{P,sol}$ was used for the inhibition term instead of s_{met}
[b] The signum of the x_i/K_{Smet} term was reversed to create an enhancement term

For two kinetics the inhibition term was also altered to use the specific soluble product $x_{P,sol}$ following the same principle as in eq. 3.1, instead of using s_{met}. It was theorized that the aggregation of the product into IBs was favored when the concentration of P_{sol} inside the cell reached higher levels. Furthermore for the CompBoost kinetic the Levenspiel inhibition was turned into an enhancement term by changing the signum of x_i/K_{Smet}, to test if the age of the cell causes more miss-folding.

The results for the optimised parameter sets of the several tested equations for IB formation are shown in fig. 3.19 and 3.20. For the validation set again the none-adjusted R^2 is compared, since the low number of measurements would skew the results too much. The analysis yielded that several kinetics performed equally well for the training and the validation data sets, where Haldane had the lowest NRMSE score. Since most of them differed only little from the optimal one, the top performers were all analysed for their structural identifiability before deciding on the final model.

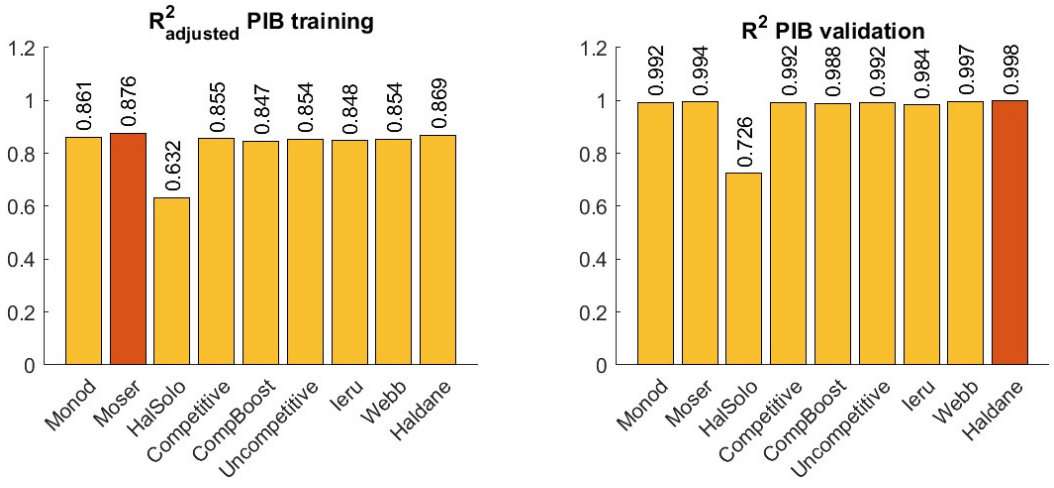

Figure 3.19: R^2_{adj} score for the training data and R^2 results for the validation data for the IB formation $q_{P,IB}$. Highest score was marked orange.

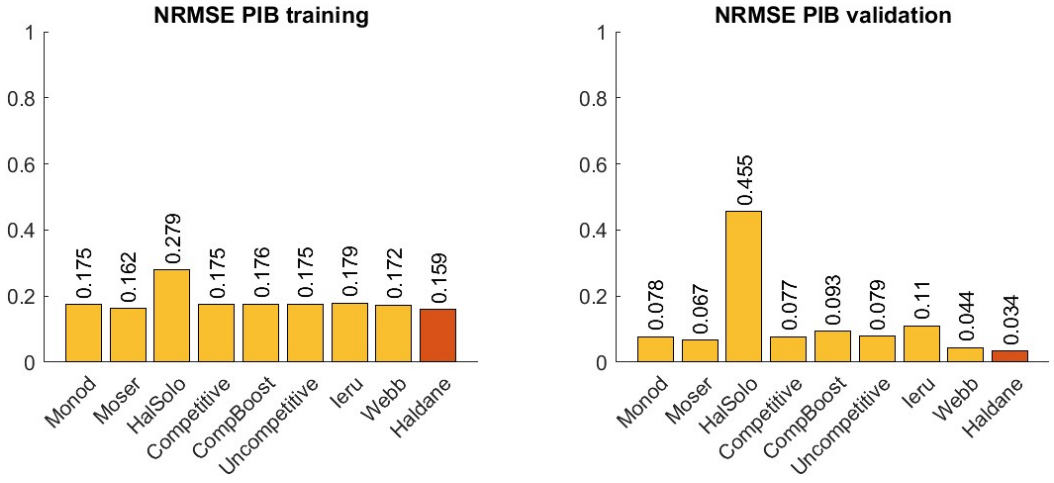

Figure 3.20: NRMSE scores for IB formation $q_{P,IB}$ on the training and validation data. Lowest score was marked orange.

The outcome of the cross validation can be seen in fig. 3.21, which happened to be the same data set separation as identified for the total specific productivity q_P.

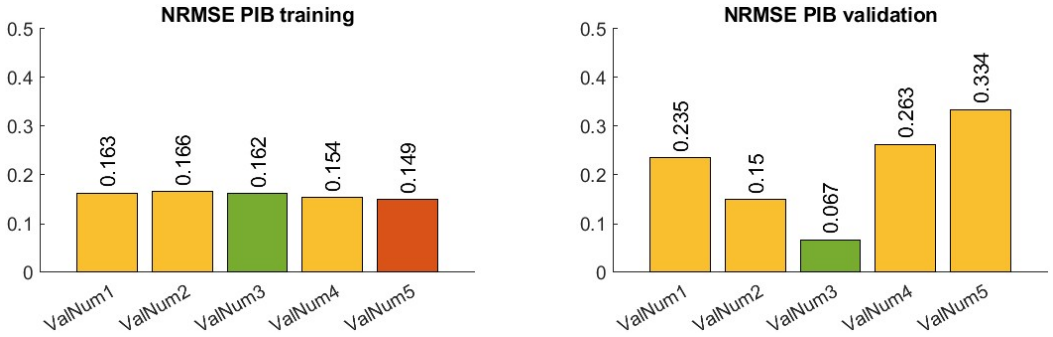

Figure 3.21: NRMSE score for the IB formation $q_{P,IB}$ of different data set combinations. Best individual (orange) and combined score (green) for training and validation were marked.

3.4.2 Identifiability analysis - $q_{P,IB}$

The structural and practical identifiability was determined for the five best scores of the parameterized IB kinetics. The results for the respective conditional numbers can be seen in fig. 3.22 on a logarithmic scale. The score for kinetics that did not have full rank and were therefore unidentifiable was set to zero in the practical identifiability analysis utilizing the **FIM**.

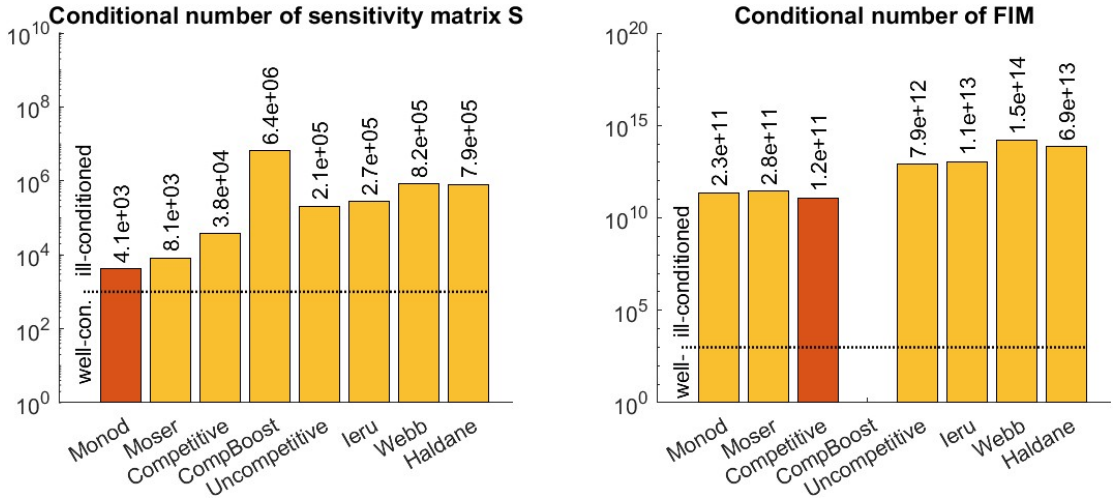

Figure 3.22: Conditional numbers for the structural (left) and practical (right) identifiability analyses of the model incorporating the productivity and IB formation, including the border between well and ill-conditioned models. Lowest score emphasized in orange.

The achieved conditional numbers from the sensitivity matrices for the IB formation qualified all kinetics as ill-conditioned as before for the total specific productivity. Several model possibilities scored similar on the lower spectrum. The Haldane kinetic, which described the data with the lowest deviations was deemed as unidentifiable by the practical identifiability analysis. Between the three options with the lowest conditional number, the Moser kinetic was chosen to describe the IB

formation of the process. It scored slightly better than the Monod term in the parameterization and although making the model structure more complex with an additional parameter, the influence was considered significant enough to be implemented.

3.4.3 Parameter importance of chosen $q_{P,IB}$ kinetic

The parameter significance δ_p^{msqr} for the model including the newly introduced kinetic is shown in 3.23. Due to the boxed approach of making the total specific productivity q_P a variable of the IB formation, the three new parameters describing $q_{P,IB}$ possessed the lowest impact on the model outputs. Nevertheless they were of comparable magnitude as the other product related parameters and thereby the choice for the kinetic equation was confirmed.

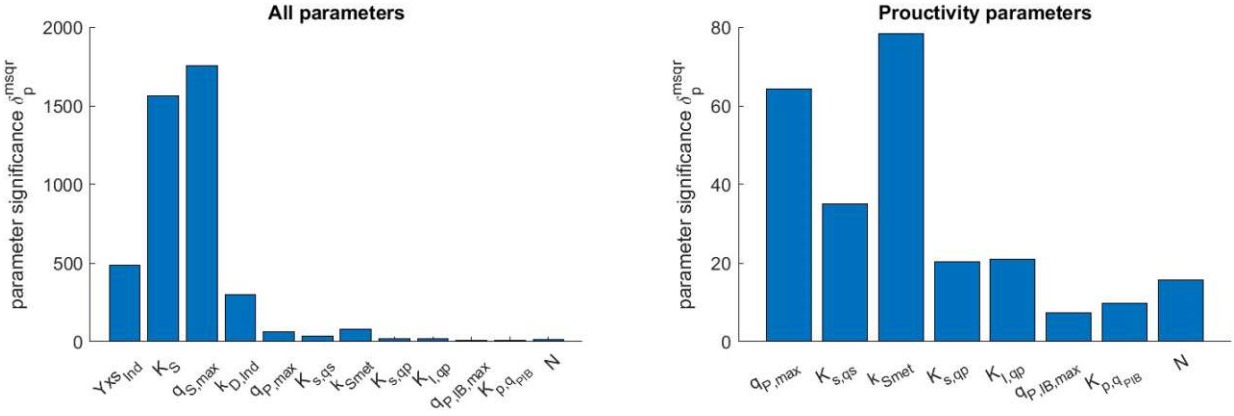

Figure 3.23: Parameter significance for all parameters associated with IB formation (left) and for the total productivity and IB model (right)

3.5 Final model & Discussion

The final model is able to describe the overall production and differentiation into soluble protein and the IB fraction, by implementing the equations shown in eq. 3.11 and 3.12. The construction of the total specific productivity q_P as a function of the specific substrate uptake q_S connects the enzyme expression rate with the overall metabolic activity. The combination with an inhibition term of Haldane type incorporates the decline of the cells performance over the course of the production phase. The specific IB formation rate $q_{P,IB}$ was best described with a simple Moser kinetic that utilizes the specific product formation rate q_P directly, making it a function of the metabolic stress due to product expression of the cell.

$$q_P = q_{P,max} \cdot \frac{q_S}{q_S + K_{s,qs}} \cdot \frac{s_{met}}{\frac{s_{met}^{k_{Smet}}}{K_{I,qp}} + s_{met} + K_{s,qp}} \tag{3.11}$$

$$q_{P_{IB}} = q_{P_{IB},max} \cdot \frac{q_P^N}{q_P^N + K_{p,q_{PIB}}^N} \tag{3.12}$$

The parameter values for the full model were obtained through the fitting experiments and are listed together with their standard deviations σ in tab. 3.8. The standard deviation was calculated as the square root of the parameter variance from the $\mathrm{Cov}_{min}(\hat{\theta})$ matrix.

Table 3.8: Parameter values for the productivity kinetics after parameterization

Symbol	Parameter name	Values	σ	Units
$q_{P,max}$	max specific productivity	$5.38 \cdot 10^{-2}$	$\pm 1.79 \cdot 10^{-4}$	$\mathrm{g\,g^{-1}\,h^{-1}}$
$K_{s,qs}$	productivity Monod constant	$14.5 \cdot 10^{-2}$	$\pm 1.13 \cdot 10^{-4}$	$\mathrm{g\,g^{-1}\,h^{-1}}$
k_{Smet}	Haldane exponent	6.50	$\pm 1.06 \cdot 10^{-2}$	1
$K_{I,qp}$	Haldane constant 1	6.03	$\pm 5.61 \cdot 10^{-2}$	$\mathrm{g}^{k-1}\mathrm{g}^{-(k-1)}$
$K_{s,qp}$	Haldane constant 2	$70.1 \cdot 10^{-2}$	$\pm 3.64 \cdot 10^{-3}$	$\mathrm{g\,g^{-1}}$
$q_{P_{IB},max}$	max specific IB formation rate	$2.14 \cdot 10^{-2}$	$\pm 2.78 \cdot 10^{-4}$	$\mathrm{g\,g^{-1}\,h^{-1}}$
$K_{p,q_{PIB}}$	IB Moser constant	$3.742 \cdot 10^{-2}$	$\pm 4.23 \cdot 10^{-4}$	$\mathrm{g\,g^{-1}\,h^{-1}}$
N	IB Moser exponent	1.496	$\pm 1.51 \cdot 10^{-3}$	1

The **FIM** calculated for the practical identifiability analysis of the complete model was also used to evaluate the covariance and further the correlation between the parameters, which yielded the results in tab. 3.9. Values close to either plus or minus one indicate a strong correlation of the two parameters.

The maximum substrate uptake rate $q_{S,max}$ and the Monod constant K_S are strongly correlated, which can cause problems when they are fitted together. Furthermore, the parameters for the total productivity are also significantly correlated with each other and even more so the three of the IB formation. The high values were connected to the ill-conditioned character of the model structure and stem from the non-linearity of the kinetic equations. The resulting issues were reduced by choosing meaningful initial conditions and evaluating the optimized set for its plausibility regarding the ranges for the estimated parameter values.

3.5.1 Model performance on the measured data

The goal of this work was to develop a model for LDH production with the examined *E. coli* strain, that describes the measured values of the conducted processes. In fig. 3.24 the observed measurements are compared to the predicted values by the obtained model. The accuracy for the prediction of the three different product confirmations was overall well, but differences in their predictive performance were evident. The model for the total product matches the experimental data the best with a NRMSE of 0.094, while the points for the IBs deviate stronger from the ideal line for perfect predictions, resulting in 0.153 NRMSE. The accuracy for the soluble product ranks in between with a remaining NRMSE of 0.126.

Table 3.9: Rounded correlation values for the parameters used for the model and fitted with the experiments. Significant correlations are marked red ($\|corr\| > 0.6$)

	$Y_{X/S,Ind}$	$K_{S,Ind}$	$q_{S,max}$	k_D	$q_{P,max}$	$K_{s,qs}$	k_{Smet}	$K_{s,qp}$	$K_{I,qp}$	$q_{P,IB,max}$	$K_{s,qs,PIB}$	N
$Y_{X/S,Ind}$	1	0.04	0.04	-0.06	-0.04	-0.01	0.04	-0.03	0.02	0.01	0.01	-0.01
$K_{S,Ind}$	0.04	1	1	0	0	-0.01	0	0	0	0	0	0
$q_{S,max}$	0.04	1	1	0	0	-0.01	0	0	0	0	0	0
k_D	-0.06	0	0	1	0.18	0.03	-0.22	0.17	-0.16	-0.08	-0.07	0.04
$q_{P,max}$	-0.04	0	0	0.18	1	0.28	-0.7	0.99	-0.9	-0.7	-0.7	0.65
$K_{s,qs}$	-0.01	-0.01	-0.01	0.03	0.28	1	-0.12	0.13	-0.16	-0.03	-0.01	-0.11
k_{Smet}	0.04	0	0	-0.22	-0.7	-0.12	1	-0.7	0.93	0.41	0.42	-0.43
$K_{s,qp}$	-0.03	0	0	0.17	0.99	0.13	-0.7	1	-0.9	-0.72	-0.72	0.69
$K_{I,qp}$	0.02	0	0	-0.16	-0.9	-0.16	0.93	-0.9	1	0.6	0.6	-0.59
$q_{P,IB,max}$	0.01	0	0	-0.08	-0.7	-0.03	0.41	-0.72	0.6	1	1	-0.95
$K_{s,qs,PIB}$	0.01	0	0	-0.07	-0.7	-0.01	0.42	-0.72	0.6	1	1	-0.97
N	-0.01	0	0	0.04	0.65	-0.11	-0.43	0.69	-0.59	-0.95	-0.97	1

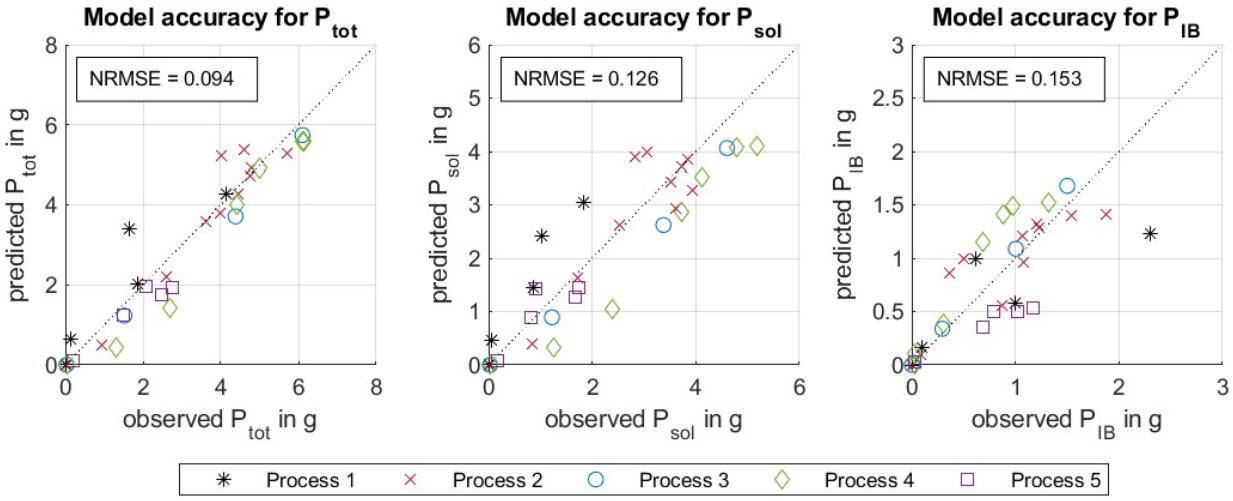

Figure 3.24: Accuracy of the models output for the different product conformations

The applied model and respective measurements for the training and validation data sets are shown collectively in fig. 3.25. Both models achieve sufficiently high scores in the goodness of fit analysis, which are summarized by the two bar graphs on the right and track the process dynamic well. Again, it can be seen that the simulated IB formation performs slightly worse than the total productivity.

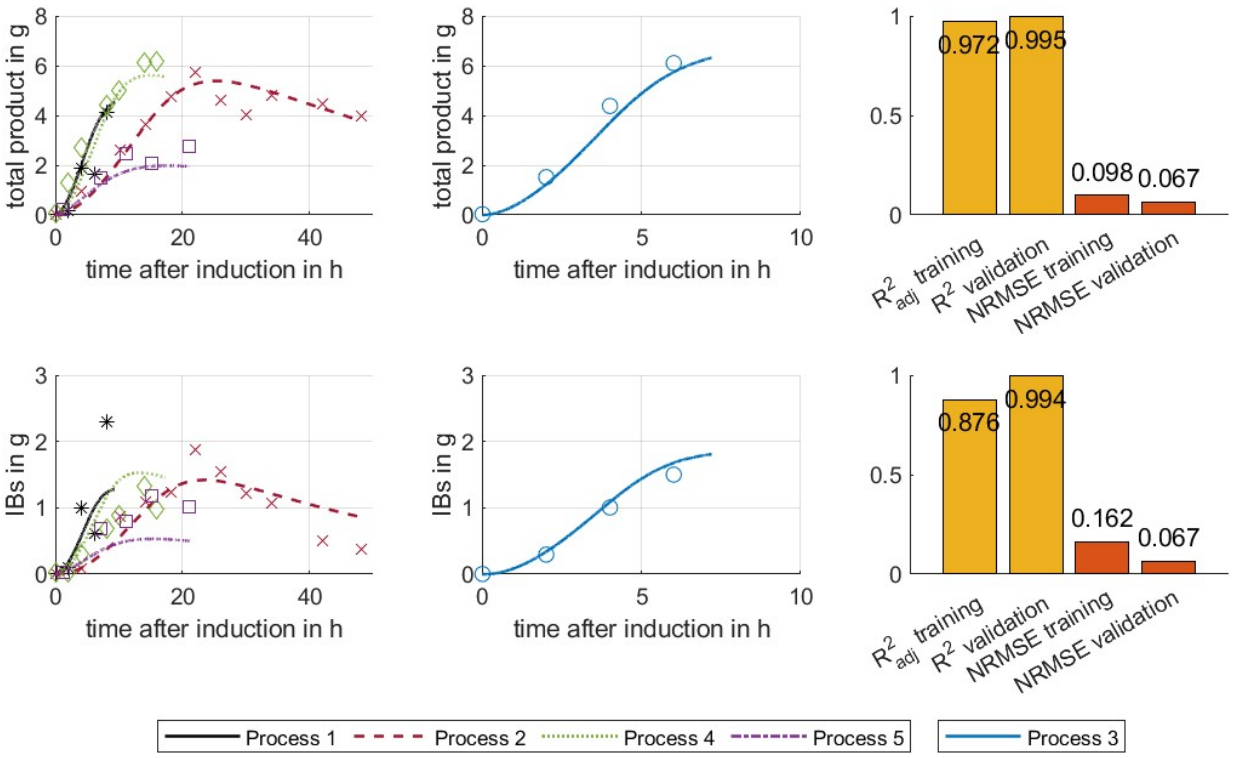

Figure 3.25: Model building results for the total productivity (upper) and IB formation (lower). The left hand graphs show the training data sets measured and simulated, the middle depicts the validation set and on the right the performance scores are given.

The model was parameterized with the data collected from the 5 performed experiments and therefore the validation ranges given in tab. 3.10 have to be kept in mind when applying it. An extrapolation outside of those boundaries can lead to unexpected behaviour or a significant decrease of the model performance. In order to expand this range it is recommended to carry out additional experiments that surpass these limits and repeat the parameterization procedure.

Table 3.10: Validity ranges of the verified model

Symbol	Name	Range	Units
X_{tot}	total biomass (DCW)	0 - 100.75	g
L	broth volume	0.81 - 2.30	L
$s_{met}{}^*$	specific metabolised substrate	0 - 2.94	$g\,g^{-1}$
$q_S{}^*$	specific substrate uptake rate	0 - 0.372	$g\,g^{-1}\,h^{-1}$
P_{tot}	total product	0 - 6.15	g
P_{sol}	soluble product	0 - 5.17	g
P_{IB}	product in IBs	0 - 2.30	g

* Values after induction

3.5.2 Model behaviour

After building and parameterizing the model, its final structure was investigated to find process optimization possibilities. A focus point was the relation between the soluble and IB content with the goal to manipulate the production in either direction. Fig. 3.26 shows the behaviour for the total product expressed by the cells in relation to the two model variables, depicted as 3-dimensional mesh graphs. The total specific productivity q_P increases quickly with the specific substrate uptake rate q_S, but the specific metabolised substrate s_{met} has a strong inhibiting effect. In order to retrieve information on the efficiency of the organism in producing the targeted enzyme, the productivity was related to the substrate uptake, effectively forming the substrate-to-product yield, which can be seen on the right graph in fig 3.26. Here the opposite can be observed, as the highest ratio is achieved at very low metabolic loads.

The same was also done for the IB formation and similar trends were observed, as seen in 3.27. The specific rate is higher for higher substrate uptake rates q_S, but the yield is inversely connected to it. As a difference to the total product formation q_P it was noted, that the maximum is slightly removed from the origin.

The last observation was explored further by creating a graph that shows the relation of the specific IB formation rate $q_{P,IB}$ to the total specific productivity q_P in fig. 3.28. This directly describes how much of the expressed enzyme stays soluble inside the cells or forms the discussed aggregations. The graph shows that high substrate uptake rates result in an enhanced IB ratio during LDH expression. This observation could be exploited to shift an industrial production into the preferred direction regarding the IB content.

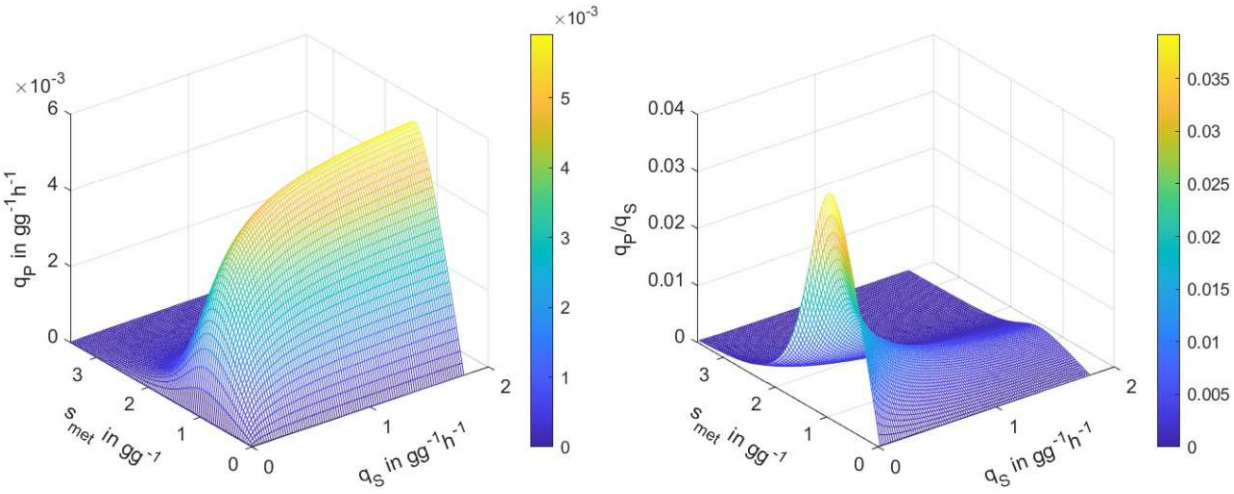

Figure 3.26: Model dynamics for the total productivity q_P in relation to its variables (left) and the yield of product formation per substrate uptake q_P/q_S

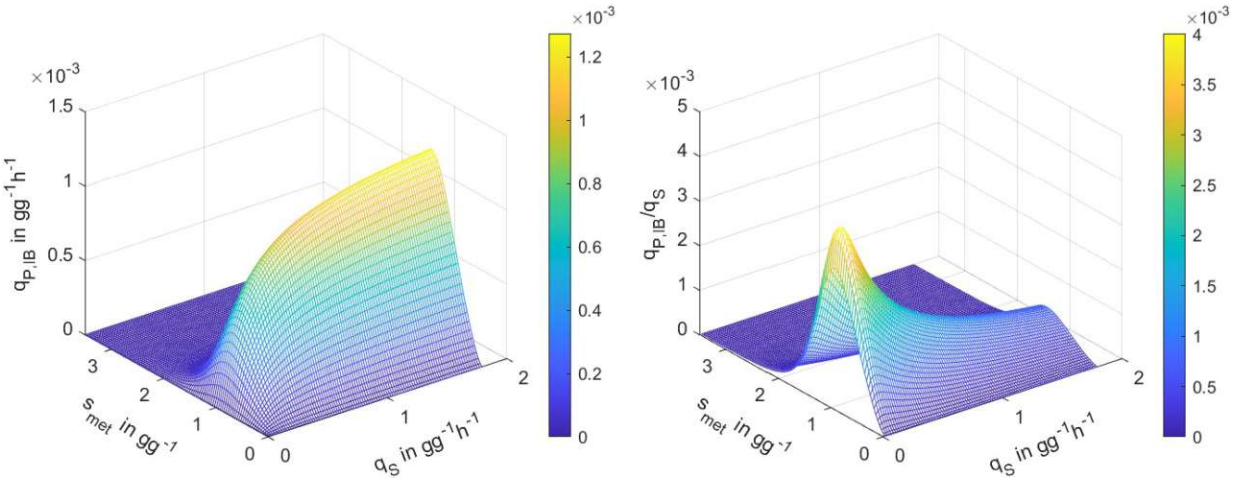

Figure 3.27: Model dynamics for the IB formation $q_{P,IB}$ in relation to the variables influencing q_S (left) and the yield of IB formation per substrate uptake q_P/q_S

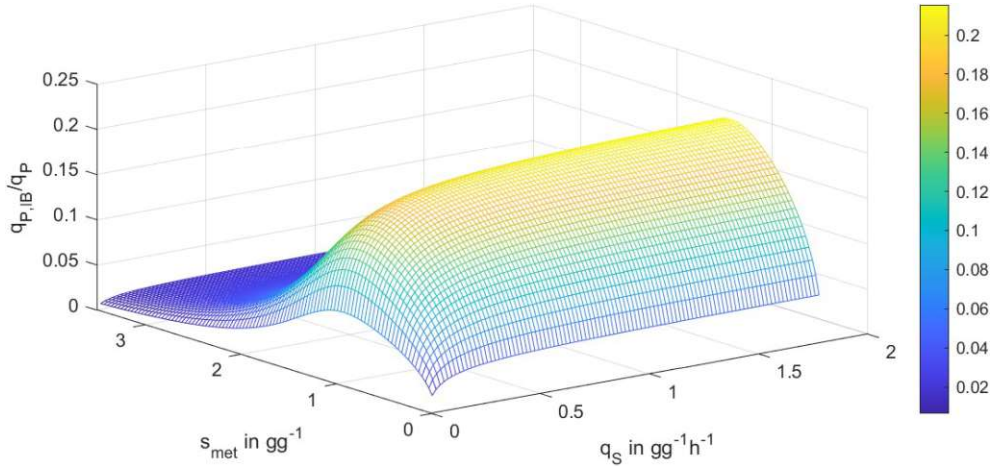

Figure 3.28: Specific IB formation $q_{P,IB}$ to total productivity q_P dependent on q_S and s_{met}

3.5.3 Discussion of the model results

The metabolised substrate, which is an indicator for the age of the cells in the broth, has a strong influence on the productivity. All graphs show a start up phase of the specific rates at the beginning, followed by a similarly fast decline. The productivity of the cells dies off completely after they metabolised approximately thrice their weight in substrate, marking what seems to be a natural barrier for this process setup.

A key variable in bioprocesses is the feedrate, since it is one of the few present opportunities to influence the course of the fermentation actively. When the process is run as a fedbatch with substrate limitation, it is directly connected to the specific substrate uptake rate q_S, an important physiological indicator. The connection with the specific total product and IB formation showed two essential dynamics. On the one hand, a high specific substrate uptake translates to higher specific productivity. On the other hand it has the opposite effect on the substrate-to-product and IB yield, meaning that lower feedrates result in a more resource efficient expression of LDH. Furthermore the share of enzymes aggregating into IBs drops also in connection with the specific substrate uptake rate q_S. This points to a connection of the metabolic stress for the cells and their ability to operate the enzyme expression mechanism efficiently.

To assess the plausibility of the resulting model parameter values, they were compared with values published in scientific literature. The presented model structure for LDH production in *E. coli* could not be identified in other resources, therefore different recombinant proteins were considered and the maximum productivity values of the different product forms were compared.

Table 3.11: Maximum productivity of different proteins from literature in $\mathrm{g\,g^{-1}\,h^{-1}}$

Product	$q_{P,max}$	$q_{P,IB,max}$	Size	Reference
LDH	0.0538	0.0214	34.2 kDa [100]	this work
soluble GFP[a]	0.07	-	26.8 kDa [99]	[35]
GIP[a] IBs	-	0.08	17.1 kDa [97]	[37]
N-pro fused model protein[b] IBs	-	0.038	28.8 kDa	[89]
scaf* of IgG[a]	0.0066	-	49.3 kDa (IgG) [98]	[44]
scaf* of IgY[b]	0.01016	0.00062	21.9 kDa (IgY) [96]	[116]

[a] established model parameter or ANN-model design space, [b] measured value
* single chain antibody fragment, size from full antibody molecule

The overall maximum productivity of the established model for LDH is located within the range of the Green Fluorescent Protein (GFP), the Gastric Inhibitory Polypeptide (GIP) and the N-pro fused model protein listed in tab. 3.11. The last two products were only recovered as IBs, but all of these proteins are of comparable size. The productivity values for the presented antibody fragments were much lower, which points towards a less efficient expression system implemented in the organism. When looking at the productivity of aggregated protein, the maximum specific rate for the Moser kinetic obtained in this work was lower but also of similar magnitude as the one for the N-pro fused model protein, while the value for the GIP is significantly greater. One reason for the reduced maximum productivity of the examined LDH process could be the smaller share of total protein accumulating into IBs, indicated by fig. 3.28.

Overall the upper limits for the specific rates for LDH productivity in this model were found to be of valid magnitudes when compared with other recombinant proteins of similar size. Together with the performance of the model on the measured data, they were therefore deemed as plausible and effective estimations for the observed bioprocess. The researched productivity values for the antibody production processes were much lower, suggesting a possibility for further optimization of their expression system and process conditions. Here, an experiment-based productivity model utilizing physiological descriptors can play an integral role in achieving this goal.

3.6 Model based process optimization

After evaluating the properties of the obtained model in the previous chapter, it was used to find optimal feeding trajectories for different process objectives in a simulation based study. The implemented algorithm numerically adjusted a targeted specific substrate uptake rate $q_{S,stp}$ expressed through a polynom. A design case was defined that was inspired by the laboratory setup used for the data acquisition. All simulation experiments were conducted with the same starting conditions and carried out a batch phase, which automatically ended when the substrate was depleted. This was followed by a fed-batch with a constant q_S setpoint to achieve a predefined biomass after exactly one day of cultivation. Then the induction agent was added and the production of LDH started. The applied setpoints for the specific substrate uptake rate $q_{S,stp}$ were then either predefined or numerically optimized through an algorithm minimizing a cost-function. The initial conditions and process parameters are summarized in tab. 3.12.

Table 3.12: Initial conditions and process parameters for simulation experiments

Name	Value
Volume at t_0	1.5 L
Substrate at t_0	20 g
Biomass for inoculation	0.25 g
Feed concentration	440 g L^{-1}

Symbol	Value
Start of induction	24 h
Biomass at induction	100 g
Volume at induction	2.07 L
Maximum volume	3 L

3.6.1 Design space exploration and optimization

In total 5 different objective functions were investigated with the simulation experiment, which are listed in tab. 3.13. The maximisation was implemented by using an unconstrained minimization algorithm and including the targeted output multiplied by minus one in the cost-function. Furthermore, absolute product formation rates r_P or specific substrate uptake rates q_S below zero were penalized.

For the simulation experiment a range of constant specific substrate uptake rates q_S were investigated to see the overall effect on the output. Then the previously mentioned objective functions were used to obtain new optimal setpoints. The Space-Time Yield (STY) was calculated with the final values of the simulation by eq. 3.13, where the process was ended right after reaching the highest product mass in order to avoid loosing protein just to meet the final reactor volume.

Table 3.13: Objective functions implemented for simulation experiments

ID	0	3	4	5	6
function	$max(P_{tot,end})$	$max(\frac{P_{sol,end}}{P_{tot,end}})$	$max(\frac{P_{IB,end}}{P_{tot,end}})$	$max(\eta_{P,tot})$	$max(\eta_{P/S})$

The same was done for the product-substrate yield $\eta_{P/S}$, which was built by dividing the produced total product P_{tot} by the metabolised substrate S_{met} at the end. The trajectories as well as their performance can be seen in fig. 3.29.

$$STY = \frac{P_{tot}}{\Delta t \cdot V_{end}} \tag{3.13}$$

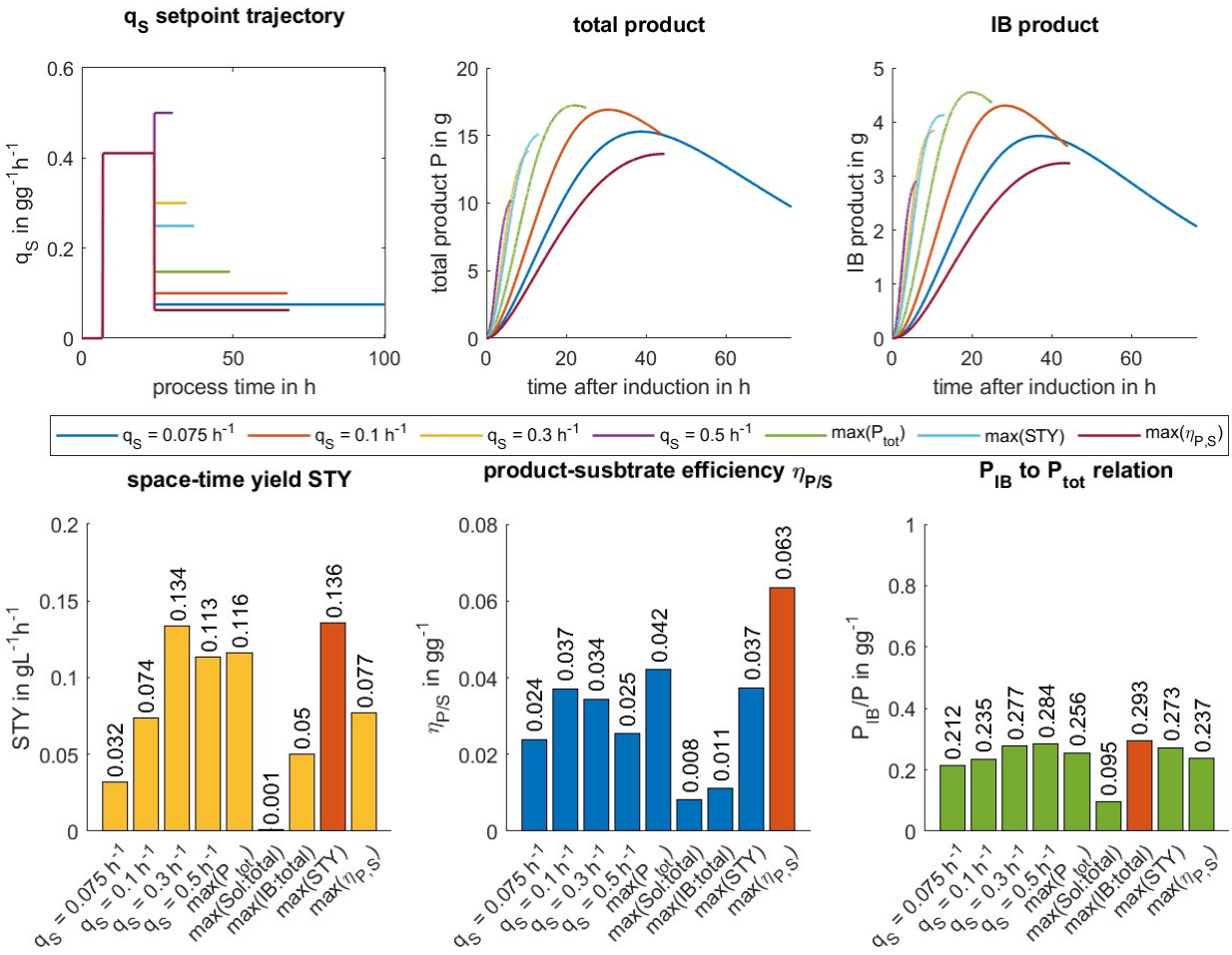

Figure 3.29: Top: Trajectories for the static specific substrate uptake rate q_S and the resulting product content in the reactor. Bottom: Performance indicators for the different trajectories.

The comparison showed an optimum for the obtained model can be found that is located in-between the predefined specific substrate uptake rates. As discussed in section 3.5.2 high q_S values result in a lower product-substrate yield and therefore in less product produced. This would suggest that

keeping the specific substrate uptake rate as low as possible would achieve the highest enzyme levels in the broth. But apart from immensely increasing the production time and therefore lowering the STY, also the maximum product amount in the tank started to fall again. This is due to the increasing product loss over time caused by the modelled cell death. The optimal solutions for the different optimization goals are listed in tab. 3.14. When striving for the highest possible total product titer 17.04 g could be achieved in the 3 L setup. Since in industrial applications economic aspects play a significant role, the constant q_S setpoints were optimized for the STY and $\eta_{P/S}$ as well. Compared to the maximal product titer these trajectories resulted in lower levels of total product mass but managed to increase the STY by 17 % and the $\eta_{P/S}$ by 50 % respectively.

Table 3.14: Optimization results with constant q_S for different maximization functions

max()	q_S	P_{tot} titer	IBs	V_{end}	t_{ges}	STY
P_{tot}	$0.1475\,h^{-1}$	17.04 g	25.6 %	3.00 L	49.02 h	$0.116\,g\,L^{-1}\,h^{-1}$
P_{sol}/P_{tot}*	$0.0100\,h^{-1}$	0.325 g	9.5 %	2.17 L	168 h	$8.9 \cdot 10^{-4} g\,L^{-1}\,h^{-1}$
P_{IB}/P_{tot}	$1.788\,h^{-1}$	3.877 g	29.3 %	3.00 L	25.71 h	$0.050\,g\,L^{-1}\,h^{-1}$
STY	$0.2494\,h^{-1}$	15.10 g	27.3 %	3.00 L	37.14 h	$0.136\,g\,L^{-1}\,h^{-1}$
$\eta_{P/S}$	$0.0626\,h^{-1}$	13.64 g	23.7 %	2.57 L	68.64 h	$0.077\,g\,L^{-1}\,h^{-1}$

* Optimization penalized q_S values below $0.01\,h^{-1}$

The trajectories for the maximal soluble and IB share of the product were excluded from fig. 3.29 since they would skew the graph strongly. Their results are represented in the performance part of the graph and in tab. 3.14. The numerical procedure penalized setpoints for q_S below $0.01\,h^{-1}$ heavily, since the unconstrained maximization algorithm would have yielded even lower specific substrate uptake rates and impractical titers for the objective to maximize soluble product. With the given constraints the simulation study showed that the IB fraction could be adjusted between 9 - 29 %. In order to achieve high soluble content the substrate uptake rate should be kept as low as possible, which also results in long process times and low product titers. For enhanced aggregation into IBs the optimization recommended to increase the specific substrate uptake rate close to the maximum $q_{S,max}$ from the Monod kinetic. The process time is drastically reduced, since the feedrate fills up the tank faster, but also the metabolised substrate increases rapidly causing a low overall titer. Additionally, the high substrate levels in the broth could give rise to unwanted overflow metabolism effects in the form of acetate formation.

Finally, the complexity of the setpoint trajectory for the specific substrate uptake rate q_S was increased by using a polynomial instead of a constant. The goal was to further improve the objective functions by adding degrees of freedom to the $q_{S,stp}$ curve, as seen in eq. 3.14. Several different orders of polynomials were subjected to the optimization algorithm and their results were compared.

$$q_{S,stp}(t) = q_{S,0} + a_1 \cdot t + a_2 \cdot t^2 + ... + a_i \cdot t^i = \sum_{i=0}^{n} a_i \cdot t^i \tag{3.14}$$

In tab. 3.15 and fig. 3.30 the improvements through a more complex setpoint trajectory are presented exemplary for the objective function max(P_{tot}). The adjusted curve shows that the product titer could be further increased by 5.8 % to 18.03 g while also benefiting the STY slightly.

This suggests that the potential of improving the production process through mathematical optimization with the obtained model can be raised further when using a more sophisticated setpoint trajectory compared to constant values. Nevertheless, the computational effort also surged the more polynomial terms were added, which has to be minded when implemented in an advanced control algorithm.

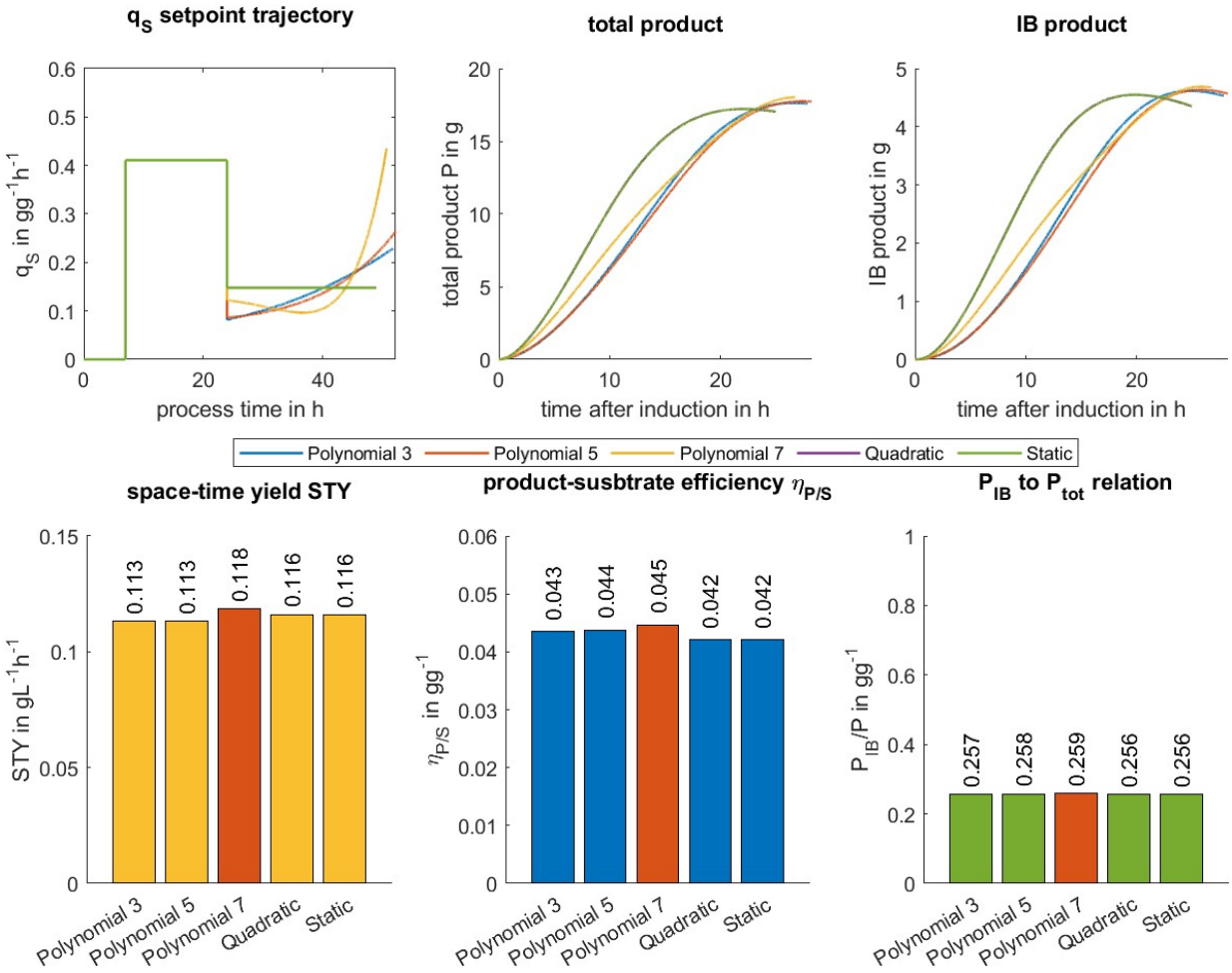

Figure 3.30: Top: Optimized trajectories for the specific substrate uptake rate q_S with varying polynomial degrees and the resulting product content in the reactor. Bottom: Performance indicators for the different trajectories.

Table 3.15: Results for maximizing P_{tot} with different polynomial degrees for q_S

ID	poly. degree	P_{tot} titer	IBs	V_{end}	t_{ges}	STY
constant	0	17.04 g	25.6 %	3.00 L	49.02 h	0.116 g L^{-1} h^{-1}
quadratic	2	17.06 g	25.6 %	3.00 L	48.92 h	0.116 g L^{-1} h^{-1}
polynomial 3	3	17.60 g	25.7 %	3.00 L	51.86 h	0.113 g L^{-1} h^{-1}
polynomial 5	5	17.73 g	25.8 %	3.00 L	52.24 h	0.113 g L^{-1} h^{-1}
polynomial 7	7	18.03 g	25.9 %	3.00 L	50.73 h	0.118 g L^{-1} h^{-1}

3.6.2 Discussion and outlook of simulations

The simulation experiments showed, that the model can be used for optimization of the specific substrate uptake rate setpoints $q_{S,stp}$. This can be utilized to calculate the feedrate during the production of LDH after induction in a feed-forward control implementing declared objective functions for the fermentation. Although the shift of the production towards soluble enzymes or IBs is possible in principle, only marginal effects were observed in the optimized simulations, which were also accompanied by a significant reduction of the overall productivity. Nevertheless, the model proved to show potential as an application to increase the product titer or the space-time yield, which can translate into an economical impact on the recombinant production of LDH in *E. coli*.

It has to be noted that the biggest driver for the expression of the product is the amount of biomass, which was not subject to optimization in this section. In order to achieve the overall best results the setpoints during the uninduced fed-batch phase should be considered as well. The biggest product titers are then achieved with the highest possible biomass present at the start of the production phase, considering also other limitations such as the oxygen transfer rate or foam build-up.

3.7 Model based monitoring with CER and OUR

The CER and OUR were already introduced in section 2.4.4. They pose as two real-time observable rates through the online measurements of the gas in- and outflows and were computed with simple elemental balances. Therefore, they are available during the process and can be utilized as indicators for the metabolic activity inside the reactor. This makes them perfect candidates to be included into model based control algorithms for example as a indirect measurement for the biomass and the substrate uptake rate q_S, when the yield coefficients are assumed to be constants. Therefore the model quality and usability for future advanced control strategies was additionally assessed by comparing the calculated and the simulated rates in fig. 3.31.

The simulation predicted the CER and OUR very well until the induction point. Then an offset can be observed and the simulation started to over-predict the rates to some extent. Nevertheless, the trajectories were generally similar which points at a good match between the model and the actual metabolic processes. The deviation shortly after induction hints at a transition phase where the cells adjusted to the new circumstances and the actual yields were lower than anticipated in the model. A potential offset between viable biomass in the process and the value predicted by the model contributes to the explanation. Over the whole data set the model resulted in a NRMSE of 0.0702 for the CER and 0.0397 for the OUR.

Altogether, the general tracking of the calculated rates with the model give reason to believe that a control strategy utilizing only the online measurements is worth pursuing. The performance of models using the CER and OUR for online soft-sensor algorithms can further be improved when implementing the time-delayed offline measurements as shown by Kager et al. [45]. Whenever a new result was ready it was fed into the algorithm, which then recalculated the states from the sample time of the provided offline measurement up until the live estimator. The reevaluation in

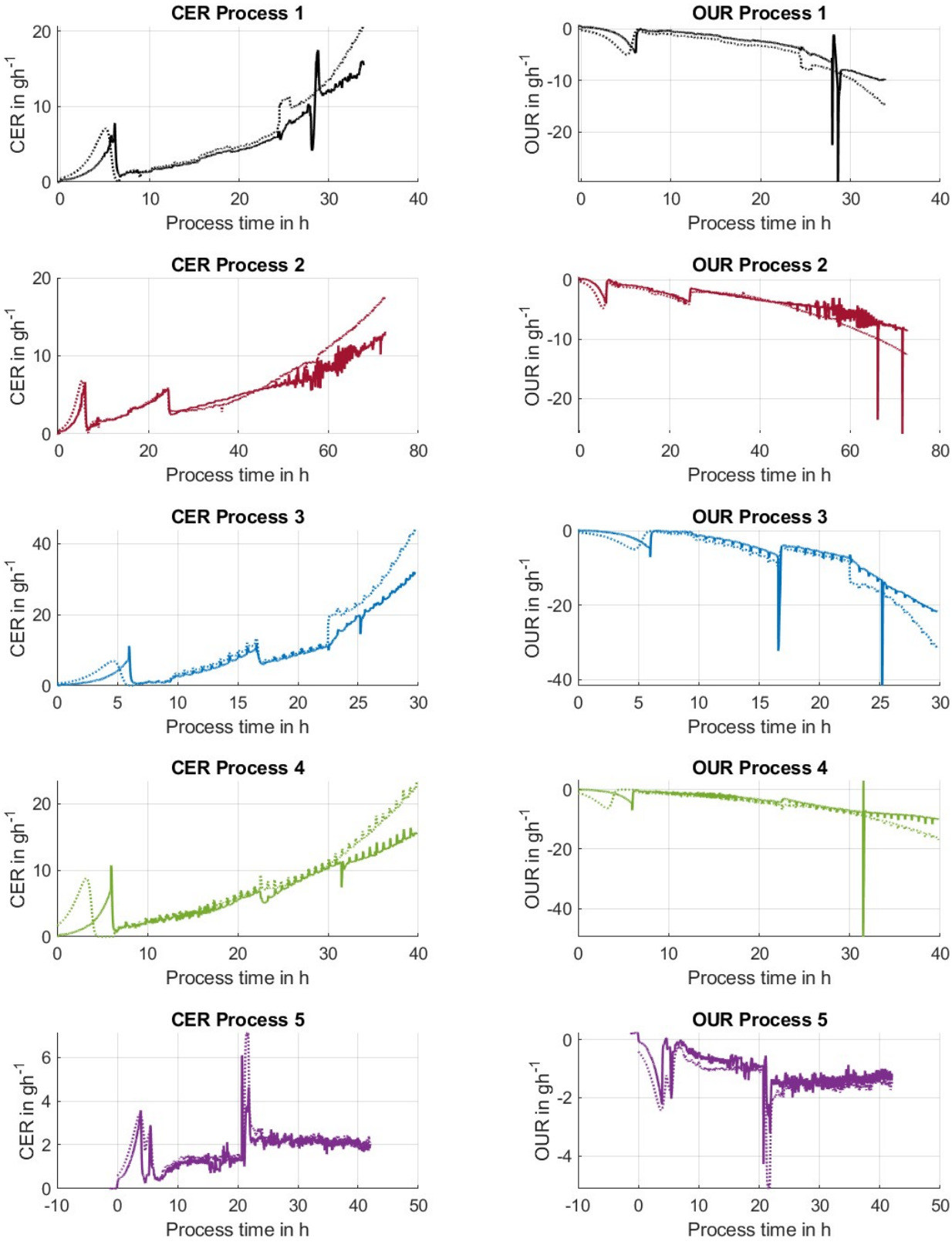

Figure 3.31: CER and OUR simulated (dashed line) and calculated from the off-gas measurements (full line) for the 5 conducted processes.

hindsight cost computing power and required offline analysis methods to provide results within reasonable time, but improved the quality of the state estimations. When implementing a closed loop feedback control with the developed model, further analyses towards stability and step response behaviour, as well as the influence of disturbances on the system have to be conducted to ensure its controlability.

4 | Conclusion

This work set out to build a mechanistic model for the production of the recombinant protein LDH in *E. coli* and its IB formation dynamics. The model should combine kinetic functions for both mechanisms using physiological descriptors and incorporate the cell viability in its structure, to provide a novel and holistic tool for this bioprocess. Experiments were carried out to find suitable correlations, parameterize the model equations and validate the results. For the latter, statistical methods were applied to analyse the identifiability of the parameters in the chosen structure. Finally, the model was used in a simulation study to explore process optimization opportunities.

Considering several different methods to quantify the product amount, the samples were analysed through an enzymatic assay and correlated to corresponding results from a RP-HPLC to yield the mass concentrations. The aggregated share of the protein was successfully measured after conducting a refolding procedure with the separated IB pellets. With the obtained data a positive correlation between the specific substrate uptake rate and the maximum productivity of the overall product expression was identified. Furthermore, a decline of the specific product formation was observed and connected to the specific metabolised substrate. These two physiological variables are indicators of the metabolic load and age of the cells. Finally, the IB formation was correlated to the total productivity to describe the localisation of the protein in the organism. These findings were subsequently used to adapt an existing model describing cell growth and death and extend it for the product formation dynamics of the investigated strain.

As a prerequisite for modeling the productivity, the equations for substrate uptake, cell growth and cell death were parameterized with the given data set, resulting in overall NRMSE values of 0.0502 for the total biomass, 0.0538 for the viable and 0.0992 for the dead cell mass. In the next step, a number of kinetics were tested with the identified physiological variables and their performances on a training and a validation data set were compared. For the total protein productivity a Monod kinetic utilizing the specific substrate uptake rate was combined with a Haldane inhibition term as a function of the specific metabolised substrate, similar to the approaches of Kager et al. and Reichelt et al. [44, 75]. The resulting model achieved a R^2 of 0.995 and 0.067 NRMSE on the validation data. For the specific IB formation a simple Moser kinetic was implemented that used the specific productivity as a variable. This links the aggregation directly to the metabolic stress caused by the protein expression and the validation data confirmed a good fit with 0.994 R^2 and 0.067 NRMSE. The final model structure was then examined for its structural and practical identifiability, which it passed as ill-conditioned. The analysis further showed significant correlations between several parameters of the kinetic equations, which hints at possible struggles for numerical procedures. Nevertheless, the parameterized model showed high accuracy over the complete set of data for all product conformations, although the performance for

the aggregated protein was poorer. The respective NRMSE values were 0.094 for the total product, 0.126 for the soluble fraction and 0.153 for IBs. Contrary to the works of the previously mentioned authors, this model combined kinetics for the total productivity and the IB formation and considers the influence of the biomass viability due to increased cell death. This comprehensive description of the product dynamics serve as a novel approach for modeling this induced *E. coli* bioprocess.

The model showed promising results for the optimization of the protein production process. During investigation of the kinetic equations it was shown that the specific substrate uptake rate can be effectively used to impact the protein expression rate but also its efficiency, as well as the proportion of IBs formed. The findings support the theory that a high metabolic stress on the cells causes the organism to work less efficient and gives rise to miss-folding of produced protein chains, which coincides with the observations by Slouka et al. [88]. The model was then applied in a simulation study which showed its potential to optimize the substrate feeding trajectory for defined objectives. The maximal total protein titer was identified as 17.04 g in the given experimental setup of 3 L, using a constant setpoint for the substrate uptake rate. It was further increased by implementing a polynomial function for the setpoint trajectory to 18.03 g. For industrial applications other process indicators such as the space-time yield STY or the product-substrate efficiency $\eta_{P/S}$ are more relevant. Optimizing the constant setpoints for these goals individually, the STY was improved by 17 % and $\eta_{P/S}$ by 50 %, compared to the maximum product titer case. The simulation study further revealed that the IB content can be manipulated within a range of 9 - 29 % of the produced protein in this setup. But maximizing or minimizing its share caused low overall titers and yields and therefore more sophisticated objective functions should be implemented here.

Apart from optimizing the feeding trajectory for a feed-forward control system, the model can be implemented as a digital twin for online prediction and supervision. The CER and OUR were also described by the model and fit the measured data well with NRMSE values of 0.0702 and 0.0397. As demonstrated by Müller et al. [64], this can be used as an online measurement for the biomass and the specific substrate uptake rate. When combined with the developed model, the product content can be estimated in real-time and thereby valuable process information is provided during the fermentation. Furthermore, these soft-sensor applications enable the optimization of the productivity trajectory through advanced control strategies such as Model Predictive Control (MPC), as suggested by Mears et al. [57]. The performance of such an algorithm could be further increased by incorporating time-delayed offline measurements of biomass or product to recalculate the state variables or improve certain parameter estimations, as implemented by Kager et al. [45]. This opens up the possibility to not only steer between soluble protein and IB formation, but also optimize the process from economic viewpoints, by adjusting the feed rate on the basis of the outputs of the developed model. Applying these concepts in an industrial setup carries the potential to significantly impact the future production of recombinant proteins in *E. coli*.

For the implementation of a model-based control algorithm the influence of model uncertainties as well as measurement errors has to be explored thoroughly to guarantee controlability of the resulting system. Regarding the transition of the bioprocess during scale-up or when implementing different equipment, the implementation of physiological variables in the model should minimize the impact on the its performance. Nevertheless, it is advised to run additional experiments on different equipment and scales to test this hypothesis and expand the validity of the developed

model, but also to verify the derived in-silico optima. In order to apply this model structure for the optimization of production processes with a defined objective, the influence of technical parameters like temperature and strength of induction should be investigated further to find global optima. In case the product or the microbial organism is changed, the systematic modelling approach described in this work should be repeated, to find and parameterize an appropriate choice for the kinetic equations.

This thesis successfully developed a mechanistic model of an induced *E. coli* process utilizing physiological descriptors, which effectively described the five conducted experiments. In doing so, it showcased a methodical approach for model development and structural analysis using statistical methods to justify the design choices. The incorporation of seperate kinetics for the overall production of the recombinant protein as well as for the aggregation into IBs and the explicit consideration of the cell viability resulted in a holistic model, that poses as a novelty within this field. The capabilities and importance of such models for bioprocesses was showcased by putting these results to use in a simulation study. Finally, the discussed applications within monitoring and control should encourage future investigations in this field and motivate the industry to increase their implementation.

List of Variables

Symbol	**Name** [Unit]
Analytical equations	
m_X	Cell mass [g]
Act	Enzymatic activity [UL^{-1} = µmol $min^{-1}\,L^{-1}$]
V_t	Total volume per well [µL]
V_s	Sample volume per well [µL]
ε_{340}	Molar extinction factor of NADH at 340 nm [L $mmol^{-1}$ cm^{-1}]
l	Path length through well [cm]
k_{abs}	Slope of the absorbance regression [Abs min^{-1}]
f_d	Sample dilution factor [1]
CER & OUR equations	
CER_{model}	Modeled carbon evolution rate [$g\,h^{-1}$]
CER_{meas}	Measured carbon evolution rate [$g\,h^{-1}$]
OUR_{model}	Modeled oxygen uptake rate [$g\,h^{-1}$]
OUR_{meas}	Measured oxygen uptake rate [$g\,h^{-1}$]
M_i	Molar mass of compound i (normalized to one carbon) [$g\,C\text{-}mol^{-1}$]
F_{AIR}	Gas input flow from in-house compressed air system [$L\,h^{-1}$]
F_{O2}	Gas input flow from oxygen bottle [$L\,h^{-1}$]
V_M	Molar volume of ideal gas [$L\,M^{-1}$]
$x_{j,out}$	Mole fraction of compound j in exhaust gas [1]
$x_{j,in}$	Mole fraction of compound j in input flow [1]
$R_{a,inert}$	Inert gas fraction [1]
y_{wet}	Humidity of gaseous phase [1]
γ_i	Degree of reduction of compound i [1]
Model analysis equations	
x_i	Measured data [-]
$\hat{x}_i$	Simulated data [-]

$\bar{x}$	Data mean [-]
SSE	Sum of squares error [-]
SSE	Sum of squares regression [-]
SSE	Sum of squares total [-]
n	Number of measurements [1]
p	Number of parameters [1]
R^2	Coefficient of determination [1]
$R^2_{adjusted}$	Adjusted coefficient of determination [1]
NRMSE	Normalized root-mean-square error [1]
$\mathbf{S}$	Discrete-time sensitivity matrix [-]
$S_{i,p}$	Sensitivity of state i towards parameter p [-]
$\bar{S}_{i,p}$	Normalized sensitivity of state i towards parameter p [1]
c_i	State i [-]
θ_i	Parameter i [-]
δ_p^{msqr}	Importance factor of parameter p [1]
λ	Entry of the discrete-time sensitivity matrix [-]
FIM	Fischer Information matrix [-]
$\mathbf{W}$	Covariance matrix of the state variables [-]
$\mathbf{S}_{sum}$	Sum of the discrete-time sensitivity matrices [-]
$\sigma_{i,l}$	Variance of state variable i at time incident l [-]
$\mathbf{Cov}_{min}(\hat{\theta})$	Minimum parameter covariance matrix [-]
$\mathbf{Cor}_{min}(\hat{\theta})$	Minimum parameter correlation matrix [-]
ρ_{pq}	Correlation of parameter p and q [-]
State-space equations	
$\mathbf{x}$	State vector [-]
V	Volume of broth in reactor [L]
X_V	Cell mass viable [g]
X_D	Cell mass dead [g]
S_1	Mass of substrate 1 [g]
S_2	Mass of substrate 2 [g]
P	Mass of product [g]
I	Mass of induction agent [g]
$CO2$	Cummulated CO_2 produced [g]
$O2$	Cummulated O_2 produced [g]
S_{met}	Metabolised substrate [g]
P_{sol}	Mass of soluble product [g]
P_{IB}	Mass of IBs [g]

X_{tot}	Total mass [g]
s_{met}	Specific metabolised substrate [$g\,g^{-1}$]
$x_{P,sol}$	Specific soluble product [$g\,g^{-1}$]
q	Specific rate [$g\,g^{-1}\,h^{-1}$]
r	Gravimetric rate [$g\,h^{-1}$]
Y_X	Biomass yield matrix [1]
Y_P	Product yield matrix [1]
$Y_{i/j}$	State i to state j yield [1]
μ	Specific growth rate [$g\,g^{-1}\,h^{-1}$]
q_S	Specific substrate uptake rate [$g\,g^{-1}\,h^{-1}$]
q_D	Specific cell death rate [$g\,g^{-1}\,h^{-1}$]
q_P	Specific total productivity [$g\,g^{-1}\,h^{-1}$]
$q_{P,sol}$	Specific soluble product formation [$g\,g^{-1}\,h^{-1}$]
$q_{P,IB}$	Specific IB formation [$g\,g^{-1}\,h^{-1}$]
q_{CO2}	Specific CO_2 production [$g\,g^{-1}\,h^{-1}$]
q_{O2}	Specific O_2 consumption [$g\,g^{-1}\,h^{-1}$]
F_{in}	Total input flow rate [$L\,h^{-1}$]
F_{RS}	Substrate input flow rate [$L\,h^{-1}$]
F_{RI}	Induction agent input flow rate [$L\,h^{-1}$]
F_S	Outgoing flow rate [$L\,h^{-1}$]
c_{SR}	Substrate concentration feed [$g\,L^{-1}$]
c_{IR}	Induction agent concentration feed [$g\,L^{-1}$]
C_{in}	Concentration of the substances in the input flow [$g\,L^{-1}$]
c	Concentration of the states in the outgoing flow [$g\,L^{-1}$]
Productivity model equations	
$q_{S,max}$	Maximum specific substrate uptake rate [$g\,g^{-1}\,h^{-1}$]
K_S	Monod constant for substrate uptake [$g\,g^{-1}\,h^{-1}$]
c_S	Substrate concentration in the broth [$g\,L^{-1}$]
k_D	Cell death constant [$g\,g^{-1}\,h^{-1}$]
$q_{P,max}$	Maximum specific productivity [$g\,g^{-1}\,h^{-1}$]
$K_{s,qs}$	Productivity Monod constant [$g\,g^{-1}\,h^{-1}$]
k_{Smet}	Haldane exponent [1]
$K_{I,qp}$	Haldane constant 1 [$g^{k-1}g^{-(k-1)}$]
$K_{s,qp}$	Haldane constant 2 [$g\,g^{-1}$]
$q_{P_{IB},max}$	Maximum specific IB formation rate [$g\,g^{-1}\,h^{-1}$]
$K_{p,q_{PIB}}$	IB Moser constant [$g\,g^{-1}\,h^{-1}$]
N	IB Moser exponent [1]

Simulation experiment equations	
STY	Space-time yield [$\mathrm{g\,L^{-1}\,h^{-1}}$]
$\eta_{P/S}$	Product-substrate efficiency [$\mathrm{g\,g^{-1}}$]
$q_{S,stp}$	Setpoint for substrate uptake rate [$\mathrm{g\,g^{-1}\,h^{-1}}$]
a_i	Coefficient for polynomal term i [1]
t	Process time after induction [h]

List of Figures

List of Tables

Glossary

DiBAC$_4$(3) Bis-(1,3-Dibutylbarbituric Acid)Trimethine Oxonol, a membrane potentail-sensitive dye that enhances damaged and dead cells [50]. 17

RH414 N-(3-Triethylammoniumpropyl)-4-(4-(4-(diethylamino)phenyl)butadienyl) pyridinium dibromide, a fluorescent dye for staining of plasmamembranes that is used in FCM to stain all cellular particles and differentiate it from non-cellular background [50]. 17

Acronyms

AcN Acetonitrile. 19, 20, 75

FIM Fisher Information Matrix. 26, 27, 48, 52, 54

E. coli *Escherichia coli*. 1–3, 7–9, 11, 13, 14, 17, 20, 22, 24, 41, 54, 59, 64, 67–69, 75

STY Space-Time Yield. 60, 62

ANN Artificial Neural Network. 8

BCA Bicinchoninic Acid. 18, 19, 35

BSA Bovine serum albumin. 19

CER Carbon Evolution Rate. 22, 27–29, 64, 65, 68, 76, 77

Cor$_{min}(\hat{\theta})$ Parameter corrrelation matrix. 27

Cov$_{min}(\hat{\theta})$ Minimum parameter covariance matrix. 26, 49, 54

CPP critical process parameter. 7

DCW Dry Cell Weight. 15–17, 31–33, 57, 75

DI water Deionized water. 17, 18

DO Dissolved Oxygen. 12, 14

DoR Degree of Reduction. 28, 29, 77

EMA European Medicine Agency. 5

FCM Flow Cytometry. 13, 16, 33, 34, 41, 75, 79

FDA U.S. Food and Drug Administration. 5

GFP Green Fluorescent Protein. 59

GIP Gastric Inhibitory Polypeptide. 59

GMoP Good Modeling Practices. 6, 20, 21, 75

IB Inclusion Body. 3–5, 8, 9, 17–19, 31, 35, 36, 38–40, 45, 48, 50–54, 56–59, 62–64, 67–69, 72, 73, 75–77

IPTG Isopropyl β-D-1-Thiogalactopyranoside. 2, 11, 13, 14, 39, 41, 75

LDH Lactate dehydrogenase. 9, 11, 14, 16–19, 29, 34, 35, 41, 45, 49, 54, 57, 59, 60, 64, 67, 75, 77

MPC Model Predictive Control. 68

MW Molar mass (C-mol based). 27–29, 77

NADH Nicotinamide-Adenine-Dinucleotide-Hydrate. 18, 19, 71

NRMSE Normalized Root-Mean-Square Error. 25, 41, 42, 45–47, 51, 52, 54, 64, 67, 68, 72, 76

OD_{600} Optical Density at 600 nm. 15, 16

OUR Oxygen Uptake Rate. 22, 27–29, 64, 65, 68, 76, 77

PAT Process Analytical Technology. 5, 6

PPG Polypropylene glycol. 13, 32

QbD Quality by Design. 5, 7

R^2 coefficient of determination. 24, 35, 67, 72

$R^2_{adjusted}$ adjusted coefficient of determination. 24, 25, 46, 72

RP-HPLC Reversed-Phase High-Performance Liquid Chromatography. 19, 34, 35, 67

SDS-PAGE Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis. 20, 35

SSE Sum of Squares Error. 24, 72

SSR Sum of Squares Regression. 24

SST Sum of Squares Total. 24

TFA Trifluoroacetic acid. 19, 20, 75

Bibliography

[1] Chapter 1: The Biorefinery Concept–An Integrated Approach. In J. H. Clark and F. E. I. Deswarte, editors, *Introduction to chemicals from biomass*, Wiley series in renewable resources, pages 1–20. Wiley, Chichester, U.K, 2008. OCLC: 213375885.

[2] Chapter 2: Chemicals from Metabolic Pathways. In J. Villadsen, J. Nielsen, and G. Lidén, editors, *Bioreaction Engineering Principles*, pages 7–62. Springer US, Boston, MA, 2011.

[3] Chapter 3: Elemental and Redox Balances. In J. Villadsen, J. Nielsen, and G. Lidén, editors, *Bioreaction Engineering Principles*, pages 63–118. Springer US, Boston, MA, 2011.

[4] Chapter 9: Design of Fermentation Processes. In J. Villadsen, J. Nielsen, and G. Lidén, editors, *Bioreaction Engineering Principles*, pages 383–458. Springer US, Boston, MA, 2011.

[5] Recombinant Protein Expression in E. coli : A Historical Perspective. In N. A. Burgess-Brown, editor, *Heterologous Gene Expression in E.coli: Methods and Protocols*, volume 1586 of *Methods in Molecular Biology*. Springer New York, New York, NY, 2017.

[6] Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: The European Green Deal., 2019.

[7] M. Aehle, A. Kuprijanov, S. Schaepe, R. Simutis, and A. Lübbert. Simplified off-gas analyses in animal cell cultures for process monitoring and control purposes. *Biotechnology Letters*, 33(11):2103–2110, Nov. 2011.

[8] M. Alibolandi and H. Mirzahoseini. Chemical Assistance in Refolding of Bacterial Inclusion Bodies. *Biochemistry Research International*, 2011:1–6, 2011.

[9] A. Ashoori, B. Moshiri, A. Khaki-Sedigh, and M. R. Bakhtiari. Optimal control of a nonlinear fed-batch fermentation process using model predictive approach. *Journal of Process Control*, 19(7):1162–1173, July 2009.

[10] N. A. Baeshen, M. N. Baeshen, A. Sheikh, R. S. Bora, M. M. M. Ahmed, H. A. I. Ramadan, K. S. Saini, and E. M. Redwan. Cell factories for insulin production. *Microbial Cell Factories*, 13(1):141, Dec. 2014.

[11] R. H. Baltz. Bacteriophage-resistant industrial fermentation strains: from the cradle to CRISPR/Cas9. *Journal of Industrial Microbiology and Biotechnology*, 45(11):1003–1006, Nov. 2018.

[12] F. Baneyx and M. Mujacic. Recombinant protein folding and misfolding in Escherichia coli. *Nature Biotechnology*, 22(11):1399–1408, Nov. 2004.

[13] H. U. Bergmeyer. [New values for the molar extinction coefficients of NADH and NADPH for the use in routine laboratories (author's transl)]. *Zeitschrift Fur Klinische Chemie Und Klinische Biochemie*, 13(11):507–508, Nov. 1975.

[14] A. Berlec and B. Štrukelj. Current state and recent advances in biopharmaceutical production in *Escherichia coli* , yeasts and mammalian cells. *Journal of Industrial Microbiology and Biotechnology*, 40(3-4):257–274, Apr. 2013.

[15] F. R. Blattner, G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. The Complete Genome Sequence of *Escherichia coli* K-12. *Science*, 277(5331):1453–1462, Sept. 1997.

[16] G. Bokinsky, P. P. Peralta-Yahya, A. George, B. M. Holmes, E. J. Steen, J. Dietrich, T. Soon Lee, D. Tullman-Ercek, C. A. Voigt, B. A. Simmons, and J. D. Keasling. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 108(50):19949–19954, Dec. 2011.

[17] G. A. Bowden, A. M. Paredes, and G. Georgiou. Structure and Morphology of Protein Inclusion Bodies in Escherichia Coli. *Nature Biotechnology*, 9(8):725–730, Aug. 1991.

[18] R. Brun, M. Kühni, H. Siegrist, W. Gujer, and P. Reichert. Practical identifiability of ASM2d parameters—systematic selection and tuning of parameter subsets. *Water Research*, 36(16):4113–4127, Sept. 2002.

[19] M. Carrió, N. González-Montalbán, A. Vera, A. Villaverde, and S. Ventura. Amyloid-like Properties of Bacterial Inclusion Bodies. *Journal of Molecular Biology*, 347(5):1025–1037, Apr. 2005.

[20] M. Carrió and A. Villaverde. Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Letters*, 489(1):29–33, Jan. 2001.

[21] M. M. Carrió, J. L. Corchero, and A. Villaverde. Proteolytic digestion of bacterial inclusion body proteins during dynamic transition between soluble and insoluble forms. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1434(1):170–176, Sept. 1999.

[22] S. Chatterjee and A. S. Hadi. *Regression analysis by example*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, N.J, 4th ed edition, 2006. OCLC: ocm66527281.

[23] E. D. B. Clark. Protein refolding for industrial processes. *Current Opinion in Biotechnology*, 12(2):202–207, Apr. 2001.

[24] J. Corchero, R. Cubarsı, S.-O. Enfors, and A. Villaverde. Limitedin VivoProteolysis of Aggregated Proteins. *Biochemical and Biophysical Research Communications*, 237(2):325–330, Aug. 1997.

[25] S. Daume, J. Kager, and C. Herwig. Time Resolved Sensitivity & Identifiability Analysis for Directed Parametrization of Highly Dynamic Models. In *Computer Aided Chemical Engineering*, volume 46, pages 1111–1116. Elsevier, 2019.

[26] S. Daume, S. Kofler, J. Kager, P. Kroll, and C. Herwig. Generic Workflow for the Setup of Mechanistic Process Models. In R. Pörtner, editor, *Animal Cell Biotechnology: Methods and Protocols*, volume 2095 of *Methods in Molecular Biology*, pages 189–211. Springer US, New York, NY, 2020.

[27] D. J. W. De Pauw and P. A. Vanrolleghem. Avoiding the Finite Difference Sensitivity Analysis Deathtrap by Using the Complex-step Derivative Approximation Technique. page 7, July 2006.

[28] M. P. DeLisa, J. Li, G. Rao, W. A. Weigand, and W. E. Bentley. Monitoring GFP-operon fusion protein expression during high cell density cultivation of Escherichia coli using an on-line optical sensor. *Biotechnology and Bioengineering*, 65(1):54–64, Oct. 1999. Publisher: John Wiley & Sons, Ltd.

[29] H. Dhamankar, Y. Tarasova, C. H. Martin, and K. L. Prather. Engineering E. coli for the biosynthesis of 3-hydroxy--butyrolactone (3HBL) and 3,4-dihydroxybutyric acid (3,4-DHBA) as value-added chemicals from glucose as a sole carbon source. *Metabolic Engineering*, 25:72–81, Sept. 2014.

[30] J. J. DiStefano. *Dynamic systems biology modeling and simulation*. Elsevier, Academic Press, Amsterdam, first edition edition, 2013. OCLC: ocn862090686.

[31] X.-Y. Dong, G.-Q. Shi, W. Li, and Y. Sun. Modeling and Simulation of Fed-Batch Protein Refolding Process. *Biotechnology Progress*, 20(4):1213–1219, Aug. 2004.

[32] E. García-Fruitós, N. González-Montalbán, M. Morell, A. Vera, R. M. Ferraz, A. Arís, S. Ventura, and A. Villaverde. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microbial Cell Factories*, 4(1):27, Dec. 2005.

[33] K. V. Gernaey, A. E. Lantz, P. Tufvesson, J. M. Woodley, and G. Sin. Application of mechanistic models to fermentation and biocatalysis for next-generation processes. *Trends in Biotechnology*, 28(7):346–354, July 2010.

[34] S. Gleizer, R. Ben-Nissan, Y. M. Bar-On, N. Antonovsky, E. Noor, Y. Zohar, G. Jona, E. Krieger, M. Shamshoum, A. Bar-Even, and R. Milo. Conversion of Escherichia coli to Generate All Biomass Carbon from CO2. *Cell*, 179(6):1255–1263.e12, Nov. 2019.

[35] S. Gnoth, M. Jenzsch, R. Simutis, and A. Lübbert. Product formation kinetics in a recombinant protein production process. *IFAC Proceedings Volumes*, 40(4):201–206, 2007.

[36] S. Gnoth, M. Jenzsch, R. Simutis, and A. Lübbert. Control of cultivation processes for recombinant protein production: a review. *Bioprocess and Biosystems Engineering*, 31(1):21–39, Jan. 2008.

[37] S. Gnoth, M. Jenzsch, R. Simutis, and A. Lübbert. Product formation kinetics in genetically modified E. coli bacteria: inclusion body formation. *Bioprocess and Biosystems Engineering*, 31(1):41–46, Jan. 2008.

[38] B. Hahn-Hägerdal, M. Galbe, M. Gorwa-Grauslund, G. Lidén, and G. Zacchi. Bio-ethanol – the fuel of tomorrow from the residues of today. *Trends in Biotechnology*, 24(12):549–556, Dec. 2006.

[39] F. S. Hartmann, I. A. Udugama, G. M. Seibold, H. Sugiyama, and K. V. Gernaey. Digital models in biotechnology: Towards multi-scale integration and implementation. *Biotechnology Advances*, 60:108015, Nov. 2022.

[40] F. Hoffmann, C. Posten, and U. Rinas. Kinetic model of in vivo folding and inclusion body formation in recombinantEscherichia coli. *Biotechnology and Bioengineering*, 72(3):315–322, Feb. 2001.

[41] C.-J. Huang, H. Lin, and X. Yang. Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *Journal of Industrial Microbiology and Biotechnology*, 39(3):383–399, Mar. 2012.

[42] D. Humer and O. Spadiut. Wanted: more monitoring and control during inclusion body processing. *World Journal of Microbiology and Biotechnology*, 34(11):158, Nov. 2018.

[43] B. Jürgen, H. Y. Lin, S. Riemschneider, C. Scharf, P. Neubauer, R. Schmid, M. Hecker, and T. Schweder. Monitoring of genes that respond to overproduction of an insoluble recombinant protein inEscherichia coli glucose-limited fed-batch fermentations. *Biotechnology and Bioengineering*, 70(2):217–224, Oct. 2000.

[44] J. Kager, J. Bartlechner, C. Herwig, and S. Jakubek. Direct control of recombinant protein production rates in E. coli fed-batch processes by nonlinear feedback linearization. *Chemical Engineering Research and Design*, 182:290–304, June 2022.

[45] J. Kager, C. Herwig, and I. V. Stelzer. State estimation for a penicillin fed-batch process combining particle filtering methods with online and time delayed offline measurements. *Chemical Engineering Science*, 177:234–244, Feb. 2018.

[46] J. Kager, A. Tuveri, S. Ulonska, P. Kroll, and C. Herwig. Experimental verification and comparison of model predictive, PID and model inversion control in a Penicillium chrysogenum fed-batch process. *Process Biochemistry*, 90:1–11, Mar. 2020.

[47] J. Kaur, A. Kumar, and J. Kaur. Strategies for optimization of heterologous protein expression in E. coli: Roadblocks and reinforcements. *International Journal of Biological Macromolecules*, 106:803–822, Jan. 2018.

[48] J. Kopp, C. Slouka, D. Strohmer, J. Kager, O. Spadiut, and C. Herwig. Inclusion Body Bead Size in E. coli Controlled by Physiological Feeding. *Microorganisms*, 6(4):116, Nov. 2018.

[49] J. Kopp, F. B. Zauner, A. Pell, J. Hausjell, D. Humer, J. Ebner, C. Herwig, O. Spadiut, C. Slouka, and R. Pell. Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles. *Journal of Pharmaceutical and Biomedical Analysis*, 188:113412, Sept. 2020.

[50] T. Langemann, U. B. Mayr, A. Meitz, W. Lubitz, and C. Herwig. Multi-parameter flow cytometry as a process analytical technology (PAT) approach for the assessment of bacterial ghost production. *Applied Microbiology and Biotechnology*, 100(1):409–418, Jan. 2016.

[51] C. Ling, J. Zhang, D. Lin, and A. Tao. Approaches for the generation of active papain-like cysteine proteases from inclusion bodies of Escherichia coli. *World Journal of Microbiology and Biotechnology*, 31(5):681–690, May 2015.

[52] H. Liu, S. Shah, and W. Jiang. On-line outlier detection and data cleaning. *Computers & Chemical Engineering*, 28(9):1635–1647, Aug. 2004.

[53] S. Liu. Chapter 9: Cell Metabolism. In *Bioprocess Engineering*, pages 451–533. Elsevier, 2017.

[54] C. Lizak, Y.-Y. Fan, T. C. Weber, and M. Aebi. *N* -Linked Glycosylation of Antibody Fragments in *Escherichia coli*. *Bioconjugate Chemistry*, 22(3):488–496, Mar. 2011.

[55] A. Luitjens and E. van Corven. Chapter 4: Production and Purification of recombinant proteins. In D. J. A. Crommelin, R. D. Sindelar, and B. Meibohm, editors, *Pharmaceutical Biotechnology: Fundamentals and Applications*, pages 57–82. Springer International Publishing, Cham, 2019.

[56] S. C. Makrides. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiological Reviews*, 60(3):512–538, Sept. 1996.

[57] L. Mears, S. M. Stocks, M. O. Albaek, G. Sin, and K. V. Gernaey. Mechanistic Fermentation Models for Process Design, Monitoring, and Control. *Trends in Biotechnology*, 35(10):914–924, Oct. 2017.

[58] L. Mears, S. M. Stocks, G. Sin, and K. V. Gernaey. A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. *Journal of Biotechnology*, 245:34–46, Mar. 2017.

[59] F. Mergulhão, D. Summers, and G. Monteiro. Recombinant protein secretion in Escherichia coli. *Biotechnology Advances*, 23(3):177–202, May 2005.

[60] A. P. Middelberg. Preparative protein refolding. *Trends in Biotechnology*, 20(10):437–443, Oct. 2002.

[61] T. Miyamoto, S. Okano, and N. Kasai. Inactivation of *Escherichia coli* Endotoxin by Soft Hydrothermal Processing. *Applied and Environmental Microbiology*, 75(15):5058–5063, Aug. 2009.

[62] A. Mogk, M. P. Mayer, and E. Deuerling. Mechanisms of Protein Folding: Molecular Chaperones and Their Application in Biotechnology. *ChemBioChem*, 3(9):807–814, Sept. 2002.

[63] D. F. Müller, K. Lagoda, D. Wibbing, C. Herwig, and J. Kager. Incorporation of error propagation into an elemental balancing based soft-sensor for improved online monitoring of microbial fed-batch processes. In *Computer Aided Chemical Engineering*, volume 51, pages 1177–1182. Elsevier, 2022.

[64] D. F. Müller, D. Wibbing, C. Herwig, and J. Kager. Simultaneous real-time estimation of maximum substrate uptake capacity and yield coefficient in induced microbial cultures. *Computers & Chemical Engineering*, 173:108203, May 2023.

[65] H. Narayanan, M. F. Luna, M. Stosch, M. N. Cruz Bournazou, G. Polotti, M. Morbidelli, A. Butté, and M. Sokolov. Bioprocessing in the Digital Age: The Role of Process Models. *Biotechnology Journal*, 15(1):1900172, Jan. 2020.

[66] R. S. Oosting. Chapter 1: Molecular Biotechnology: From DNA Sequence to Therapeutic Protein. In D. J. A. Crommelin, R. D. Sindelar, and B. Meibohm, editors, *Pharmaceutical Biotechnology: Fundamentals and Applications*, pages 1–18. Springer International Publishing, Cham, 2019.

[67] S. Pan, M. Zelger, A. Jungbauer, and R. Hahn. Integrated continuous dissolution, refolding and tag removal of fusion proteins from inclusion bodies in a tubular reactor. *Journal of Biotechnology*, 185:39–50, Sept. 2014.

[68] J. N. Pauk, J. Raju Palanisamy, J. Kager, K. Koczka, G. Berghammer, C. Herwig, and L. Veiter. Advances in monitoring and control of refolding kinetics combining PAT and modeling. *Applied Microbiology and Biotechnology*, 105(6):2243–2260, Mar. 2021.

[69] J. V. Pham, M. A. Yilma, A. Feliz, M. T. Majid, N. Maffetone, J. R. Walker, E. Kim, H. J. Cho, J. M. Reynolds, M. C. Song, S. R. Park, and Y. J. Yoon. A Review of the Microbial Production of Bioactive Natural Products and Biologics. *Frontiers in Microbiology*, 10:1404, June 2019.

[70] M. Pianta and M. Lucchese. Rethinking the European Green Deal: An Industrial Policy for a Just Transition in Europe. *Review of Radical Political Economics*, 52(4):633–641, Dec. 2020.

[71] S. A. Pizarro, R. Dinges, R. Adams, A. Sanchez, and C. Winter. Biomanufacturing process analytical technology (PAT) application for downstream processing: Using dissolved oxygen as an indicator of product quality for a protein refolding reaction. *Biotechnology and Bioengineering*, 104(2):340–351, Oct. 2009.

[72] X. Ponsoda, R. Jover, J. V. Castell, and M. J. Gómez-Lechón. Measurement of intracellular LDH activity in 96-well cultures: A rapid and automated assay for cytotoxicity studies. *Journal of Tissue Culture Methods*, 13(1):21–24, Mar. 1991.

[73] R. Pörtner, editor. *Animal Cell Biotechnology: Methods and Protocols*, volume 2095 of *Methods in Molecular Biology*. Springer US, New York, NY, 2020.

[74] A. S. Rathore, P. Bade, V. Joshi, M. Pathak, and S. K. Pattanayek. Refolding of biotech therapeutic proteins expressed in bacteria: review: Refolding of biotech therapeutic proteins expressed in bacteria. *Journal of Chemical Technology & Biotechnology*, 88(10):1794–1806, Oct. 2013.

[75] W. N. Reichelt, M. Brillmann, P. Thurrold, P. Keil, J. Fricke, and C. Herwig. Physiological capacities decline during induced bioprocesses leading to substrate accumulation. *Biotechnology Journal*, 12(7):1600547, July 2017.

[76] W. N. Reichelt, F. Haas, P. Sagmeister, and C. Herwig. Bioprocess development workflow: Transferable physiological knowledge instead of technological correlations. *Biotechnology Progress*, 33(1):261–270, Jan. 2017.

[77] W. N. Reichelt, P. Thurrold, M. Brillmann, J. Kager, J. Fricke, and C. Herwig. Generic biomass estimation methods targeting physiologic process control in induced bacterial cultures. *Engineering in Life Sciences*, 16(8):720–730, Nov. 2016.

[78] U. Rinas, E. Garcia-Fruitós, J. L. Corchero, E. Vázquez, J. Seras-Franzoso, and A. Villaverde. Bacterial Inclusion Bodies: Discovering Their Better Half. *Trends in Biochemical Sciences*, 42(9):726–737, Sept. 2017.

[79] D. Rodrik. Green industrial policy. *Oxford Review of Economic Policy*, 30(3):469–491, Sept. 2014.

[80] K. E. Royle, I. Jimenez del Val, and C. Kontoravdi. Integration of models and experimentation to optimise the production of potential biotherapeutics. *Drug Discovery Today*, 18(23-24):1250–1255, Dec. 2013.

[81] S. Sahdev, S. K. Khattar, and K. S. Saini. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry*, 307(1-2):249–264, Nov. 2007.

[82] M. Sauer and D. Mattanovich. Construction of microbial cell factories for industrial bioprocesses. *Journal of Chemical Technology & Biotechnology*, 87(4):445–450, Apr. 2012.

[83] A. Savitzky and M. J. E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8):1627–1639, July 1964.

[84] K. Schügerl and K.-H. Bellgardt, editors. *Bioreaction Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.

[85] L. L. Simon, H. Pataki, G. Marosi, F. Meemken, K. Hungerbühler, A. Baiker, S. Tummala, B. Glennon, M. Kuentz, G. Steele, H. J. M. Kramer, J. W. Rydzak, Z. Chen, J. Morris, F. Kjell, R. Singh, R. Gani, K. V. Gernaey, M. Louhi-Kultanen, J. O'Reilly, N. Sandler, O. Antikainen, J. Yliruusi, P. Frohberg, J. Ulrich, R. D. Braatz, T. Leyssens, M. von Stosch, R. Oliveira, R. B. H. Tan, H. Wu, M. Khan, D. O'Grady, A. Pandey, R. Westra, E. Delle-Case, D. Pape, D. Angelosante, Y. Maret, O. Steiger, M. Lenner, K. Abbou-Oucherif, Z. K. Nagy, J. D. Litster, V. K. Kamaraju, and M.-S. Chiu. Assessment of Recent Process Analytical Technology (PAT) Trends: A Multiauthor Review. *Organic Process Research & Development*, 19(1):3–62, Jan. 2015.

[86] R. Simutis and A. Lübbert. Bioreactor control improves bioprocess performance. *Biotechnology Journal*, 10(8):1115–1130, Aug. 2015.

[87] G. Sin, K. V. Gernaey, and A. E. Lantz. Good modeling practice for PAT applications: Propagation of input uncertainty and sensitivity analysis. *Biotechnology Progress*, 25(4):1043–1053, July 2009.

[88] C. Slouka, J. Kopp, S. Hutwimmer, M. Strahammer, D. Strohmer, E. Eitenberger, A. Schwaighofer, and C. Herwig. Custom made inclusion bodies: impact of classical process parameters and physiological parameters on inclusion body quality attributes. *Microbial Cell Factories*, 17(1):148, Dec. 2018.

[89] C. Slouka, J. Kopp, D. Strohmer, J. Kager, O. Spadiut, and C. Herwig. Monitoring and control strategies for inclusion body production in E. coli based on glycerol consumption. *Journal of Biotechnology*, 296:75–82, Apr. 2019.

[90] K.-H. Song, H.-H. Lee, and H.-H. Hyun. Characterization of salt-tolerant mutant for enhancement of l -threonine production in Escherichia coli. *Applied Microbiology and Biotechnology*, 54(5):647–651, Nov. 2000.

[91] E. Soriano, N. Borth, H. Katinger, and D. Mattanovich. Flow Cytometric Analysis of Metabolic Stress Effects Due to Recombinant Plasmids and Proteins in Escherichia coli Production Strains. *Metabolic Engineering*, 1(3):270–274, July 1999.

[92] E. J. Steen, Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. del Cardayre, and J. D. Keasling. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, 463(7280):559–562, Jan. 2010.

[93] L. Strandberg and S. O. Enfors. Factors influencing inclusion body formation in the production of a fused protein in Escherichia coli. *Applied and Environmental Microbiology*, 57(6):1669–1674, June 1991.

[94] F. Studier and B. A. Moffatt. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, 189(1):113–130, May 1986.

[95] K. Talmadge and W. Gilbert. Cellular location affects protein stability in Escherichia coli. *Proceedings of the National Academy of Sciences*, 79(6):1830–1833, Mar. 1982.

[96] The UniProt Consortium. J9VF44 - J9VF44_trase. IgY(Delta Fc)5. Last viewed on the 21.02.2023 at https://www.uniprot.org/uniprotkb/J9VF44/entry#sequences.

[97] The UniProt Consortium. P09681 - GIP_human. Gastric inhibitory polypeptide. Last viewed on the 21.02.2023 at https://www.uniprot.org/uniprotkb/P09681/entry#sequences.

[98] The UniProt Consortium. P0DOX5 - IGG1_human. Immunoglobulin gamma-1 heavy chain. Last viewed on the 21.02.2023 at https://www.uniprot.org/uniprotkb/P0DOX5/entry#sequences.

[99] The UniProt Consortium. P42212 - GFP_aeqvi. Green fluorescent protein. Last viewed on the 21.02.2023 at https://www.uniprot.org/uniprotkb/P42212/entry#sequences.

[100] The UniProt Consortium. P56512 - LDH1_lacpl. L-lactate dehydrogenase 1. Last viewed on the 21.02.2023 at https://www.uniprot.org/uniprotkb/P56512/entry.

[101] The UniProt Consortium, A. Bateman, M.-J. Martin, S. Orchard, M. Magrane, S. Ahmad, E. Alpi, E. H. Bowler-Barnett, R. Britto, H. Bye-A-Jee, A. Cukura, P. Denny, T. Dogan, T. Ebenezer, J. Fan, P. Garmiri, L. J. da Costa Gonzales, E. Hatton-Ellis, A. Hussein, A. Ignatchenko, G. Insana, R. Ishtiaq, V. Joshi, D. Jyothi, S. Kandasaamy, A. Lock, A. Luciani, M. Lugaric, J. Luo, Y. Lussi, A. MacDougall, F. Madeira, M. Mahmoudy, A. Mishra, K. Moulang, A. Nightingale, S. Pundir, G. Qi, S. Raj, P. Raposo, D. L. Rice, R. Saidi, R. Santos, E. Speretta, J. Stephenson, P. Totoo, E. Turner, N. Tyagi, P. Vasudev, K. Warner, X. Watkins, R. Zaru, H. Zellner, A. J. Bridge, L. Aimo, G. Argoud-Puy, A. H.

Auchincloss, K. B. Axelsen, P. Bansal, D. Baratin, T. M. Batista Neto, M.-C. Blatter, J. T. Bolleman, E. Boutet, L. Breuza, B. C. Gil, C. Casals-Casas, K. C. Echioukh, E. Coudert, B. Cuche, E. de Castro, A. Estreicher, M. L. Famiglietti, M. Feuermann, E. Gasteiger, P. Gaudet, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz, C. Hulo, N. Hyka-Nouspikel, F. Jungo, A. Kerhornou, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, V. Muthukrishnan, S. Paesano, I. Pedruzzi, S. Pilbout, L. Pourcel, S. Poux, M. Pozzato, M. Pruess, N. Redaschi, C. Rivoire, C. J. A. Sigrist, K. Sonesson, S. Sundaram, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, and J. Zhang. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1):D523–D531, Jan. 2023.

[102] M. Theisen and J. C. Liao. Industrial Biotechnology: Escherichia coli as a Host. In *Industrial Biotechnology*. Wiley-VCH Verlag GmbH & Co., Weinheim, 2017.

[103] J. G. Thomas and F. Baneyx. Divergent Effects of Chaperone Overexpression and Ethanol Supplementation on Inclusion Body Formation in RecombinantEscherichia coli. *Protein Expression and Purification*, 11(3):289–296, Dec. 1997.

[104] S. Ulonska, D. Waldschitz, J. Kager, and C. Herwig. Model predictive control in comparison to elemental balance control in an E. coli fed-batch. *Chemical Engineering Science*, 191:459–467, Dec. 2018.

[105] N. A. Valdez-Cruz, L. Caspeta, N. O. Pérez, O. T. Ramírez, and M. A. Trujillo-Roldán. Production of recombinant proteins in E. coli by the heat inducible expression system based on the phage lambda pL and/or pR promoters. *Microbial Cell Factories*, 9(1):18, Dec. 2010.

[106] S. Ventura and A. Villaverde. Protein quality in bacterial inclusion bodies. *Trends in Biotechnology*, 24(4):179–185, Apr. 2006.

[107] C. E. Vickers, D. Klein-Marcuschamer, and J. O. Krömer. Examining the feasibility of bulk commodity production in Escherichia coli. *Biotechnology Letters*, 34(4):585–596, Apr. 2012.

[108] J. Villadsen, J. Nielsen, and G. Lidén. *Bioreaction Engineering Principles*. Springer US, Boston, MA, 2011.

[109] A. F. Villaverde, A. Barreiro, and A. Papachristodoulou. Structural Identifiability of Dynamic Systems Biology Models. *PLOS Computational Biology*, 12(10):e1005153, Oct. 2016.

[110] G. Walsh and E. Walsh. Biopharmaceutical benchmarks 2022. *Nature Biotechnology*, 40(12):1722–1760, Dec. 2022.

[111] D. S. Weaver, I. M. Keseler, A. Mackie, I. T. Paulsen, and P. D. Karp. A genome-scale metabolic flux model of Escherichia coli K–12 derived from the EcoCyc database. *BMC Systems Biology*, 8(1):79, 2014.

[112] P. Wechselberger, P. Sagmeister, and C. Herwig. Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes. *Bioprocess and Biosystems Engineering*, 36(9):1205–1218, Sept. 2013.

[113] T. Wheeler and J. von Braun. Climate Change Impacts on Global Food Security. *Science*, 341(6145):508–513, Aug. 2013.

[114] D. J. Wurm, L. Marschall, P. Sagmeister, C. Herwig, and O. Spadiut. Simple monitoring of cell leakiness and viability in *Escherichia coli* bioprocesses-A case study. *Engineering in Life Sciences*, 17(6):598–604, June 2017.

[115] D. J. Wurm, J. Quehenberger, J. Mildner, B. Eggenreich, C. Slouka, A. Schwaighofer, K. Wieland, B. Lendl, V. Rajamanickam, C. Herwig, and O. Spadiut. Teaching an old pET new tricks: tuning of inclusion body formation and properties by a mixed feed system in E. coli. *Applied Microbiology and Biotechnology*, 102(2):667–676, Jan. 2018.

[116] D. J. Wurm, L. Veiter, S. Ulonska, B. Eggenreich, C. Herwig, and O. Spadiut. The E. coli pET expression system revisited—mechanistic correlation between glucose and lactose uptake. *Applied Microbiology and Biotechnology*, 100(20):8721–8729, Oct. 2016.

[117] X. Zhang, K. Jantama, J. C. Moore, L. R. Jarboe, K. T. Shanmugam, and L. O. Ingram. Metabolic evolution of energy-conserving pathways for succinate production in Escherichia coli. *Proceedings of the National Academy of Sciences*, 106(48):20180–20185, Dec. 2009.

[118] K. J. Åström and R. M. Murray. *Feedback systems: an introduction for scientists and engineers*. Princeton University Press, Princeton, 2008. OCLC: ocn183179623.