Technische Universität München
TUM School of Engineering and Design

Model-based investigation of population heterogeneity induced variations in process performance

Dieu Thi Doan

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung einer

Doktorin der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz: Prof. Dr.-Ing. Dirk Weuster-Botz

Prüfer*innen der Dissertation:

1. Prof. Dr.-Ing. Andreas Kremling
2. Prof. Jérôme Morchain

Die Dissertation wurde am 15.06.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 23.10.2023 angenommen.

Acknowledgements

First of all, I want to thank my Doktorvater, Prof. Dr.-Ing. Andreas Kremling, for providing me the opportunity of spending the last few years on interesting research questions and all the experiences I have gained during this time. Thank you for the patience and the weekly discussions in the beginning, when I had little knowledge in the field of biotechnology, and continuous confidence in me to continue with the path of this PhD. You always had an open ear every time I was stuck with a problem and gave me your suggestions.

Furthermore, I want to thank Jérôme Morchain for helping me find my way to understand the complex nature of population heterogeneities and the very enthusiastic monthly hour-long discussions we had the last two years. These discussions were really fruitful, even though they sometimes created more open questions than answering some. But I think that is part of research and makes it so interesting and wonderful. I was very happy to be welcomed to Toulouse for a month to collaborate with you and you showed me around your research group with my very limited French. During this very short time I met a lot of wonderful nice and warm people.

I also want to thank my project partners Anna-Lena Heins and Manh Dat Hoang for our collaboration and all the interesting project meetings. You gave me a better understanding of the experimental side of the project and what I am trying to find a theoretical answer for.

Next, I want to thank Prof. Dr. Johannes Müller and Prof. Dr. Christina Kuttler for awakening my interest in mathematical modeling of biological systems and supervising my theses during my studies. Without you I would not be where I am today.

Thanks to all my colleagues Marleen Beentjes, Martina Cantone, Karina Hobmeier, José Garcia Lima, Carina Meiners and Ana-Sofia Ortega-Arbulú for making work additionally enjoyable. My special thanks to Franziska Kratzl for being the best work buddy for going to all the project meetings with, sharing sushi and Sauerkram with me, accompanying me through all the ups and downs of the "Phasen einer Doktorarbeit" and a lot more things that would take to long to mention. Additionally, I want to thank Alberto Marin-Sanguino and Hannes Löwe for listen to all my questions in the beginning of my journey, when I was trying to understand the basics of modeling in biotechnology, and your effort to help me. For proof reading my dissertation and all the questions regarding English syntax and "Kommasetzung", I want to give thanks to you, José. Here, I also want to thank Igor for proof reading parts of my dissertation as well.

Finally, I want to, especially, thank my family and friends who are always there for me and are never tired of their never-ending support for me. Thank you, Florian, for being there all the way with me during my journey of this PhD and being so supportive.

Zusammenfassung

Heterogenität in Populationen wird in allen biologischen Systemen in der Natur und Bioprozessen beobachtet. Sie resultiert durch zum einen genetischer Diversität auf Populationsebene oder phenotypischer Differenzen zwischen genetisch identischen Zellen. Zusätzlich können stochastische Schwankungen in biochemischen Prozessen (intrinisches Rauschen) oder Fluktuationen in den Umgebungsbedingungen (extrinsisches Rauschen), zu Unterschieden in Genexpression führen. Aufgrund verbesserter Fitness und Stresstoleranz kann Diversität einen selektiven Vorteil verschaffen. Andererseits wird Populationsheterogenität als unerwünschte Folge bei der Skalierung eines Bioprozesses auf Industriemaß erachtet, da diese mit verringerter Produktivität, Ausbeute und vermehrter Nebenproduktbildung assoziiert wird. Aufgrund der vielen Ursachen und Faktoren, sind die Mechanismen vom besonderen Interesse. Das Ziel dieser Arbeit ist es ein tieferes Verständnis von dem Verhalten einer Population unter wechselnden Umgebungsbedingungen zu gewinnen und eine quantitative Beschreibung durch mathematische Modelle und Simulationen zu erhalten. Zuerst wurde ein grobkörniges Modell als Ansatz für massenerhaltende Modelle zur Beschreibung von bakteriellem Wachstums in einem biotechnologischen Produktionsprozess erarbeitet. Hierfür wurde ein Zusammenhang zwischen der Wachstumsrate und den Aufnahme- und Produktionsraten mit dem intrazellulären Netzwerk und der Zellkomposition basierend auf Massenbilanzen präsentiert. Ferner wurden Eigenschaften des Modells, wie Gleichgewichtslagen und Relaxationszeiten untersucht. Das Konzept von Zellen im und außerhalb ihres Gleichgewichtzustands wurde eingeführt, wobei Zellen im Gleichgewicht eine Zellkomposition im Equilibrium mit ihrer Umgebung besitzen. Die Analyse der Relaxtionszeiten offenbarte zwei Zeitskalen der Zellreaktion auf extrinsische Veränderungen für Zellen außerhalb ihres Gleichgewichtszustands und eine Zeitskala für Zellen im Gleichgewicht. Die Beschreibung des intrazellulären Netzwerks, welches aus dem grobkörnigen Modells resultiert, bildet eine Basis für die Konzeption eines Populationsbilanzmodelles. Dies ermöglicht eine segregierte Sichtweise auf die Population und die Einbindung der Zellteilung. Die Neuartigkeit des Modells liegt in der Fähigkeit eine Population mit strukturierter Zellkomposition und Zellteilung in einem Bioreaktor zu betrachten. Aufgrund des hohen Rechenaufwands von Populationsbilanzgleichungen, wurde ein zweiter Ansatz, welcher ein Euler-Lagrange-Kompartimenten-Ansatz ist, entwickelt. Dieser beschreibt die biologische Phase als einzelne Teilchen, welche mit dem grobkörnigen Modells zur Beschreibung der intrazellulären Prozesse ausgestattet sind und dementsprechend auf ihre Umgebung reagieren. Die Erkenntnis von zwei unterschiedlichen Relaxtionszeiten für Zellen außerhalb ihres Gleichgewichtzustandes motiviert die Einbindung von intrinsischem Rauschen, um den Zellzustand von der lokalen Substratkonzentration in der Flüssigphase zu entkoppeln. Durch das Simulieren von verschiedenen Bioreaktorsystemen wurde gezeigt, dass das Ausmaß und die Frequenz der externen Perturbationen zusammen mit der zufälligen Zuordnung der Zellkomposition bei Zellgeburt die Entstehung von Subpopulationen begünstigt. Experimentelle Daten eines L-Phenylalanin-Produktionsprozesses wurden als exemplarischer Bioprozess verwendet und eine gute Übereinstimmung von Simulation und Daten wurde erzielt.

Abstract

Heterogeneity within a population is universally observed in all natural systems and bioprocesses. It is a result from either genetic diversity at population level or phenotypic differences between genetically identical cells and may be driven by stochastic fluctuations within biochemical processes (intrinsic noise) or environmental changes perceived by the cell population (external noise), leading to differential gene expression. In many cases, diversity may provide a selective advantage due to an improved fitness and higher stress tolerance during changes in environmental conditions. On the other hand, population heterogeneity is often perceived as an undesirable consequence during upscaling of a bioprocess as it may result in reduced productivity, yield and increased by-product formation. Because of the large variety of sources and factors affecting heterogeneity, its mechanisms are of particular interest, providing motivation for this work. The aim is to gain deeper insight into the behavior of a population exposed to environmental changes and develop a quantitative description using mathematical models and simulations.

As first step, a framework in form of a coarse-grained model was established that presents a strict mass conservative model for bacterial growth during a biotechnological production process. Conventionally, the growth rate is determined by an empirical function, but in this study a relationship between the growth rate and the exchange rate of the entire network and cell composition, which guarantees mass conservation at all time points, is presented. Furthermore, the key properties of the model were analyzed, including the steady state behavior. The concept of balanced and unbalanced cells was introduced, where balanced cells exhibit a cell composition in steady state with their environment, while unbalanced cells are out of equilibrium. The investigation of the time scales revealed two different time scales in the cell response for unbalanced cells, and only one time scale when balanced cells respond to external fluctuations.

The description of the intracellular network obtained from the coarse-grained model was integrated into the framework of a population balance equation model. This enabled a segregated view of the population and incorporation of the cell division process. The novelty of the proposed model lies in its ability to depict a population with structured biomass composition and cell division in the environment of a well-mixed homogeneous bioreactor. Due to the high computational effort of solving a population balance equation, a second segregated model framework was proposed, which is a Euler-Lagrange compartment approach, where the biological phase is represented by trackable particles. Each particle was equipped with the coarse-grained model to describe the intracellular response to extracellular conditions. The finding of two time scales in case of unbalanced cells showed the necessity to include intrinsic noise, in order to maintain different time scales in cellular response and decouple the cellular state from the local concentrations in the liquid phase. Through studying various examples of scale-down bioreactor systems, it was revealed that the magnitude and frequency of external perturbations, combined with randomization of properties at birth, favors the emergence of subpopulations. Experimental data from an L-phenylalanine production process were taken as an exemplary bioprocess, which showed a good agreement between simulation and experimental data.

Contents

1 Introduction: Population heterogeneity in bioprocesses

Independent of the studied organism, population heterogeneity is observed in all bioprocesses even in homogeneous environments. It is especially associated to be a consequence during the upscaling of a bioprocess to an industrial-scale bioprocess, where concentration gradients of process parameters arise from mixing insufficiencies due to higher viscosities and higher volumes. Therefore, single cells are exposed to environmental changes [32, 38] and increased expression of genes associated with stress have been observed [113]. An originally isogenic monoculture can then evolve to a culture with different phenotypes to cope with this dynamic environment [171], leading to possibly reduced yield, productivity, growth and increased by-product formation [57]. But on the other hand, population heterogeneity can also result in cells with improved fitness due to a higher stress tolerance from adapting to the fluctuating environment. Even though the phenomenon of population heterogeneity is well-known, the origins and mechanisms require a deeper understanding. Gaining more knowledge about the changes of the individual cells and their interaction with the environment, can therefore be beneficial to the optimization of bioprocesses [32, 17].

The origins and influences of population heterogeneity can be divided into two classes: intrinsic and extrinsic. Intrinsic heterogeneity has its sources in the stochasticity of gene expression and biochemical reactions. Due to low numbers in cellular components, especially reactants like DNA structures, promoters, mRNA and some proteins, fluctuations occur and largely impact the dynamics of the population [69]. Furthermore, cellular processes such as growth, replication and cell division contribute to differences between cells as well [65]. During cell division the molecules are separated randomly, resulting in so-called partitioning errors [168, 22, 99]. Thus, observed noise in RNAs and proteins is a consequence of both noise in gene expression and partitioning errors and the origins are difficult to retrace [65]. As intrinsic heterogeneity, regardless of its source, is not affected by conditions of the environment, it occurs even in homogeneous environments [29, 2]. Extrinsic heterogeneity originates from changes in the environment including gradients in substrate, temperature and oxygen. In industrial-scale bioreactors the substrate concentration can range between the concentration of the feeding solution to absolute substrate exhaustion. While the former arises mainly from insufficient mixing, the latter is formed because of a disparity between mixing and nutrient uptake. Therefore, cells present in the bioreactor may move between one extreme to another and back leading to different cell compositions of the population as response and ultimately to fluctuating growth rates [32, 73, 57, 84].

It is assumed that cell populations in bioreactors use the same strategies as in natural habitats to cope with a dynamic environment, as the population distributions resemble each other. The strategies can be divided into categories.

On one hand, noise in gene expression can result in population heterogeneity, often manifesting in multimodal distributions of cell characteristics, and can appear even in stable environments [63].

The sources of noise can be divided into intrinsic noise, originating from the randomness of gene expression due to low amount of reacting molecules, and extrinsic noise, which arise from the interaction of the cell with global cellular and environmental factors [57, 48]. In *E. coli*, at low levels of gene expression, intrinsic noise is the driving force in distribution of gene expression, whereas at high expression levels extrinsic noise dominates [160]. It is suggested, that gene expression noise is used as a regulatory mechanism to enhance population robustness [148, 120].

Additionally, in case of bet-hedging population heterogeneity occurs prior to a change in the environment. This ensures that some individuals of a cell population can survive future environments even though it may be disadvantageous for the current environment. Thus, rather than being a response to a certain environmental change, it spreads the risk in order to minimize the temporal variance of offspring numbers per individual to maximize the mean fitness of the whole population over various environments for many generations [57, 25, 150, 2]. This is achieved by stochastically developed phenotypes with reduced fitness in a homogeneous environment and is predicted to be successful in case of sudden environmental changes with severe costs for maladaptation or for infrequently changing environments [48, 80].

One special case of bet-hedging is persistence, where upon stress exposure in a fluctuating environment the population splits into persister and non-persister cells [62]. Normally the subpopulation of persister cells is smaller and performs poorer in an environment without stress. During stress exposure, persister cells are unaffected, whereas non-persister cells go dormant or grow poorly [2, 90]. Interestingly, after stress removal, persister cells in a yeast population are able to revert to fast growing cells, enabling rapid growth in the surviving cells [157]. Extrinsic and intrinsic noise is also an important factor of persistence [159, 90].

Another special case of bet-hedging is bistability, where due to noise in gene expression the population is divided into two stable subpopulations. Cells then belong to distinct "ON" and "OFF" subpopulations with different protein content, where the former state has a high and the latter state a low protein level and are able to switch between these gene expression states [116]. The time cells spent in the two states is referred to as memory and depending on the history of the population, cells are more likely to switch to the "ON" state in a stress recurring environment to improve their fitness [172, 32]. Additionally, bistability also means that cells are either in one or the other state and cannot rest in an intermediate state [35].

The last category of population heterogeneity is division of labor which involves cell-to-cell-interaction where diffusive molecules are interchanged. In this process, one phenotype sacrifices some of its fitness to improve the fitness over the whole population. Thus, this strategy depends on the neighboring cells in the environment and is especially advantageous when multiple cellular processes are present, but not compatible or efficient in the same cell [2, 17]. During a bioprocess, cells can employ more than one of the strategies during different environments and thus the effects of the strategies can be superimposed.

Investigating the gradients occurring in bioreactors is a challenging task, as the complex interaction between the flow field in the bioreactor and the physiology of the cell population has to be understood. Arising population heterogeneity are mostly analyzed through lab-scale experiments simulating the conditions of large-scale reactors in scale-down reactors that simulate the fluctuating environments, also called multi-compartment reactors. The advantages of scale-down reactors are that they are an easy and cheap alternative to full large-scale experiments and enable investigations in a well-controlled environment [57]. There are many different set-ups for scale-down reactors: a stirred tank bioreactor (STR) with separated internal compartments,

a STR with randomly fluctuating substrate addition, a cascade of two or more STRs, tubular loop bioreactors and a STR coupled to one or two plug-flow reactors (PFR), where the latest set-up is most frequently used. It simulates the gradient zone in the PFR and the well-mixed environment in the STR [122]. With such setups, the influence of gradients of different process parameters on population heterogeneity can be studied using reporter strains, which carry fluorescent proteins and can directly be linked to specific cell characteristics [85, 87]. Reporter strains are the most common tools to visualize single cell dynamics in real time, since they are fast and non-invasive. The growth of single cells has been monitored for different growth phases and in pulse experiments [49, 56, 149]. Reporter strains have also been used to detect cellular stress response to various stresses, e.g. starvation and nutrient limitation, and cell robustness for different process conditions [30]. Furthermore, they can sense oxygen limitation [43]. It is possible to introduce several fluorescent proteins into one strain and measure different cell characteristics at the same time [32].

In addition to insight gained through experimental work, a better understanding of the interplay between different variables can be achieved by employing mathematical models, which has become a standard in process engineering. The classical approach in biotechnology is to combine models describing the cellular environment in a typically well-mixed bioreactor with models describing all the reactions related to biomass. But for the purpose of investigating population heterogeneity, more sophisticated modeling approaches are required to capture the differences between the cells of the population, since former models typically average the values over the whole population, thus making it impossible to characterize the processes in single cell [4]. This can be accomplished by employing segregated models [86, 130]. Previous studies have proposed modeling approaches to investigate the influence of gradients present in a bioreactor on population heterogeneity [154, 51, 112]. Here, black-box kinetics for the biomass were often assumed, which means that intracellular reactions were not considered, only the substrate uptake and output of the desired product [87]. The downside of a majority of those investigations is the disregard of the different time scales of cell adaption to the environment, as they assume instant equilibrium between the extracellular and intracellular conditions.

Additionally, we established that intrinsic noise is another important factor resulting in population heterogeneity, which was often neglected in bioreactor studies as well [87]. Due to random fluctuations and switches between different cell states, a deterministic approach is not suitable and a stochastic approach has to be used. Stochastic approaches are not commonly used, as deterministic models are of moderate complexity and can often be analyzed directly. This does not always apply to stochastic approaches. As intrinsic noise occurs during biochemical reactions, models which describe the system at single-species resolution are more advantageous leading to a higher complexity than models on population level [47, 63]. Because of their restriction and high complexity, the analysis of the mechanisms of intrinsic noise is challenging by itself and a combination with the influence of dynamic fluid pattern in the environment is not feasible.

The points raised motivates this work to develop a framework for studying population heterogeneity induced by process variations and find a middle ground in complexity to study extrinsic and intrinsic noise. In large-scale bioreactors cells may be exposed to zones with nutrient excess or scarcity affecting the cell state. The goal is to apply this framework to a population in a bio-reactor system with two distinguished zones consisting of a well-mixed region and a zone exhibiting gradients to simulate two extreme scenarios a cell encounters.

The upcoming chapters are structured as follows: In Chapter 2 the theoretical background and methods, including the most important mathematical concepts and tools, used in this work

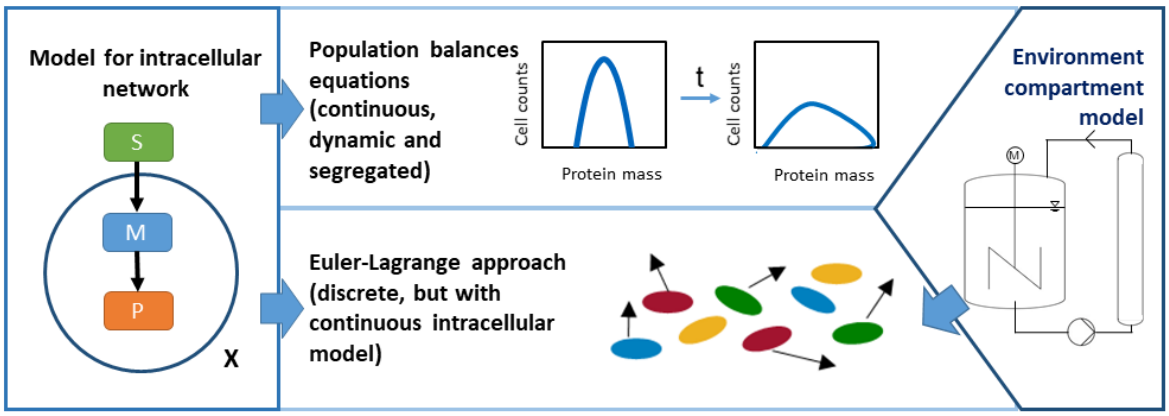

Figure 1.1: Model approaches to describe population heterogeneity proposed in this work: A model for the intracellular network is introduced and integrated into two separate segregated approaches: a population balance equation and an discrete approach, which is coupled with an environment compartment model, resulting in an Euler-Lagrange approach.

are explained. As a foundation for segregated models, a deterministic model describing microbial growth in form of a coarse-grained model based on strict mass conservation is established in Chapter 3. Several properties of the coarse-grained model are addressed and the model is then applied to a L-phenylalanine production process, which serves as an exemplary biotechnological production process. The description of the intracellular network is integrated into a population balance equation (PBE) model, which is a dynamic, continuous and segregated approach, in Chapter 4. Due to the challenging nature of a PBE, we focus on cell division as source for intrinsic noise. The influence of cell division is investigated for a fed-batch process and compared to a model based on average values. A second segregated approach is proposed in Chapter 5 which is an Euler-Lagrange approach. In contrast to the PBE, this approach is discrete, as the cells are considered to be discrete entities, which are equipped with an intracellular network established in Chapter 3. Here, the influence of extrinsic and intrinsic noise due to cell division is investigated. Extrinsic noise is implemented through the simulation of the hydrodynamics of a bioreactor using a compartment model (CM) approach. Then, various examples of scale-down bioreactor studies are described and the resulting cell property distribution analyzed. The presented model approaches are pictured in Figure 1.1. Chapters 3-5 close off with a short summary and discussion and Chapter 6 provides an overall discussion of the results and an outlook on future directions.

2 Theoretical background and methods

In this chapter, the most important definitions, terminology and methods used in this work are introduced to describe population heterogeneity in a dynamic environment of a bioreactor. The focal point of this study lies in the description of bacterial growth and, in particular, the interplay between the environment perceived by the cells and the cellular state. Mathematical models can take many different forms such as dynamical systems, differential equations and statistical models. As we are especially interested in the behavior of a system over time, we have to utilize dynamic models, which are typically represented by differential equations. Deterministic models form a common starting point for describing a system of interest. Thus, the chapter starts with an introduction into deterministic models and continues with the formulation of a mass conservating model approach for microbial growth. The set of mass balance equations is then complemented by equations describing the environment of a bioreactor.

In general, one can classify mathematical models for cell populations into unsegregated/segregated and unstructured/structured models [153]. The simplest case of an unstructured and unsegregated model considers a homogeneous population of averaged cells, while the most complex case, segregated and structured models, describe a heterogeneous population of single cells taking into account a cellular network with more than one variable [45]. There are two general approaches that can quantify the degree of heterogeneity with special focus on dynamic changes: PBE models [107, 86, 4, 109, 97], and Euler-Lagrange models [82, 81, 27, 79, 51], where the first approach is a continuous and the latter a discrete approach [58]. Both approaches are introduced here as well. Additionally, the concept of a CM approach for the hydrodynamics of a bioreactor is presented. An overview of the approaches is depicted in Figure 1.1.

2.1 Definition of a deterministic model

A deterministic model is based on a set of assumptions and rules used to describe the relationships between the variables of the system. For a given set of input, it produces the same output repeatedly, thus it is a non-probabilistic approach. ODEs are the most prominent deterministic models, which describe the evolution of a variable with respect to a second variable such as time or space. If the dynamics of several variables are considered, one speaks of an ODE system. A first-order ODE system has the following general description

$$\dot{c} = f(c), \tag{2.1}$$

where $c \in \mathbb{R}^r$ is a vector containing the variables of interest and f describes its rate of change with respects to a quantity of interest. Here, this variable is the time t. Thus, we use the common notation of $\dot{c}$ to indicate the time derivative. The term "first-order" indicates, that the function f is only a function of c and t and does not contain derivatives of c, which would make the ODE system a higher-order problem.

2.1.1 Steady states

Often times, one is not interested in the exact solution, but in the long-time behavior or in the so-called steady states or stationary solution. A steady state s is given if the state fulfills the equation

$$\dot{c} = f(s) = 0. \tag{2.2}$$

If this state is reached, the system does not leave this state, as the rate of change is zero at this point. In the following, we want to study the solution space of the steady states for a linear ODE system, which reads

$$\dot{c} = A\ c, \tag{2.3}$$

where $c \in \mathbb{R}^r$ and $A \in \mathbb{R}^{m \times r}$. The steady state space of this system is given by the kernel or null space of A. In the scope of this work, we use the term "kernel". It is formally defined as

$$\ker(A) = \{v \mid A\ v = 0\}. \tag{2.4}$$

The kernel is a linear subspace and therefore has the following three properties:

1. $\ker(A)$ always contains the zero vector.
2. If $v \in \ker(A)$ and $w \in \ker(A)$, then $v + w \in \ker(A)$.
3. If $v \in \ker(A)$ and a scalar k, then $k\ v \in \ker(A)$.

Additionally, we introduce the term "linear independence". A set of vectors v_1, v_2,...,v_k is said to be linear dependent if $\exists\ a_1$, a_2,..., $a_k \in \mathbb{R}$, but not all zero, such that

$$a_1\ v_1 + a_2\ v_2 + ... + a_k\ v_k = 0. \tag{2.5}$$

If no scalars can be found, the set of vectors is linearly independent. The concept of linear independent vectors is central for the definition of dimension of a vector space V. The dimension of a vector space is given by the number of vectors of its basis. A basis B of vector space V contains a set of vectors, where each element of V can be written as a finite linear combination of vectors of the basis B. The elements of a basis are linearly independent and are said to span the vector space V. A vector space V may have several bases, but the dimension of the vector space remains the same.

Going back, to the ODE system in Eq. (2.3), the matrix A is a map that transforms the vector c from the vector space V, called domain, into a vector $A\ c = c' \in W$. We have introduced the term "kernel" which is a linear subspace of the domain V. A subspace of W is called image of map A such that

$$\mathrm{Im}(A) = \{w \mid A\ v = w,\ v \in V\}. \tag{2.6}$$

Now that the solution space of the steady states have been characterized, further insight into the behavior of the system can be gained by analyzing the stability of a steady state. Arnold [6] provides a few definitions:

Steady states are called stable if $\forall \epsilon > 0$: $\exists \delta > 0$, which is independent from t, such that $\forall c_0$ with $\|c_o\| < \delta$[1] the solution ϕ of system (2.1) with initial condition $\phi(0) = c_0$ can be

[1]Euclidean norm: For $c = (c_1, ..., c_n)^T$: $\|c\|^2 = c_1^2 + ... + c_n^2$.

extended for all $t > 0$ and satisfies $\|\phi(t)\| < \epsilon \ \forall t > 0$. Thus in other words, the solutions whose initial values tend to the steady state, converge uniformly for $t \to \infty$ to a constant solution.

A steady state is called asympotically stable if it is stable and

$$\lim_{t\to\infty} \phi(t) = 0, \tag{2.7}$$

for every solution with $\phi(0) = 0$ lying in a sufficiently small neighborhood of zero. The existence of asympotically stable steady states therefore provides us some knowledge about the long-term behavior of a system. If the steady state is not stable, it is called unstable. In case of a linear ODE system (2.3) with quadratic A, a steady state $c = 0$ is asymptotically stable if all eigenvalues λ_i of A have negative real parts $\mathrm{Re}(\lambda) < 0$.

In general, the system of interest is not linear. In this case, the system can be linearized in the close neighborhood of a steady state c_0 using the Jacobian matrix J_0

$$J_0 = \frac{\partial f}{\partial c}\Big|_{c=c_0} = \begin{pmatrix} \frac{\partial f_1}{\partial c_1} & \frac{\partial f_1}{\partial c_2} & \cdots & \frac{\partial f_1}{\partial c_r} \\ \frac{\partial f_2}{\partial c_1} & \frac{\partial f_2}{\partial c_2} & \cdots & \frac{\partial f_2}{\partial c_r} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_r}{\partial c_1} & \frac{\partial f_r}{\partial c_2} & \cdots & \frac{\partial f_r}{\partial c_r} \end{pmatrix}\Bigg|_{c=c_0}, \tag{2.8}$$

The linearization is valid only in the close neighborhood of the steady state c_0. For the stability analysis, the eigenvalues of J_0 have to be computed. Again, all eigenvalues have to exhibit negative real parts, in order for c_0 to be asymptotically stable. Even if one eigenvalue has a positive real part, the steady state is unstable. There are other possibilities instead of calculating the steady states such as the Routh-Hurwitz criterion, but we will not got further in to these methods.

2.1.2 Time scales of a model

In biological systems, the dynamics of involved components may occur on different time scales. Reaction rates involving macromolecular components, such as proteins, are much slower, as their molecular weights are much higher, than reaction rates involving lowmolecular components, such as metabolites [40, 125, 110]. In order to measure this, different concepts of time scales can be employed. One possibility is to investigate the relaxation time of a variable. Relaxation is defined as the return of a system to its equilibrium, if it exists, after being perturbed. Each process can be then categorized by a time scale indicating its relaxation time τ_r. The simplest function describing a relaxation is

$$f(t) = e^{-\frac{t}{\tau_r}}. \tag{2.9}$$

In this case, τ_r indicates the time at which f reduces its value to $1/e$ times of its initial value. This function is the solution of the following ODE

$$\dot{x} = -\lambda\ x. \tag{2.10}$$

For a general ODE with an asymptotically stable steady state $s_0 \in \mathbb{R}^n$ the rate of change from a point s towards the steady state is given by

$$\frac{\mathrm{d}(s - s_0)}{\mathrm{d}t} = J_0(s - s_0), \tag{2.11}$$

where J_0 is the Jacobian matrix, introduced earlier.

Let λ_i for $i \in \{1, ..., n\}$ be the eigenvalues of J_0. Thus, the solution of system (2.11) for the components of $(s - s_0)$ is given by

$$(s_i - s_{0,i})(t) = \sum_{j=1}^{n} C_j e^{\lambda_j t}. \tag{2.12}$$

Since s_0 is an asymptotically stable steady state, the eigenvalues λ_i are negative and (2.12) is an exponentially declining function. If the components of s are independent from each other, the relaxation time of each component is given by $\tau_r = 1/\lambda_i$, otherwise the relaxation time is determined by the superposition of the exponential functions [162].

In this work, we examine two scenarios and recover the relaxation times for these scenarios. In the first scenario, the relaxation time of the variables of the system towards a steady state after a process parameter is perturbed, e.g. substrate concentration in the environment, is investigated. The system is in equilibrium, while the environment is perturbed. In the second scenario, the environment stays constant and the system is perturbed to exhibit an arbitrary state and its relaxation time towards the steady state is recovered. These relaxation times give us insight into the behavior of the system upon perturbations of different origins, external and internal.

2.2 A mass conservative model for microbial growth

Growth of microbial populations can be modeled in a various ways. In the simplest case, the growth can be described by a deterministic ordinary differential equation (ODE)

$$\mu = \frac{\dot{m_X}}{m_X}. \tag{2.13}$$

where μ is the specific growth rate at which the population of cells increase by means of its expanding biomass m_X in grams (g).

2.2.1 Description of a intracellular network

Eq. (2.13) does not describe the biochemical processes which lead to cellular growth and a structured, but not segregated, approach can be utilized. The biomass is represented by an averaged cell population and the metabolism of an average cell is characterized by a small network. For this network, the biomass of the cells m_X is structered into N_C molecular components C_i with weights m_{C_i} in grams (g)

$$m_X = \sum_i m_{C_i} \qquad i \in \{1, ..., N_C\}. \tag{2.14}$$

Via the biochemical reactions, which form the intracellular network of the cells, the molecular components C_i are continuously produced and consumed, leading to an increase of the total biomass m_X. As the biochemical reactions are based on physical encounters in the cells [55], it is more appropriate to express the molecular components C_i by their intracellular concentration

$$C_i = \frac{n_{C_i}}{m_X}, \tag{2.15}$$

where n_{C_i} is the molar mass of the component C_i. The relation between the mass m_{C_i} and the molar mass n_{C_i} is given by the molecular weight of the component w_{C_i}, which is a fixed number, leading to

$$1 = \sum_i w_{C_i}\ C_i. \tag{2.16}$$

The temporal change of the intracellular concentrations are given by means of a mass balance. They are dependent on the structure of the intracellular network, which is given by a stoichiometric matrix $\mathcal{N}$, and the reaction rates r, indicating the rates at which a component is produced or consumed. Therefore, the dynamics can be written as

$$\dot{c} = \mathcal{N}\ r - \mu\ c, \tag{2.17}$$

where c is the vector containing all N_C components, r defines the vector of N_r reaction rates and $\mathcal{N}$ is a stoichiometric matrix with the dimension $N_C \times N_r$. $\mathcal{N}_{ij}$ corresponds to the net change number of component i through the reaction j, meaning, for positive $\mathcal{N}_{ij}$ component i is produced by reaction j, while for negative $\mathcal{N}_{ij}$ component i is consumed. Thus, the row vector $\mathcal{N}_i$ contains all stoichiometric coefficients of all reactions of the intracellular network concerning component i. Therefore, the first term of Eq. (2.17) describes the change in component i due to the intracellular network. The second term is a dilution term caused by the growth of the cells. From Eqs. (2.13)-(2.17) another expression for the growth rate can be obtained

$$\mu = w^T\ \mathcal{N}\ r, \tag{2.18}$$

where w is a vector containing the molecular weights w_i from Eq. (2.16) of all components C_i. The reaction rates r may be dependent on the components C_i. In this case, it would be more appropriate to write $r(c)$. It is important to note here, that in this approach all cells are averaged and the reaction rates do not differ between the cells. From Eq. (2.18) it is apparent that the growth rate is the net rate of accumulation of the intracellular components C_i based on the structure of the metabolic network scaled with their molecular weight w_i.

2.2.2 Description of the environment of a bioreactor

The microbial population does not grow independently, but in an environment, which means in the context of biotechnology that it corresponds to a bioreactor with a volume V_R, filled with various extracellular concentrations $\mathcal{S}$, such as substrate and products. These concentrations affect the metabolic network or are produced by it, leading to an interaction between the microbial population and the environment. The dynamics of the extracellular concentrations $\mathcal{S}$ are described by

$$\dot{\mathcal{S}} = r_{\mathcal{S},ex} - n_{\mathcal{S}}\ \phi(c, \mathcal{S})\ X, \tag{2.19}$$

where $r_{\mathcal{S},ex}$ is the term containing the net rate of change in $\mathcal{S}$, which is independent of the interaction with the cells, such as dilution and feed, depending on the bioprocess. $n_{\mathcal{S}}$ is a stoichiometric vector, containing the net change numbers of $\mathcal{S}$. $X = m_X/V_R$ is the biomass concentration in the bioreactor and its temporal change is

$$\dot{X} = (\mu - r_{X,ex})\ X, \tag{2.20}$$

where $r_{X,ex}$ accounts for changes in the biomass concentration unrelated to cellular growth. $\phi(c, \mathcal{S})$ is the consumption or production rate of $\mathcal{S}$ due to the metabolic network. It is a subset of the reaction rate vector r in Eq. (2.17), containing those reactions where intracellular components and extracellular components are exchanged. Due to this relation the sign of $\phi(c, \mathcal{S})$ is chosen.

Through Eqs. (2.19) and (2.20), different process designs can be described. For batch-processes, where the reactor volume stays constant over the process ($\dot{V}_R = 0$) without additional substrate feed, the terms $r_{\mathcal{S},ex}$ and $r_{X,ex}$ are zero. In case of a fed-batch process with a feeding rate of q_{in}, feed concentration S_{in}, the equations for the environment read

$$\begin{aligned} \dot{V}_R &= q_{in} \\ r_{\mathcal{S},ex} &= \frac{q_{in}}{V_R}(\mathcal{S}_{in} - \mathcal{S}) \\ r_{X,ex} &= \frac{q_{in}}{V_R}. \end{aligned} \tag{2.21}$$

In a chemostat the culture is supplied through a feed with feeding rate q_{in}, while culture liquid is removed at the same rate keeping the volume of the reactor constant. The equations in this case are the same as in (2.21), except for the equation for V_R, which derivative is zero. The term $\frac{q_{in}}{V_R}$ is usually denoted as D and represents the dilution term.

2.2.3 Coarse-grained models

Modeling of microbial cells can vary in complexity, ranging from whole-cell models with hundreds of individual reactions and components, to pathway models, which describe specific pathways in detail to so-called coarse-grained models [3]. Coarse-grained models only comprise a low number of components and reactions, thus individual molecular components are lumped together to form a few classes of components with their respective macroreactions. The kinetic parameters are obtained by estimation from literature data or by regression from experimental data.

Typically, coarse-grained models involve a self-replicator system. The simplest self-replicator system contains a precursor and a second component, e.g. ribosomes, which synthesizes itself from the precursor, providing the name for the system [106]. Thus, the simplest microbial growth model involves the conversion of external substrate into a metabolic precursor and subsequent synthesis of a macromolecule as seen in Figure 2.1A. Cellular processes are carried out by proteins and are accelerated with increasing protein levels. Protein translation are dependent on the level of ribosomes, which have a self-replicating nature. Therefore, the synthesis of ribosomes competes with protein production, providing a limit for the increase in protein production [70]. For this reason, the simplest coarse-grained model should be extended to account for ribosomal and non-ribosomal proteins (Figure 2.1B). In order to investigate additional

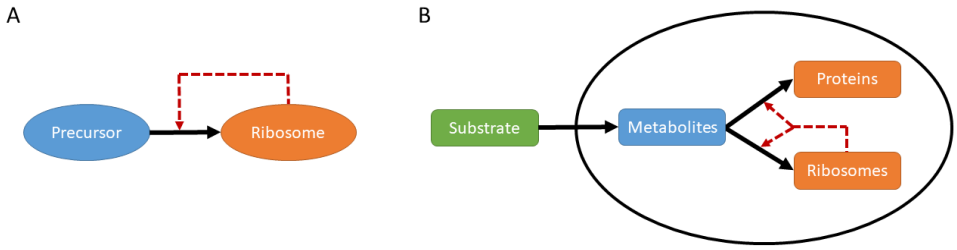

Figure 2.1: Simplest self-replicator system (A), simplest coarse-grained model of a cell including a self-replicator system with cells comprising metabolites, ribosomes and other proteins (B).

factors affecting cellular growth, the simple coarse-grained model can easily be adapted to suffice the additional requirements. Self-replicator models have been employed under the aspect of overflow metabolism [106, 10], a closer investigation into different growth conditions [19, 66, 163] and the adaptation to environmental changes [44, 111, 166].

A common way to incorporate more detail into the coarse-grained model is to consider different protein fractions depending on their cellular function. The first fraction has been already mentioned: the ribosomal proteins R. Scott et al. [140] showed that the rate for protein synthesis is linearly correlated to the ribosomal fraction R. Another protein fraction T is dedicated to the central metabolism, which we simply call transporter proteins in the course of this work. This protein fraction affects the transport of substrate into the cell. Several works consider models with only these two protein fractions, while some also consider an additional fixed protein sector, representing the remaining proteins, for the sake of completeness. Those models were already able to show interesting dynamics and behavior with this simple structure [174, 140, 19, 44, 5]. This model structure already poses an allocation problem, as the cells have to regulate where to allocate their resources in order to achieve maximal growth

$$\begin{aligned} \max_{T,R} \quad & \mu \\ \text{subject to} \quad & \\ & 0 = \mathcal{N}\, r - \mu\, c \\ & \mu = w^T\, \mathcal{N}\, r \\ & 0 = P - R - T \\ & r = f(T, R). \end{aligned} \tag{2.22}$$

In order to investigate specific conditions and scenarios, the fractions or cell components can be refined further to include fractions reacting to different limitations, such as catabolic and anabolic limitations [66], a component for ATP to investigate the energy fluxes [96] or a more detailed cell description to include rRNA [158].

2.2.4 General solution of a system of equations

In the course of this work, the steady states of the intracellular network are investigated. Usually, in classic flux balance analysis, the dilution term in Eq. (2.17) is neglected and only consists of the term with the stoichiometric matrix $\mathcal{N}$. It is a constraint-based model approach, where the reaction vector r, called flux vector, is calculated at steady-state under given constraints, e.g. enzymatic capacity limitations or experimentally determined maximal uptake or production

rates [119]. Thus, the reaction rates of interest fulfill the equation

$$\mathcal{N}\ r = 0. \tag{2.23}$$

The stoichiometrix matrix $\mathcal{N}$ is not typically quadratic, as it usually contains more unknown reaction rates than equations. Without neglecting the dilution term, the steady state solutions are not fully covered by Eq. (2.23), this will be investigated in the scope of this work in Section 3.2.1. For this, we need the concept of pseudoinverses. We consider a system of linear equations

$$A\ x = b \tag{2.24}$$

with $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^n$. In this case, the solutions can be calculated by

$$x = A^+\ b + (\mathrm{Id}_n - A^+A)s \qquad s \in \mathbb{R}^n, \tag{2.25}$$

where Id_n is the identity matrix of dimension n, s and arbitrary scalar and A^+ is the pseudoinverse of A. Pseudoinverses are often referred to as Moore-Penrose inverse. It is the most widely known generalization of the inverse matrix [12]. In case that A is invertible, the pseudoinverse is its inverse A^{-1}. A pseudoinverse $A^+ \in \mathbb{R}^{n\times m}$ of $A \in \mathbb{R}^{m\times n}$ has to satisfy the following conditions, also known as Moore-Penrose conditions [124]:

1. $AA^+A = A$ (A^+ maps all column vectors of A onto themselves)
2. $A^+AA^+ = A^+$ (A^+ acts like a weak inverse)
3. $(AA^+)^T = AA^+$ (Hermitian)
4. $(A^+A)^T = A^+A$ (Hermitian)

If A has linearly independent columns, A^+ can be expressed as

$$A^+ = (A^TA)^{-1}A^T \tag{2.26}$$

and if A has linearly independent rows, as

$$A^+ = A^T(AA^T)^{-1}. \tag{2.27}$$

In the first case, the pseudoinverse is a left inverse, meaning $A^+A = \mathrm{Id}$, and in the second it is a right inverse, meaning $AA^+ = \mathrm{Id}$. The pseudoinverse is often used to find a "best fit", using least squares solution, in case the system is not solvable, or the minimum Euclidean-norm solution, in case of infinite solutions, as

$$\forall x \in \mathbb{R}^n\colon \quad \|Ax - b\| \geq \|Az - b\|, \qquad z = A^+b, \tag{2.28}$$

where equality only holds if $x = A^+\ b + (\mathrm{Id}_n - A^+A)s$ for any vector $s \in \mathbb{R}^n$ [127]. Solutions of Eq. (2.24) exist if and only if $AA^+b = b$ and is unique if and only if A has full column rank.

2.3 Population balance equation models

A PBE describes of the evolution of a population of particles, such as droplets or cells, which interact with their environment. As the name suggests, it is based on a balance on the number of particles of a particular state. These models take into account the various processes

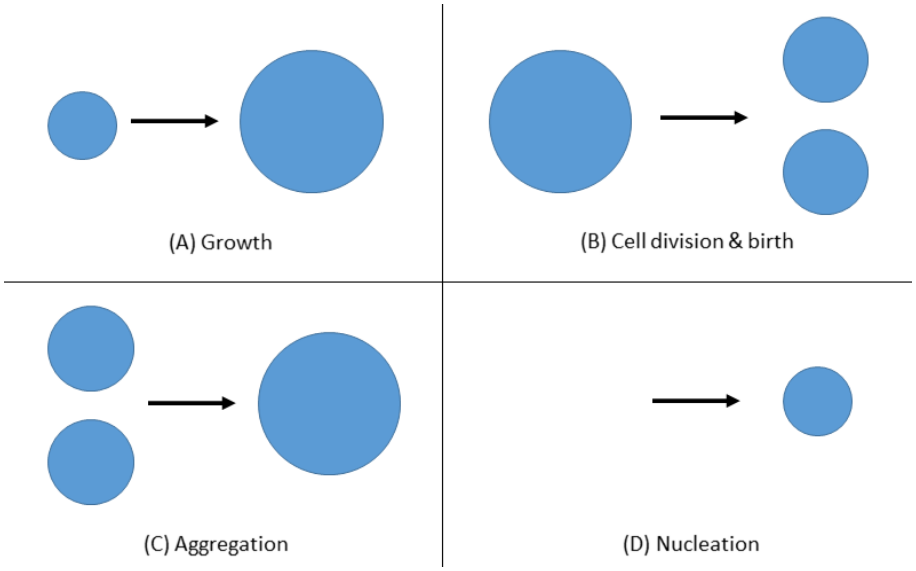

Figure 2.2: Processes involved in a PBE

that can affect the population, such as nucleation, aggregation, cell division, birth, death and growth (Figure 2.2). In this work, we are investigating bacterial cell populations and therefore use the term "cells". For bacterial cell populations only cell division, birth, death and growth processes are relevant. Nucleation and aggregation can occur e.g. in droplets, but not in bacterial cell populations. The main idea of PBE is that individual cells are differentiated by a state vector of variables that characterize the cell (age [11], size [93, 24], mass [92, 105, 86], intracellular concentration of a certain cell component representing a cell's state [126, 145], etc.), allowing a segregated view on the population. The state vector generally contains continuous variables, marking this a continuous modeling approach to address population heterogeneity. Depending on the number of cell properties, population balance equation models are classified into single- or multivariate models. In general, PBE are integro-partial differential equations, which are equations that involve both integrals and derivatives of a function, making this approach a continuous modeling approach to investigate population heterogeneity.

2.3.1 General formulation

The state vector of a cell can be divided into internal properties $\xi \in \Omega_\xi$, which represent the state of the cell, and external properties $x \in \Omega_x$, which describe the physical location of the particle in the environment [133, 132]. Ω_ξ is the domain of the internal properties with $\dim(\Omega_\xi) = d$ and Ω_x the domain of external variables with $\dim(\Omega_x) = d_x$. These domains can be bounded or have infinite boundaries. In general, the PBE for a population of cells has the following formulation proposed by Ramkrishna [131]

$$\frac{\partial N(\xi,x,t)}{\partial t} + \nabla_\xi[r_\xi(\xi,x)N(\xi,x,t)] + \nabla_x[r_x(\xi,x)N(\xi,x,t)], = h(\xi,x,t) \tag{2.29}$$

where $N(\xi,x,t)$ is the number density function, meaning $N(\xi,x,t)\mathrm{d}V_\xi\mathrm{d}V_x$ is the number of cells at time t with states between (ξ,x) and $(\xi,x)+\Delta(\xi,x)$. $\mathrm{d}V_\xi\mathrm{d}V_x$ is the infinitesimal volume in the cell state space. The second and third term on the left account for the loss of cells in state (ξ,x)

due to change in the internal property ξ with rate $r_\xi(\xi, x)$ and due to transport in the spatial space x with velocity $r_x(\xi, x)$. The function $h(\xi, x, t)$ represents the net rate of appearance of new cells with state (ξ, x). Thus, $h(\xi, x, t)$ can include terms related to cell division, birth, death and loss of cells due to dilution. In case of cell division of a cell, or breakage, as it is usually referred to in the field of PBE models, and consequently birth of two daughter cells, the net term reads

$$h(\xi, x, t) = 2 \int_{\xi}^{\xi_{max}} \Gamma(\xi', x) p(\xi, \xi', x) N(\xi', x, t) \, \mathrm{d}V_{\xi'} - \Gamma(\xi, x) N(\xi, x, t), \tag{2.30}$$

where $\Gamma(\xi, x)$ is the division rate and $p(\xi, \xi', x)$ the partition probability density where a cell with state vector (ξ', x) produces two daughter cells with state (ξ, x). The partition probability function is symmetric $p(\xi, \xi', x) = p(\xi' - \xi, \xi', x)$ since there is no loss of "mass" during cell division. Both functions have to be non-negative. Under the assumption that cells divide into two identical daughter cells, the partition probability density is

$$p(\xi, \xi', x) = \prod_{i=1}^{d} \delta \left(\xi_i - \frac{\xi_i'}{2} \right). \tag{2.31}$$

Here, $\delta(\cdot)$ is the Dirac delta function, which assumes the value ∞ for 0 and 0 otherwise. With this partition probability density, the net term (2.30) simplifies to

$$h(\xi, x, t) = 2^{d+1} \, \Gamma(2\xi, x) N(2\xi, x, t) - \Gamma(\xi, x) N(\xi, x, t). \tag{2.32}$$

This assumption has been compared to unequal partitioning for a single-variable case as well as a three-variable test case by Matzaris et al. [101, 100]. For equal partitioning the number density function exhibits a periodic behavior, which has been shown mathematically and explained for a single-variable case by Diekmann et al. [33] and further verified numerically [92, 101, 100]. The different moments of the particle density function play an important role, since they correspond to physical properties such as the total cell number, total cell mass, etc. [39, 75, 152]. The p-th order moment of $N(\xi, x, t)$ is given by

$$\mathcal{M}_{p_1,\ldots,p_d}(t) = \int_{\Omega_\xi} \int_{\Omega_x} \prod_{i=1}^{d} \xi_i^{p_i} N(\xi, x, t) \, \mathrm{d}V_x \mathrm{d}V_\xi, \tag{2.33}$$

where p_i is a non-negative integer. The total number of cells $N_t(t)$ at time t, which is the zeroth moment of $N(\xi, x, t)$, is given by

$$N_t(t) = \int_{\Omega_\xi} \int_{\Omega_x} N(\xi, x, t) \, \mathrm{d}V_x \mathrm{d}V_\xi. \tag{2.34}$$

The temporal change of the total number of cells is given by

$$\frac{\mathrm{d}N_t(t)}{\mathrm{d}t} = \int_{\Omega_{\xi'}} \Gamma(\xi', x) N(\xi', x, t) \, \mathrm{d}V_{\xi'}. \tag{2.35}$$

For a single-variable PBE with the cell mass as descriptor variable, the first moment corresponds to the total cell mass of the population.

In case of a multi-variable PBE, where each entry of ξ represent a different cell property, the first moment of a specific property ξ_i is the total content of that particular property

$$\mathcal{M}_{0,\dots,1,\dots,0}(t) = \int_{\Omega_\xi} \int_{\Omega_x} \xi_i N(\xi, x, t) \, \mathrm{d}V_x \mathrm{d}V_\xi, \tag{2.36}$$

where 1 is in the i-th position.

In order for the model (2.29) to be well-defined, initial and boundary conditions are necessary. The initial condition is given by the initial population distribution, which has to be specified for the internal properties as well as the external variables at $t = 0$

$$N(\xi, x, 0) = N_0(\xi, x). \tag{2.37}$$

Depending on the structure of the PBE, several types of boundary conditions can be applied. Dirichlet boundary conditions hold when fixed values are assigned to $\partial\Omega_\xi$ and $\partial\Omega_x$. If the fluxes at the boundary of the state vector space are defined, then the solution satisfies a Neumann boundary condition. For unbounded state vector spaces the fluxes at infinity vanish, as they do not represent cell sources or sinks

$$\begin{aligned} \nabla_\xi [r_\xi(\xi, x) N(\xi, x, t)] &\to 0 \qquad \text{for} \quad |\xi| \to \infty, \\ \nabla_x [r_x(\xi, x) N(\xi, x, t)] &\to 0 \qquad \text{for} \quad |x| \to \infty. \end{aligned} \tag{2.38}$$

In a closed system, the fluxes at the boundaries are zero

$$\begin{aligned} \nabla_\xi [r_\xi(\xi, x) N(\xi, x, t)] &= 0 \qquad \xi \in \partial\Omega_\xi, \\ \nabla_x [r_x(\xi, x) N(\xi, x, t)] &= 0 \qquad x \in \partial\Omega_x. \end{aligned} \tag{2.39}$$

An open system is given if at parts of the boundary cells can originate. This can be the case, if cells enter the system as part of an particle-fluid mixture for example.

In this work, we consider a cell population interacting with the environment of a bioreactor. Thus, the PBE (2.29) has to be rewritten to account for the dependence on the environment with nutrient and extracellular concentrations $\mathcal{S}$

$$\frac{\partial N(\xi, x, t)}{\partial t} + \nabla_\xi [r_\xi(\xi, x, \mathcal{S}) N(\xi, x, t)] + \nabla_x [r_x(\xi, x, \mathcal{S}) N(\xi, x, t)] = h(\xi, x, \mathcal{S}, t) \tag{2.40}$$

The equation of the external concentrations $\mathcal{S}$, which has been introduced for a averaged cell model (2.19), is adjusted to be

$$\dot{\mathcal{S}} = r_{\mathcal{S},ex} + \int_{\Omega_\xi} \int_{\Omega_x} \phi(\xi, x, \mathcal{S}) N(\xi, x, t) \, \mathrm{d}V_x \mathrm{d}V_\xi \tag{2.41}$$

where $\phi(\xi, x, \mathcal{S})$ is the production or consumption rate of the external concentrations $\mathcal{S}$ and $r_{S,ex}$ again the net rate of change in $\mathcal{S}$. In this case, $\phi(\xi, x, \mathcal{S})$ also has to include a factor for the cell's mass.

2.3.2 Numerical methods

One challenge of PBE models is that analytical solutions are rarely available and numerical methods have to be employed to solve the PBE. Possible methods include the method of successive approximations [132], the method of moments, which is useful if one is only interested in the evolution of moments leading to a better trade-off between accuracy and numerical effort [9, 94, 37], finite element methods [117, 95, 136], finite volume schemes [20, 137, 24] and sectional methods [77]. Finite element methods are based on the discretization of the domain into smaller elements, which are finite, and the employment of Ansatzfunctions to solve the problems for the finite elements [175]. Finite volume methods solve the problem by using volume integrals and evaluating the fluxes at the surface of each finite volume, resulting in a method based on conservation of mass, momentum and energy within the finite volumes. Thus, the solution is approximated through the usage of the average values of the solution over the finite volumes [89]. Sectional methods divide the population into smaller sections and solve the PBE independently for each section. They have been mostly employed to approximate PBE for aggregration-breakage processes [59, 164]. The choice of method depends on the processes which are considered by the PBE model. In the scope of this work, we consider a population of cells that are growing and dividing in an environment S without external variables x. Therefore, we neglect the spatial coordinates x of the cells and assume a homogeneous environment. For these type of models, the finite volume method with the inclusion of a cell average technique has been proposed in literature [76, 137, 24]. Finite volume methods have the advantage of being conservative and require less computation time than e.g. the fixed pivot technique, which has been proposed by Kumar and Ramkrishna [77]. They are often used to solve conservation laws.

Finite volume method for a bivariate PBE

In the following, a finite volume method for a bivariate PBE is introduced. The domain $\Omega_\xi = [0, \xi_{1,\max}] \times [0, \xi_{2,\max}]$ is considered to be finite for the computation and is discretized into a equally spaced 2D-grid, where each grid cell has the size $\Delta\xi_1 \times \Delta\xi_2$. Instead of an equally spaced grid space, there are also other possibilities to discretize the space, e.g. geometric spacing and even moving boundaries [78]. A geometric grid space can be an appropriate choice for breakage and aggregation processes. But as cell growth is considered, an equally spaced grid is a better fit. Each grid cell is represented by its grid cell center $\chi_{ij} = (\xi_{1,i}, \xi_{2,j})$ for $i \in \{1, ..., I\}$ with $I = \frac{\xi_{1,\max}}{\Delta\xi_1}$ and $j \in \{1, ..., J\}$ with $J = \frac{\xi_{2,\max}}{\Delta\xi_2}$, whose cell faces in ξ_1 direction are given by $\xi^{\pm}_{1,i} = \xi_{1,i} \pm \frac{\Delta\xi_1}{2}$ and in $\xi^{\pm}_{2,j} = \xi_{2,j} \pm \frac{\Delta\xi_2}{2}$ in ξ_2 direction. As we only discretize the state space and not the temporal space, the numerical scheme is semi-discrete. The cell average value of the number density $\widehat{n}_{ij}(t)$ of the grid cell represented by χ_{ij} at time t is defined as

$$\widehat{n}_{ij}(t) = \frac{1}{\Delta\xi_1\Delta\xi_2} \int_{\xi^-_{1,i}}^{\xi^+_{1,i}} \int_{\xi^-_{2,j}}^{\xi^+_{2,j}} N(\xi_1, \xi_2, t)\ \mathrm{d}\xi_1\mathrm{d}\xi_2 \tag{2.42}$$

and the total number of cells in each grid cell as

$$\widehat{N}_{ij}(t) = \Delta\xi_1\Delta\xi_2\ \widehat{n}_{ij}(t) = \int_{\xi^-_{1,i}}^{\xi^+_{1,i}} \int_{\xi^-_{2,j}}^{\xi^+_{2,j}} N(\xi_1, \xi_2, t)\ \mathrm{d}\xi_1\mathrm{d}\xi_2. \tag{2.43}$$

For simplicity reasons $\widehat{N}_{ij}(t)$ will be denoted as $\widehat{N}_{ij}$ and $\widehat{n}_{ij}(t)$ as $\widehat{n}_{ij}$.

The two processes of increase in the space of the internal variables ξ and cell division can be split and handled as two separate processes

$$\frac{\partial N(\xi,t)}{\partial t} + \text{Increase in } \xi = \text{Cell division.} \tag{2.44}$$

Increase in the space of internal variables ξ

For the rate of change in the internal space a first-order upwind scheme is presented [128, 16]. The term "upwind scheme" is used for a class of numerical schemes for the solution of hyperbolic partial differential equations and was introduced by Courant et al. [23]. It involves estimating the derivatives by taking backward differences, thus resulting in estimations biased to more "upwind" points suitable for positive advection [88]. The accuracy of this scheme is determined by the estimation of the fluxes at the grid-cell faces [128]. The scheme is said to be first-order accurate because the accuracy of the solution is determined by the first derivative of the solution with respect to time. The PBE (2.29), considering only positive rate of change in ξ using a cell average finite volume technique, can be rewritten as

$$\frac{\mathrm{d}\widehat{N}_{ij}}{\mathrm{d}t} + \int_{\xi^-_{1,i}}^{\xi^+_{1,i}} \int_{\xi^-_{2,j}}^{\xi^+_{2,j}} \frac{\partial}{\partial \xi_1} \left(r_{\xi_1}(\xi_1,\xi_2,\mathcal{S})N(\xi_1,\xi_2,t)\right) + \frac{\partial}{\partial \xi_2} \left(r_{\xi_2}(\xi_1,\xi_2,\mathcal{S})N(\xi_1,\xi_2,t)\right) \mathrm{d}\xi_1 \mathrm{d}\xi_2 = 0. \tag{2.45}$$

The first term can be written as

$$\int_{\xi^-_{1,i}}^{\xi^+_{1,i}} \int_{\xi^-_{2,j}}^{\xi^+_{2,j}} \frac{\partial}{\partial \xi_1} \left(r_{\xi_1}(\xi_1,\xi_2,\mathcal{S})N(\xi_1,\xi_2,t)\right) \mathrm{d}\xi_1 \mathrm{d}\xi_2 = \Delta\xi_2 \left((r_{\xi_1}N)_{i^+j} - (r_{\xi_1}N)_{i^-j}\right), \tag{2.46}$$

where the first term $(r_{\xi_1}N)_{i^+j}$ corresponds to the flux leaving the grid cell ij in ξ_1 direction, while the second term $(r_{\xi_1}N)_{i^-j}$ is the influx of the grid cell in ξ_1 direction . Using a first-order upwind scheme the flux terms at the cell faces are expressed as

$$(r_{\xi_1}N)_{i^+j} = r_{\xi_1,i^+j}\widehat{n}_{ij} \qquad (r_{\xi_1}N)_{i^-j} = r_{\xi_1,i^-j}\widehat{n}_{i-1,j}. \tag{2.47}$$

One has to note that in Eq. (2.47) $\widehat{n}_{ij}$ is the cell average number density, while the values of the rate of change term r_{ξ_1,i^+j} are computed at the point $(\xi^+_{1,i}, \xi_{2,j})$. Analogously, the fluxes in ξ_2 direction can be expressed as

$$(r_{\xi_2}N)_{ij^+} = r_{\xi_2,ij^+}\widehat{n}_{ij} \qquad (r_{\xi_2}N)_{ij^-} = r_{\xi_2,ij^-}\widehat{n}_{i,j-1}. \tag{2.48}$$

In order to obtain higher order accuracy, the fluxes can be interpolated piecewise polynomialy instead of a first-order upwind scheme

$$(r_{\xi_1}N)_{i^+j} = r_{\xi_1,i^+j}\left(\widehat{n}_{ij} + \frac{1+\kappa}{4}(\widehat{n}_{i+1,j} - \widehat{n}_{ij}) + \frac{1-\kappa}{4}(\widehat{n}_{ij} - \widehat{n}_{i-1,j})\right) \qquad \kappa \in [-1,1], \tag{2.49}$$

where we obtain an second-order one-sided upwind scheme for $\kappa = -1$ and a second-order central scheme for $\kappa = 1$.

Cell division

Multidimensional systems pose a challenge when dealing with cell division and preserving its moments. In a one-dimensional case, the first moment corresponds to the total mass of the cell system. In a multidimensional system, the d-th order moment $\mathcal{M}_{1,\dots,1}(t)$ represents the hypervolume of the cells if the properties are sizes, as length, width, etc. This concept cannot be applied to general cell properties, which can include internal cell concentrations. In order to ensure conservation of the total content, the sum of first-order moments has to be conserved. Depending on the choice of division rate $\Gamma(\xi, x)$ and partition probability density function $p(\xi, \xi', x)$, cell division processes can lead to mass-loss in the system, which is also known as "shattering" [21, 151]. Therefore, Saha et al. [137] proposed a weighted finite-volume scheme, which is able to predict the zeroth moment and first-order moments with good accuracy. Discretizing the PBE (2.29) without the growth term, with cell division into two daughter cells and inclusion of two weight functions for a bivariate case leads to

$$\begin{aligned}\frac{\mathrm{d}\widehat{N}_{ij}(t)}{\mathrm{d}t} =& \sum_{l=i}^{I}\sum_{k=j}^{J}\Gamma(\xi_{1,l},\xi_{2,k},\mathcal{S})\ \widehat{N}_{ij}\ \mathcal{P}(i,j|l,k)\ W_{lk}^{b} \\ &- \Gamma(\xi_{1,l},\xi_{2,k})\ \widehat{N}(\xi_{1,i},\xi_{2,j})\ W_{ij}^{d},\end{aligned} \tag{2.50}$$

where

$$W_{lk}^{b} = \frac{\xi_{1,l}+\xi_{2,k}}{\sum_{i=1}^{l}\sum_{j=1}^{k}\left((\xi_{1,l}+\xi_{2,k})-(\xi_{1,i}+\xi_{2,j})\right)\ \mathcal{P}(i,j|l,k)} \tag{2.51}$$

$$W_{lk}^{d} = \frac{W_{lk}^{b}}{\xi_{1,l}+\xi_{2,k}}\sum_{i=1}^{l}\sum_{j=1}^{k}\left((\xi_{1,i}+\xi_{2,j})\right)\ \mathcal{P}(i,j|l,k) \tag{2.52}$$

and

$$\mathcal{P}(i,j|l,k) = \int_{\xi_{1,i}^{-}}^{\rho_{1,il}}\int_{\xi_{2,j}^{-}}^{\rho_{2,jk}} p((\xi_1,\xi_2)^T,(\xi_{1,l},\xi_{2,k})^T)\mathrm{d}\xi_1\ \mathrm{d}\xi_2, \tag{2.53}$$

as well as

$$\rho_{m,il} = \begin{cases} \xi_{m,l} & \text{for } l = i, \\ \xi_{m,i}^{+} & \text{for } l \neq i \end{cases} \tag{2.54}$$

for $m \in \{1,2\}$. W_{lk}^{b} and W_{lk}^{d} are the weight functions related to the birth and death terms, respectively. It can be shown, that the scheme is number-preserving as well as conservative of the total first-order moments. To combine the two processes of change in the internal space and cell division, the schemes can simply be added as seen in Eq. (2.44).

2.4 Hydrodynamic model of a bioreactor

In order to understand the behavior of a microbial population in a bioreactor, the hydrodynamics and mixing in a bioreactor have to be investigated and described. Computational fluid dynamics (CFD) can provide a very detailed description of the liquid phase for different bioreactor

scales and give us more insight into the differences when scaling up a process. CFD simulations are also able to help improve the bioreactor design and estimate parameters such as shear stress, oxygen concentration and mass-transfer coefficients [72, 134, 52]. One downside of CFD simulations is that they are computationally time consuming and simulations of a few hundered seconds of process time can take hours. The incorporation of complex biological systems as well as multiphase (gas-liquid, liquid-solid or gas-liquid-solid) simulations pose a very difficult challenge. Due to this problem, CM approaches, also called multi-zone or network-of-zone models, are used to predict the hydrodynamics [165, 173, 28, 27]. Here, the volume of interest is divided into a finite number of interconnected volumes which are considered to be homogeneous. They were first developed in 2D [173] and then extended to 3D [64, 71]. In order to predict the mixing behavior accurately, CFD simulations are combined with CMs, where CFD are used to obtain the turbulent flow and developing the compartment models based on the CFD results. The flow rates between the compartments can be obtained from CFD velocity fields [28]. CMs require less computational effort than CFD simulations and, therefore, it is possible to include more complex descriptions for the biophase later on [68]. This way, they form a middle ground between ideal well-mixed bioreactor descriptions and full CFD models. In the course of this work, CFD simulations were not necessary, as the flow between the compartments is known through a set flow rate circulating the liquid phase in the bioreactor system.

2.4.1 Flow field of a bioreactor

In this section, we present the CM proposed by Delafosse et al. to describe the transport of the liquid phase in a bioreactor [28]. The bioreactor with volume V_R is divided into n_z compartments, where each compartment is a volume of fluid V_i with $i \in \{1, ..., n_z\}$ and therefore $\sum_i V_i = V_R$. Let $\mathcal{F}$ be a square matrix with dimension $n_z \times n_z$, such that $\mathcal{F}_{i,j}$ contains the exit flow rate from compartment i to neighboring compartment j as seen in Figure 2.3 and are set positive. The diagonal entries $\mathcal{F}_{i,i}$ correspond to the cumulative exit flow rates of compartment i and are set negative. In this manner, the matrix conserves the total mass in each compartment of the bioreactor

$$\mathcal{F} = \begin{pmatrix} \mathcal{F}_{1,1} & \mathcal{F}_{1,2} & \dots & \mathcal{F}_{1,n_z} \\ \mathcal{F}_{2,1} & \mathcal{F}_{2,2} & \dots & \mathcal{F}_{2,n_z} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{F}_{n_z,1} & \mathcal{F}_{n_z,2} & \dots & \mathcal{F}_{n_z,n_z} \end{pmatrix} \qquad \mathcal{F}_{i,i} = -\sum_{j \neq i}^{n_z} \mathcal{F}_{i,j}. \tag{2.55}$$

In general, the matrix $\mathcal{F}$ is not constant, but is periodically updated from previously stored flow fields, which can be obtained from CFD. This way, unsteadiness of the flow, which occur in the impeller swept zone in the bioreactor, can be considered as well.

2.4.2 Scalar transport in the liquid phase

Let $s_l \in \mathbb{R}^{n_s}$ be a dissolved species (hereafter named scalar) in the liquid phase, where $s_{l,i}$ is the concentration of this scalar over a compartment i, and n_s is the total number of dissolved scalars. From the definition of the compartments and the flow rates between the compartments, we can set up the mass balance for the scalar $s_{l,i}$ describing the transport of the dissolved scalar

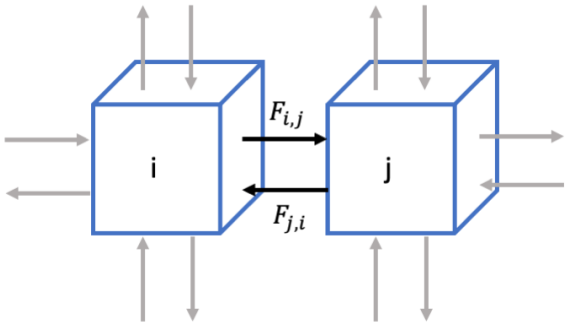

Figure 2.3: Flow rates between compartment i and compartment j

in the bioreactor

$$\frac{\mathrm{d}(V_i s_{l,i})}{\mathrm{d}t} = \sum_{k=1}^{n_z} s_{l,k}\mathcal{F}_{k,i} - s_{l,i}\sum_{j=1}^{n_n}\mathcal{F}_{i,j} + \Phi_{l,i}V_i \qquad i \in \{1,..,n_z\},\, l \in \{1,..,n_s\}, \tag{2.56}$$

where n_n is the number of neighbors of compartment i. In a 3D case, the number of neighbors is 6, coinciding with the number of faces of the compartment, while in a 2D case the number reduces to 4. The first term represents the total flow into compartment i originating from compartment k, while the second term represents the sum of all fluxes leaving the compartment i and the third term is the sum of volumetric source terms due to the interfacial mass transfer. From a rigourous perspective, the multiphase nature of the bioreactor requires refering to mass transfer between the liquid and biotic phases. Let $\mathcal{S}$ be a $n_s \times n_z$ matrix, where $\mathcal{S}_{l,i}$ contains the concentration of scalar s_l in compartment i, and Φ be a matrix with the dimension $n_s \times n_z$, where $\Phi_{l,i}$ contains the overall transfer rate for scalar s_l in the compartment i. The mass balance (2.56) can be rewritten using the matrix formulation (2.55) as

$$\frac{\mathrm{d}(\mathcal{S}V)}{\mathrm{d}t} = \mathcal{S}\mathcal{F} + \Phi V, \tag{2.57}$$

where V is a diagonal matrix with the dimension of the number of compartments n_z containing the volume of each compartment V_i on its diagonal. It is worth noting, that there is no spatial information associated to the compartments. The index indicating the compartment of interest is embedded in the matrix F. The non-zero terms of $\mathcal{F}$ through the column indices provide the knowledge about the neighbors of any compartment i. With this information, a visual representation resembling the physical fluid domain can be recreated. The distribution of the scalar s_l in the bioreactor can be characterized by the following quantities:

The mean volumetric concentration

$$\bar{s}_l(t) = \frac{1}{V_R}\sum_{i=1}^{n_z} s_{l,i}(t)\ V_i(t). \tag{2.58}$$

The variance of the concentration distribution

$$\sigma^2_{s_l}(t) = \frac{1}{n_z}\sum_{i=1}^{n_z}(s_{l,i}(t) - \bar{s}_l(t))^2. \tag{2.59}$$

While the mean volumetric substrate concentration indicates the mean environment a particle encounters, the variance in concentration distribution provides information about the degree of gradients in the substrate concentration.

2.5 Euler-Lagrange Models

Microbial growth and product formation is the result of a complex interplay between the extracellular environment and intracellular variables. In order to investigate this, models linking the metabolic network and description of the bioreactor have been developed. First approaches have been based on the assumption that liquid, gas and biophase are continuous, marking these Euler-Euler-approaches [135, 138]. In segregated approaches of these, the biotic phase is usually described with a PBE, which was introduced in Section 2.3, and an Euler approach for the liquid phase. Because of the nature of the PBE, it is very challenging to incorporate a complex system for the intracellular network involving multiple variables even for the description of a population in a well-mixed homogeneous bioreactor [101, 58, 36]. The liquid phase can also be described using CFD simulations. But the approaches combining a PBE and CFD simulations have the same limitations [109]. One resort is to divide the environment into compartments, as presented earlier in Section 2.4, and formulate PBE within these compartments with an unstructured description of the liquid phase [165, 15]. Besides Euler-Euler models employing PBEs, another possibility to quantify the degree of heterogeneity is the Euler-Lagrange approach, which is a discrete. Lapin et al. [81, 82] introduced these category of models, the term "Euler" refers to the continuous liquid phase, while the biotic phase is described by discrete virtual particles, which are coupled with a state vector describing its intracellular state ξ, thus employing a "Lagrangian" approach. The particles and their movement through the bioreactor, the so-called "lifelines" can be tracked and investigated [50]. The cellular state of a single particle at time t^* is given by their initial cell state at $t = 0$ and their particular trajectory during the time period $0 < t < t^*$ [31]. With a sufficiently high number of particles the continuum limit represented by a PBE model can be approximated.

2.5.1 Transport of particles

The movement of the particles and their encounter of different environmental condition can be considered as stochastic processes. A deterministic approach as used in Euler-Euler approaches is not appropriate since the approach here follows the perspective of a particle and a decision process for each particle has to be defined. First attempts to simulate this includes a multiple change point model, which accounts for stochasticity in growth rates due to a fluctuating environment [31]. Another approach employs a Gillespie's stochastic simulation algorithm to calculate the intracellular reaction rates based on their probibility for certain events such as a change in position in the bioreactor or a cell division event [118]. In recent years, the particles are considered as Lagrangian point masses moving in the bioreactor. This process can be described as a discrete random walk and the obtained lifelines are evaluated statistically [79, 50].

In the framework here, the use of a compartmentalized description of the reactor, as presented in Section 2.4, allows the replacement of particle trajectory calculation by a procedure changing the location index of a particle in the space of compartments. Mayorga Espinoza [103] presented a stochastic transport algorithm, adapted from Delafosse et al. [27], based on the residence time

distribution to describe the movement of the particles. It is a continuous-time Markov chain, which is a stochastic process in which a quantity moves from one state to another. A key property of a Markov chain is that the probability of each event is only dependent on the very last event. Thus, the movement from compartment i to j occurs with a probability that is independent of all the previous states and the resident time in the compartment. We consider N_p particles, present in the bioreactor, and each particle p is assigned a location index x, corresponding with the index of the compartment in which it resides. As the particle moves in the bioreactor over the course of time, the location index x is updated after each time step. Let P_{out} be the probability that a particle that resides in compartment i leaves its current compartment during the time interval Δt

$$P_{out} \begin{cases} = 1 - e^{-\Delta t/\tau_r(i)} \\ \approx \Delta t/\tau_r(i), \qquad \text{if} \quad \Delta t << \tau_r(i) \end{cases} \tag{2.60}$$

where $\tau_r(i)$ is the mean residence time in the compartment i

$$\tau_r(i) = \frac{V_i}{\sum_j \mathcal{F}_{i,j}}. \tag{2.61}$$

The probability for a particle leaving the compartment i to reach the compartment j is given by

$$P_{i,j} = \frac{\mathcal{F}_{i,j}}{\sum_k \mathcal{F}_{i,k}} \tag{2.62}$$

The transport scheme is a iterative process that is performed in two steps. First, we determine if a particle leaves its compartment and secondly, the destination compartment of the particle is set. In order to determine if the particle leaves its compartment, a random number α from a uniform distribution is drawn and particle p leaves if $\alpha < P_{out}$. To select the destination compartment by drawing a second random number β from a uniform distribution, a step-wise function is defined

$$\mathcal{C}P_{i,n} = \sum_{k=1}^{n} P_{i,k} \qquad n \in 1, ..., n_z, \tag{2.63}$$

which exhibits as many non-zero entries as neighbors of compartment i and the destination compartment is chosen by identifying the interval β falls into using Eq. (2.63)

$$\mathcal{C}P_{i,n} < \beta < \mathcal{C}P_{i,n+1}. \tag{2.64}$$

The location of the particle is updated to $n + 1$. In order to save storage space, the matrix $\mathcal{C}P$ can be shortened to only exhibit as many columns as the maximal number of possible neighbors. The remaining columns are zero vectors otherwise and do not contain any useful information. Two examples of the procedure to identify the destination compartment index are presented in Figure 2.4.

2.5.2 Formulation of the mass transfer between biotic and liquid phase

Due to the biological phase present in the compartment i, the scalars s_l in the liquid phase are produced or consumed. Strictly speaking, the source terms, which have been introduced in

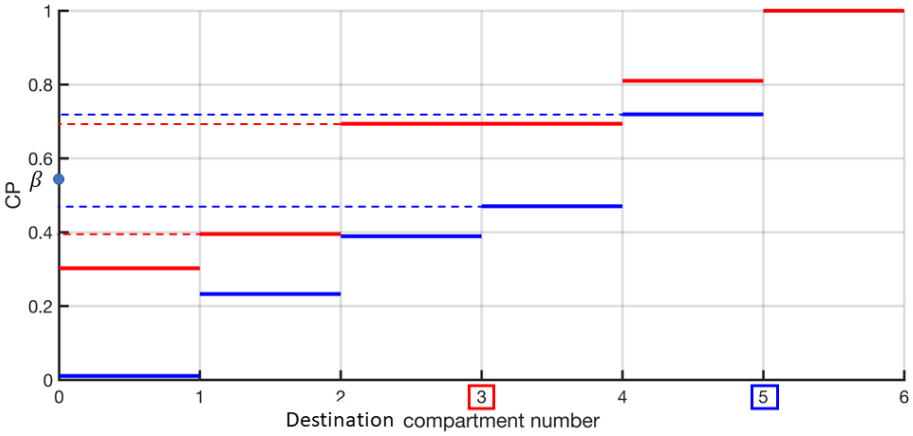

Figure 2.4: Two examples of the identification of the destination compartment index for a given random number β in case of 6 neighboring compartments using the cummulative probability function $\mathcal{C}P$ of Eq. (2.63). The index of the destination compartment is indicated by a red or blue square respectively.

Section 2.4.2, correspond to mass transfer between the liquid and biotic phases or between gas and liquid phases. As the latter aspect is not treated in this work, we will not consider this. The overall consumption or production rates result from the contribution of all the cells present in the bioreactor considering their particular physiological state and the availability of nutrients in the environment. With the approach of a CM, the source term $\Phi_{l,i}$ is computed through the summation of elementary transfer rates ϕ over the particles present in the compartment i

$$\Phi_{l,i} = \frac{1}{V_i} \sum_{p}^{N_{p,i}} \phi_l(\xi_p, \mathcal{S}_i), \tag{2.65}$$

where $N_{p,i}$ denotes the number of particles in compartment i. $\phi_l(\xi_p, \mathcal{S}_i)$ is a vector of specific transfer rates achieved by particle p exploiting its physiological properties ξ_p in an environment characterized by the vector of concentrations $\mathcal{S}_i$. For unstructured descriptions of the biotic phase, where the cells are only described by their biomass $m_p = \xi_p$, the transfer rates $\phi_l(\xi_p, \mathcal{S}_i)$ are only dependent on $\mathcal{S}_i$ and m_p [154]. For structured cells, the transfer rates $\phi_l(\xi_p, \mathcal{S}_i)$ can be obtained from the description for the intracellular network, which can easily be incorporated in this framework as seen in recent works [112, 51].

3 A minimal description of a intracellular network

In order to investigate a cell population in a bioreactor and its response to the dynamic environment, a description of the cellular network is required. For this, we employ a coarse-grained model. These types of models are useful choice to get a better understanding on cellular control strategies, gene expression, and resource allocation [18, 140]. Characteristics of coarse-grained models have already been presented in Section 2.2.3.

Typically, coarse-grained models comprise a set of differential equations for the components of the model. However, mass balance equations must fulfill the conservation of mass as dictated by the first fundamental theorem of thermodynamics, and often, a consistent transfer from mass balance equations to differential equations for the concentrations of the model components is faulty or inadequate. Therefore, we introduce a minimal coarse-grained model, where a new relationship for the specific growth rate in dependence on the exchange reactions of the entire network is given. This equation is fundamental since it guarantees strict mass conservation for the complete system. Conventionally, the growth rate is an empirical function and, therefore, strict mass conservation is not ensured.

Next, we perform flux analysis of these types of models and show interesting properties of the steady state behavior of the reaction rates. We provide a general steady state solution for biochemical networks and for the minimal coarse-grained model. Furthermore, under consideration of the regulation of the cells metabolism as resource allocation problem to achieve maximal growth, relations between the allocation of the proteome and the growth rate are determined. The linear relation between the ribosomal protein fraction and the growth rate, which is well-established knowledge [140, 8], is recovered as result of the optimization problem using the minimal coarse-grained model. Additionally, limitations on the substrate rate, which may occur in a changing environment, where the substrate uptake is controlled by the transport of substrate in the liquid phase rather than by the cell uptake capacity itself, are addressed.

With the coarse-grained model approach, a L-phenylalanine production process is simulated and compared to experimental data. The deterministic behavior of the coarse-grained model is characterized by the steady states of the cellular components and a time scale analysis for cellular response to environmental changes and internal noise is proposed. Here, the concept of balanced and unbalanced cells, which gains importance in Chapter 5, is introduced and discussed. The results of this chapter have partially been published in [34].

3.1 Model formulation

At first, we present a typical minimal reaction system as shown in Figure 3.1, which we have briefly mentioned in Section 2.2.3. The cellular network represents the entire biomass and

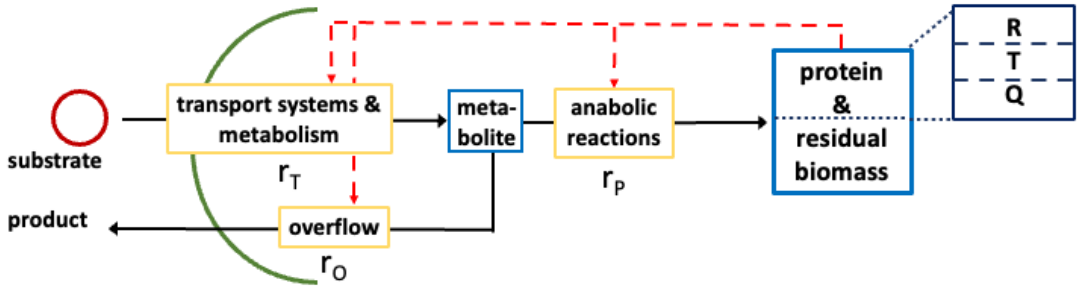

Figure 3.1: General scheme of a coarse-grained model with partitioned proteome (ribosomal proteins R, proteins linked with the central metabolism T, and residual protein fraction Q) as the self-replicator system [142]; it consists of two components, indicated as blue boxes; (metabolite, low molecular weight), protein; and residual biomass (high molecular weight; protein is assumed to be 50% of total biomass). The components are connected by a minimal set of reactions, indicated by yellow boxes, for substrate uptake, overflow metabolism, and protein synthesis.

contains only one anabolic reaction. The scheme of the system is given by

$$\begin{aligned} (S) &\rightarrow \alpha\ M \\ \gamma\ M &\rightarrow \beta\ P \\ M &\rightarrow \text{(by-product)}. \end{aligned} \tag{3.1}$$

An extension to two or more anabolic reactions can easily performed since, in general, the mass fractions of the macromolecules are well-known. The systems consist of a pool of metabolites M, the precursor of the minimal system, proteins P, and reactions r_i which connect the intracellular components with each other and the environment. The transport rate of substrate S into the cell is noted as r_T, while the overflow metabolism is described by r_O. Proteins P are synthesized with rate r_P [96, 44, 121]. The stoichiometric matrix for this system, therefore, reads

$$\mathcal{N} = \begin{pmatrix} \alpha & -\gamma & -1 \\ 0 & \beta & 0 \end{pmatrix}. \tag{3.2}$$

Thus, the intracellular network for this minimal model can be written as

$$\begin{pmatrix} \dot{M} \\ P \end{pmatrix} = \mathcal{N} \begin{pmatrix} r_T \\ r_P \\ r_O \end{pmatrix} - \mu \begin{pmatrix} M \\ P \end{pmatrix}, \tag{3.3}$$

where the vector of intracellular components is $c = (M, P)^T$ and represent the concentrations of these components in the cell. With Eq. (2.18) the growth rate for this specifc system is

$$\mu = (\alpha\ r_T - r_O)\ w_M + r_P\ (\beta\ w_P - \gamma\ w_M), \tag{3.4}$$

where w_M and w_P are the molecular weights of metabolites M and proteins P, respectively. In case, that no mass is lost in the synthesis of proteins P from metabolites M, the second term in Eq. (3.4) becomes zero and the growth rate is only determined by the reaction rates connecting the cells with the environment.

3.2 Flux analysis

The construction of kinetic models rely on extensive knowledge of the functional formulation of the enzymatic reaction rates and their associated parameters. The exchange rates between the microbial population and the environment can easily be obtained from experiments, but knowledge about the intracellular rates is limited, as they are more difficult to be recovered.

3.2.1 Flux analysis of a general intracellular network

In this section, we investigate metabolic networks which are given through a mass balance based on stoichiometry. Those systems are described by Eqs. (2.17) and can be rewritten by plugging Eq. (2.18) into (2.17)

$$\dot{c} = \mathcal{N}\ r - w^T\ \mathcal{N}\ r\ c = \left(\mathrm{Id} - c\ w^T\right) \mathcal{N}\ r = W\ \mathcal{N}\ r, \tag{3.5}$$

where W is a matrix determined by molecular properties of the components C_i, which includes the concentration values c and the molecular weights w. Thus, W is a matrix describing the cell composition. The analysis of this system provides useful information for the estimation of the reaction kinetics. The concentrations of the intracellular components can be measured or estimated using data from literatue. Under the assumption that the stoichiometry $\mathcal{N}$ of the system and the intracellular concentrations c are known, the solution space of the reaction rates r in steady state can be analyzed

$$0 = W\ \mathcal{N}\ r, \tag{3.6}$$

which is obtained by setting Eq. (3.5) to zero. The steady states r_0 are given by the kernel of $W\mathcal{N}$. The kernel of $W\mathcal{N}$ is determined by two vector spaces

$$r_0 = r_{w,0} + r_{n,0}. \tag{3.7}$$

One subset of the kernel of $W\mathcal{N}$ is given by the kernel of $\mathcal{N}$, as in classic flux balance analysis, and the elements are defined by

$$r_{n,0} = (\mathrm{Id} - \mathcal{N}^{+}\mathcal{N})\ a, \tag{3.8}$$

where $a \in \mathbb{R}$ is an arbitrary vector with the same dimension as r_0. Here, we make usage of the pseudoinverse, which was introduced in Section 2.2.4, as $\mathcal{N}$ is not a square matrix in general. The remaining elements of $\ker(W\mathcal{N})$, denoted as $r_{w,0}$, satisfy the following equations

$$W\ \mathcal{N}\ r_{w,0} = 0 \quad \text{and} \quad \mathcal{N}\ r_{w,0} = c_{w,0} \neq 0. \tag{3.9}$$

To put conditions given by Eq. (3.9) into words, this means $c_{w,0}$ is given by the kernel of W. The cell composition matrix W is nonsingular, as the determinant of W is given by

$$\det(W) = 1 - \sum_{i=1}^{N_C} w_i\ c_i, \tag{3.10}$$

where the addends of the second term are the mass fractions of the intracellular components C_i and with Eq. (2.16) sum up to one, and $\det(W)$ is, therefore, always zero. Additionally, the

kernel of W is in fact one-dimensional, which is shown in the Appendix, Section A.1.1, and can be formulated as

$$c_{w,0} = c\ s, \qquad s \in \mathbb{R}, \tag{3.11}$$

leading to

$$r_{w,0} = \mathcal{N}^+\ c_{w,0}. \tag{3.12}$$

Thus, plugging Eqs. (3.8) and (3.12) into (3.7), the steady states of (3.5) are given by

$$r_0 = \mathcal{N}^+\ c_{w,0} + (\mathrm{Id} - \mathcal{N}^+\mathcal{N})\ a. \tag{3.13}$$

Alternatively, under the assumption that the growth rate μ and cellular composition c is known, e.g. from experimental data, the steady state condition of the system is given by

$$\mu\ c = \mathcal{N}\ r. \tag{3.14}$$

Comparing the solution of this Eq. (3.14), using Eq. (2.25), and Eq. (3.13) reveals that both solutions correspond to each other with

$$c_{w,0} = \mu\ c. \tag{3.15}$$

The principle of the found solution space, defined by Eq. (3.13) holds true for all types of metabolic networks independently of its size and form, which can range from whole-cell models to coarse-grained models.

3.2.2 Flux analysis of a minimal coarse-grained cell model

The procedure for the search of the solution space of the rate vector can be applied to the minimal coarse-grained model given by Eqs. (3.2) and (3.3). The pseudoinverse of $\mathcal{N}$ in this case is

$$\mathcal{N}^+ = \begin{pmatrix} \frac{\alpha}{\alpha^2+1} & \frac{\gamma\alpha}{\beta+\beta\alpha^2} \\ 0 & \frac{1}{\gamma} \\ -\frac{1}{\alpha^2+1} & -\frac{\gamma}{\beta+\beta\alpha^2} \end{pmatrix}, \tag{3.16}$$

and with the formular obtained in Eq. (3.13) the solution space of the reaction rates is

$$r_0 \ = \ \begin{pmatrix} \frac{1}{\alpha^2+1} \\ 0 \\ \\ \frac{\alpha}{\alpha^2+1} \end{pmatrix} a \ + \ \mathcal{N}^+ c\ s \qquad s \in \mathbb{R}, \quad a \in \mathbb{R}, \tag{3.17}$$

with the first term representing the solution from the stoichiometry of the system, usually also obtained from flux balance analysis, and the second term the solution determined by the molecular composition of the cell. A closer look at the solution given by Eq. (3.17) reveals that the rate connecting only the intracellular components of the system, which is the protein synthesis rate r_P, is defined by the second term solely. Thus, for the assumption of the known specific growth rate μ and molecular composition of the cell, the second term is fixed to $\mathcal{N}^+ c\ \mu$ and the protein synthesis rate r_P is uniquely defined. On the other hand, the remaining rates, meaning the substrate uptake rate r_T and the overflow metabolism rate r_O, are coupled through

one degree of freedom a. If one of these rates is known, the degree of freedom a can be determined and, therefore, the last remaining rate.

3.3 Resource allocation as optimization problem

In the last section, the solution space of reaction vector r was investigated. Now, we assume fixed reaction kinetics for the three reaction rates and analyze the allocation of resources to the different protein fractions. This is treated as optimization problem (introduced in Section 2.2.3), as the cells have to regulate its metabolism to achieve maximal growth and using Eq. (2.22) this problem can be written as

$$\begin{aligned} &\max_{0\leq x\leq 1} && \mu(x) \\ &\text{subject to} && \\ & && 0 &&= \mathcal{N}\ r - \mu\ c \\ & && \mu &&= w\ \mathcal{N}\ r \\ & && 0 &&= 1 - Q - T - R, \end{aligned} \tag{3.18}$$

where x is a fraction of P, such that $R = x\ \theta\ P$ and $T = (1-x)\ \theta\ P$ with Q forming a constant protein fraction, meaning $Q/P = 1 - \theta$. We assume the residual biomass U to form a constant biomass fraction of $\rho = w_U\ U$. The constraints of this optimization problem correspond to cells, which exhibit a balanced cell state and are in steady state. In the following, different sets of kinetic rate laws are investigated. The first set of reaction rates includes linear dependencies on the protein fractions T and R and Michaelis-Menten kinetics depending on M for the protein synthesis r_P

$$\begin{aligned} r_T &= k_T\ \frac{S}{S\ +\ K_T}\ T \\ r_P &= k_P\ \frac{M}{M\ +\ K_P}\ R \\ r_O &= k_O\ M\ T. \end{aligned} \tag{3.19}$$

The optimal solution for the partitioning of the proteome is obtained for different substrate concentrate values S ranging from $S = 0.01$ g/L to $S = 1$ g/L and can be seen in Figure 3.2. The parameters such as the maximal reaction rates k_i are chosen to achieve a growth rate of around $\mu = 0.5$ 1/h for $S = 1$ g/L. The residual biomass fraction is assumed to occupy 40 % of the total biomass and the metabolite concentration is adjusted to occupy 3-10 % of the total biomass as seen in Figure 3.2B. The optimal solution for the protein fractions T/P and R/P are plotted as functions of μ in Figure 3.2C. With increasing growth rate, the concentration of transporter proteins T decrease while ribosomal proteins R increase, which is a well-established relation [142]. As the sum of both protein fractions add up to half of the proteome, the optimal solution differs from the protein fraction values obtained from experimental data taken from the study by Schmidt et al. [139]. In comparison with the experimental data, the slopes are steeper with fraction $T/P \in [0.21, 0.34]$, $R/P \in [0.16, 0.29]$ and a point of intersection at around $\mu \approx 0.41$ 1/h as seen in Figure 3.2C. In the following, only the protein fractions T/R and R/P are visualized and compared in Figure 3.3, as the parameters are chosen to achieve a maximal growth rate of $\mu = 0.5$ 1/h and the cell component composition does not differ between the different models. The optimal solution of the first set of reaction rates is depicted in Figure 3.3A for comparison. The parameter sets for the kinetic rate laws are given in Appendix, Section A.1.3.

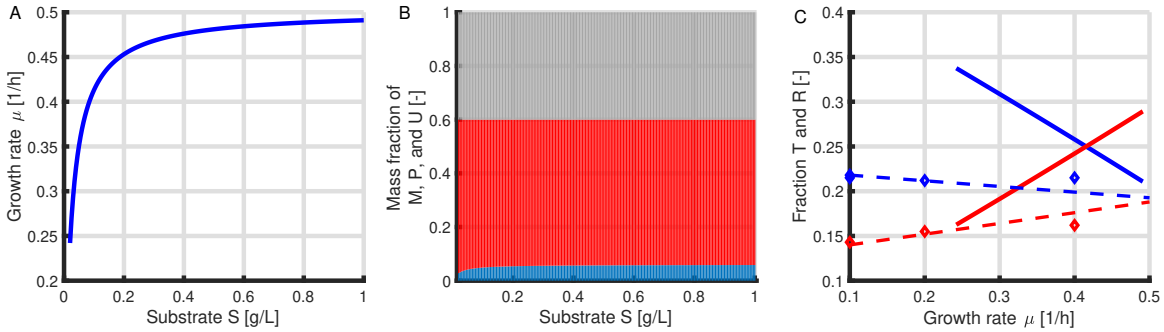

Figure 3.2: Solution of the optimization problem for a model with reaction rates given by Eq. (3.19): Specific growth rate μ as a function of substrate S (A). Mass fractions of the intracellular components for different substrate concentrations: metabolites M (blue), proteins P (red), and residual biomass U (grey) (B). Mass fractions of T (blue) and R (red) and experimental data (dashed lines) are taken from the study by Schmidt et al. [139] (C).

In Figure 3.3B, the proteome allocation of a model with linear dependencies on the protein fractions T and R and Michaelis-Menten kinetics for M in both reaction rates r_P and r_O is shown. Here, a very shallow slope with a greater distance between the fraction values is observed. The values of the protein fractions vary for different kinetic parameters, as e.g. higher parameter values for k_T or k_P lead to a smaller T fraction or R fraction, respectively. A higher value for k_T reduces the impact of T on the reaction, which can shift the protein allocation in favor of R to maximize the growth rate. This relation holds true for all model variation, considered in this section. Nevertheless, the shallow slope remains for all parameter variations. In the third variation, we consider Michaelis-Menten kinetics for T in both reactions r_T and r_O, which contain a dependency of T. Figure 3.3C, shows a non-linear relation between the optimal allocation into T and R and the growth rate μ, as the fraction T first increases for growth rate values $\mu \in [0.1, 0.2]$ and then decreases, with the fraction R the other way round. This non-linear relation can not be seen in the experimental data. A similar behavior is obtained in Figure 3.3D for a model which assumes Michaelis-Menten kinetics for all reactions concerning T and M. The turning point in this case with the set parameter values occurs at a higher growth rate value

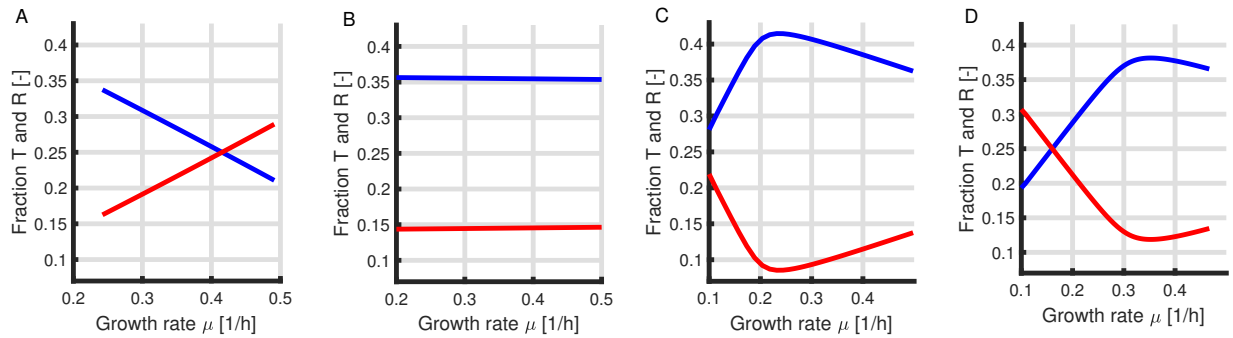

Figure 3.3: Mass fractions of T (blue) and R (red) over growth rate μ for different sets of kinetic rate laws: Michaelis-Menten kinetics for M in r_P (A), Michaelis-Menten kinetics for M in r_P and r_O (B), Michaelis-Menten kinetics for T in r_T and r_O (C) and Michaelis-Menten kinetics for T in r_T and r_O and M in r_P and r_O (D).

$\mu = 0.34$ 1/h. It is worth to note, that even though the parameter values are chosen to satisfy $\mu_{max} \approx 0.5$, the biomass yield is not maintained constant over the 5 different model variations. In order to keep the biomass yield constant in addition, a more elaborate approach has to be employed to determine the appropriate kinetic parameters. Nevertheless, the first three model variations all reproduce the linear relation between R and μ, while the last two with the chosen parameter sets do not. In the latter sections, we chose the first set of reaction rate formulations given by Eq. (3.19), as this formulation is able to reproduce the relations obtained experimental data with best approximation.

3.4 Application: L-phenylalanine production with Escherichia coli

In this section, we consider a L-phenylalaine producing *E. coli* strain with glycerol as substrate S as an example for a biotechnological production process to which we can employ the proposed coarse-grained approach. The process exhibits a decoupling between the biomass and product formation due to L-tyrosine auxotrophism, meaning biomass is only formed if L-tyrosine is available in the environment [155, 169]. The considered strain and the considered process are characterized in the following works [34, 60, 61].

3.4.1 Model formulation

The minimal model in Section 3.1 is extended to include an additional rate r_F describing the L-phenylalanine (F) production formation with a corresponding protein sector F_P and a rate r_C proportional to the rate of respiration. Furthermore, we split the macromolecules into the proteins and residual biomass U, which is synthesized with rate r_U. The extension of the scheme can be seen in Figure 3.4.

The stoichiometric matrix for this system is

$$\mathcal{N} = \begin{pmatrix} \alpha & -\gamma & -1 & -\delta & -1 & -1 \\ 0 & \beta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \epsilon & 0 & 0 \end{pmatrix} \tag{3.20}$$

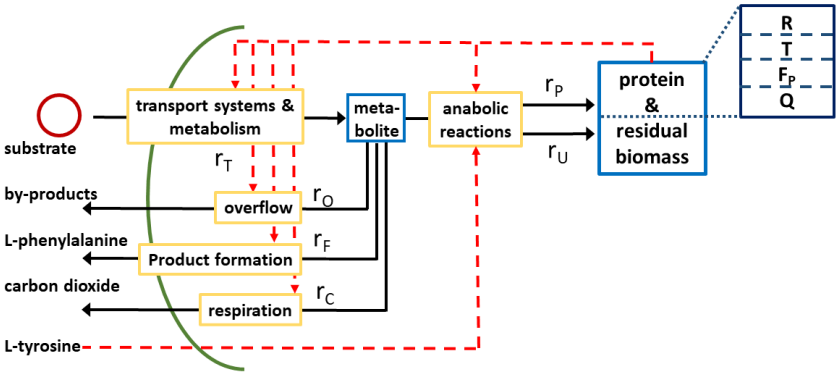

Figure 3.4: Scheme of the coarse-grained model expanded to include the formation of L-phenylalanine with rate r_F, respiration r_C, and residual biomass U and corresponding synthesis rate r_U.

and the differential equations describing the system are

$$\begin{pmatrix} \dot{M} \\ P \\ U \end{pmatrix} = \mathcal{N} \begin{pmatrix} r_T \\ r_P \\ r_O \\ r_U \\ r_F \\ r_C \end{pmatrix} - \mu \begin{pmatrix} M \\ P \\ U \end{pmatrix}. \tag{3.21}$$

In the next step the reaction rates of the system are defined. The rates which are involved in the central metabolism are dependent on the proteome fraction T. The rates in question are: substrate uptake rate r_T, overflow metabolism r_O and respiration rate r_C. The anabolic reactions, r_P and r_U, and the L-phenylalanine production rate r_F are dependent on R and F_P respectively. The dependency of the rates on the the respective protein fractions is a common approach resulting in self-replicator systems, as introduced in Section 2.2.3 [142, 121, 44]. The L-tyrosine auxotrophism is incorporated through a factor $\tau(A)$ which is multiplied with the protein synthesis rate r_P and the synthesis rate of the residual biomass r_U. This factor $\tau(A)$ is set to 1 if L-tyrosine A is available and otherwise 0.1, leading to a lower biomass production for a absence of L-tyrosine, which occurs during the L-phenylalanine formation phase. Thus, the reaction rates are

$$\begin{aligned} r_T &= k_T \frac{S}{S + K_T} T \\ r_P &= k_P \frac{M}{M + K_P} R\, \tau(A) \\ r_O &= k_O\, M\, T \\ r_U &= k_U \frac{M}{M + K_U} R\, \tau(A) \\ r_F &= k_F\, M\, F_P \\ r_C &= k_C\, M\, T, \end{aligned} \tag{3.22}$$

where

$$\tau(A) = \begin{cases} 1, & A > 0 \\ 0.1, & A \leq 0. \end{cases} \tag{3.23}$$

The partitioning of the proteome into the fractions T and R is determined by an estimated linear function of M [34] obtained from experimental data [139]

$$\begin{aligned} T &= (a_T\, M + b_T)\, P \\ R &= (a_T\, M + b_R)\, P. \end{aligned} \tag{3.24}$$

where a_T and a_R are the slope at which the fractions T/P and R/P increase or decrease with M and b_T and b_R are the fractions at $M = 0/\mathrm{mol/g_{DW}}$. After induction of the L-phenylalanine production, a part of the proteome is allocated to fraction F_P, leading to a shift of carbon flux from biomass production to product formation due to lack of L-tyrosine in the feed at time t_{ind} with a time delay as follows

$$F_P = \psi(t)\, F_{\max}\, P, \tag{3.25}$$

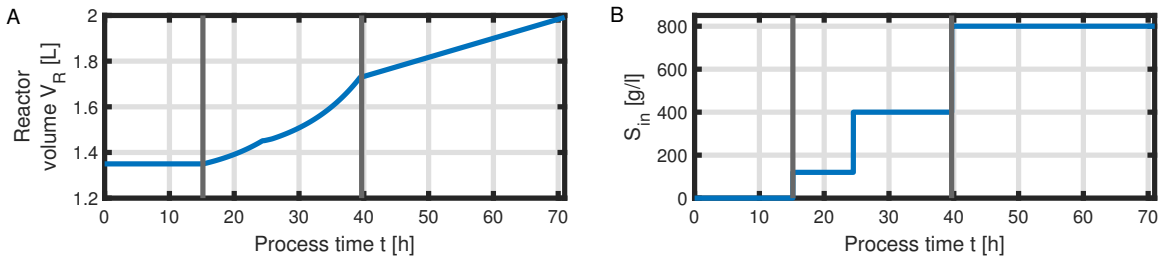

Figure 3.5: Feeding profile of the process. Reactor volume V (A) and substrate concentration S of the feed S_{in} (B) over the time course of the process, where vertical lines indicate the three process phases (batch phase, fed-batch phase, and production phase with constant feeding).

where

$$\psi(t) = \frac{t - t_{\text{ind}}}{(t - t_{\text{ind}}) + t_{\psi}} \tag{3.26}$$

with t_{ψ} a constant indicating the time where half amount of $F_{\max}$ is attained. The remaining fraction form fraction Q, which is not of further interest

$$1 = T + R + F_P + Q. \tag{3.27}$$

In the following, we can set equations to model a complete bioprocess consisting of two process phases: a biomass production phase, which comprises a initial batch phase. This is followed by two fed-batch phases with two corresponding feeding solutions containing different substrate concentrations, and which was initiated with induction of the cells with IPTG. For now, we assume ideal mixing conditions in a bioreactor of volume $V_R = 1.35$ L with a feeding profile, as seen in Figure 3.5, and feeding rate q_{in}. The fed-batch phases exhibit exponential feeding to accompany a growth rate of $\mu_{\text{set}} = 0.1$ 1/h and a constant feeding is chosen for the L-phenylalanine production phase. Thus, according to the already presented equations in Section 2.2.2, the dynamics of V_R are described by

$$\dot{V}_R = q_{\text{in}} \tag{3.28}$$

and we consider only one feeding substrate S (glycerol) with feed concentration S_{in}

$$\dot{S} = \frac{q_{\text{in}}}{V_R}\,(S_{\text{in}} - S) - r_T\; X\; w_S\; n_S. \tag{3.29}$$

The biomass concentration is described by

$$\dot{X} = \mu\; X - \frac{q_{\text{in}}}{V_R}\; X \tag{3.30}$$

and the product equations are

$$\begin{aligned} \dot{F} &= n_F\; r_F\; X\; w_F - \frac{q_{\text{in}}}{V_R}\; F \\ \dot{O} &= r_O\; X\; w_O - \frac{q_{\text{in}}}{V_R}\; O, \end{aligned} \tag{3.31}$$

where w_F and w_O are the molecular weights of F and O, respectively.

Here, O represents acetate as an exemplary by-product of the process. Experimental data provide the variables q_{in}, S_{in}, A and initial values for Eqs. (3.28)-(3.31). After around process time $t = 71$ h, experimental data suggest a stop in biomass and product formation and a increased production of acetate. The behavior at the end of the process has not been investigated at this point and the mechanism of this behavior is not depicted in the reaction system.

3.4.2 Flux analysis of a metabolic model including product formation

Flux analysis can also be performed on the metabolic model with L-phenylalanine production. The pseudo-inverse of $\mathcal{N}$ in this model is

$$\mathcal{N}^{+} = \frac{1}{D}\begin{pmatrix} \alpha & \alpha\frac{\gamma}{\beta} & \alpha\frac{\delta}{\epsilon} \\ 0 & \frac{D}{\beta} & 0 \\ -1 & -\frac{\gamma}{\beta} & -\frac{\delta}{\epsilon} \\ 0 & 0 & \frac{D}{\epsilon} \\ -1 & -\frac{\gamma}{\beta} & -\frac{\delta}{\epsilon} \\ -1 & -\frac{\gamma}{\beta} & -\frac{\delta}{\epsilon} \end{pmatrix}, \tag{3.32}$$

where $D = \alpha^2 + 3$. The kernel of $W\mathcal{N}$ is the image of the matrix

$$Id - \mathcal{N}^{+}\mathcal{N} = \frac{1}{D}\begin{pmatrix} D-\alpha^2 & 0 & \alpha & 0 & \alpha & \alpha \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha & 0 & D-1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha & 0 & -1 & 0 & D-1 & -1 \\ \alpha & 0 & -1 & 0 & -1 & D-1 \end{pmatrix} \tag{3.33}$$

Analogous to the minimal model described by Eqs. (3.2) and (3.3), the anabolic reaction rates, which connect internal cell components, are determined by only the first term in solution (3.13). The solution space for reaction rates, which connect intracellular pools with the environment, is expanded by the second term, thus by the image of (3.33). The non-zero column vectors in (3.33) are linearly independent and therefore the kernel of $W\mathcal{N}$ is 4-dimensional, resulting in four degrees of freedom.

3.4.3 Numerical simulation and comparison with experimental data

A numerical simulation of Eqs. (3.20)-(3.31) with process time up to $t = 75$ h was performed using MATLAB 2022a with ode15s as ODE solver, which is a multistep solver with variable step sizes. The numerical solution, which is divided into the two process phases analogous to the experiments, is compared with experimental data as seen in Figure 3.6. Experimental data show positive substrate concentration during the batch phase and no substrate excess afterwards in Figure 3.6A. The cell population grows exponentially during the biomass production phase and its growth slows down after the induction of the product formation with a complete stop in biomass increase coinciding with the absence of L-tyrosine in Figure 3.6B. Since the cultivation is exposed to different process conditions and is affected by its auxotrophism due to the absence of L-tyrosine during the L-phenylalanine production phase, different sets of parameters were

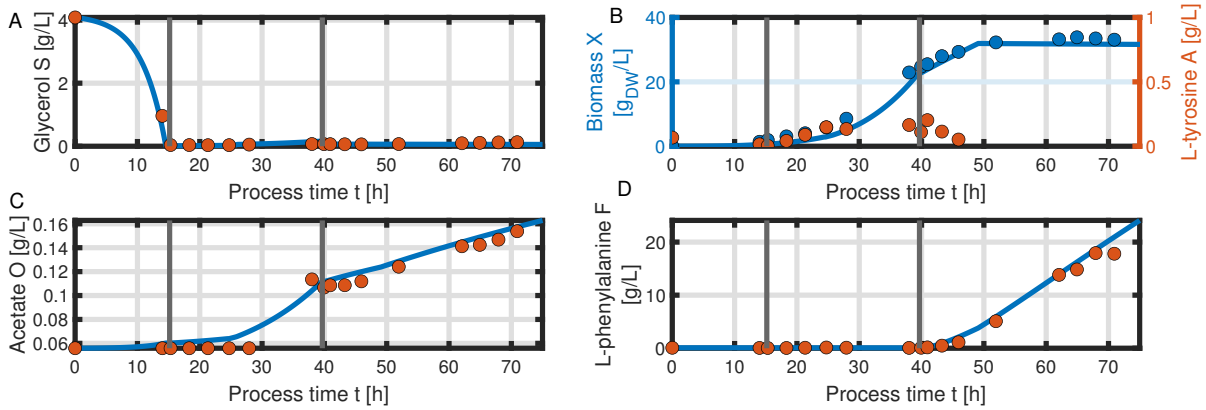

Figure 3.6: Comparison of the simulated quantities (blue) against the experimental data (orange points) of the L-phenylalanine production process up to $t = 75$ h. Time course of the following concentrations: glycerol S (A), biomass X, L-tyrosine A (B), L-phenylalanine F (C), and acetate O (D).

used for the different process phases. The L-phenylalanine production phase exhibits a higher by-product formation as seen in Figure 3.6C. With this in mind, the reaction constants were set higher for by-product formation and respiration for the production phase, and the used parameters can be found in Appendix, Section A.1.4. The solution at the end of the first process phase is used as initial value for the production phase. Figure 3.6 shows good agreement between the simulated concentrations of substrate S (Figure 3.6A), biomass X (Figure 3.6.B), acetate O (Figure 3.6C), L-phenylalanine F (Figure 3.6D) and the experimental data. Thus, for the given parameter set, the model is able to reproduce the dynamics of the L-phenylalanine production process.

Besides the measurable data, the solution of the intracellular concentrations can be investigated as well, as seen in Figure 3.7. The metabolite concentration M is high during the time of the fed-batch coinciding with higher substrate levels in the environment, decrease over the course of the fed-batch phases and remains at constant level during the product formation phase (Figure 3.7A). A look at the mass fractions of the intracellular components in Figure 3.7B reveals that the macromolecules P and U occupy most of the biomass, where the first makes up

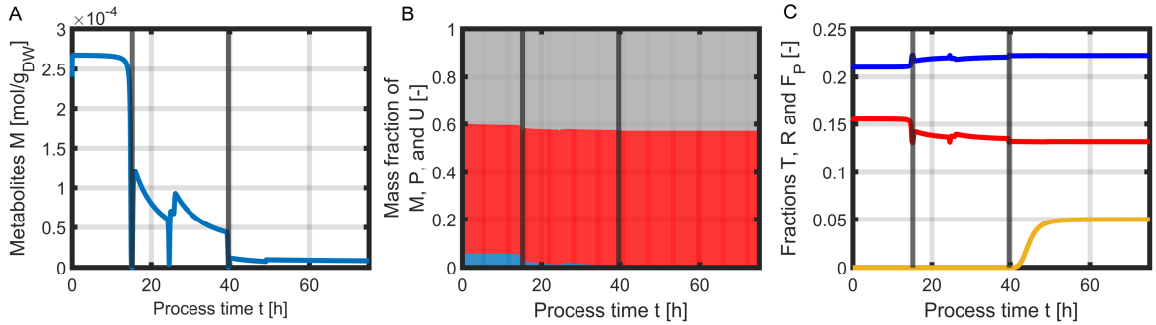

Figure 3.7: Time course of intracellular concentrations of the L-phenylalanine process: metabolite concentration M (A). Mass fractions over the course of the process: metabolites M (blue), proteins P (red), and residual biomass U (grey) (B). Mass fractions of T (blue), R (red), and F (yellow) over the course of the process (C).

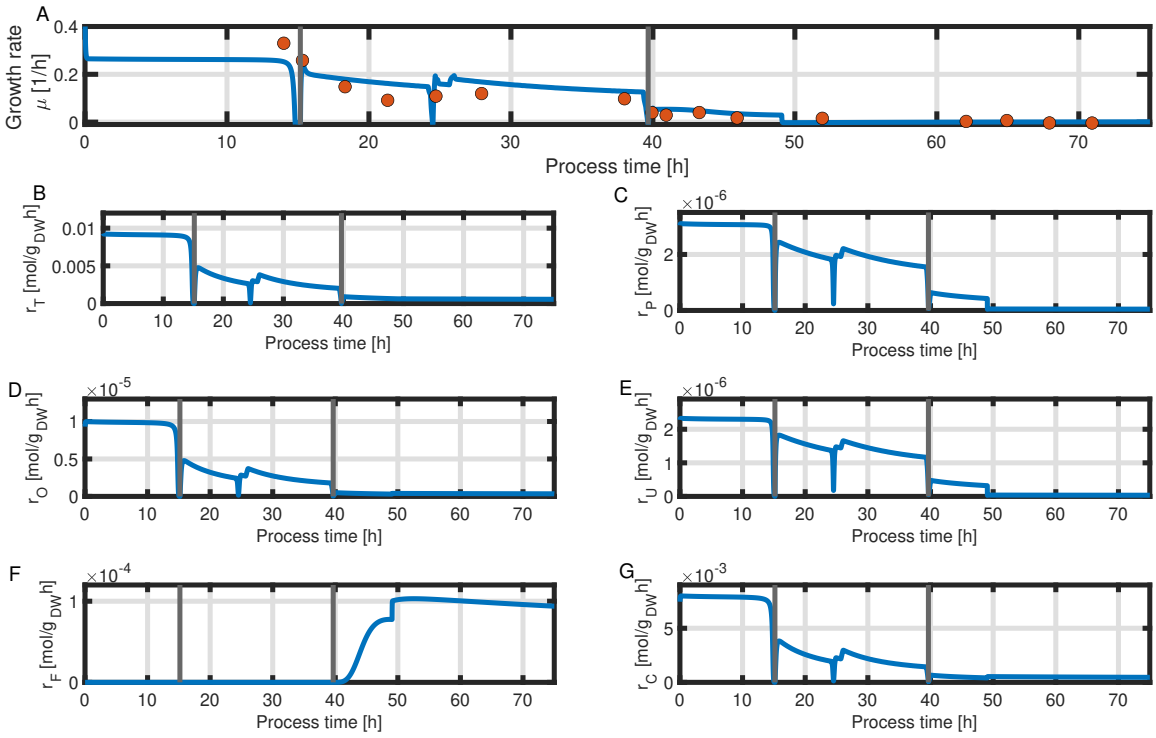

Figure 3.8: Time course of the simulated specific growth rate μ (blue) and the point-wise calculated specific growth rate obtained from experimental data indicated as orange dots (A), substrate transport rate r_T (B), protein synthesis rate r_P (C), overflow metabolism rate r_O (D), residual biomass synthesis rate r_U (E), product formation rate r_F (F), and respiration rate r_C (G).

around 60% of the biomass, while the mass fraction of the metabolites M is negligible, especially after the initial batch phase. With decreasing metabolite concentration M over the course of the biomass production phase, the concentration of ribosomal proteins R decreases as well, while the concentration of transporter proteins T increases, as the protein fractions R and T follow the behavior of M (Figure 3.7C). During the L-phenylalanine production phase, the levels of ribosomal and transporter protein concentrations remain constant. Protein fraction F_P, which is dedicated to the product formation, follows the description in Eq. (3.25) and (3.26).

Additionally, the growth rate μ and the reaction rates of the intracellular network can be obtained from the model and are depicted in Figure 3.8. During the batch phase, the growth rate starts at a higher value and decreases during the fed-batch phase as long as L-tyrosine is available in the environment, roughly following the trend of the growth rate, which is obtained by pointwise calculation from the experimental data of the biomass (Figure 3.8A). We note that the pointwise calculated growth rate has to be considered with caution, as it is calculated from two consecutive points with large time differences which can result in large deviations from the actual growth rate. Over the course of the biomass production phase, the reaction rates, except for r_F, follow the dynamics of the metabolites M since all of them are dependent on M and the concentration of T and R hardly changes as seen in Figure 3.8B-G. In the product formation phase, the anabolic reaction rates are dominated by the L-tyrosine auxotrophism and therefore nearly zero (Figure 3.8C and E).

3.5 Consideration of substrate uptake limitations

In this section, we want to introduce a rate limiting concept, where the substrate uptake rate is limited by two factors, which was presented by Morchain et al. [109, 110]. This refinement is necessary to reflect heterogeneous systems in a changing environment. Here, cells may encounter the situation in which the substrate uptake rate is limited by the transport of the substrate in the liquid phase (physical regime) rather than by the the substrate uptake capacity of the cell itself (biological regime). Thus, in order to obtain the actual substrate uptake, both limiting factors (external and internal) have to be considered. The internal biological uptake capacity ϕ_S^b is determined by the availability of transporter proteins T

$$\phi_S^b = k_T \; T. \tag{3.34}$$

The biological uptake rate ϕ_S^b can be compared to the substrate transport rate by micro-mixing. The shift between these two regimes is based on the characteristic time scales for micro-mixing τ_{mix}, which is related to properties of the fluid and local turbulent energy dissipation rate, and biological uptake τ_u. The micro-mixing time is defined as

$$\tau_{\text{mix}} \approx 17 \sqrt{\frac{\mu_L}{\rho_L \epsilon}}, \tag{3.35}$$

where μ_L is the dynamic viscosity and ρ_L the density of the liquid phase, and ϵ the local turbulent-energy dissipation rate [7]. In this work, the micro-mixing time τ_{mix} remains constant. The characteristic time for substrate uptake writes

$$\tau_u = \frac{S}{w_S \; X \; k_T \; T_{\max}}, \tag{3.36}$$

where $T_{\max}$ is the maximal possible transporter protein concentration. This time scale is determined by the local substrate concentration and the denominator represents the substrate the cells are able to take up. Thus, τ_u is the time the cells require to consume the local substrate concentration. With Eqs. (3.34)-(3.36) the actual specific substrate uptake rate ϕ_S is

$$\phi_S = \phi_S^b \; \left(1 - e^{-\tau_u / \tau_{\text{mix}}}\right) = k_T \; T \; \left(1 - e^{-\tau_u / \tau_{\text{mix}}}\right). \tag{3.37}$$

Therefore, the substrate uptake rate depends on both the biological substrate uptake capacity, represented by the first factor, and the relative supplying capacity of the environment, represented by the second factor. If the environment is sufficiently rich to feed the cells, the actual substrate uptake rate is determined by its own internal uptake capacity. Otherwise, the uptake rate is limited by the environmental supply. It can be shown that the simulation of a bioprocess for the refined substrate uptake rate (3.37) compares to the reaction rate with Michaelis-Menten kinetics given by Eq. (3.22) as seen in Appendix, Section A.1.5.

3.6 Deterministic behavior of the metabolic model

We consider the metabolic system given by Eqs. (3.20)-(3.22). Product formation left aside, it is a very generic model that is able to reproduce the dynamics of an *E. coli* population in a bioreactor. In the following sections, we use the same parameter set as in the numerical

simulation of Section 3.4. In order to understand the response of the model to different sources of perturbations in the system, we investigate the deterministic behavior of this model.

3.6.1 Steady states

First, we study a simplified metabolic model containing the four reactions: substrate uptake rate r_T, protein synthesis rate r_P, residual synthesis rate r_U and overflow metabolism r_O. The rate of respiration is left aside here, but can be assumed to be incorporated in r_O. For the purpose of calculating an analytical solution for the steady states, we do not consider feedback through protein fractions T and R, leading to following system of ODEs

$$\begin{aligned}
\dot{M} &= \alpha\ k_T - \gamma\ k_P \frac{M}{M+K_P} - k_O\ M - \delta\ k_U \frac{M}{M+K_U} - w_M\ (\alpha\ k_T - k_O\ M)\ M \\
\dot{P} &= k_P \frac{M}{M+K_P} - w_M\ (\alpha\ k_T - k_O\ M)\ P \\
\dot{U} &= k_U \frac{M}{M+K_P} - w_M\ (\alpha\ k_T - k_O\ M)\ U.
\end{aligned} \tag{3.38}$$

In order to study the stability of steady states, we can linearize the ODE system and set up the Jacobian matrix by differentiating the differential equation system

$$J = \begin{pmatrix}
-\gamma\ k_P \frac{K_P}{M+K_P} - k_O - \delta\ k_U \frac{K_U}{M+K_U} - w_M\ (\alpha\ k_T - 2\ k_O\ M) & 0 & 0 \\
\gamma\ k_P \frac{K_P}{M+K_P} - w_M\ k_O\ P & -w_M\ (\alpha\ k_T - k_O\ M) & 0 \\
\delta\ k_U \frac{K_U}{M+K_U} - w_M\ k_O\ U & 0 & -w_M\ (\alpha\ k_T - k_O\ M)
\end{pmatrix} \tag{3.39}$$

The Jacobian matrix is a triangular matrix and, therefore, the eigenvalues are the elements of the main diagonal. For positive parameter values and concentrations the steady states are stable if $\alpha\ k_T > k_O\ M$. We set parameter values as in the simulation in Section 3.4, the stoichiometric coefficients, the molecular weights of the cellular components and $S = 0.1$ g/L. This leads to three steady states, where only one solution is biologically relevant, as the other two have negative values for M. With the observation in (3.39), we can conclude that this steady state is stable if $\alpha\ k_T > k_O\ M$.

The steady states of the metabolic model including feedback for both substrate uptake rates, described in the last Section 3.5, are simulated for different substrate concentration values S. The solutions for the different substrate concentration values are indicated by dots and linearly interpolated. For simplicity reason, the substrate rate defined in Section 3.5 is simply called substrate uptake rate with micro-mixing. For this case, the steady states are calculated with a constant value of the biomass value of $X = 18$ g/L, as it equals the biomass concentration of the L-phenylalanine production process at $t = 38$ h. This time point coincides with the time at which the distribution of cell properties is investigated in Chapter 5. The results of the simulated steady states are depicted in Figure 3.9. The content in P ranges from $P_{\min} = 11.75\ \times\ 10^{-6}$ mol/g_{DW} to $P_{\max} = 12.45\ \times\ 10^{-6}$ mol/g_{DW} (Figure 3.9A), whereas the metabolite content ranges from $M_{\min} = 0$ mol/g_{DW} to $M_{\max} = 25\ \times\ 10^{-5}$ mol/g_{DW} (Figure 3.9B). Thus, we can note that the steady states of P are closer together than the steady states of M. For large substrate concentrations, the steady states of both models are equivalent, since the substrate uptake rate is determined by the biological regime for a substrate rich environment. For lower substrate concentrations, the substrate uptake rate with micro-mixing (orange curve) is dominated by the

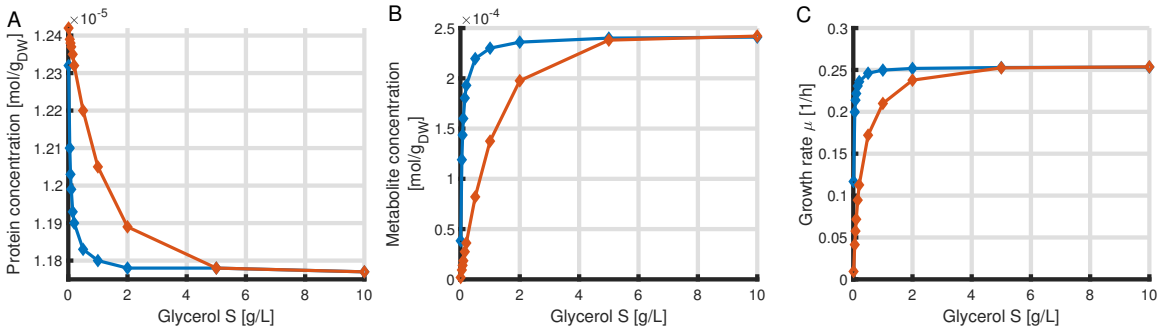

Figure 3.9: Steady state values for the protein concentration P (A), metabolite concentration M (B) and growth rate μ (C) for different substrate concentrations (solutions indicated by dots) for the model with substrate uptake rate with Michaelis-Menten kinetics (blue) and with micro-mixing (orange)

physical regime, leading to a lower metabolite concentration M and therefore higher protein concentration P. This reflects that at low substrate concentration and high biomass concentration, the environmental supply might become insufficient to satisfy the intracellular demand, resulting in a lower growth rate compared to the model with the standard Michaelis-Menten formulation (Figure 3.9C). The difference between the two models become less prevalent or non existent when the biomass concentration is low. The values M, P and μ (also U, but this variable is not shown here and can be obtained from M and P), at steady state form a set of consistent values, which correspond to cells with balanced metabolism.

3.6.2 Time scale for cellular response to environmental changes

As mentioned in Section 2.1.2, we investigate two scenarios. First, the response to external noise and later, in Section 3.6.3, the response to intrinsic noise is examined. For the first case, a homogeneous cell population in steady state is initialized and we investigated the adaption to a new steady state due to a change in the environmental variables like substrate concentration. The relaxation time of the variables is recovered through determination of the time needed to reach 95 % of the steady state when the substrate concentration changes, while starting in a steady state. The relaxation time for M is $\tau = 0.015$ h as seen in Figure 3.10A. The relaxation time for P depends on the transition distance between the steady states. For steady states which are closer together as pictured in Figure 3.10B (light blue), the relaxation time is long. Here, the initial change in metabolite concentration are fast, but then require a long time to converge to the final steady state. The relaxation time is determined by a superposition of at least two exponential functions, one with a fast time scale analogous to the metabolites M and one with a longer time scale, driving the protein concentration slowly to its new steady state. It is important to note that the difference between these two steady states in this case is very small and therefore can be neglected.

The relaxation time for the growth rate μ is in the same range as for M and P. But we observe that for the growth rate the values vary a lot for the first time steps until it reaches its new steady state (Figure 3.11). During the transition time, the growth rate exhibits higher values before converging to a lower growth rate if the substrate concentration increases. If the cells are adapted to an environment without substrate excess and the substrate concentration decreases

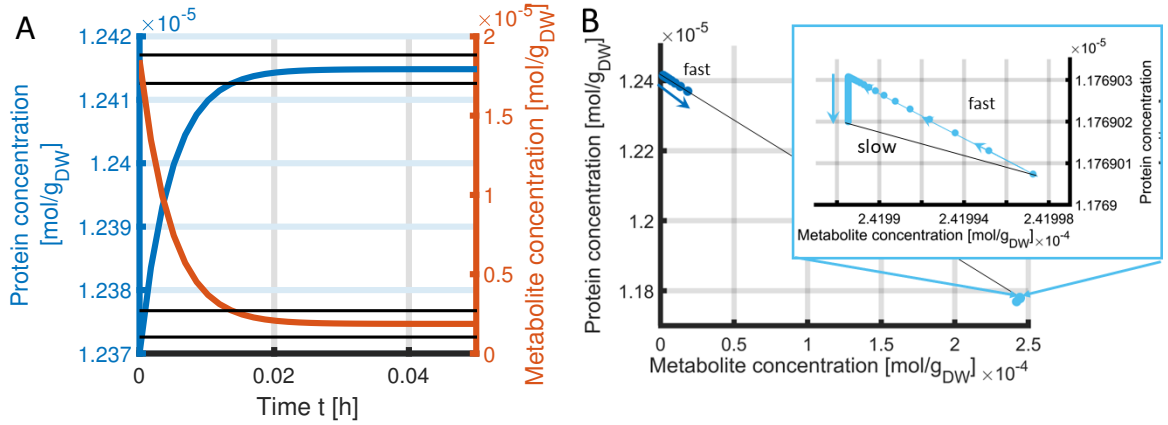

Figure 3.10: Investigation of the relaxation time for the transition from the the steady state at $S = 0.1$ g/L to $S = 0.01$ g/L for the following variables: P (blue), M (orange), where the black lines indicate the bands, where the distance to the new steady state is smaller than 5% of its distance to its original steady state (A), and the phase portrait over the M-P plane for the transition from the the steady state at $S = 0.1$ g/L to $S = 0.01$ g/L (blue) and $S = 5$ g/L to $S = 7$ g/L (light blue), where the black line indicates the line of the steady states for different substrate concentration (B).

further, then the growth rate of the cells even take negative values, otherwise we only observe positive growth rate during the transition phase. Negative growth rate values indicate a loss in biomass, which can occur if the substrate uptake is smaller than the respiration rate and overflow metabolism. Thus, if the substrate concentration S increases in the environment, the cells experience an accumulation of metabolites on short term, leading to a higher growth rate during transition. In case of a decrease of substrate S, the metabolite pool M of the cells are exhausted, leading to a short term loss of biomass (negative growth rate).

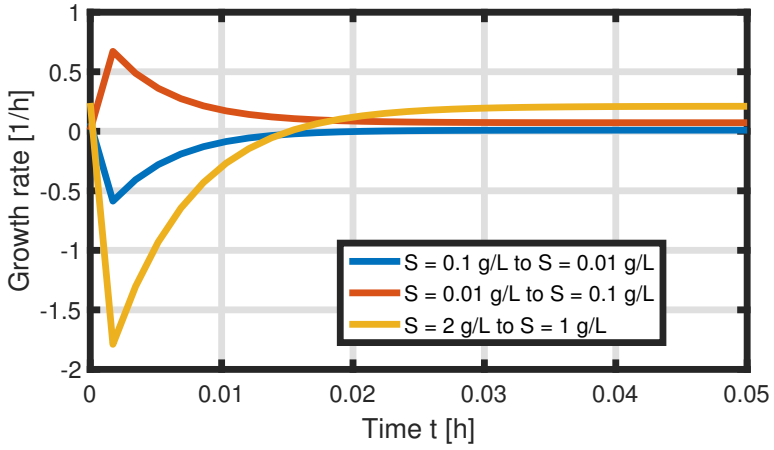

Figure 3.11: Investigation of the relaxation time for the growth rate μ for different substrate changes.

3.6.3 Time scale for cellular response to intrinsic noise

During cell division, unbalanced cells, whose system variables are not in steady state, are formed. This corresponds with an perturbation of a cell from steady state to a randomly chosen numerical value. In this section, the system dynamics and its time scales are analyzed for this scenario. Particles are initialized in a constant environment with random properties at $t = 0$ and their evolution is tracked in time. Figure 3.12 shows the initialized particles converging to a steady state. A closer look at the time axis reveals that the protein concentration converges to its steady state in Figure 3.12A a lot slower taking hours than metabolite M in Figure 3.12B, which converges in seconds. The numerical solution of the steady state can be compared with the analytical solution from the simplified model in Section 3.6.1. The different time scale for P compared to the time scale obtained in the last section does not pose a contradiction, as the difference lies in the initial state of the cells: Since the initial state is randomly assigned, the cells do not exhibit a consistent set of properties P, M and μ, meaning that the cell are unbalanced. The time scales do not change for different substrate concentration values in the environment, which remain constant during the simulation.

The time scales of different modifications of the metabolic model can be obtained and compared with the full model. Interestingly, the time scales do not change in Figure 3.13A and B for a metabolic model, which does not contain the feedback loops through T or R. When modifying the reaction kinetic $r_i = k_i\ f(c)$ to exclude the dependency of a protein fraction, the maximal reaction rate constant k_i is multiplied with a constant value of the corresponding protein fraction P_i equal to the value of this protein fraction in steady state of the full metabolic model. Thus, the modified reaction rate $\hat{r}_i$ is determined by

$$\hat{r}_i = k_i\ P_i\ f(c). \tag{3.40}$$

Only considering feedback in the substrate uptake rate r_T, couples the time scales for M to P and M now converges after hours, not after seconds, as seen in Figure 3.13C and D. In the very first time steps, the values of M change a lot within seconds and slow down until they converge with the same time scale as P in Figure 3.13D. An even larger changes in the first time steps and slow convergence afterwards are observed for M in a model with only feedback through R or only T in Figure 3.13F and H, while the behavior of convergence for P stays the same for

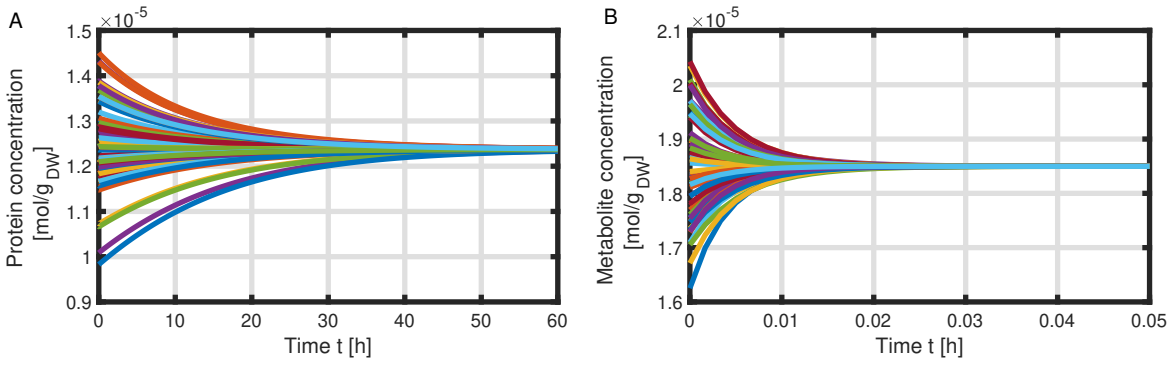

Figure 3.12: Time scale of P (A) and M (B) to converge to their respective steady-state in an environment with a substrate concentration of $S = 0.1$ g/L and random initial system variables P and M indicated by different colors.

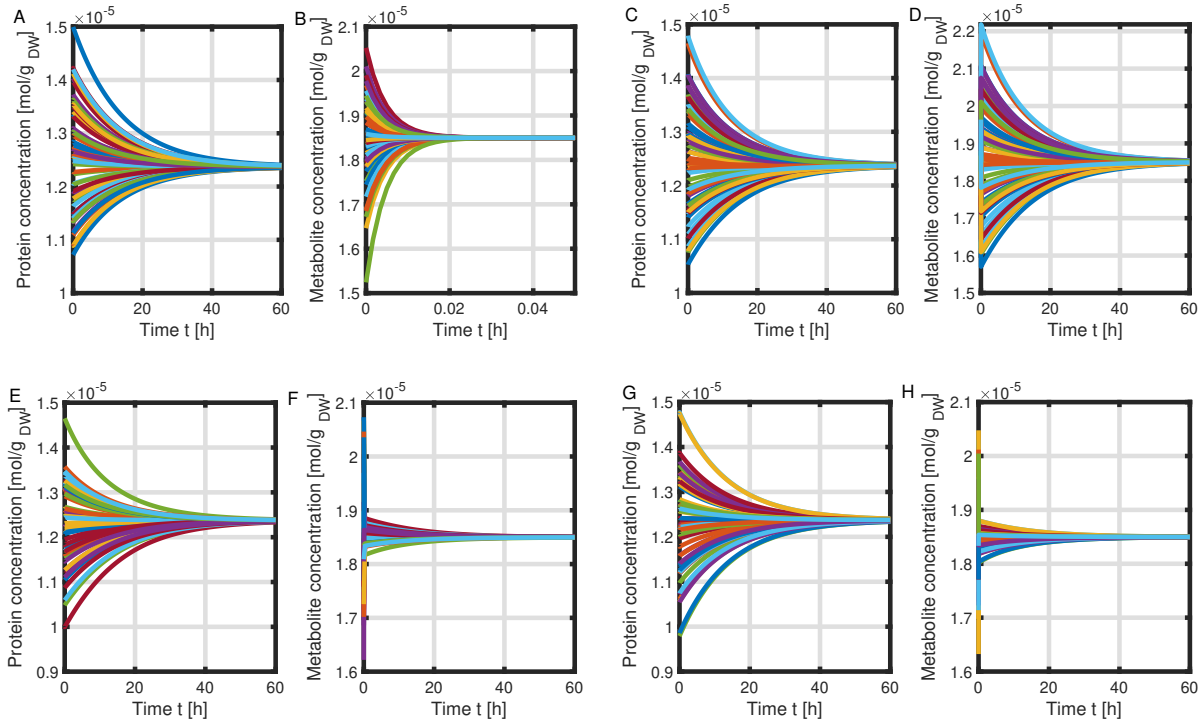

Figure 3.13: Time scale of P and M to converge to their respective steady-state in an environment with a substrate concentration of $S = 0.1$ g/L and random initial system variables P and M for different modified models: no feedback through R nor T (A,B), only feedback in r_T (C,D), only feedback through R (E,F) and only feedback through T (G,H).

all model modifications. Thus, two different time scales for M and P can only be observed if we consider feedback through T and R or none at all. The consideration of only one protein fraction results in a coupling of M to the dynamics of P and the exhibition of only one time scale.

3.7 Modified minimal coarse-grained model

For the purpose of incorporating a description of the intracellular network of the cells in a PBE, a modified minimal coarse-grained model is presented, which is shown schematically in Figure 3.14. Again, the model comprises two components M and P, but this time, the interpretation of the components differ. P represents a part of the proteome consisting of transporter proteins T and ribosomal proteins R. Component M represents the remaining cell components, in terms of former cell component classes: metabolites, other proteins and residual biomass. Both cell components have equal molecular weight w, thus the stoichiometric matrix is given by

$$\mathcal{N} = \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \end{pmatrix} \tag{3.41}$$

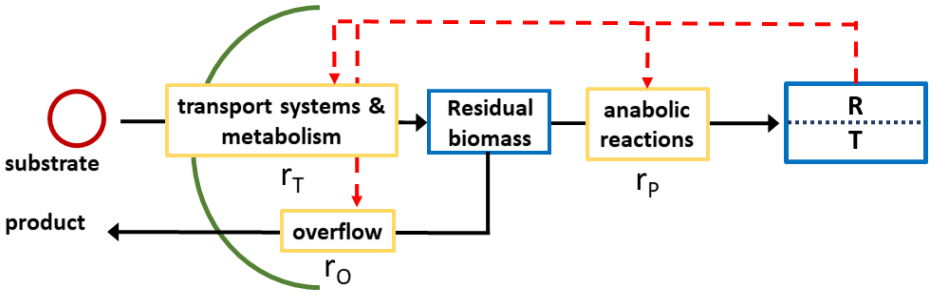

Figure 3.14: Scheme of a coarse-grained model with partitioned proteome (ribosomal proteins R, proteins linked with the central metabolism T); it consists of two components, indicated as blue boxes; residual biomass, proteins (transporter and ribosomal proteins). The pools are connected by a minimal set of reactions, indicated by yellow boxes, for substrate uptake, overflow metabolism, and protein synthesis.

and the reaction rates are described by Eq. (3.19). This model is later used in a well-mixed environment and thus we do not consider limitations through the liquid phase, as described in Section 3.5. Since the biomass only contains two cell components, the concentration of one component can be expressed by the other, if the latter is known

$$M = \frac{1}{w} - P. \tag{3.42}$$

M represents a different pool of cellular components in this model, therefore another function of M has to be employed to describe the allocation of proteins to fractions T and R. We assume that compartment P consist of $58 - 63\%$ transporter proteins T and the remaining fraction is allocated to R. Furthermore, M is assumed to take values between $M_{\text{min}} = 1$ mol/g_{DW} and $M_{\text{max}} = 2.5$ mol/g_{DW}. With these assumptions, the concentrations of R and T are determined by

$$T = \left(0.63 - \frac{0.63 - 0.58}{(2.5 - 1) \times 10^{-5}} M\right) P \tag{3.43}$$

$$R = \left(0.37 + \frac{0.63 - 0.58}{(2.5 - 1) \times 10^{-5}} M\right) P. \tag{3.44}$$

As the component M in this model is macromolecular, the higher molecular weight has to be considered when setting up the equation of the substrate concentration S. Thus, the Eq. (3.29) is adapted to

$$\dot{S} = \frac{q_{in}}{V_R} (S_{in} - S) - r_T \; X \; w_S \; \gamma \; n_S. \tag{3.45}$$

The equations for biomass and the reactor volume remain the same.

In the following, the biomass production phase, as described in Section 3.4, is simulated using the modified model (orange) and compared with the simulation obtained in Section 3.4.3 (blue). Figure 3.15A shows the simulated substrate concentration S, which reveal almost no difference between both models. A slightly higher biomass production is seen for the modified model in Figure 3.15B, while still being in good agreement with the experimental data. The concentrations of intracellular components M and P are depicted in Figure 3.16A. Similar to the full model,

compartment M shows higher values during the batch phase and decreases over the course of the fed-batch phase. This is also apparent, when investigating the mass fractions of components M and P in Figure 3.16B. The mass fraction of P varies from $10-50\%$ of the biomass, thus showing a much higher range than in the full model. It is worth noting that the full model assumes a fixed mass fraction for the sum of transporter proteins T and R and exhibits a protein concentration with little variation, leaving a very narrow range for T and R. In comparison, this modified model allows a larger variation for protein fractions T and R. The ratio between transporter proteins T and ribosomal proteins R in Figure 3.16C shows the same tendencies and similar values as the ratio between T and R in the full model, which are obtained by calculating the fraction $T/(T+R)$ and $R/(T+R)$. With decreasing values for M, the fraction T increases, while R decreases, as more transporter proteins are needed to transport components M into the cell to counter the decrease.

Figure 3.17 shows a comparison of reaction rates between the modified model and the full model. Despite of a very high growth rate in the beginning of the batch phase, the difference between the values of the growth rate μ is small in Figure 3.17A. The reaction rates r_T, r_P and r_O cannot be compared side-by-side, as M has higher molecular weight in the modified model. To conclude, this modified model reproduces the dynamics of the concentrations in the liquid phase, but with lower detail in the description of the metabolic network. The relation between the protein fractions T and R are not based on experimental data as in the full model and have to be taken with caution.

3.8 Short summary and discussion

While simple growth models taking into account only biomass, substrate and product often are insufficient to describe observed dynamics in a bioprocess [46], whole-cell models are complex and difficult to calibrate [102]. A good compromise represents coarse-grained models because of their simple model structure and their ability to take the most important cellular processes into account. In this chapter, we proposed a coarse-grained modeling approaches based on a strict mass conservation to model bacterial growth as a basis for metabolic engineering applications. In this way, classical flux analysis could be extended to take into account fluxes into macromolecules such as the proteome. In classic flux balance analysis, the system of equations only consists of the stoichiometric matrix $\mathcal{N}$ as mentioned in Section 2.2.4. Thus, the term describing the

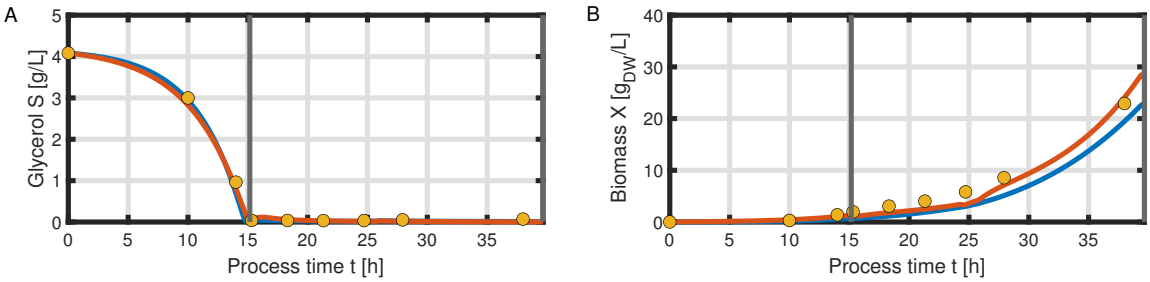

Figure 3.15: Comparison of the simulated quantities using the full model (blue) and the modified model (orange) against the experimental data (yellow points) of the L-phenylalanine production process up to $t = 39$ h. Time course of the following concentrations: glycerol S (A) and biomass X (B)

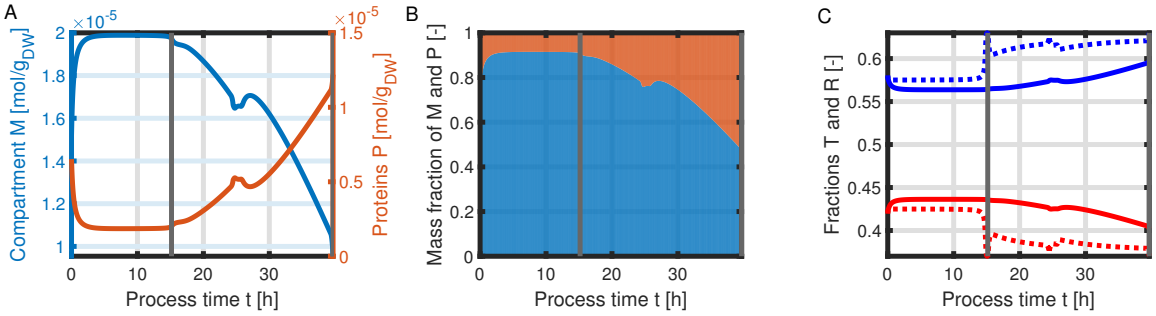

Figure 3.16: Time course of intracellular concentrations of the modified model: compartment M (blue) and proteins P (orange) (A). Mass fractions over the course of the process: compartment M (blue) and proteins P (orange) (B). Mass fractions of T (blue) and R (red) for the modified model (solid line) and full model (dashed line) (C).

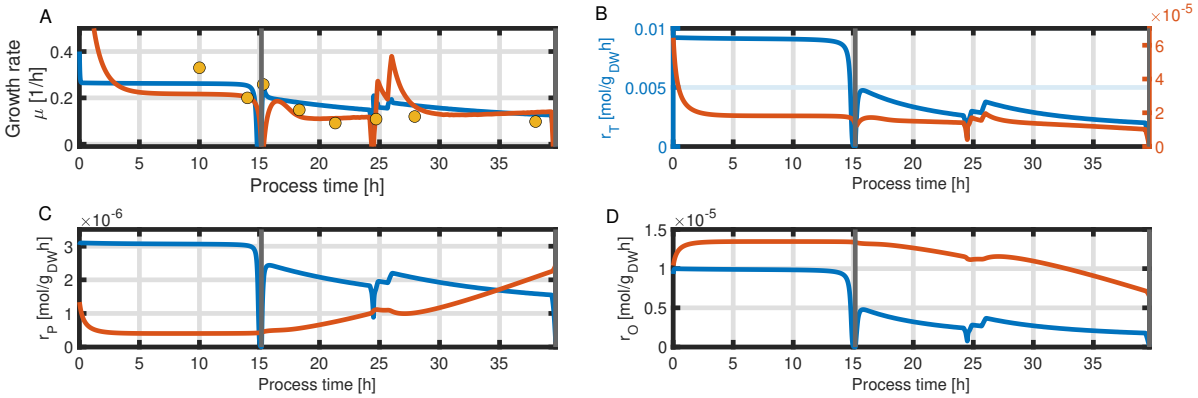

Figure 3.17: Time course of the simulated specific growth rate μ (A), substrate transport rate r_T (B), protein synthesis rate r_P (C) and overflow metabolism rate r_O (D) for the modified model (orange) compared to the full model (blue).

dilution due to cellular growth is absent. In the approach presented, additional solutions of the reaction rate state space are provided by the null space of the mass matrix W that requires information on the mass fraction of the components of the model. Therefore, the solution space is a superposition of the kernel of $\mathcal{N}$, which is typically also obtained in flux balance analysis, and the kernel of W. This principle holds for the minimal coarse-grained model as well as for a general cellular network independently of its internal structure under the condition of strict mass conservation. One downside of this approach is the required knowledge of the cell composition to compute W. The reaction rates obtained from the steady state analysis were not computed in particular, but if one would continue with the approach, the challenge has to be faced. While the molecular weights of the components can easily be obtained from data bases, the concentrations of the components are more difficult to recover. With the minimal coarse-grained model, the low number of components is a additional advantage, as the mass fractions can be estimated in this case. Macromolecular components are measurable and given the condition of mass conservation, the concentration of the low molecular can be determined through procedure of exclusion. In case of more complex networks, especially multiple low molecular components, the clear identification of the mass fraction can pose a difficult challenge due to low abundance and wide range of the metabolites [13, 144]. On the other hand, the reaction kinetic laws are often times unknown.

Thus, if the cellular composition can be determined, the approach provides the opportunity to estimate the reaction kinetics through the solution space obtained.

As the focus of this research is coarse-grained models, we provided a formulation for a minimal model whose structure is in accordance with that in previous studies [143, 14]. Many studies have dealt with the derivation of growth laws under various conditions [74, 19, 66]. We considered two main protein fractions, an R fraction representing the transcription and translation apparatus and a T fraction, taking into account metabolic and transport enzymes. In contrast to the previous approach, the reaction kinetics were fixed, while the cellular composition is unknown. An optimization program, similar to other works [8, 123], under the assumption that the sum of the R fraction and T fraction is taken as constant, was set up and the results for different sets of reaction rates were compared. In the process, we obtained a solution that leads to a comparable behavior of the R and T fraction as a function of the specific growth rate, and we conclude that the measured data are in good agreement with the expectation of an efficient and optimal acting organism.

Based on the structure of the minimal model, we have expanded the model to include the dynamic environment of a bioreactor system that allows us to realize also different process design strategies such as feeding or continuous culture. For the reaction kinetics, the set of equations with the best agreement from the optimization program was chosen. Experimental data from an L-phenylalanine production process are taken as an example for parameter identification and estimation, and a good agreement between simulation and experimental data is obtained. The end of the bioprocess, indicated by a stop of product and increased by-product formation, has not been investigated in the scope of this work. In order to describe this phase, mechanistics behind the shift from product to by-product have to be understood. Here, additional cell process regulations might be necessary.

Furthermore, the concept of different limitations on the substrate uptake was introduced. As we aim to simulate population heterogeneity induced by process variations, a model with the appropriate response to a wide range of substrate concentrations in different environments has to be formulated. Often times, the parameters of the biological model had to be adjusted according to the reactor scale since either lower substrate uptake than expected or an over-prediction of biomass yield was obtained [84, 109]. This emphasizes the requirement to incorporate the limitations originating from the substrate supply provided by the environment affected by the dynamics of the liquid phase and substrate uptake capacity of the biological phase. In a nutrient rich well-mixed environment the substrate uptake rate is determined by the biological regime and the refinement would not be necessary, but with the possibility of cells competing for the available resources, the substrate uptake is limited by the supplying capacity of the environment. This was observed when comparing the steady state solution of the models with and without the incorporation of micro-mixing. While the dynamics remain unchanged for an substrate rich environment, the growth rate is decreased in environments with lower substrate concentrations. The severity of slowing down of the substrate uptake depends on the parameter value assumed for the micro-mixing time scale and on the biomass concentration present, but does not change the overall behavior of the simulation. This justifies the higher parameter value, as usually assumed in other studies [109, 103], chosen for the simulations in the upcoming Chapter 5.

Due to the possibility of population heterogeneity originating from external and internal sources, the time scales of cellular response to these two kind of perturbations have been analyzed. The investigation revealed two different time scales in the cell response for the appropriate choice of reaction kinetics when the cell are out of equilibrium (uptake rate, internal composition

and growth rate are not balanced), and only one time scale when cells at equilibrium respond to external fluctuations. A deciding factor for the two time scales stem from the different magnitudes in molecular weight of the cellular components. While the formation of metabolites is a fast process, the synthesis of proteins happens over longer time-scale due to the high molecular weight and the unbalanced state of the cells. In unbalanced cellular state, the cell's composition is not adapted to the current environment leading to a non-optimal amount of T and R proteins, which are responsible for the regulation of the cellular processes. This result shows the necessity to include intrinsic noise, which result in randomly distributed cells regarding their cell state, when simulating a population in an changing environment, the aim in Chapter 5, in order to maintain different time scales in cellular response and decouple the cellular state from the local concentrations in the liquid phase.

In order to utilize a coarse-grained model as a basis for a PBE modeling approach, the current coarse-grained model has to be simplified further to decrease the number of cellular components to two. Additionally, a higher possible range of mass fractions of the components is required to ensure the necessary flexibility in cell state. A detailed reasoning for the requirements of the biological model is given in the upcoming Section 4.1. The modified coarse-grained model, which is characterized by a different interpretation of the cell components, is able to reproduce similar time courses for the substrate and biomass concentration and, additionally, the growth rate is in agreement with the experimental data. Due to the components representing different intracellular pools, the proteome allocation into the protein fractions in dependency of M is more challenging and difficult to compare with experimental data. Here, the allocation into T and R is based on assumption about the range of M, T and R. While it overall does have similar tendencies as the behavior in the original coarse-grained model, the dynamics of T and R are much slower upon changes in the environment. The precursor M represents a much larger pool from which P is synthesized, which is not biologically sound and questionable. The limitations have to be kept in mind and caution should be exercised, when utilizing this model.

To summarize, coarse-grained models are a sound basis for the development of bioprocesses due to their simple structure with only a minor number of parameters and the flexibility to simulate and optimize different biotechnological process designs. The structure of the minimal coarse-grained model can easily be extended to include other cellular processes of interest such as the dependence of oxygen availability on growth through the addition of reaction rates or protein fractions.

4 A population balance equation approach for modeling microbial growth

In the previous chapter, an unsegregated coarse-grained model based on mass conservation was established. To account for the fact that cell populations are segregated into discrete cells with different cell states, a common choice are PBE models. The characteristics of a PBE and its general formulation have been presented in Section 2.3. Early examples of these models were based on the age [11, 42] or the mass [54, 101] as an indicator for the cell state. The problem with the former as variable for a PBE is that it is impossible to obtain data about the age of a microbial population. Cell size or cell mass as state variable overcomes this problem. Additionally, cell size strongly depends on the environmental conditions as the population leans towards higher cell sizes in a nutrient rich environment. But with this interdependence, the downside is that the cell size is primarily an outcome of the perceived environment and cannot predict the cell's response towards dynamical changes in the environment [41]. Therefore, we introduce a PBE model which includes the cell mass and the protein content of a cell to describe its cellular state. The cellular processes are defined by the coarse-grained model established in the last section. With the combination of the two variables a good indicator about the cell's history through the cell mass is given and through the protein content as variable the cellular response to the environment is determined. The outcome of the PBE model is compared to the results of the unsegregated coarse-grained model and the influence of intrinsic noise is investigated. The latter is accomplished by varying the cell division rate function and the degree of unequal distribution of the cell mass and protein content during cell division. Furthermore, it is shown that the numerical scheme used for the simulation conserves the first moments.

4.1 Model formulation

In order to employ a PBE to model microbial growth, we have to define the processes involved in the description of the population. For a cell population, we consider a PBE with growth and cell division. Additionally, descriptor variables of the model have to be defined. To keep the computational effort when solving the PBE numerically, the number of descriptor variables have to be minimized. Coming from the set-up of coarse-grained models, one could think about using the protein concentration as descriptor variable. But the usage of the concentration of a cell component is not a suitable choice, as cell division is not a compatible concept with this descriptor choice. A cell with a given cell component concentration divides into two daughter cells, which are usually assumed to have the same cell component concentration as the mother cell. Thus in this instance, the cell division term in the PBE would vanish. Another possibility is to consider the biomass concentration or mass of a cell as descriptor variable. The disadvantage of this choice is the lack of knowledge about the cell composition especially the proteome of the cell. Therefore, knowledge about the cell composition is needed. In Section 3.7, we introduced a modified coarse-grained model with only two cell components, where knowledge of one

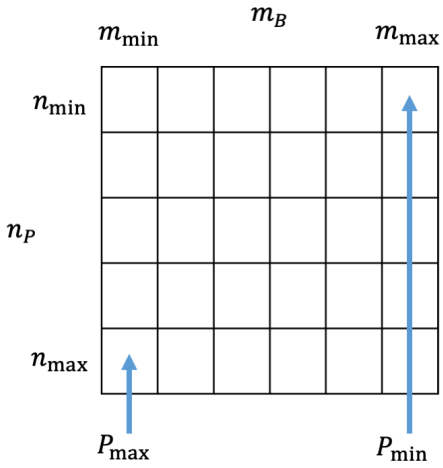

Figure 4.1: State space of a population balance equation with two descriptor variables: biomass of a cell m_B and protein content of a cell n_P.

component is sufficient to describe the cell state through Eq. (3.42). The reaction rates of the model are determined by the concentration of each cell component. Furthermore, to compare the result of the PBE with the non-segregated simulation in Section 3.7, the biomass concentration of the population is required. On that account, we require two descriptor variables in order to employ the model presented in Section 3.7 for the description of the metabolic network of the cell population: the mass of a cell $m_B \in [m_{\min}, m_{\max}]$ given in g_{DW} and the amount of protein substance in a cell $n_P \in [n_{\min}, n_{\max}]$ given in mol. Through both variables the protein concentration of a given cell can be calculated

$$P = \frac{n_P}{m_B}. \tag{4.1}$$

The question arises why the coarse-grained model which was presented in Section 3.1 had to be modified in the first place. The choice of the proteome consisting of only the fractions T and R has the following reasoning: If we take a look at the two-dimensional state space for the two descriptor variables and consider the protein content in y-direction and biomass in x-direction, the smallest possible concentration is given by $P_{\min} = n_{\min}/m_{\max}$, which is the upper right corner, and the largest by $P_{\max} = n_{\max}/m_{\min}$, which is the lower left corner. One has to keep in mind, that in order for cell division to take place $m_{\max}$ and $n_{\max}$ have to be multiples of $m_{\min}$ and $n_{\min}$, respectively. Additionally, the largest possible protein concentration, which e.g. can be assumed as proteins forming all of the biomass of the cell, is constraint by

$$P_{\max} \leq \frac{1}{w}. \tag{4.2}$$

This imposes a constraint on how $m_{\min}$ and $n_{\max}$ can be chosen. The smaller $m_{\min}$ is, the smaller $n_{\max}$ has to be as well, as $P_{\max}$ is bounded. Therefore, we have to set $m_{\min} > 0$. If we want to look at a population, the cells ideally should mainly reside in the upper left part of the state space to be able to grow in both directions. The division process should also dominate over the growth for larger cells, meaning that larger cells are more likely to divide than to further increase their biomass and cell content. The necessity for the cell division process is the problem if we consider a model with two components, where e.g. M represents only the metabolites and

P all the macromolecules, which would make up nearly all of the biomass, as the cells in question would mainly reside in the lower left part of the state space due to the protein concentration being near to the maximal possible concentration. We need the protein concentration to be able to vary and therefore to be considerably smaller than $P_{\max}$. The same problem occurs if we assign a fixed biomass fraction to the compartment U and keep the intracellular components M and P.

According to Eq. (2.29), the PBE for the density number function $N(n_P, m_B, t)$ with the two defined descriptor variables is given by

$$\frac{\partial N(n_P, m_B, t)}{\partial t} + \frac{\partial}{\partial n_P}\left(r_n(n_P, m_B, S)N(n_P, m_B, t)\right) + \frac{\partial}{\partial m_B}\left(r_m(n_P, m_B, S)N(n_P, m_B, t)\right) \\ = 8\ \Gamma(2n_P, 2m_B)N(2n_P, 2m_B, t) - \Gamma(n_P, m_B)N(n_P, m_B, t) \tag{4.3}$$

under the assumption that cells divide into two identical daughter cells. Furthermore, we neglect the spatial coordinate of the cells, as we assume a well-mixed homogeneous environment. The first term accounts for the accumulation of cells with the state (n_P, m_B), the second and third term represent the loss cells with state (n_P, m_B) due to a change in the protein content and cell growth. The right hand-side comprise the terms due to cell division. For simplicity reasons, we will use $n_p = n$ and $m_B = m$.

The rate of change functions are given by

$$r_n(n, m, S) = \dot{n} = \gamma_P\ m\ r_P(n, m) \tag{4.4}$$

and

$$r_m(n, m, S) = \dot{m} = \mu\ m = w^T\ \mathcal{N}\ r(n, m, S)\ m \tag{4.5}$$

where $r(n, m)$ is the vector containing all the reaction rates r_T, r_P and r_O, which are defined in the coarse-grained model by Eq. (3.19).

Following the works of Mantzaris [98, 99] and Quedeville et al. [129], the division rate is given by

$$\Gamma(n, m) = \left(\frac{m}{m_c}\right)^{\alpha}\left(\frac{n}{n_c}\right)^{\beta}, \tag{4.6}$$

where m_c and n_c are critical values above which cells are more likely to divide and two scalars α and β. The higher the exponents α and β are, the sharper is the division function. The advantage of this division rate function over the one used in some works [53, 101] is that it is defined on the whole state space and does not require an additional assumption for the distribution used, but only the critical values m_c and n_c. In the following, both exponents in the division rate are set to 5, analogous to the work of Mantzaris [98]. Additionally, the critical values used are $m_c = 500$ fg and $n_c = 1.3 \times 10^{-3}$ fmol. Figure 4.2 shows the division rate function in logarithmic scale for variable protein amount n and constant biomass m (Figure 4.2A) and variable biomass m and constant protein content n (Figure 4.2B), as well as the function dependent on both variables (Figure 4.2C). Thus, the probability for a cell to divide increases exponentially with its cell mass m and protein content n.

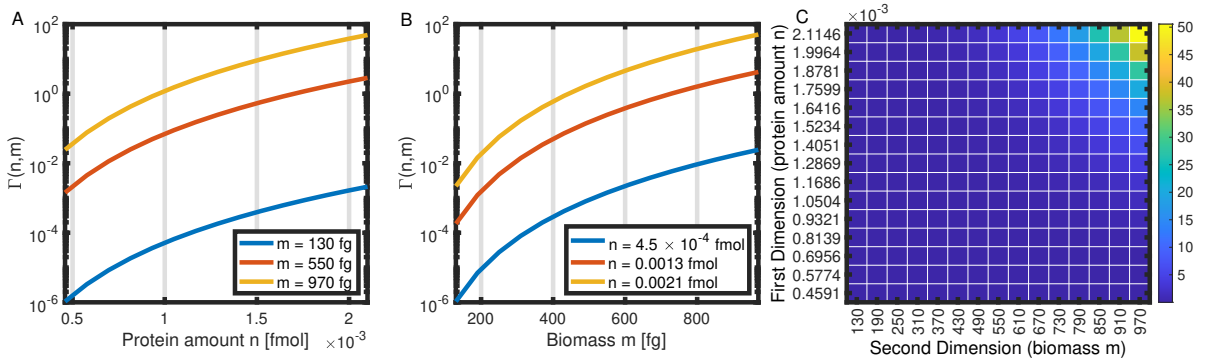

Figure 4.2: Division rate functions Γ as functions of protein amount n for constant biomass $m = 130$ fg (blue), $m = 550$ fg (red), and $m = 970$ fg (yellow) in logarithmic scale (A), as functions of biomass m for constant protein content $n = 4.5 \cdot 10^{-4}$ fmol (blue), $m = 0.0013$ fmol (red), and $m = 0.0021$ fmol (yellow) in logarithmic scale (B), as well as function of both variables m and n.

Furthermore, boundary conditions have to be set. Here, the system is closed and the fluxes at the boundaries are zero. The cells do not grow when they reach their maximal mass

$$r_m(n, m_{\max}) = 0. \tag{4.7}$$

This is not entirely true, as cells do not in fact stop growing, but they never reach infinite mass, as for high mass values the division process dominates over the growth process. To avoid dealing with unbounded state spaces, the maximal mass of the cells is set to a very high values. The fluxes at the other side of the domain are also zero, since there are no cells outside of the domain which grow to reach a mass of $m_{\min}$

$$r_m(n, m_{\min}) = 0. \tag{4.8}$$

The same applies for the protein content of a cell

$$r_n(n_{\max}, m) = 0 \qquad r_n(n_{\min}, m) = 0. \tag{4.9}$$

In order to get a full set of equations to describe the dynamics of a well-mixed bioreactor. The liquid phase is described by the equation for the substrate

$$\dot{S} = r_{S_{in}} - \int_{m_{min}}^{m_{max}} \int_{n_{min}}^{n_{max}} r_T(n, m, t)\ m\ N(n, m, t)\ w_S\ \gamma\ n_S\ \mathrm{d}n\ \mathrm{d}m \tag{4.10}$$

4.2 Numerical simulation

Using the Finite Volume method introduced in Section 2.3.2, the biomass production phase of the L-phenylalanine process can be simulated for the PBE model given by Eqs. (4.4)-(4.10) using MATLAB 2022a with ode23s as ODE solver, which is based on an explicit Runge-Kutta method. For the process involving change in the internal variables we employ a first-order upwind scheme, as the rate r_m is always positive and r_n is positive for the majority of the time. For the cases where $r_n < 0$, we use a first-order backward scheme instead. The state space of Eq. (4.4) is

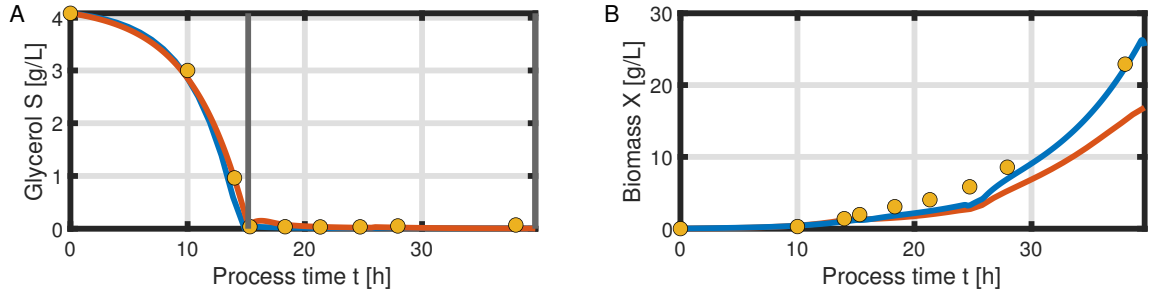

Figure 4.3: Comparison of the simulated quantities using the unsegregated model (blue) and the PBE model (orange) against the experimental data (yellow points) of the L-phenylalanine production process up to $t = 39\ h$. Time course of the following concentrations: glycerol S (A) and biomass X (B)

given by $n \in [4 \times 10^{-4}\ \text{fmol}, 2.2 \times 10^{-3}\ \text{fmol}]$ and $m \in [100\ \text{fg}, 1000\ \text{fg}]$ and the domain is divided into a 15×20 grid. For the initial population distribution, we assume every cell to have the same cell mass $m_0 = 250$ fg and the protein content of the cells to be normally distributed with $\mu = 1.2 \times 10^{-4}$ fmol and $\sigma = 5 \times 10^{-5}$ fmol. The total number of cells is set to correspond to the initial biomass concentration of the bioprocess and with Eq. (2.36) given by

$$X_0 = \frac{\mathcal{M}_{0,1}(0)}{V_R(0)} = \frac{1}{V_R(0)} \int_{m_{min}}^{m_{max}} \int_{n_{min}}^{n_{max}} m\ N(n, m, 0)\ \mathrm{d}n\ \mathrm{d}m \tag{4.11}$$

The simulation of the PBE model is compared with the simulation of the coarse-grained model from Section 3.7, which is a unsegregated model. The time course of the substrate concentration S shows no big differences in both approaches in Figure 4.3A. The biomass yield during the fed-batch phase in the PBE model is not as high as in the unsegregated model as seen in Figure 4.3B. The dynamics for batch phase, on the other hand, do not differ. A look at the internal properties of the cell population in Figure 4.4A reveal a higher mean protein concentration P during the fed-batch phase and therefore a lower concentration for M, leading to a small difference in the allocation into protein fractions T and R in Figure 4.4C. The reduced biomass yield is also apparent in the growth rate, which reveals lower values for the PBE model in Figure 4.5A. This

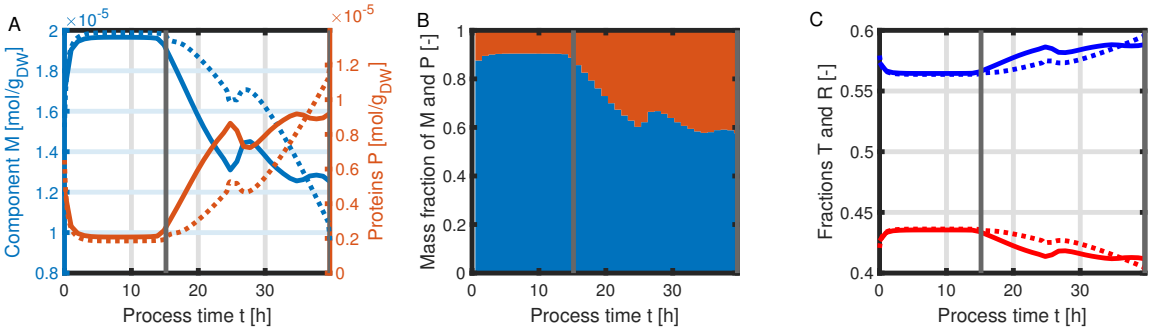

Figure 4.4: Time course of intracellular concentrations: compartment M (blue) and proteins P (orange) (A). Mass fractions over the course of the process: compartment M (blue) and proteins P (orange) (B). Mass fractions of T (blue) and R (red) for the PBE model (solid line) and unsegregated model (dashed line) (C).

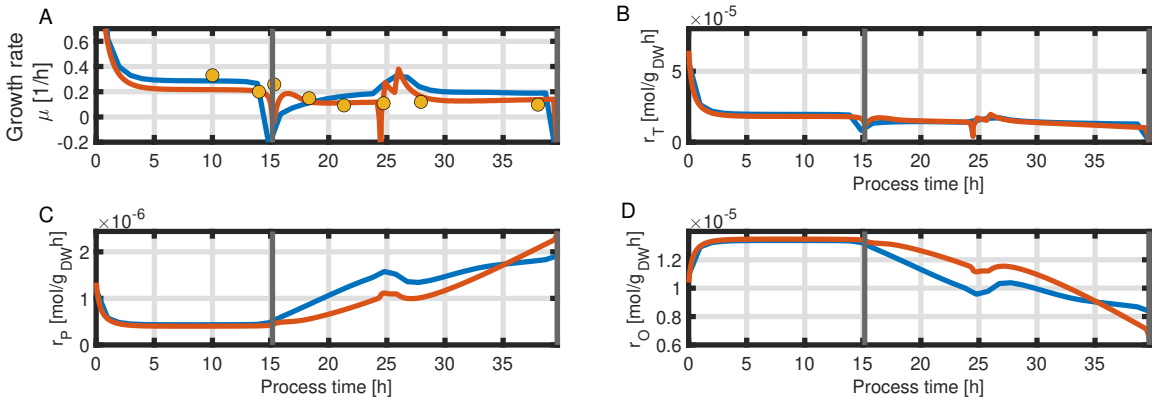

Figure 4.5: Time course of the simulated specific growth rate μ (A), substrate transport rate r_T (B), protein synthesis rate r_P (C) and overflow metabolism rate r_O (D) for the PBE model (orange) compared to the unsegregated model (blue). The experimental data points are indicated by yellow dots.

is a result of a slightly lower substrate uptake rate, lower protein synthesis rate and a higher overflow metabolism over the majority of the fed-batch phase seen in Figure 4.5B to D.

With the approach of a segregated model, we can investigate the distribution of the population over its internal properties n and m. Figure 4.6 shows the number density function over the state space for different time points: the beginning of the process ($t = 0$ h), the middle of the

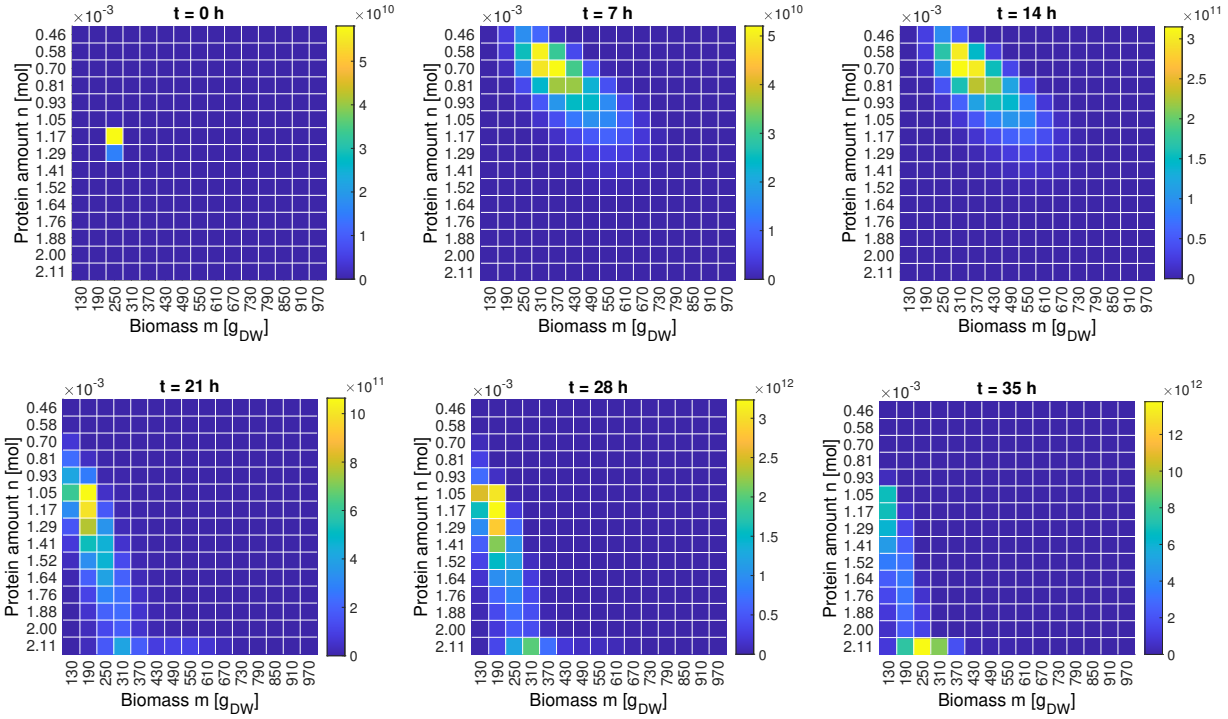

Figure 4.6: Number density function over the n-m-plane for different time points of the process.

batch phase ($t = 7$ h), the end of the batch phase ($t = 14$ h),the middle of the first fed-batch phase ($t = 21$ h), a point in the middle of the second fed-batch phase ($t = 28$ h) and a point near to the end of the fed-batch phase ($t = 35$ h). The distribution during the batch phase at $t = 7$ h and $t = 14$ h shows a peak at around $n \approx 0.6 \times 10^{-3}$ fmol and $m \approx 330$ fg. As the mean protein concentration P increases for the fed-batch phase at $t = 21$ h, the distribution moves to higher protein content values at around $n \approx 1.1 \times 10^{-3}$ fmol, but lower biomass at $m \approx 190$ fg. A small subpopulation forms at the upper boundary of the protein content domain. This subpopulation continues to increase as seen for time points $t = 28$ h and $t = 35$ h.

A distribution over the protein concentration is obtained by Eq. (4.1) and sorting the resulting values in ascending order. Figure 4.7A shows the normalized protein concentration distributions for different time points, such that the bars add up to 1. These time points coincide with the beginning of the process ($t = 0$ h), a time point at the end of the batch phase ($t = 14$ h), the middle of the first fed-batch phase ($t = 21$ h) and a time point near to the end of the fed-batch phase ($t = 35$ h). It is worth noting that sorting the protein concentrations corresponding to the mid grid cell points does not result in a protein concentration domain with evenly spaced values. This leads to a highly noisy protein concentration distribution. To deal with this problem, the resulting distribution is binned into evenly spaced intervals. Since the population distribution does not change much over the batch phase, only one time point ($t = 14$ h) is shown here. The same applies for the fed-batch phase. As seen in the mean values of the protein concentration in Figure 4.4, the protein concentration distribution exhibits smaller values during the batch phase (orange) and shifts to higher protein concentration values in the fed-batch phase (yellow). Additionally, the distribution broadens over the course of the fed-batch phase, which can also be observed when calculating the variance of the protein concentration in Figure 4.4B. The variance shows a peak at $t \approx 25$ h, which coincides with the switch to the second fed-batch phase with a higher substrate concentration feed. After the switch the variance drops, but increases again

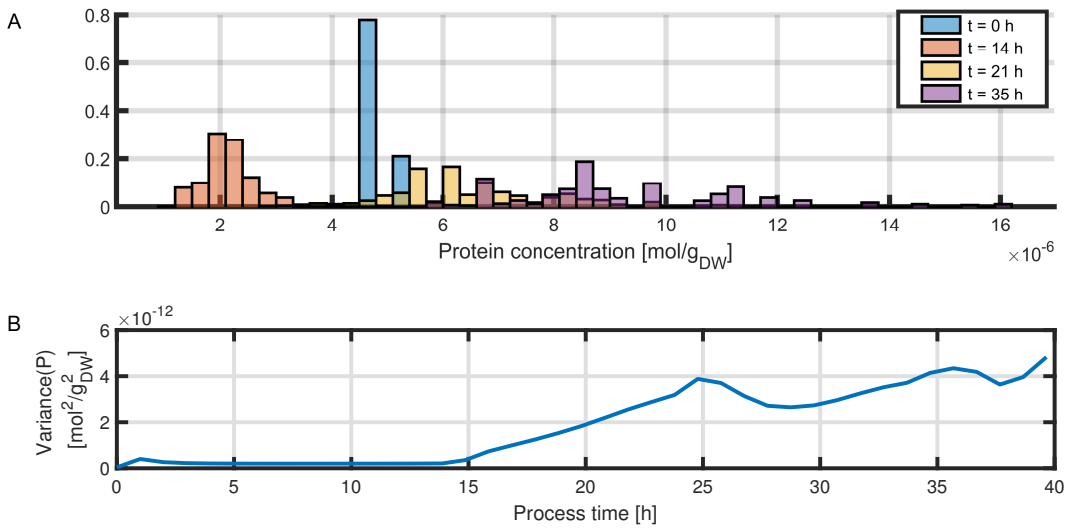

Figure 4.7: Evolution of the protein concentration distribution: Protein concentration distribution normalized to 1 for different time points of the bioprocess: the beginning of the process ($t = 0$ h), a time point at the end of the batch phase ($t = 14$ h), the middle of the first fed-batch phase ($t = 21$ h) and a time point near to the end of the fed-batch phase ($t = 35$ h) (A), Variance of protein concentration over the course of the bioprocess (B).

until $t \approx 35$ h, where it drops on the short-term. During these peaks, a significant part of the population forms at the boundary as seen in Figure 4.6 at $t = 35$ h.

4.2.1 Influence of cell division rate function

The critical values m_c and n_c have an impact on the simulation, as they determine the values at which the cells are more likely to divide. Therefore, the resulting protein concentration distributions are compared for different sets of critical values: base case with $m_c = 500$ fg and $n_c = 1.3 \times 10^{-3}$ fmol (blue), higher critical values $m_c = 800$ fg and $n_c = 1.7 \times 10^{-3}$ fmol (orange) and lower critical values $m_c = 400$ fg and $n_c = 1.3 \times 10^{-3}$ fmol (yellow). The different critical values have a large effect on the resulting protein distribution especially during the fed-batch phase, as seen in Figure 4.8. Two time points during the fed-batch phase ($t = 21$ h and $t = 35$ h) are shown in Figure 4.8A and B, which reveal a narrower distribution for higher critical values and broader distribution for lower critical values. This observation is confirmed when calculating the variance of the protein concentration in Figure 4.8D. These results also apply for the batch phase. The protein distributions during the batch phase are not shown here, but can be found in Appendix, Figure A.3. The mean protein concentration shifts to lower values for higher critical values and higher values for the other case (Figure 4.8C). This relation is not obvious and requires a deeper investigation. During the fed-batch phase a part of the population exhibits internal property values at the boundary of the state space, which effects the mean value of the population and the process of cell division dominates the dynamics of the population. Therefore, the question arises if the differences between the simulations arise mainly from the subpopulation at boundaries of the state space. Thus, the influence of the different critical cell division values is investigated by analyzing the behavior of an exponentially growing population in a constant environment. Here, the differences between the different protein distributions are negligibly small, which is shown in Appendix, Figure A.3. The population shifts on a diagonal for higher critical values, but the mean protein concentration remains the same. This implies that the

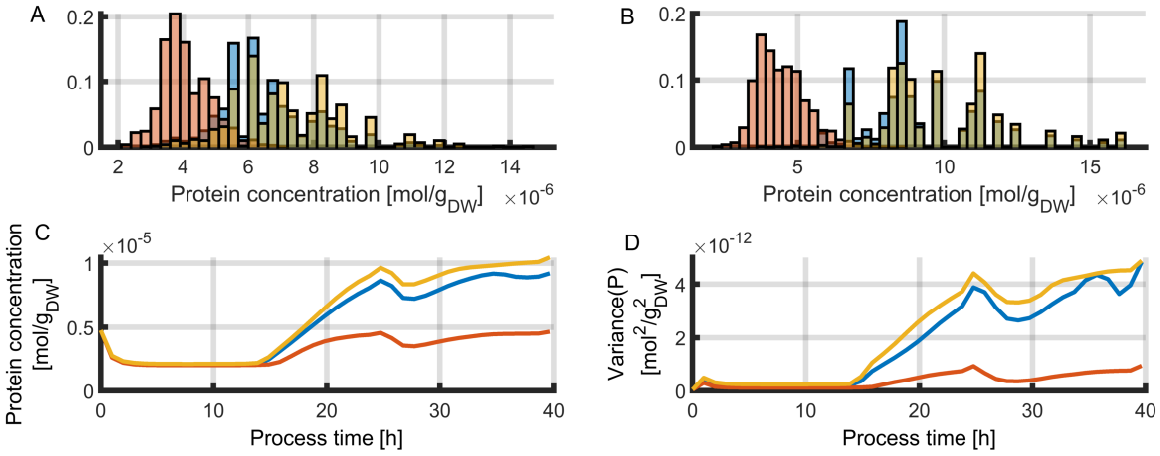

Figure 4.8: Comparison of the protein distribution for division rate functions with different critical values m_c and n_c: base case with $m_c = 500$ fg and $n_c = 1.3 \times 10^{-3}$ fmol (blue), higher critical values $m_c = 800$ fg and $n_c = 1.7 \times 10^{-3}$ fmol (orange) and lower critical values $m_c = 400$ fg and $n_c = 1.3 \times 10^{-3}$ fmol. Protein distribution at time point $t = 21$ h (A) and $t = 35$ h. Mean protein concentration (C) and variance of the protein concentration (D) over the course of the bioprocess.

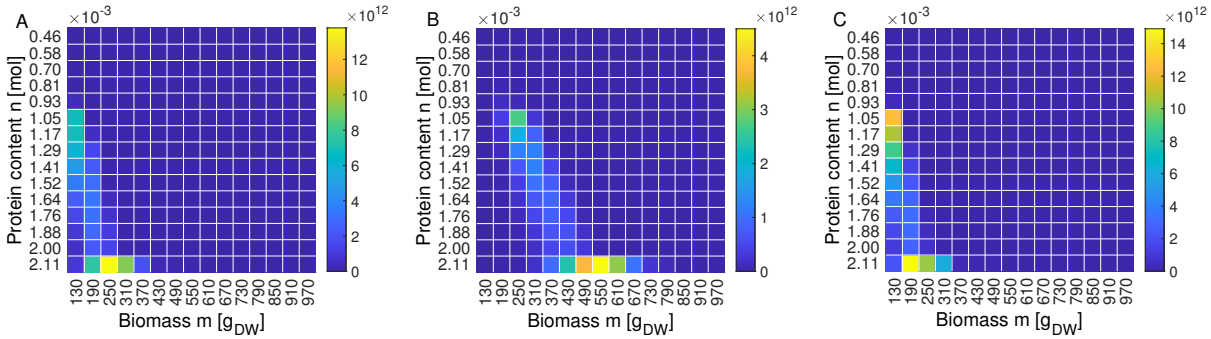

Figure 4.9: Comparison of the number density function over the n-m-plane for division rate functions with different critical values m_c and n_c at time point $t = 35$ h: base case with $m_c = 500$ fg and $n_c = 1.3 \times 10^{-3}$ fmol (A), higher critical values $m_c = 800$ fg and $n_c = 1.7 \times 10^{-3}$ fmol (B) and lower critical values $m_c = 400$ fg and $n_c = 1.3 \times 10^{-3}$ fmol (C).

critical cell division values affect the solution of the distribution only if part of the population shows values at the boundary of n. Figure 4.9 shows the number density functions over the n-m-plane for the time point $t = 35$ h for the three different cell division functions, where all three cases show the mentioned subpopulation. For higher m_c this population is shifted to the right as seen in Figure 4.9B, while it is shifted to the left for lower m_c. We can observe that the subpopulation for higher critical values forms the majority of the cell population and explains the smaller variance in protein concentration in Figure 4.8D. The reasoning behind this is that the cells are only able to grow in m direction, if they have reached the maximal value in protein content n, and since m_c is higher, these cells are able to increase their biomass for a longer time, resulting in more cells at the boundary of n. With increasing biomass the protein concentration of the cells decreases as well, explaining the observation in Figure 4.8C. If the critical value m_c is smaller, the cells are more likely to divide when reaching the boundary, resulting in a more evenly split population as seen in Figure 4.9C. The influence of the cell division rate on the dynamics of the cell population also affects the biomass yield as seen in Figure 4.10. Lower protein concentrations lead do a larger pool of cell components M, an increased overflow metabolism and therefore smaller growth rate of the population.

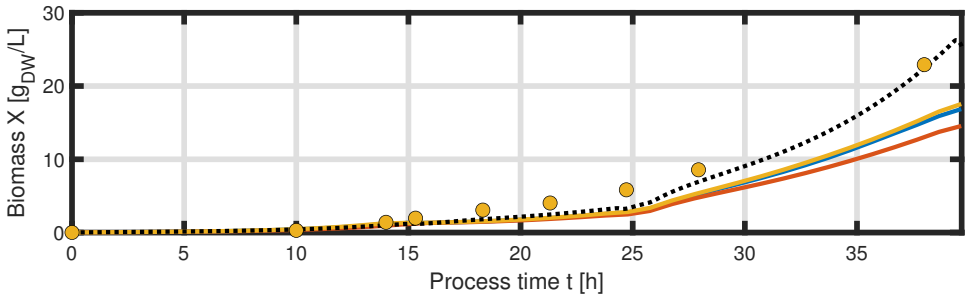

Figure 4.10: Comparison of the dynamics of the biomass for division rate functions with different critical values m_c and n_c: base case with $m_c = 500$ fg and $n_c = 1.3 \times 10^{-3}$ fmol (blue), higher critical values $m_c = 800$ fg and $n_c = 1.7 \times 10^{-3}$ fmol (orange), lower critical values $m_c = 400$ fg and $n_c = 1.3 \times 10^{-3}$ fmol and unsegregated model (dotted line). The experimental data points are indicated by yellow dots.

4.2.2 Influence of unequal cell division

For a one-dimensional PBE it has been shown that cell division into two equally sized daughter cells creates periodicity in the number density function, which vanishes when considering unequal partitioning [33, 101]. In case of unequal partitioning, the partition probability density is given by

$$p(n,m,n',m') = \frac{1}{B(q,q)^2}\frac{1}{n'}\left(\frac{n}{n'}\right)^{q-1}\left(1-\frac{n}{n'}\right)^{q-1}\frac{1}{m'}\left(\frac{m}{m'}\right)^{q-1}\left(1-\frac{m}{m'}\right)^{q-1}, \tag{4.12}$$

which is a symmetrical beta distribution B with parameter q, which is set to 20 in the simulation. This way the cells divide into two daughter cells, where its new internal properties are distributed according to the beta distribution, rather than producing two identical cells. In order to focus on the effect of the partitioning function, the number density functions of an exponentially growing population in a constant environment is considered. It is worth noting, that in the presented 2D PBE, the number density function does not show periodicity unlike the 1D PBE [101]. Comparing the number density function over the n-m-plane in Figure 4.11A, which shows equal partitioning, to the result in Figure 4.11B, reveals a slightly broader distribution for unequal partitioning. This is reasonable as unequal partitioning creates more variety in the newly formed cells. Interestingly, the protein distributions in Figure 4.11C are almost identical with a very slight shift towards smaller protein concentrations in case of unequal partitioning. Note that due to the set up of the discrete grid space, empty bins in the protein concentration space may occur as apparent in Figure 4.11C.

In the last section, we observed a higher dependency of the dynamics on the division process during the fed-batch phase. Therefore, we investigate the affect of unequal partitioning in this scenario. Figure 4.12A reveals a broader cell population distribution over the n-m-plane analogous to cell population in a constant environment in Figure 4.11. An additional subpopulation forms at the upper-left corner, which corresponds to the daughter cells of the middle cell population. A closer look at the protein concentration distribution in Figure 4.12B shows a slight shift towards lower concentration values. Contrary to the distribution over n-m, the distributions is narrower leading to lower variance values as seen in 4.12C. As the population exhibits

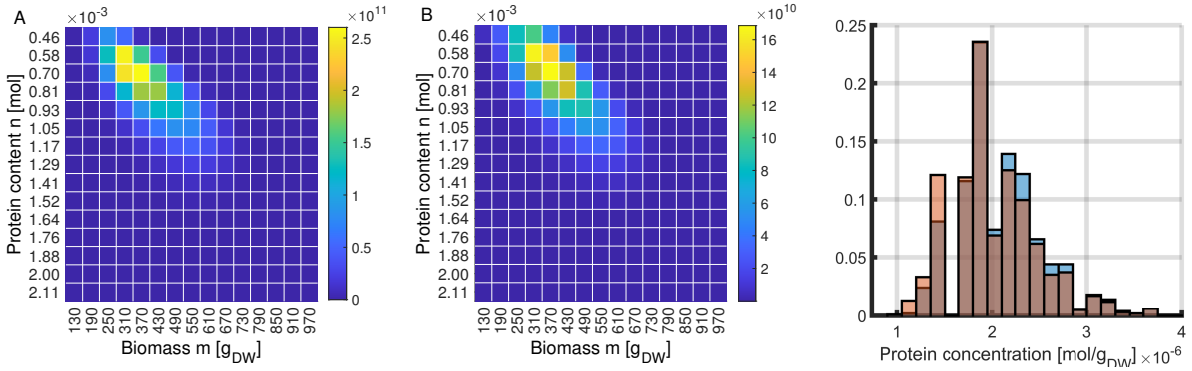

Figure 4.11: Comparison of equal and unequal partitioning: Number density function over the n-m-plane for equal partitioning (A) and unequal partitioning (B). Comparison of the protein concentration distribution, where the distribution for equal partitioning is indicated in blue and unequal partitioning is indicated in orange (C).

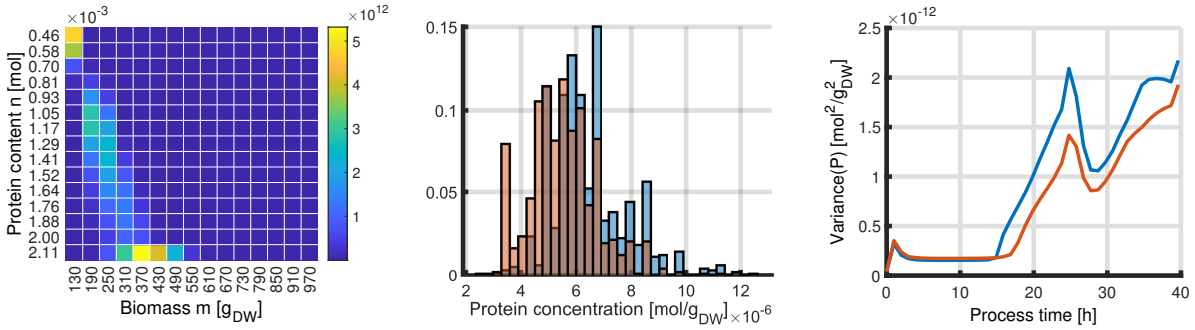

Figure 4.12: Comparison of equal and unequal partitioning for the whole biomass production phase: Number density function over the n-m-plane for unequal partitioning at $t = 35$ h (A). Comparison of the protein concentration distribution at $t = 35$ h (B) and Variance of protein content (C), where the distribution for equal partitioning is indicated in blue and unequal partitioning is indicated in orange.

values at the boundary of the state space, it can not be excluded that this affects the broadness of the protein concentration distribution. Overall, we can conclude that unequal partitioning leads to a broader distribution over the internal variables, but does not change the dynamics of the population.

4.3 Validation of conservation of the first moments

In this section, we check the numerical scheme for conservation of the first moments. First, the conservation of the first moments is investigated for the growth process. The first moment in protein content is given by $\mathcal{M}_{10}$ and its dynamics is obtained by multiplying the PBE given by Eq. (4.4) by n and integrating over the whole domain

$$\frac{\mathrm{d}\mathcal{M}_{10}}{\mathrm{d}t} = \int\int n \, \frac{\partial N(n,m,t)}{\partial t} \, \mathrm{d}m \, \mathrm{d}n = \int\int r_n(n,m,S) N(n,m,t) \, \mathrm{d}m \, \mathrm{d}n. \tag{4.13}$$

The right hand-side of the PBE is zero as cell division only changes the number of cells and does not change the total sum of protein content across the population. The step-wise increase rate in total protein content calculated using the first moment in protein content

$$\mu_{n,step}(t) = \frac{1}{\mathcal{M}_{10}(t)} \frac{\mathcal{M}_{10}(t+\Delta t) - M_{10}(t-\Delta t)}{2\Delta t}, \tag{4.14}$$

which is obtained through a central scheme, can be compared to the rate of change obtained in Eq. (4.13)

$$\mu_n(t) = \frac{\int\int r_n(n,m,S) N(n,m,t) \, \mathrm{d}m \, \mathrm{d}n}{\mathcal{M}_{10}} \tag{4.15}$$

Analogously, the growth rate in biomass can be compared for consistency. Figure 4.13 shows the increase rates of the two first moments using Eqs. (4.14) and (4.15). The simulation is performed for an exponentially growing cell population in a constant environment with $S = 1$ g/L using a 15×15 and a 20×20 grid. The increase rates concerning the protein content n are almost

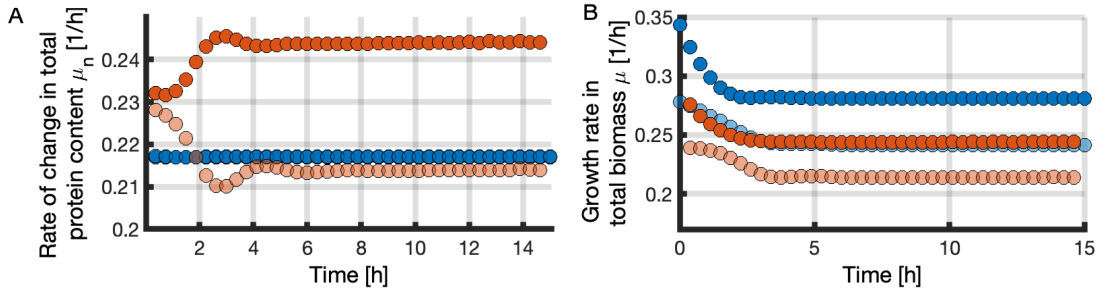

Figure 4.13: Comparison of rates of change of the first moments $\mathcal{M}_{10}$ (A) and $\mathcal{M}_{01}$ (B), where stepwise rates of change calculated using the moments are indicated by red circles and rate of change rates obtained using r_n and r_m by blue circles. The transparent circles are obtained using a 20×20 grid, while the solid circles are from a 15×15 grid

aligned (Figure 4.13A), while a larger difference is observed for the growth rate in biomass (Figure 4.13B). The difference between the rates of change can be decreased if a finer grid is employed. Additionally, it has to be noted that a second order central scheme is used to estimate the growth rate in Eq. (4.14), which can be exchanged for a higher order method. Overall, we can conclude that the first moments are conserved considering the growth dynamics of the cell population.

In the next step, we check for mass conversation during the cell division process. For this, we calculate the right hand-side after multiplying by n, and or m respectively, and integrating over the domain. For conservation of the sum of the first moments, we calculate

$$\int\int (n \ + \ m)(\text{birth} - \text{death}) \ \mathrm{d}m \ \mathrm{d}n = \mathcal{O}(10^{-1}) \tag{4.16}$$

which is comparable small to the sum of the first moments which have $\mathcal{O}(10^{14})$, thus the division process conserves sum of the two first moments. For the conservation of the total protein content, we obtain

$$\int\int n(\text{birth} - \text{death}) \ \mathrm{d}m \ \mathrm{d}n = \mathcal{O}(10^{8}) \tag{4.17}$$

while $M_{10} = \mathcal{O}(10^{10})$ and for the total biomass

$$\int\int n(\text{birth} - \text{death}) \ \mathrm{d}m \ \mathrm{d}n = \mathcal{O}(10^{8}) \tag{4.18}$$

while $M_{10} = \mathcal{O}(10^{15})$, thus we do have some numerical error, but it is negligible compared to the order of magnitude on the left hand-side. Overall, it has been shown that the numerical scheme preserves the first moments and the numerical errors are sufficiently small.

4.4 Short summary and discussion

Past approaches to model microbial growth using PBEs either considered only the cell mass as variable or multiple variables for the intracellular components of a cell. While the former is not sufficient to fully capture a populations growth and its cell states, the latter is to complex and

no numerical solution is attempted if the number of variables exceed two [156, 41] or extensive computational effort is required [67]. Other cases with several state variables did not consider cell division, but incorporated stochastic intracellular behavior. Here, the PBE model is solved numerically through a transformation into a Fokker-Planck type differential equation using an Euler algorithm [145] or stochastic algorithm [146, 147].

In this chapter, a compromise between simplification and necessary complexity has been aimed for and a bivariate PBE with the cell mass m and protein content n of a cell as state variables presented. With knowledge about the two variables, the protein concentration P in the cells can be determined as well as the concentration of compartment M since the biomass only comprise these two intracellular components. The dynamics of the components in the cell follow the coarse-grained model formulation given in Section 3.7. The presented PBE describe the evolution of the state variables and cell division in the environment of a well-mixed homogeneous bioreactor. Thus, in the current set-up intrinsic noise due to cell division is incorporated and external noise neglected. The biomass production phase of the L-phenylalanine production process has been simulated using the PBE model with a weighted Finite Volume method. While the mean values of the segregated and unsegregated model do not differ during the batch phase, the results vary significantly during the fed-batch phase leading to a lower biomass to substrate yield for the PBE model. The distribution over the state variables were accessed for different time points of the process, showing a normally distributed population during the batch phase and a split population in the n-m-state space during the fed-batch phase.

In the following, we want to discuss this behavior during the fed-batch phase. The two emerging distributions lie on a line in the n-m-plane. Thus, exhibiting roughly the same protein concentration. During the fed-batch phase, the protein concentration shifts to higher values, which are located beneath the diagonal of the state space. The higher the value the closer it is to the lower left corner of the state space. The boundaries of the state space are fixed by the maximum value of P and cannot be freely set. If one would allow higher values for n, protein concentration values that are larger than possible would result. Furthermore, a look at the range of n reveals that a cell with $n_{\max}$ can only divide two times. The larger subpopulation exhibits a protein content value of $n \approx 1.1 \times 10^{-3}$ mol and, given the upper boundary of n, cells of this subpopulation are able to grow to double their size, while keeping their protein concentration constant, forming the second subpopulation. Cells reaching the boundary of n are then only able to change their state in direction of m, leading to a decreased protein concentration, an overall mean protein concentration and higher variance of the whole population. This is a downside of the current set-up of the PBE as one normally sets the upper boundaries of the state variables to a high value which is usually not reached as cells tend to divide earlier. Thus, the emergence of the subpopulations and the different mean values during the fed-batch phase can be attributed to the upper-boundary imposed on the protein content n. This also explains the normally distributed population during the batch phase. Here, the protein concentration P is smaller and the population resides on the upper left corner of the state space. Therefore, cells tend to divide before they reach the boundaries of the state space. The time point at which cells divide is determined by the critical values in the division function and different sets of critical values were simulated. While the effects during the batch phase were negligible, larger differences were observed during the fed-batch as anticipated. Higher critical values led to a higher accumulations of cells at the boundary with higher mass, while smaller critical values forced the cells to divide earlier, but not enough prevent the formation of a subpopulation at the boundary. The incorporation of unequal partitioning, on the other hand, only increased the variance of the population, but left the overall dynamics unchanged.

The numerical scheme used was examined for conversation of the first moments and consistency resulting in sufficiently small numerical errors. Overall, we can conclude that the presented model is able to describe a segregated population in a well-mixed environment and produces good results in a nutrient rich environment as seen in the simulation of the batch phase. The results during the fed-batch phase have to be considered with caution due to the limitations of the state space of the PBE.

5 Analysis of a cell population exposed to external and intrinsic noise

As the employment of PBEs poses a big challenge to describe a segregated population with multiple properties, we utilize an Euler-Lagrange approach in this chapter. The advantages of these models were already mentioned in Section 2.5. Here, we present an Euler-Lagrange CM for a heterogeneous cell population in a scaled-down bioreactor system to study the effects of external and intrinsic noise. For this, a particle tracking algorithm, introduced in Section 2.5.1, is used to describe the movement of virtual particles within the bioreactor system. Following a large number of virtual particles allows an accurate computation of the local consumption and production terms, due to biological activity, in the mass balance equations. Also, a dynamic model describing the cellular response to external stimuli, which was presented in Chapter 3, is attached to each virtual particle which enables following cell adaptation along its trajectory inside the bioreactor system. Hence, not only the global heterogeneity at the reactor scale can be accessed (as shown in [51]), but also the local distributions of cell physiological properties at different locations of the bioreactor system. Moreover, the introduction of division processes constitutes a major originality of the present work. As randomization of cell properties with a frequency relates to the cell growth rate, cell division produces individuals whose metabolism are not balanced. Various examples of scale-down bioreactor studies are described revealing that the magnitude and frequency of external perturbations combined with randomization of properties at birth favors the emergence of subpopulations.

5.1 Model setup of an Euler-Lagrange approach

The scaled down bioreactor system in question consists of a well-mixed ideal STR and PFR in bypass. The STR is assumed to be homogeneous. For the fed-batch process, the feed can be assumed to be in the STR as well as the PFR. The coupling of the two bioreactors splits the environment into a well-mixed zone (STR) and a zone which exhibits substrate gradients, leading to possible starvation zones or zones with excess substrate availability. These two scenarios depend on the feed input location, where a feed in the STR creates the former and a feed in the PFR the latter [114]. This is due to the fact that the volume of the STR is significantly larger than the inlet of the PFR, leading to a quick dispersion of the feed in the STR and higher dilution. Thus, the substrate is quickly consumed, leaving no substrate to be transported into the PFR. For a feed at the inlet of the PFR, the substrate concentration is much higher in the PFR due to the small volume.

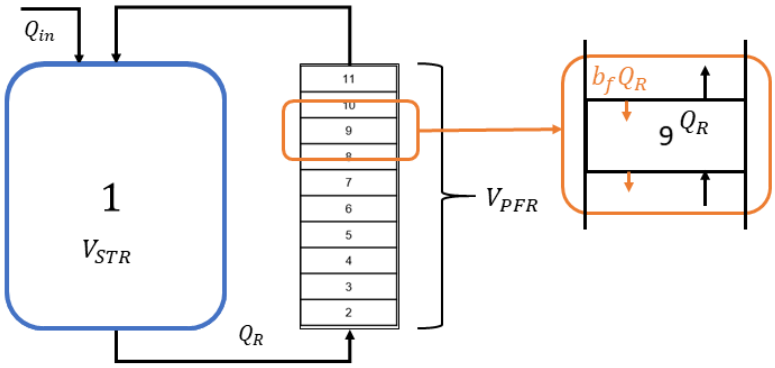

Figure 5.1: Compartmentalization of a scaled-down bioreactor system with a stirred tank bioreactor and plug flow reactor in bypass ($n_z = 11$).

5.1.1 Hydrodynamic model of the scaled-down bioreactor system

As presented in Section 2.4, we opt to employ a CM to describe the hydrodynamics of the scaled-down bioreactor system. Thus, we compartmentalize the environment of the bioreactor system into a subset of n_z volumes. Each compartment is considered to be homogeneous. The STR is assumed to be ideal and is represented by the first compartment with volume V_{STR}. The PFR is divided into n_{PFR} volumes, which we set to $n_{\mathrm{PFR}} = 10$ in the following, with a total volume V_{PFR} as seen in Figure 5.1. Thus, in this setup each compartment has exactly two neighbors. The working volume of the STR is $V_{\mathrm{STR}} = 0.9$ L and the volume of the PFR is $V_{\mathrm{PFR}} = 0.45$ L. Therefore, the ratio between the two environments is $2:1$. The liquid phase moves through the PFR with flow rate Q_R and backflow of $b_f Q_R$. These parameters of the hydrodynamic model can be varied to modify the degree of mixing in the bioreactor system, the frequency of feast and starvation events, as well as their magnitude. The transport of dissolved scalars in the bioreactor system is described using the finite volume approach presented in Section 2.4 by Eq. (2.57), while the mass transfer between the liquid an biological phase is described in Section 2.5.2 by Eq. (2.65). The environment in the bioreactor system is characterized by only one variable: the glycerol concentration S.

5.1.2 Dynamical evolution of the particles

The biological phase is represented by N discrete virtual particles, which are coupled with a state vector describing its intracellular state as shortly depicted in Section 2.5. In this work, each virtual particle represents an ensemble of biological cells, which possess the same set of physiological properties ξ and a total biomass m. Thus, it is worth noting that we track the total cell mass of the virtual particle, but do not specify the exact number of biological cells a virtual particle represents. The total biomass m can be regarded as an additional physical property of the virtual particle besides the physiological properties ξ. Furthermore, as the mass transfer between liquid and biological phase given through Eq. (2.65) sums over the number of particles present in a given compartment at time t, each particle is assigned a physical property x, which

indicates the index i of the compartment the particle resides in. Thus, the state vector of each particle is given by (ξ, m, x). Over time, the properties of the particles change as they move through the bioreactor system and equations for the evolution of the properties have to be proposed. The movement in x is characterized by a stochastic transport algorithm, which was introduced in Section 2.5.1. In contrast to the PBE model in Chapter 4, we are not bound to at best only two variables, as the computational effort to solve the PBE increases exponentially with the number of variables. In this framework of an Euler-Lagrange model approach, the dynamics of the internal properties of the particles can be described by simple ODEs. Additionally, transport of particles in the spatial space and dynamics of the internal properties are handled separately. At each time step, the virtual particle is first updated in its internal property and then moved in space. This creates the opportunity to employ more detailed descriptions for the particles forming the biological phase. Hence for the evolution in the physiological properties m and ξ, we can utilize the equations of the coarse-grained model (3.20)-(3.22), which are described in Section 3.4.1. The advantage of the coarse-grained model is that it is able to describe the intracellular network of a cell with only 6 reactions and 3 variables, while still containing the most important cellular processes. The variables are the three intracellular components: P and M, with the first representing the complete proteome, the second the residual biomass and both being macromolecular; and M representing a metabolite, which has low molecular weight. Using Eqs. (3.20)-(3.22) the physiological properties of each particle p are $\xi = (P, M, U)$ and their evolution is described by

$$\begin{aligned} \dot{\xi} &= N\ r(\xi, S)\ -\ \mu\ \xi \\ \dot{m} &= \mu\ m \\ \mu &= w\ N\ r(\xi, S). \end{aligned} \tag{5.1}$$

The vector r contains all intracellular reactions and mass transfer rates for those scalars present in the environment and is not to be mistaken with $\phi_l(\xi_p, \mathcal{S}_i)$, which only contains the latter rates. As we only consider one scalar in the environment S, the mass transfer rate of S is denoted as ϕ_S. For the mass transfer rate ϕ_S, we additionally used the rate limiting concept introduced in Section 3.5. As the cells move in the scaled-down bioreactor system, they encounter situations in which the substrate uptake is limited by either external or internal factors. The characteristic time for substrate uptake given by Eq. (3.36) varies over the different compartments and can be calculated for each compartment as

$$\tau_u = \frac{S}{w_S\ X_i\ k_T\ T_{\max}}, \tag{5.2}$$

where the total biomass X_i in each compartment i is calculated through summing over all $N_{p,i}$ particles present in compartment i

$$X_i = \frac{1}{V_i} \sum_{p}^{N_{p,i}} m_p \qquad p \in \{1, .., N_{p,i}\}. \tag{5.3}$$

In the scope of this work, we use a rather large parameter value for τ_{mix} in order to avoid negative substrate concentration values in compartments with very low substrate concentrations as well as keep the computational time reasonable. We have already shown in Section 3.5 and 3.6.1 that the two models behave similarly justifying the choice for a higher value of τ_{mix}.

5.1.3 Intrinsic noise

In the current setup, there is no distinction between growth in cell mass or growth in cell number. The virtual particles represent a large ensemble of individual cells and the growth in mass and growth in number is equivalent being described by the growth rate μ. Through cell division, however, the cellular content of a cell is randomly distributed to the daughter cells leading to noise in the cellular composition. Thus, cell division creates unbalanced cells, a concept we introduced in Section 3.6.3. As a short recap, unbalanced cells require a longer time to return to their balanced state, especially in terms of P, than balanced cells need to adapt to a change in the environment. In order to account for this phenomenon and still keep the number of virtual particles constant, it is proposed to change the physiological properties ξ of a number of virtual particles, which equals the number of cell divisions during a time interval Δt. For this purpose, the total number of cell divisions have to be calculated and the same amount of virtual particles with corresponding mass receive a new set of physiological properties ξ.

In the model, the virtual particles do not represent the same number of cells. The total biomass of the particle and its growth rate has to be taken into account in order to determine the total number of cell divisions during a time step Δt. On average, the relative change in cell number N is proportional to the relative change in mass of the virtual particles

$$\frac{\Delta N}{N} = \frac{\sum_p \mu_p \; m_p \; \Delta t}{\sum_p m_p}. \tag{5.4}$$

A more detailed derivation of Eq. (5.4) can be found in Appendix, Section A.3.1. Now that we obtained the number of cell divisions, we have to decide which virtual particles are chosen to be assigned new physiological properties. This can be done by introducing weights d for the random sampling of the cells. In later Section 5.3, we investigate the outcome of different weighting functions. One consideration when setting up a weighting function is to take the total biomass of the particle into account

$$d_j = \frac{m_j}{\sum_p m_k} \qquad j \in \{1, ..., N_p\}, \tag{5.5}$$

where $\sum_j d_j = 1$. Note that here we use the subscript p in order to emphasize the summation over all particles p. Another possibility is to consider the growth rates as particles with higher growth rates are more likely to contain cells that divide than slower growing particles leading to following weights

$$d_j = \frac{\mu_j}{\sum_p \mu_p} \qquad j \in \{1, ..., N_p\}, \tag{5.6}$$

These two considerations can be combined leading to

$$d_j = \frac{\mu_j \; m_j}{\sum_p \mu_p \; m_k} \qquad j \in \{1, ..., N_p\}. \tag{5.7}$$

When considering the growth rate to determine the sampling weights, there is the possibility of negative weights as the growth rate can be negative. In order to ensure non-negativity, we set the sampling weights of particles with negative growth rate to zero. This is also biologically sound, as particles with negative growth do not divide.

The new properties of the selected particles are sampled from a normal distribution centered around the old properties (M_0, P_0), which can be regarded as the properties of the mother cells, with a standard deviation of $\sigma_M = 0.1\ M_0$ and $\sigma_P = 0.05\ P_0$. The remaining property U is uniquely defined once M and P are chosen and has to satisfy

$$w_M\ M\ +\ w_P\ P\ +\ w_U\ U\ =\ 1. \tag{5.8}$$

It can be anticipated that cell division is a source of heterogeneity that will affect the physiological state of the particles of the population even in absence of external perturbations. Generally, both sources of noise are present and are investigated in this chapter.

5.2 Contribution of external noise on population heterogeneity

In this section, intrinsic noise, as described in the previous Section 5.1.3, is not considered in order to focus solely on extrinsic sources of population heterogeneity. With the presented framework, simulations of a fed-batch process in the scaled-down bioreactor system with $N\ \approxeq\ 10^4$ particles have been performed using MATLAB 2022a. The investigated process follows the same structure as the process investigated in Section 3.4.1 with the only difference being the bioreactor setup and the focus on the biomass production phase. Parameter values used can be found in Appendix, Section A.3.2. Additionally, the flow rate is set to $Q_R = 1$ L/h, back flow in the PFR is left aside and the time step used is $\Delta t = \min(\tau_r)/10$. The mean population growth and reaction rates, over all particles present in the bioreactor system are calculated using particle mass weighted averages

$$\tilde{v} = \frac{\Sigma_p m_p\ v_p}{\Sigma_p\ m_p} \quad v \in \{\xi, \mu, r\}. \tag{5.9}$$

In order to investigate the effects of a heterogeneous environment present in the scaled-down bioreactor system, the simulation of a cell population in such a environment with a feed in the STR is compared with the simulation of a single well-mixed STR as seen in Section 3.4.1 [115]. As mentioned earlier, the PFR creates a starvation zone in this operating condition. The difference between these two operating conditions manifest only during the fed-batch phase, as the environment over all compartments during the batch phase is nearly homogeneous and only the addition of the feed into the STR creates a substrate gradient over the compartments. This is confirmed by comparing the simulated mean quantities in Figure 5.2.

Thus, we focus on the fed-batch phase. The mean substrate consumption of the cell population in both setups does not differ (Figure 5.2A), but the biomass yield is lower in the scaled-down bioreactor (Figure 5.2B) as the cell population grows with lower specific growth rate μ compared to the population in a homogeneous environment (Figure 5.2C). The lower growth rate stems from a higher mean substrate uptake rate r_T, higher mean respiration rate r_C, but lower mean protein synthesis rate r_P. The synthesis rate of the residual biomass r_U and overflow metabolism rate r_O are not depicted in Figure 5.2 since they have the same dynamics as the protein synthesis rate r_P and respiration rate r_C, respectively. Basically, this reflects that a part of the cell population resides in a starvation zone and additionally can not optimally adapt its intracellular content due to the fluctuating environment. Hence, a segregated approach combined with a dynamic model which uncouples growth and local substrate uptake is essential to predict the well-known result that concentration gradients reduce substrate to biomass yield and growth [38, 97]. The novelty compared to the previous work of Morchain et al. [108] is

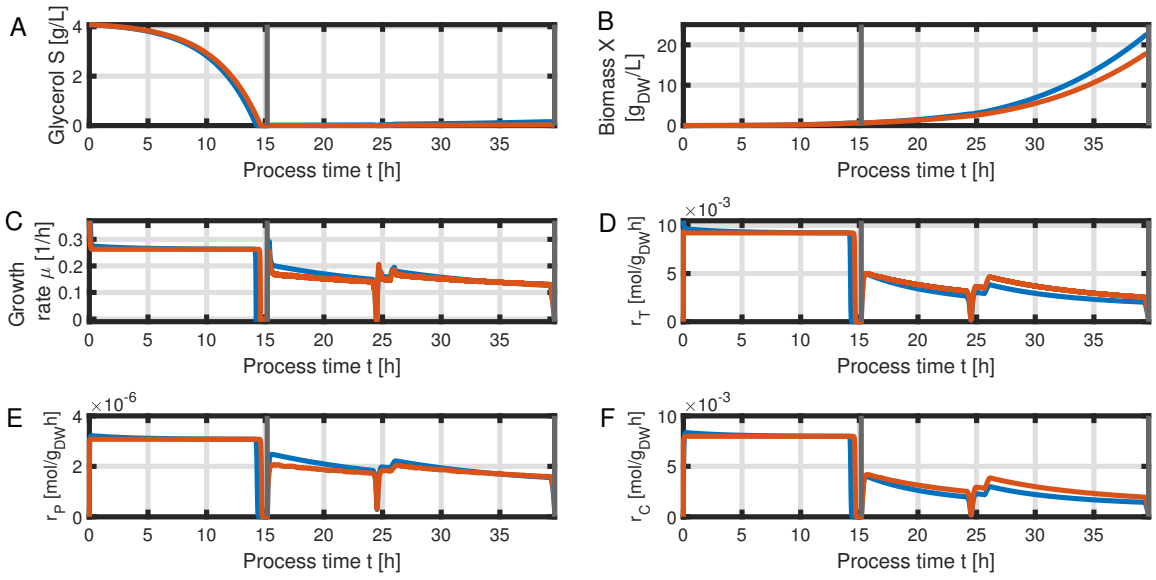

Figure 5.2: Comparison of the substrate concentration S (A), biomass concentration X (B), growth rate μ (C), substrate transport rate r_T (D), protein synthesis rate r_P (E) and respiration rate r_C (F) in a STR (blue) and in a two-compartment bioreactor (orange), where the quantities are mean values over all compartments and particles. The separation of the batch and fed-batch phase is indicated by the vertical black line.

that the two aspects (growth and uptake) are considered simultaneously as the growth rate is determined by the intracellular composition of the cell rather than being a empirical function.

With the employment of a segregated approach and the large number of virtual particles used, we can access the distribution of the internal properties for any time point of the cultivation providing insight into occurring population heterogeneity. It can be shown that the differences between the substrate and biomass concentration X are small with a slight accumulation of substrate concentration S being observable in compartment 1 and a decrease of substrate concentration over the compartments with a concentration of zero in compartment 11 (Appendix, Figure A.6). Figure 5.3C-K shows the protein, metabolite concentration and growth rate distributions of the entire population over the whole bioreactor system for different time points of the process: the end of the batch phase $t = 14$ h (C-E), the beginning of the fed-batch phase $t = 16$ h (F-H) and the end of the fed-batch phase $t = 38$ h (I-K). The distributions are normalized to 1 considering all particles of the bioreactor system. As expected, the distributions during the batch phase are unimodal in Figure 5.3A-C, since all cells experience the same environment. The protein distribution exhibits a narrow gaussian shape in Figure 5.3A, while the metabolite concentration and growth rate is equal for all particles in the system. With the beginning of the fed-batch, the bioreactor is split into two environments and Figure 5.3D and E reveal a bimodal distribution for M and P reflecting the two different substrate environments. These subpopulations can be categorized into growing and non-growing cells as seen in Figure 5.3F. At the end of the fed-batch phase, the population remains split into two subpopulations as seen in Figure 5.3G-I, Interestingly, the two subpopulations are closer together compared to the distributions at the beginning of the fed-batch. This is also evident when calculating the standard deviation of the internal properties as shown in Figure 5.3J-L. The particles start with a rather broad protein concentration distribution, which narrows over the batch phase and increase again

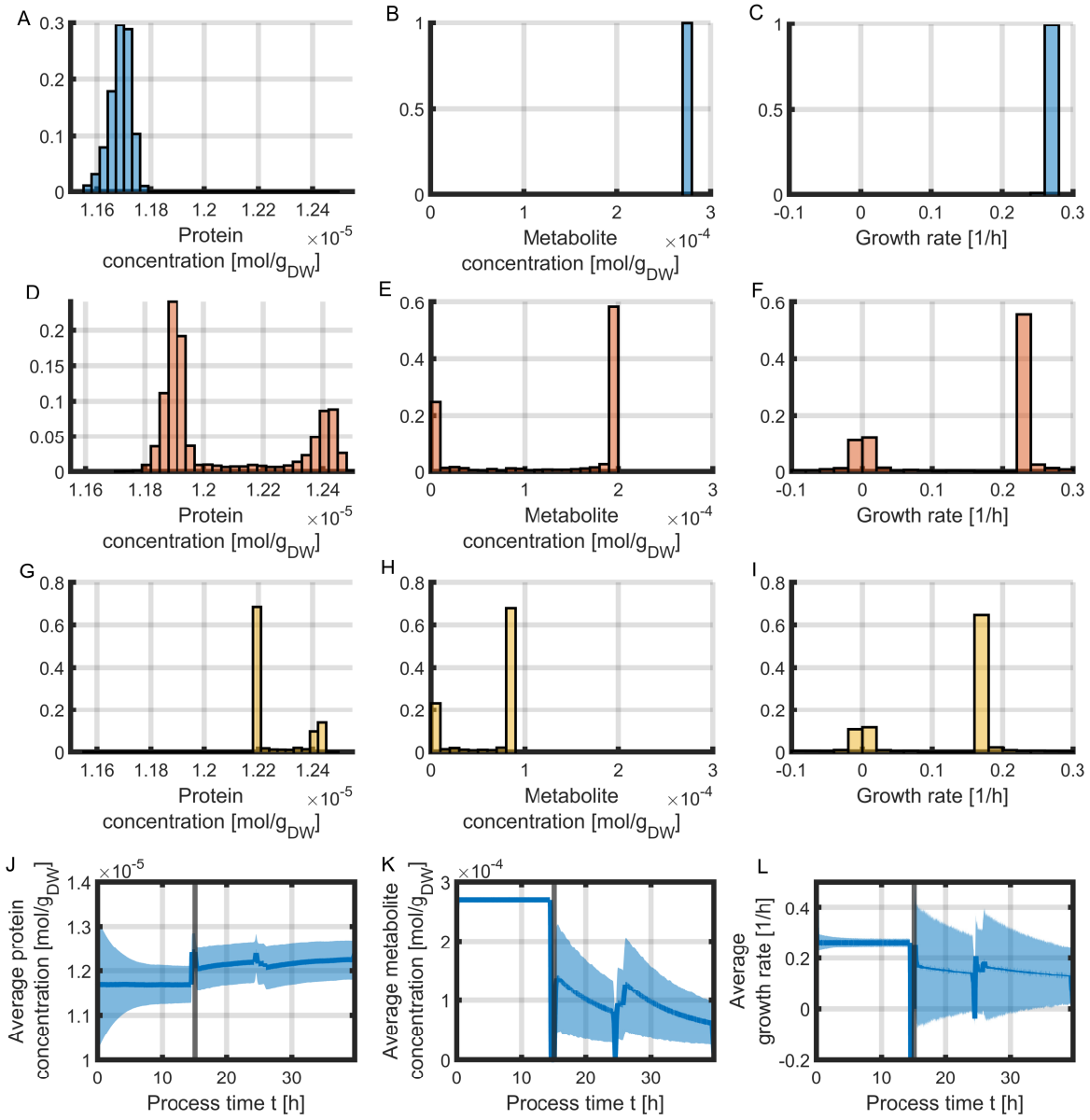

Figure 5.3: Distributions of protein concentration (A,D,G), metabolite concentration (B,E,H) and growth rate (C,F,I) at different time points $t = 14$ h (A-C), $t = 16$ h (D-F) and $t = 38$ h (G-I) without intrinsic noise normalized to 1 for all particles over the whole bioreactor system for different compartments: compartment 1 STR (blue), compartment 2 PFR inlet (orange) and compartment 11 PFR outlet (yellow). Time course of the averaged mean of the protein concentration (J), metabolite concentration (K) and growth rate (L) with its standard deviation.

with the formation of the two subpopulation in Figure 5.3J. The particles have the same metabolite concentration at the end of the batch, thus the standard deviation is zero during this time in Figure 5.3K. With the addition of the feed, the standard deviation increases instantly, decreases until the switch to the second feed solution increasing the deviation instantly again with a following decrease until the end of the fed-batch. The narrowing of the metabolite concentration

distribution is surprising as one would initially expect that with an increasing substrate gradient the distribution would widen as well. The reason behind this is explained later on when the steady states for the corresponding environments in the compartments are calculated. This behavior is mirrored by the standard deviation of the growth rate μ in Figure 5.3L.

In addition to the distributions for any time point of the process, we can access the local distribution of any compartment as well. Figure 5.4A-C shows the local distributions at the end of the fed-batch ($t = 38$ h) for the protein concentration, metabolite concentration and growth rate in three different compartments (colored according to the compartment): the STR (compartment 1), the first and last compartments of the PFR (compartment 2 and 11). A look at the local distributions of metabolite content in Figure 5.4B compared to Figure 5.3H shows that cells with high metabolite concentrations are present in the STR (compartment 1) or have just left the STR, while the subpopulation residing at the outlet of the PFR (compartment 11) has metabolite concentration close to zero. Thus, differentiating after the compartments reveals that the subpopulations correspond to their respective substrate environment. In Section 3.6.2 we have established that the relaxation times of M and P for balanced cells are both $\tau = 0.015$ h and, therefore, shorter than the smallest mean residence time in a compartment $\min(\tau_r) = 0.045$ h. The cell population at the end of the batch phase comprise balanced cells and with the short residence time the particles are able to adjust their metabolite and protein content to be in equilibrium with their local environment. Thus, the cellular state is directly correlated to its current environment and regardless of their location, the vast majority of the particles are in balanced state. The gradient of the substrate in the bioreactor system is depicted in Figure 5.4D and E through two different representations: in terms of volume distribution and

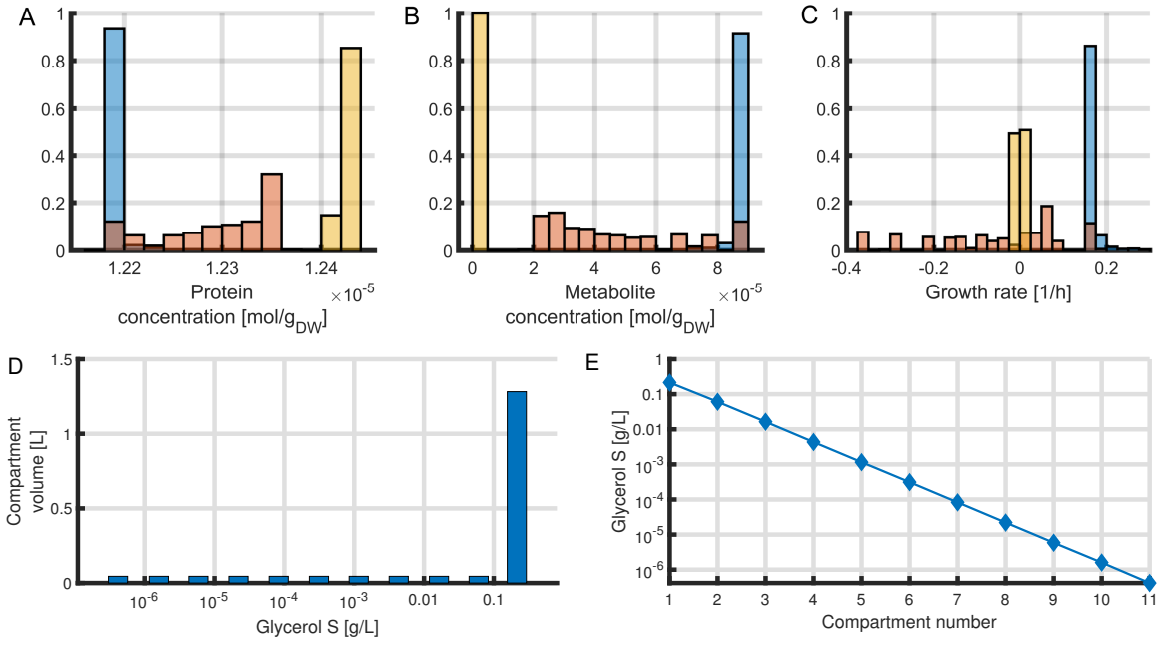

Figure 5.4: Comparison of the distributions of protein concentration (A), metabolite concentration (B) and growth rate (C) at $t = 38$ h without intrinsic noise normalized to 1 for different compartments (compartment 1 STR (blue), compartment 2 PFR inlet (orange) and compartment 11 PFR outlet (yellow)). Distribution of substrate over the compartment volumes normalized to 1 (D) and substrate concentration in log scale for all compartments (E).

more conventionally in terms of concentration profile. The concentration in the STR is around 0.2 g/L and drops suddenly to 0.06 g/L and further decreases logarithmically to 10^{-7} g/L at the end of the PFR. As result, the metabolite content decreases to zero due to the limitation by the environment. It must be noted that the concept of limitation corresponds to an insufficient supply from the environment with respect to the cell needs (largely determined by the conditions prevailing in the STR). Comparison of Figure 5.3K and Figure 5.4C indicates the origin of the non-growing subpopulation (yellow distribution in Figure 5.4C) as the PFR, which causes a prolonged exposure of the particles to starvation.

Comparing the local average values of the internal properties to the steady states, calculated using Eq. (5.9), corresponding to the local environment in Figure 5.5 confirms that the internal properties of the particles in the STR (compartment 1) and at the end of the PFR (compartment 11) are in balanced state with their respective environment. These balanced cell states can be obtained from the steady states calculated in Section 3.6.1. The steady state protein concentration value of the particles in compartment 11 cannot be obtained as the local substrate concentration is practically zero and the metabolite concentration in the cells get exhausted. Therefore, the values of the protein concentration are set to correspond to a biomass fraction of 60 %. Some of the particles which recently entered the PFR (compartment 2) are in transition and the values do not completely agree with the steady state values. This is especially apparent for the growth rate values in Figure 5.5C, which exhibits negative growth rate values as already seen in Figure 5.4C. Additionally, the global average (dotted purple line) and the steady states

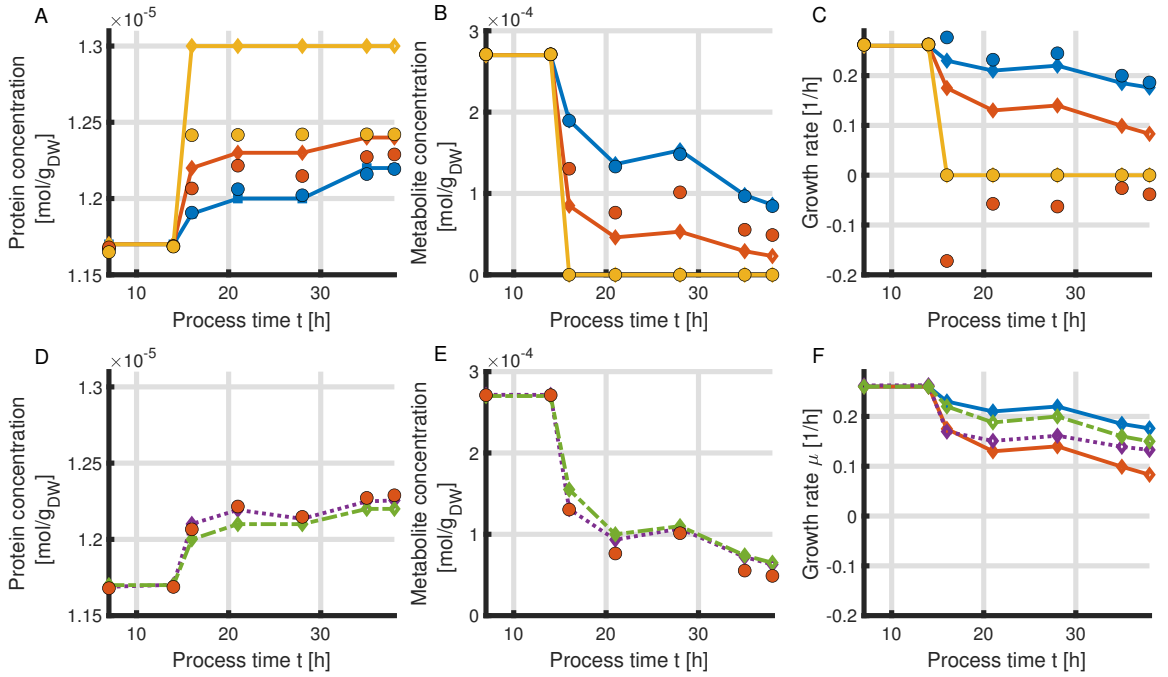

Figure 5.5: Comparison of the local average values for compartments 1 (blue), 2 (orange) and 11 (yellow), indicated by circles, steady state values corresponding to the local environment (solid lines), global average (dotted purple line) and steady state corresponding to the global average environment (dashed green line) of internal properties for different time points of the process for the protein concentration (A,D), metabolite concentration (B,E) and growth rate (C,F).

corresponding to the global average environment, meaning the average substrate concentration weighted by the respective volumes of the compartments (dashed green line), calculated using Eq. (2.58), is depicted as well in Figure 5.5D-F. Figure 5.5E shows a decrease in metabolite concentration over the course of the bioprocess as the steady state solution is affected by the limitation due to the physical regime. With increasing biomass concentration the relative supplying capacity of the environment in Eq. (3.37) decreases leading to a lower steady state value for the metabolite concentration. This explains the narrowing of the metabolite distribution in Figure 5.3. Interestingly, the mean average protein and metabolite concentration of the particles in compartment 2 are equal to the global average concentration in Figure 5.5D and E. The global average and steady state corresponding to the global average environment of the growth rate show the largest difference out of the three properties. Here, the local average value of compartment 2 does not match the global environment, but rather the steady state values corresponding to the local environment of compartment 1 and compartment 2. Therefore, particles in compartment 2 are unbalanced and its cell state is closer to the cell state of the global environment. Overall, we can conclude that in case of a populating comprising balanced cells, the cell state of the particles represent their respective local environment if the mean residence time is larger than the relaxation time.

5.3 Contribution of cell division on population heterogeneity

In the last section, we considered a population of mostly balanced cells, but now we want to incorporate intrinsic noise through cell division as well and analyze the influence of different weighting and redistribution functions during the cell division process. Cell division leads to a reshuffling of cell content. Therefore, it creates particles with non-balanced metabolism which further induces a decoupling of the time scales for protein P and metabolite M adaptation as time scale analysis in Section 3.6.3 indicates that the protein concentration P converges much slower than the metabolites M. Thus, intrinsically induced heterogeneity is most likely to interfere with external noise and enhances heterogeneity in a population. The simulated mean quantities, such as substrate concentration, biomass concentration and reaction rates do not differ from the case with only external noise. This can be seen in Appendix, Figure A.5.

First, in Figure 5.6 we consider a simulation of the bioreactor system with intrinsic noise using weights defined by Eq. (5.7). The first row depicts the distributions for the entire population, while the second row refers to the local distributions. To compare the results with a population with only extrinsic noise, the last row shows the distributions of the entire population with and without intrinsic noise. With the inclusion of the effect of cell division, the protein distribution is much wider and the two modes, present in the simulation without noise seen in Figure 5.4A, are only slightly visible in Figure 5.6A. Contrary, the metabolite concentration in Figure 5.6B is mostly unchanged and explains the two distinctive subpopulations exhibiting zero growth and a growth rate of $\mu = 0.2$ 1/h in Figure 5.6C. The overall temporal evolution of the global property distributions is the same as the case with only external noise and can be seen in Appendix, Figure A.7. Distinguishing after the local distribution, shows little variation in the protein concentration distributions over the different compartments as protein synthesis is a biologically slow process. It is implied that the majority of the particles is not able to optimally adapt their protein content to the local environment, but nevertheless, the local average protein concentration is correlated to the steady state solution corresponding to the local environment, analogous to the last section. But it is worth mentioning that the global average protein concentration

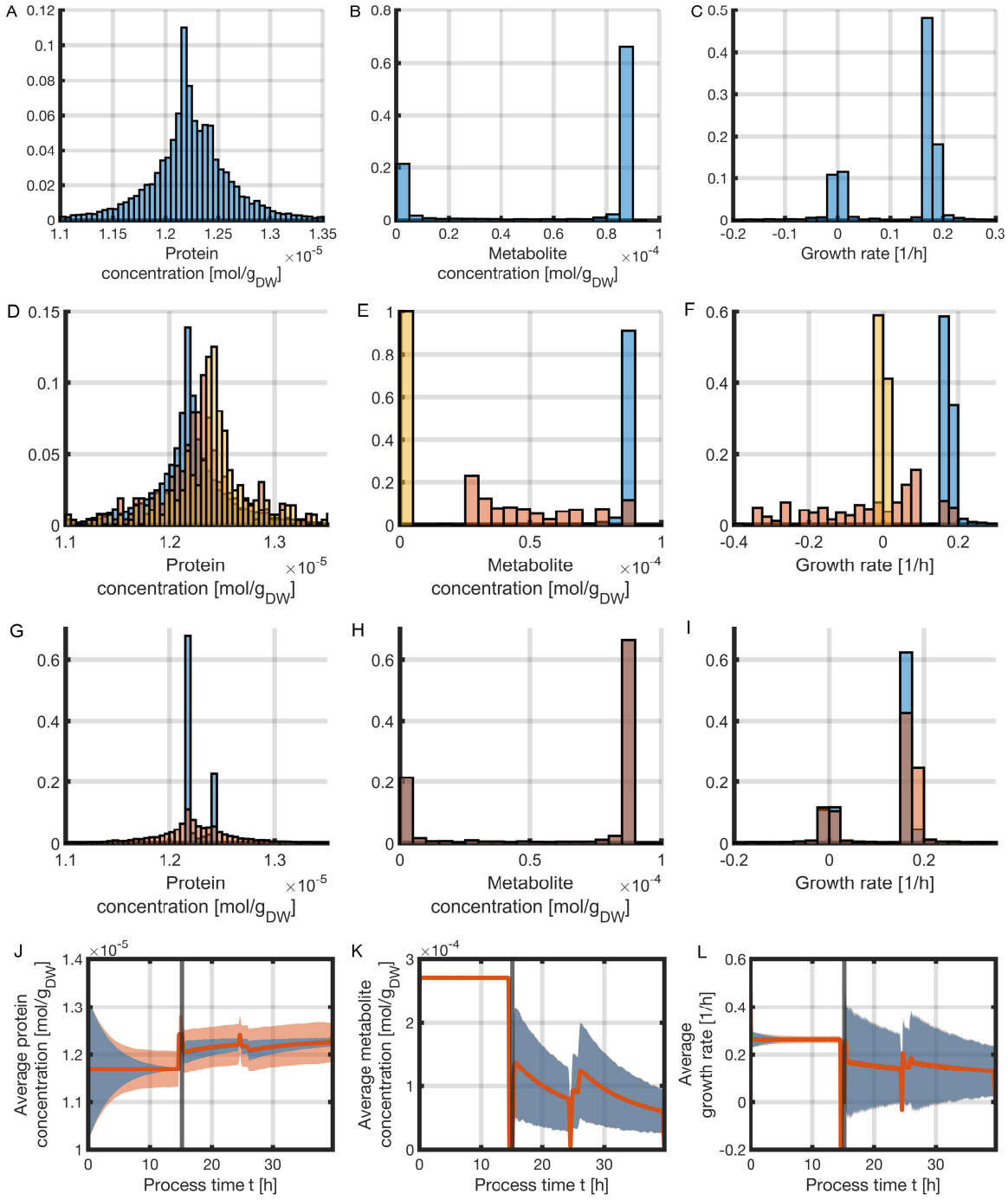

Figure 5.6: Comparison of the distributions of protein concentration (A,D), metabolite concentration (B,E) and growth rate (C,F) at $t = 38$ h with intrinsic noise. Normalized distribution for all particles over the whole bioreactor system (A, B, C) and for different compartments (D,E F; compartment 1 STR (blue), compartment 2 PFR inlet (orange) and compartment 11 PFR outlet (yellow)). Comparison of the distributions over the whole bioreactor without intrinsic noise (blue) and with intrinsic noise (orange) (G, H, I). Time course of the average of the protein concentration (J), metabolite concentration (K) and growth rate (L) and its standard deviation.

and the values exhibit by particles in compartment 1 are much closer together. This is shown in Appendix, Figure A.8.

The distributions in metabolite concentration over the different compartments in Figure 5.6E on the other hand resemble the distributions in Figure 5.4E, showing again a correlation with the local environment. This is due to the fact that the conversion of substrate into metabolites is a fast process compared to the transport in the space of concentrations by fluid motion. The substrate uptake is, on one hand, defined by the substrate availability in the environment and, on the other hand, based on the proteome fraction T, which varies on a degree depending on the already available metabolite concentration M defined by Eq. (3.24). A high amount of available metabolite concentration reduces the fraction T of the proteome and, therefore, decreases the influx of substrate into the cell, while up-regulating the protein synthesis by increasing the fraction R. The inclusion of noise does not change the mean growth rate over the whole population (Figure 5.6J-L) but creates a population of individuals that are all out-of-equilibrium, not in a balanced state. This leads to some spreading in the distribution of the growth rate as seen in the population in compartment 2 in Figure 5.6F, which is broader in comparison to the simulation without intrinsic noise. Calculating the standard deviation of the internal properties and comparing those with the standard deviations in case of only external noise confirm the observations. Figure 5.6J shows overall standard deviations in the protein concentration P due to the added intrinsic noise, while the standard deviations for the metabolite concentration M and growth rate μ are comparable with the latter case. Therefore, the additional noise in the growth rate distribution as seen in compartment 2 has no effect on the overall standard deviation.

Next, we investigate the influence of the sampling weighting function for cell division on the distribution of the internal properties. As the process of intrinsic noise through cell division is a stochastic process, we perform the simulations for each case five times to exclude differences originating from stochasticity. We focus on the fed-batch phase since population heterogeneity is more pronounced in this process phase. The three cases which are investigated in the following were defined by Eqs. (5.5)-(5.7) and we will show that introducing sampling weights favors particles residing in compartment 1.

First, we compare the outcome of random sampling with equal weights and weights considering the total biomass of the virtual particles following the formulation of Eq. (5.5). Thus, if we consider the total biomass of a particle when sampling for cell division, particles with higher mass are more likely to be selected. Particles with higher total biomass are more likely to exhibit higher metabolite concentrations. Due to this, these particles comprise faster growing cells and are able to increase their biomass faster. A more detailed analysis of the relation between the biomass and a particles' internal properties can be found in Appendix, Section A.3.3. Particles with generally higher total biomass, therefore, can be matched to the environment of compartment 1, where particles exhibit the corresponding cell state, in Figure 5.7D. Additionally, the the total biomass distribution of particles in the last compartment is shifted to lower values, leading to a lower average total biomass in comparison to compartment 1 as seen in Figure 5.7D.

Particles in compartment 1 are exposed to higher substrate concentrations compared to particles in the PFR, hence, in average they contain lower protein concentrations. The assignment of new physiological properties follows a normal distribution centered around the old properties. With a favored selection of these particles, a shift towards a lower mean protein concentration is expected. This is confirmed through comparison of the performed simulations in Figure 5.8A which shows a slight shift towards lower values. The difference is very small, but consistent

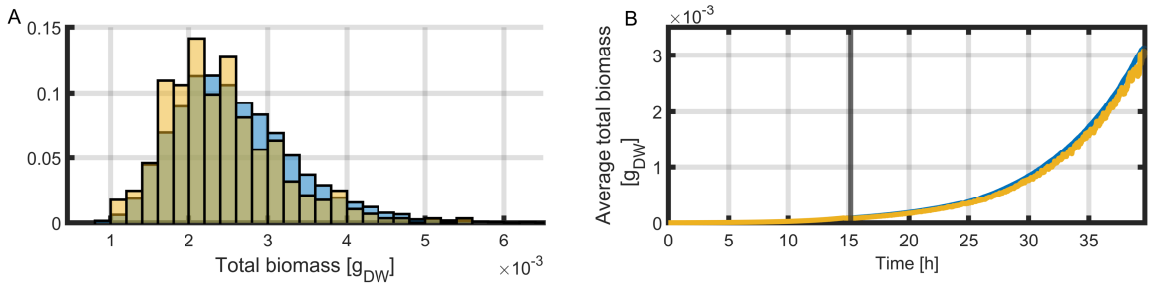

Figure 5.7: Normalized distribution of the total biomass of the virtual particles at $t = 38$ h (A) and time coarse of the average total biomass of the virtual particles (B) for compartment 1 (blue) and compartment 11 (yellow).

for the whole fed-batch phase. In the following, we opt to investigate the variation instead of the standard deviation indicated by bands around the average mean values to be able to see the differences more clearly. Figure 5.8B shows the variance of the protein concentration over the course of the fed-batch phase. The variances are comparable at the beginning of the fed-batch. After a couple of hours, the variance for the case of weights considering the particle mass is higher until $t = 32$ h, after which it is smaller than for the simulations considering equal sampling weights. A detailed description of the influence on the protein and growth rate distribution can be found in Appendix, Section A.3.3. Because of the favoring of particles in compartment 1, the distributions are affected, but the differences are small.

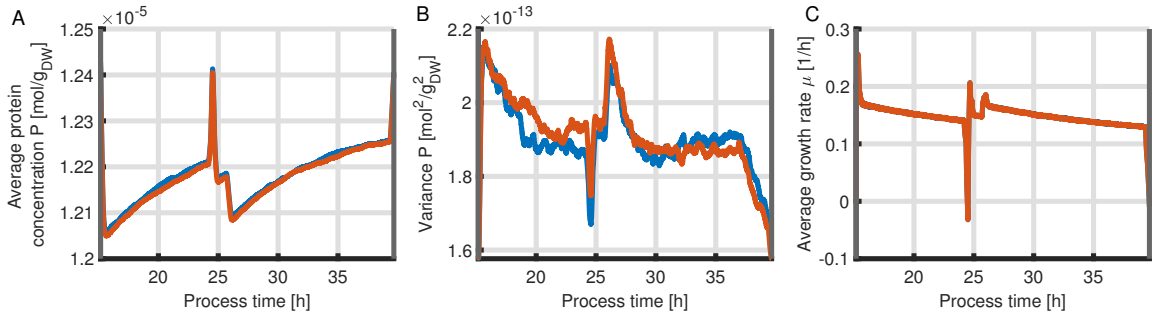

Figure 5.8: Comparison of the average protein concentration P (A), variance of protein concentration (B) and average growth rate μ (C) for sampling with equal weights (blue) and weights considering the biomass of the particles (orange) during the fed-batch phase. Each case is the average of 5 simulations.

Next, we compare random sampling with equal weights and weights considering the growth rate of each particle as described in Eq. (5.6). It is worth noting that in case of random sampling with equal weights, each particle has the same probability to be chosen. Thus, it is possible that particles with negative growth rate can be chosen to "divide" and are assigned new physiological properties. The same problem applies when considering the total biomass for the sampling weights. With the introduction of weights considering the growth rates, only particles with positive growth rates are eligible to be chosen. As faster growing particles are favored, we expect particles residing in the STR (compartment 1) are more likely to be selected again. This bias is even stronger than the last case, since particles with higher growth rate only reside

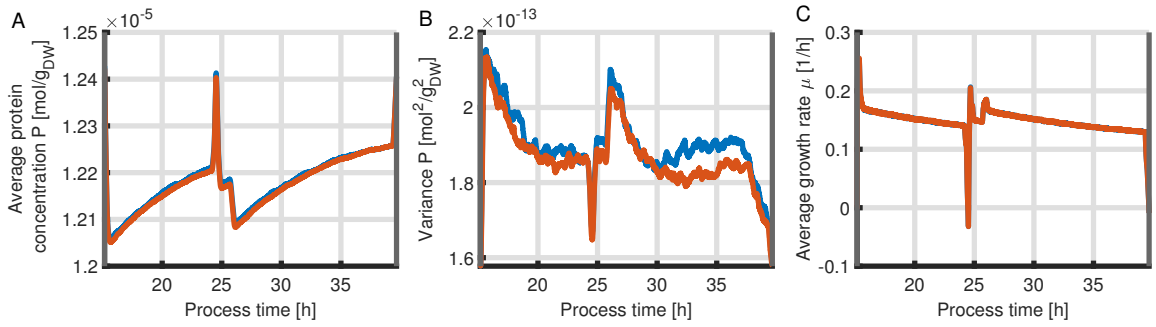

Figure 5.9: Comparison of the average protein concentration P (A), variance of protein concentration (B) and average growth rate μ (A) for sampling with equal weights (blue) and weights considering the growth rate of the particles (orange) during the fed-batch phase. Each case is the average of 5 simulations.

in compartment 1, while particles with higher total biomass can be found at other locations of the bioreactor systems as well. Analogous to the last case, a shift towards a lower mean protein concentration is expected. This is confirmed in Figure 5.9A which shows a slight shift towards lower values when considering the growth rate as sampling weights. The difference is very small again, except for a small number of short time periods near to the end of the fed-batch phase, where the average protein concentration is the same. Furthermore, a lower variance in protein concentration over all particles in the bioreactor system is observed in Figure 5.9B as well. This originates from the fact that the standard deviation of the normal distribution for the new internal properties is dependent on the protein concentration of the selected particles. As particles with lower protein concentration are favored, the standard deviation is smaller leading to a smaller variety of protein concentration. The mean growth rate μ does not differ between the two cases in Figure 5.9C.

The problem of only considering the growth rate for the sampling weights is that particles with a very high growth rate, but low total biomass are favored even though they represent a smaller amount of biological cells. Thus, it is more realistic to consider both factors: the total biomass of a particle and the growth rate. This is the reasoning behind the original choice of the weighting function given by Eq. (5.7), and we continue to use this set of weights in the following. From the previous observations, we expect a combination of the effects. The mean total biomass of a particle is higher in compartment 1 and the difference between the mean in the compartments increases over the course of the fed-batch phase. Additionally, the majority of particles with higher growth rate reside in compartment 1 as well. Thus, a lower average protein concentration, lower protein concentration variance is expected. This is confirmed in Figure 5.10. The average protein concentration and variance is smaller than for sampling with equal weights as shown in Figure 5.10A and B. The mean growth rate in Figure 5.10C is the same for both cases. Since we employ these sampling weights going forward, we analyze the effects on the distribution more in detail. Favoring particles of higher growth rate and particle mass leads to more variety in particles with lower protein concentration during the switch to a new feed solution at $t = 26.5$ h in Figure 5.10D and less spreading for particles with higher protein concentration, which tend to reside in the PFR. A higher peak for particles with lower protein concentration can be observed near to the end of the fed-batch at $t = 38$ h, accompanied by a narrower distribution in Figure 5.10F. The growth rate distribution shows a higher peak considering the subpopulation of growing cells with a right skewed distribution, while sampling

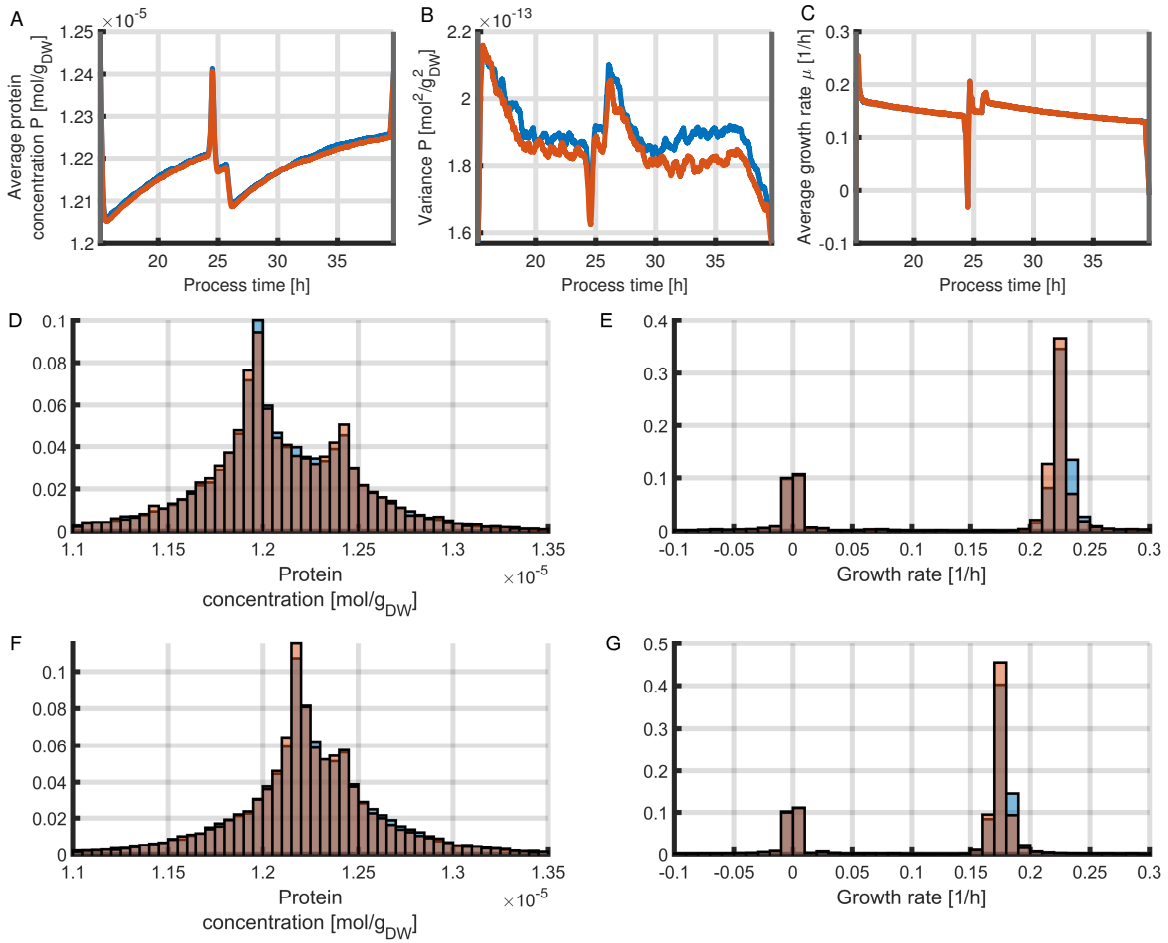

Figure 5.10: Comparison of the average protein concentration P (A), variance of protein concentration (B) and average growth rate μ (A) for sampling with equal weights (blue) and weights considering the growth rate and biomass of the particles (orange) during the fed-batch phase. Comparison of the protein concentration (D,F) and growth rate (E,G) distributions at $t = 26.5$ h (middle row) and $t = 38$ h (lower row) over the whole bioreactor normalized to 1. Each case is the average of 5 simulations.

with equal weights result in a left skewed distribution for both time points in Figure 5.10E and G. The differences between the distributions are slightly more pronounced compared to the other cases. Overall, we can conclude that the differences between the different sets of sampling weights are small and do not largely affect the behavior of the population and its properties. Nevertheless, we opt to use the most reasonable choice given by Eq. (5.7) for selection of cells to undergo cell division.

Another parameter that influences the degree of population heterogeneity is the standard deviation chosen for the normal distribution used for reshuffling of the internal properties during cell division. Thus, we compare the simulations using different values for the standard deviation by doubling ($\sigma_M = 0.2\ M_0$ and $\sigma_P = 0.1\ P_0$) and halving ($\sigma_M = 0.05\ M_0$ and $\sigma_P = 0.025\ P_0$) the factor. Figure 5.11A-C shows the distributions of the internal properties for the three different cases at the end of the fed-batch phase $t = 38$ h. The protein distribution is wider for higher and narrower for lower values as seen in Figure 5.11A. This fact is supported by calculating the

standard of P over the course of the fed-batch phase, which is provided in Figure 5.11D. Doubling the standard deviation nearly quadruples the variance in P, but remains below the factor of 4 for the majority of the time, while halving the standard deviation reduces the variance to around a third of the original case. Interestingly, for higher standard deviation the variance remains noisy over the course of the fed-batch without showing the same trends as the original case, where the variance decreases over each fed-batch phase. This can be traced back to the difference between the distributions of the split environment of the bioreactor system. For a higher standard deviation it can be shown, that bimodality in the protein concentration is less pronounced than in case of higher standard deviations (Appendix, Figure A.14A and D). Therefore, the increased intrinsic noise dominates over the dynamics of the distance between the values of the two subpopulations. But besides the shape of the protein concentration distribution, the increased or decreased standard deviation does not affect the mean protein concentration over the course of the fed-batch phase, thus the average protein concentration is indicated in black in Figure 5.11D. In contrast to the proteins, the dynamics of the metabolites are fast. Therefore, the distribution of metabolite concentration remains unaffected by the standard deviation. We have established earlier that particles undergoing cell divisions tend to reside in the STR and are growing at faster rate. Thus, the differences between the standard deviations manifest in the subpopulation of growing cells. With higher standard deviation, the variety of growth rate is larger without affecting the mean growth rate of the population. Investigating the effects of different standard deviations, we can conclude that the shape of the distributions for the protein concentration and growth rate are dependent on this parameter, but does not change the overall behavior of the population. If fluorescent data correlating to the growth rate exists, the appropriate standard deviation parameter can be found. For understanding the interplay

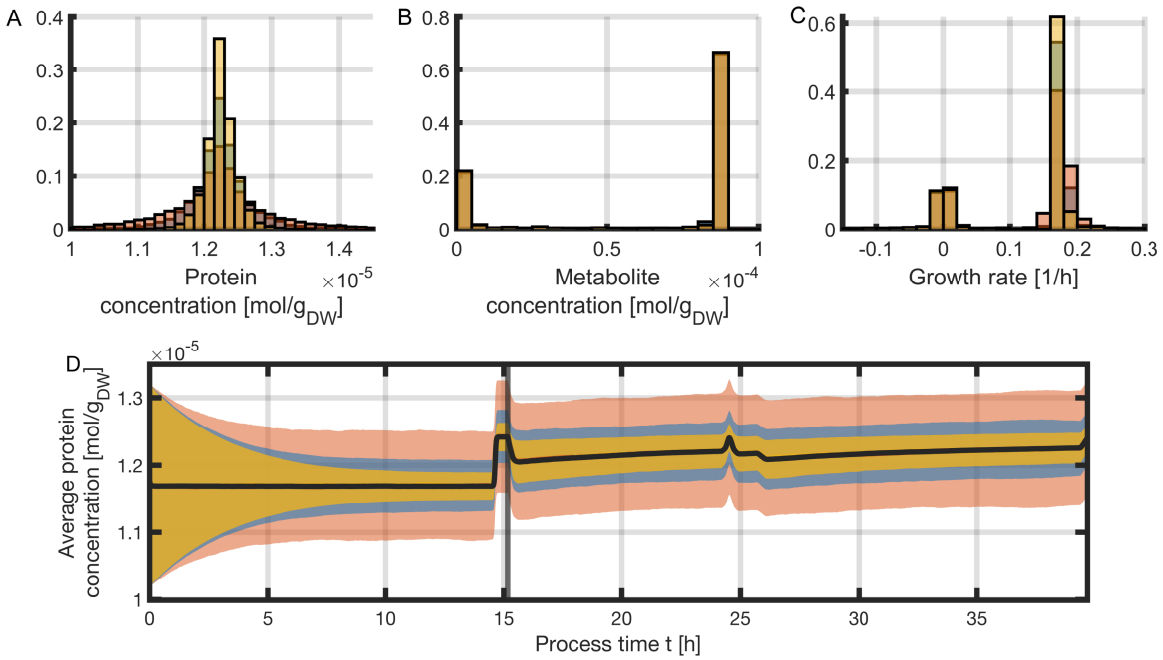

Figure 5.11: Comparison of the distributions of protein concentration (A), metabolite concentration (B) and growth rate (C) at $t = 38\ h$, as well as, variance of P (D) and average protein concentration P (black) with its standard deviation, for simulations using a normal distribution with $\sigma_M = 0.2\ M_0$ and $\sigma_P = 0.1\ P_0$ (orange), $\sigma_M = 0.1\ M_0$ and $\sigma_P = 0.05\ P_0$ (blue) and $\sigma_M = 0.05\ M_0$ and $\sigma_P = 0.025\ P_0$ (yellow) for reshuffling of the properties.

between the internal properties or the qualitative behavior of the particle properties, the value of the standard deviation is irrelevant.

Before going further, a summary of the findings is necessary. Without intrinsic noise, cells remain in their balanced state as the adaptation of P and M is fast compared to the residence time in the different compartments. The dynamics of P and M are similar. When intrinsic noise is added, the randomization of cell properties at birth creates new cells that are not in a balanced state. For those cells, the adaptation of the protein content is now much slower than the adaptation of the metabolite content. Hence, more physiological diversity is found when a reshuffling of cell properties is considered. These results show the necessity to include intrinsic noise when using a structured metabolic model such as a coarse-grained model for the biological phase. It provides a rationality of the decoupling between the growth rate and the substrate concentration, which has been observed for changing environments [1, 91]. In our model, the growth rate depends on both the cellular state, which is determined by the history of the particle, and the environmental supply. In other words, the growth rate must be regarded as a dynamic cell property, rather than simply being set by the liquid phase composition through an algebraic equation $\mu = f(S)$. The impact of cell division on cell dynamics as well as the decoupling of uptake and growth rates is of utmost importance. It is consistent with the choice of μ as an internal variable subject to a redistribution mechanisms at cell division, in biological population balance models previously developed [108, 109, 110, 125].

5.4 Contribution of process operation on population heterogeneity

In the last sections, the effects of external noise and intrinsic noise induced by cell division have been investigated. But the degree of external noise is also dependent on the operation of the bioprocess. Thus, in this section different operation settings are simulated.

5.4.1 Different flow rates Q_R

As the residence time in the compartments has an effect on the cell composition of the particles, we simulate different flow rates through the bioreactor system. A higher flow rate leads to smaller substrate gradients and smaller mean residence times in the compartments as seen in Figure 5.12A and B, which shows the substrate concentration in log scale for all compartments of the bioreactor system at $t = 38$ h as well as the variance of the substrate distribution over the whole time course of the bioprocess for two different flow rates: $Q_R = 1$ L/h and $Q_R = 20$ L/h. The substrate concentration barely drops from compartment 1 to 2 and reaches its minimal value in compartment 11 at $S = 0.06$ g/L, which is around the value of the substrate concentration in compartment 2 for a flow rate of $Q_R = 1$ L/h in Figure 5.12A. The variance of the substrate concentration distribution increases significantly after the addition of the second feed solution for both flow rates in Figure 5.12B, but the sudden increase is proportionally smaller for the higher flow rate. Near to the end of the fed-batch phase at $t = 38$ h, the variance is almost ten times lower. The mean volumetric substrate concentration remains the same in Figure 5.12C. In order to perform the simulation in reasonable computation time, the step size is increased to $\Delta t = \min(\tau_r)$. For a flow rate of $Q_R = 20$ L/h, the smallest mean residence time in a compartment $\min(\tau_r) = 0.0023$ h is smaller than the relaxation time for M, while the mean residence time of the first compartment $\tau_r(1) = 0.045$ h is larger. Results of the simulation are

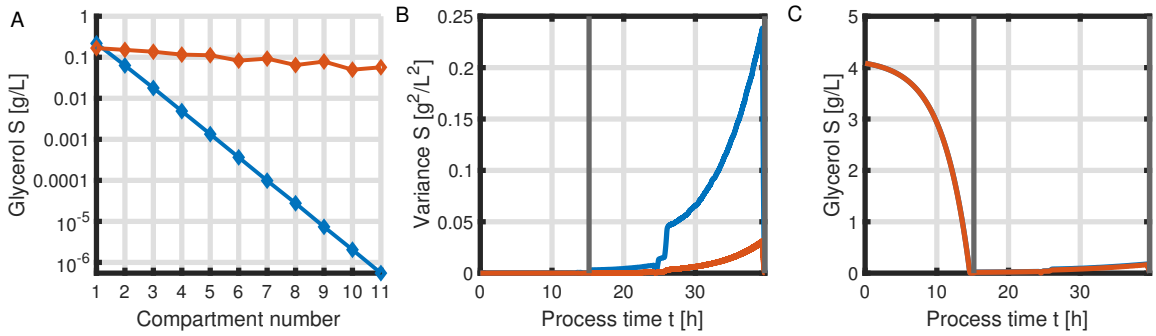

Figure 5.12: Comparison of the environment for different flow rates $Q_R = 1$ L/h (blue) and $Q_R = 20$ L/h (orange) showing the substrate concentration in log scale for all compartments at $t = 38$ h (A), variance of the substrate concentration distribution (B) and the mean volumetric substrate concentration S (C).

presented in Figure 5.13: the first row deals with the local protein, metabolite and growth rate distributions, while colors refer to the compartment (STR, inlet/outlet of the PFR). The second row compares the overall distributions for various flow rates, colors being here indicative of the flow rate.

At $Q_R = 20$ L/h, the protein distributions are again space-independent in Figure 5.13A and span the same range as in Figure 5.6D. All metabolite distributions, in Figure 5.13B, now span a narrower range, from 20 to 80 μmol/g_{DW}. The range corresponds to the values of the steady state in compartment 1 and 11. The biomodality in M is less pronounced (compared to 5.6 E) but still present. The growth rate distribution on the other hand is very broad and even exhibit negative growth rates in Figure 5.13C.

As they enter the PFR zone, cells are subject to a sudden decrease in substrate availability, which leads to a depletion in the metabolite content and results in negative growth rates. In contrast to the previous case (at $Q_R = 1$ L/h), the residence time in the PFR is not sufficient for cell to reach a local equilibrium with the environment. As a result, the change in the metabolite distribution between the first and the second compartment is smaller than at lower flow rates and metabolite distributions are wider. When they reach the last compartment in the PFR, nearly all particles show negative growth rates, as all of them have been experiencing decreasing substrate concentrations for a time, but since they had not enough time to adapt to it, their content in M is not zero. As they move back to the STR, with higher substrate concentration again, the particles rebuild their metabolite content. The mean residence time in the first compartment is larger than the relaxation time, which allow cells to reach high growth rates. To sum up, the exposure frequency to low substrate concentration is multiplied by 20 when the flow rate goes from $Q_R = 1$ L/h to $Q_R = 20$ L/h whilst the magnitude of the concentration change from STR to PFR is smaller and the duration of the exposure to adverse conditions is reduced.

Increasing the flow rate up to $Q_R = 120$ L/h, confirms that the range of the protein and metabolite distributions over the whole population decreases as the flow rate increases as seen in Figure 5.13D and E. Thus, the particle population is more homogeneous as the environment is more homogeneous, in line with the reduction of external noise. The width of the growth rate distribution is accordingly smaller since the substrate gradient is smaller at higher flow rates. Nevertheless, some heterogeneity in the growth distribution in the form of two subpopulations

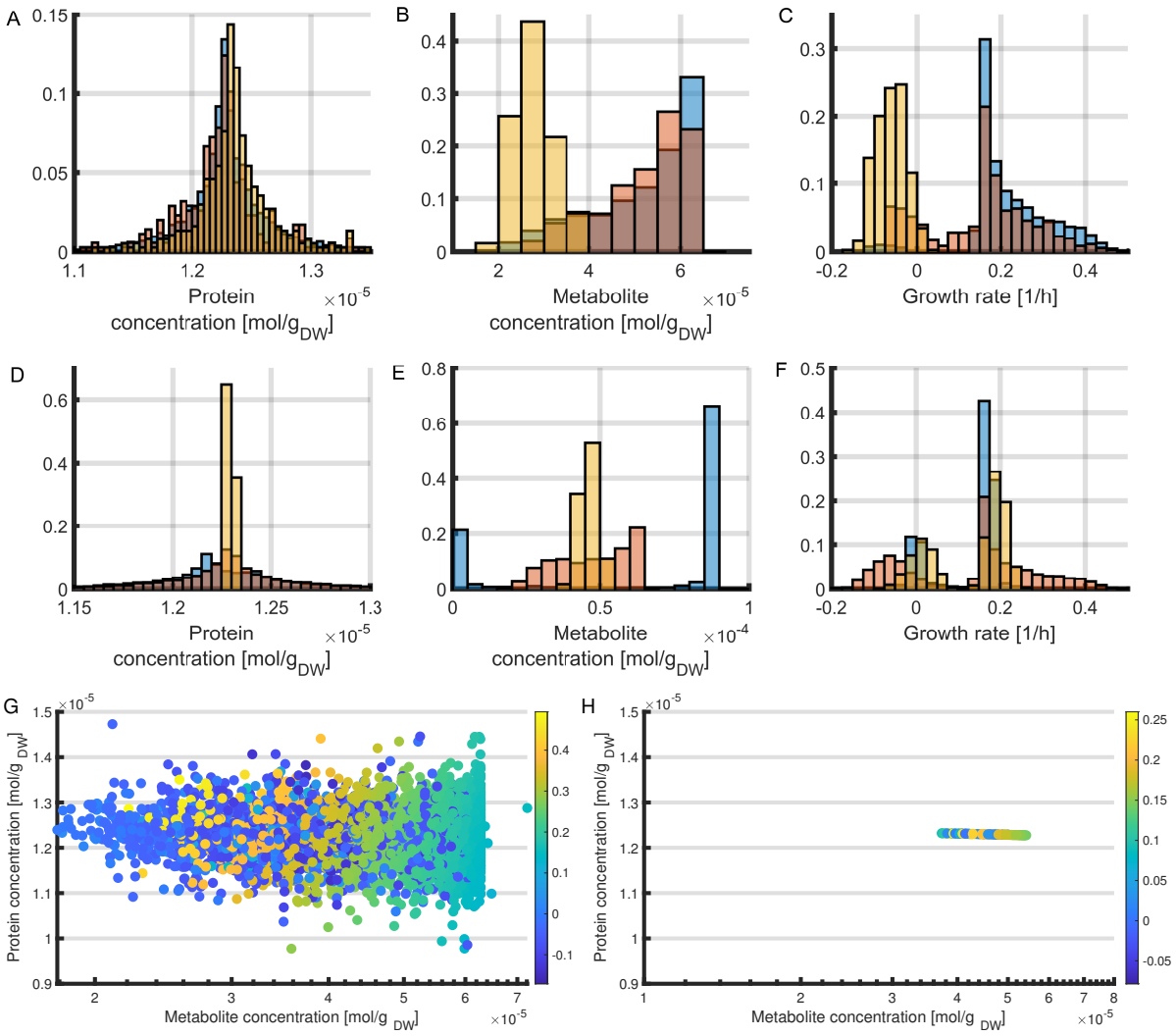

Figure 5.13: Comparison of the distributions of protein concentration (A,D), metabolite concentration (B,E) and growth rate (C,F) at $t = 38$ h with a flow rate of $Q_R = 20$ L/h normalized to 1 for different compartments (A-C; compartment 1 STR (blue), compartment 2 PFR inlet (orange) and compartment 11 PFR outlet (yellow)) and comparison of the distributions over the whole bioreactor system (D-F) for three flow rates $Q_R = 1$ L/h (blue), $Q_R = 20$ L/h (orange) and $Q_R = 120$ L/h (yellow). M-P-plane of all particles over the whole bioreactor system colored by their growth rate for $Q_R = 20$ L/h (G) and $Q_R = 120$ L/h (H)

remains because intrinsic noise keeps forming cells with unbalanced properties. These two subpopulations cannot be matched to a specific compartment as it was the case for low flow rates. This can also be observed when comparing the spreading of the particles over the M-P plane for two different flow rates, in Figure 5.13G and Figure 5.13H. Interestingly, higher growth rates in Figure 5.13G are achieved by cells with an intermediate content in M while low content is associated with negative growth rate and high content correlated with growth rates μ between 0.1 and 0.2 1/h, i.e. close to the mean population growth rate. It was further verified that the mean values of the growth rate and reaction rates at $Q_R = 120$ L/h coincides with the values obtained from a homogeneous STR which can be seen in Appendix, Section A.3.4.

The homogenization of particle content with increasing flow rate is further confirmed through calculating the variance of protein and metabolite concentration in Figure 5.14A and B. Since we are investigating several flow rates later on, we compare the variances instead of the standard deviation again. The flow rate has a much larger effect on the variance of metabolite concentration in Figure 5.14B than protein concentration in Figure 5.14A, as the dynamics of the protein concentration is much slower and is also affected by internal noise through cell division, while the metabolite concentration is only affected by external noise. For high flow rates the effect of intrinsic noise on the protein concentration dominates over the effect of external noise and the hint of the formation of subpopulations which can be associated to the split environment is not visible.

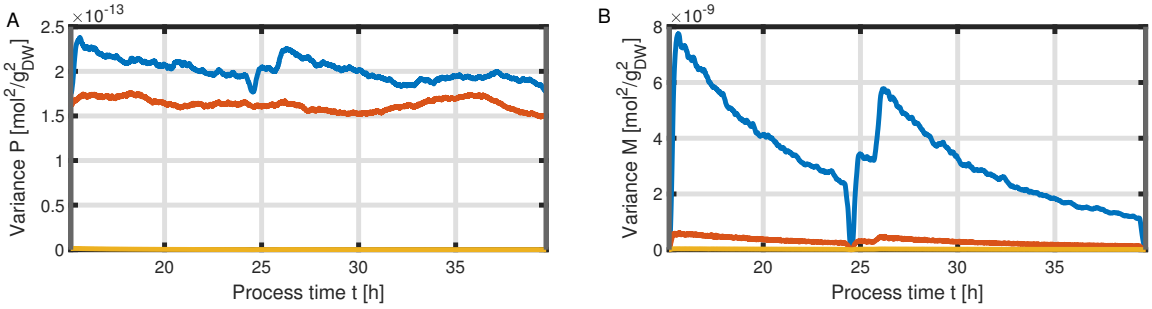

Figure 5.14: Comparison of the variance of protein concentration (A), metabolite concentration (B) for different flow rates $Q_R = 1$ L/h (blue), $Q_R = 20$ L/h (orange) and $Q_R = 120$ L/h (yellow).

Investigating the variances of growth rate in Figure 5.15, reveals the importance of the residence time compared to the relaxation time of the growth rate as well as the range of substrate concentration encountered, which was already hinted at while analyzing the property distributions. At a flow rate of $Q_R = 20$ L/h, the substrate gradient is sufficiently large and the residence time short enough to result in a higher variance compared to $Q_R = 1$ L/h. For $Q_R = 120$ L/h the substrate gradient is too small and, therefore, the particles despite of comprising unbalanced cells exhibit growth rates of smaller variety. From simulating with a flow rate of $Q_R = 20$ L/h, we observed an increased variance of the growth rate and a decreased variance for a very high flow rate of $Q_R = 120$ L/h in Figure 5.15. This poses the question, at which flow rate the variance is maximal. Performing the simulation for different flow rates reveal a maximal variance in growth rate can be achieved for $Q_R \approx 7.5$ L/h. In this case, the variance triples in comparison to the base case of $Q_R = 1$ L/h. The internal property distributions can be seen in Appendix, Section A.3.5. The growth rate distribution for $Q_R = 7.5$ L/h takes shape as mixture of the distributions of $Q_R = 1$ L/h and $Q_R = 20$ L/h, but with a wider spread between the two subpopulations of growing and non-growing cells. Thus, the substrate gradient is large enough to enable a wide range of growth rates and the residence time in each compartment is too small

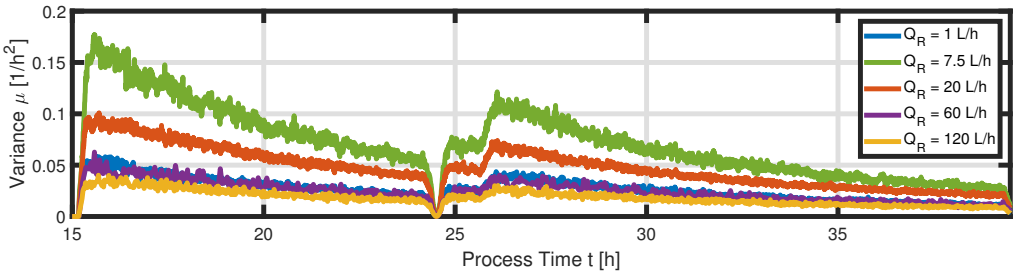

Figure 5.15: Comparison of the variance for growth rate for different flow rates: $Q_R = 1$ L/h (blue), $Q_R = 7.5$ L/h (green), $Q_R = 20$ L/h (orange), $Q_R = 60$ L/h (purple) and $Q_R = 120$ L/h (yellow).

leaving the particles in unbalanced state. With a flow rate of $Q_R = 60$ L/h, the variance of growth values are the same as for $Q_R = 1$ L/h, but the shape of the distributions clearly differ.

We can conclude that the residence time of the particles in the compartments has a significant effect on population heterogeneity, especially on the growth rate distribution. Pivotal is the mean residence time compared to the relaxation time of the variables. For small flow rates the substrate distribution in the bioreactor system is wide and the distributions of the proteins, metabolites and growth correlates with the substrate distribution, since the cell dynamics are faster than the rate of change in the external concentration. For high flow rates the residence times are lower than the relaxation time leading to a homogenization in the protein and metabolite concentration. However, intrinsic noise creates diversity, even in a spatially homogeneous reactor. Thus, particles are not optimally adapted to their environment and the repeated exposure to perturbations (even small) in the cell environment lead to a broad range of growth rates.

5.4.2 Different feeding regimes

In order to introduce even higher substrate gradients in the bioreactor system, one can feed the susbtrate into the PFR instead of the STR. This way, the particles now experience very high substrate levels in the PFR (feast zone) and a low substrate environment in the first compartment as seen in Figure 5.16D, which shows the substrate concentration levels across the different compartments. These feeding strategies were also experimentally investigated in the work of Neubauer et al. (Figure 2) [114]. The substrate gradient for a feed into the PFR at $t = 38$ h is $S_{max} - S_{min} = 11.34$ g/L with a weighted mean substrate concentration of $\bar{S} = 2.18$ g/L, while the substrate gradient for a feed into the STR only is $S_{max} - S_{min} = 0.21$ g/L with a weighted mean substrate concentration of $\bar{S} = 0.16$ g/L. The substrate gradient enhances over the course of the fed-batch process as more substrate is accumulated in the PFR. The mean residence time in the PFR in both feeding regimes is the same since the same flow rate of $Q_R = 1$ L/h is assumed. We note that the substrate distribution for a feed in the PFR is left skewed, while it is right skewed for a feed in the STR as seen in Figure 5.16.D. The exposure to high substrate concentration also creates some heterogeneity in the population but with distinct consequences from an exposure to starvation. Namely, the metabolites content never falls to zero. The metabolite concentration reflects their respective environment again and the range of its values is larger as seen in Figure 5.16B. With a higher substrate gradient, the metabolite content shows

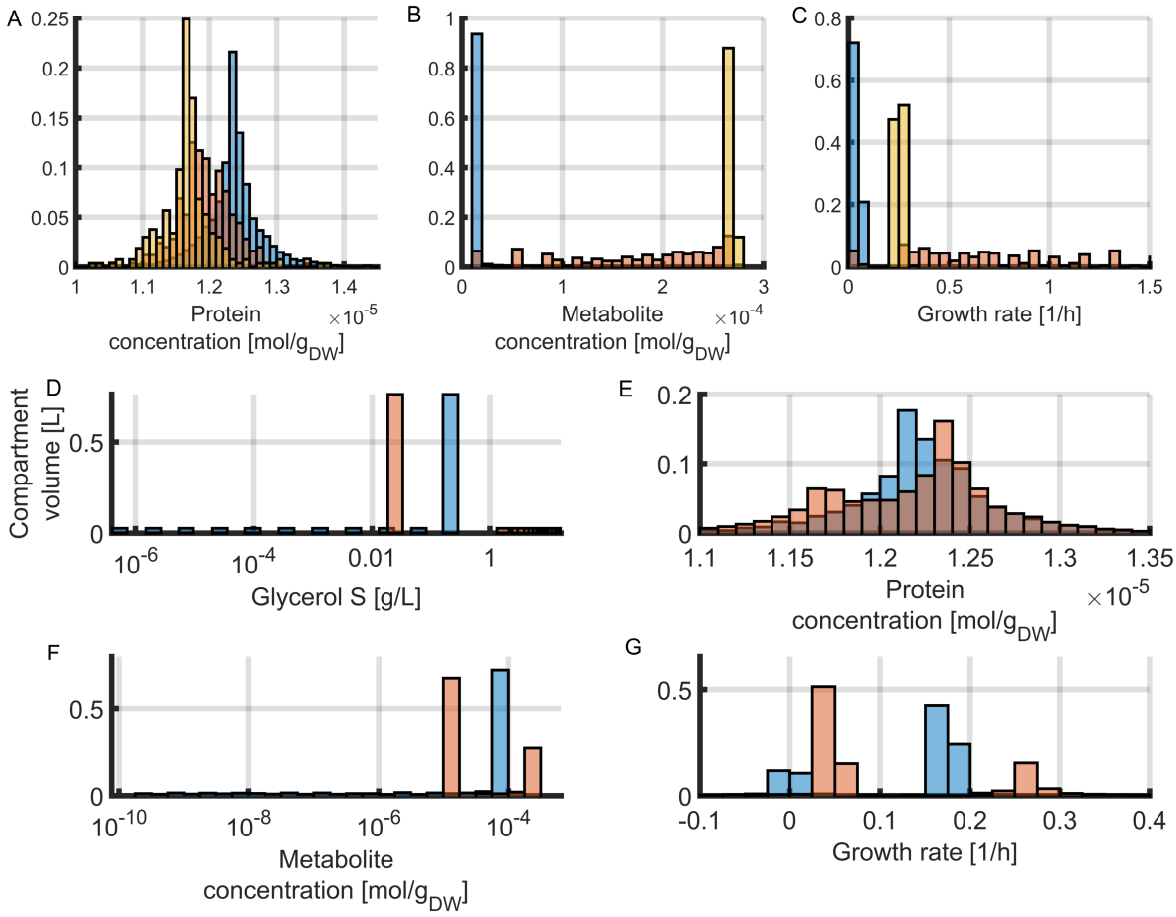

Figure 5.16: Comparison of the distributions of protein concentration (A), metabolite concentration (B) and growth rate (C) at $t = 38$ h with feed ($Q_R = 1$ L/h) into the PFR normalized to 1 for different compartments: compartment 1 STR (blue), compartment 2 PFR inlet (orange) and compartment 11 PFR outlet (yellow). Comparison of the distributions of the substrate concentration over the compartment volumes (D), protein concentration (E), metabolite concentration (F) and growth rate (G) at $t = 38$ h for two feeding regimes: feed into the STR (blue) and feed into the PFR (orange) normalized to 1.

a higher range as well. The opportunity of growing faster is offered to those cells that have the ability to increase their growth rate due to their metabolic composition when they enter the PFR zone. This is accompanied with an increase in the metabolite content, see Figure 5.16B and a shift to the left of the protein distributions along the PFR in Figure 5.16A. The particle register the environment simply as a high substrate environment and the cell demand is sufficiently covered by the environmental supply. Taking the histogram over the whole population results in a distribution with two subpopulations of slow growing and fast growing cells as shown in Figure 5.16G. The population in the STR exhibits a lower metabolite concentration, a higher protein concentration and therefore grows at lower rate, while the population in the PFR exhibits the contrary. The particles are not able to consume the substrate provided in the PFR as the substrate supply is higher than the cellular demand leading to an accumulation of substrate in the bioreactor system. Due to higher substrate concentration levels, the growth rate distribution shows higher values in Figure 5.16H, but the mean value over the whole population is smaller with $\bar{\mu}_{PFR} \approx 0.11$ 1/h compared to $\bar{\mu}_{STR} \approx 0.13$ 1/h, as the majority of the particles reside in

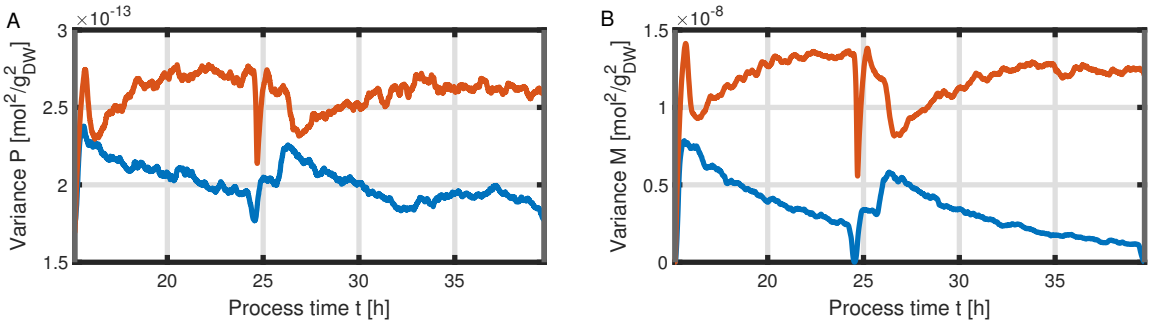

Figure 5.17: Comparison of the variance of protein concentration (A) and metabolite concentration (B) for different feeding regimes: addition of feed into the STR (blue) and PFR (orange).

a low substrate environment. In practice, such low differences in mean growth rate might not be measurable. However, with the proposed modeling approach the difference can be attributed to overflow metabolism [31], as seen in reaction rate r_C, which occurs in the PFR, leading to globally lower carbon flux towards growth as seen in Appendix, Section A.3.6.

The substrate gradient increases during each fed-batch phase leading to an amplification in heterogeneity of the environment. The substrate concentration in different compartments can be found in Appendix, Figure A.19. As the population is exposed to these changes in the environment, the internal properties show a higher variance over each fed-batch phase in Figure 5.17 as well. This holds true for both variables and contrasts with the behavior of a population in the bioreactor system with a feed into the STR. Since the population in the latter case is exposed to environment, in which substrate gradient remains nearly constant over the course of the fed-batch process, the decrease in variance of the properties is due to the effects of micro-mixing. In the current feeding regime, the environment of the bioreactor system changes, thus the variance of the properties are dominated by the effect. With a higher substrate gradient in general, the variance is larger over the whole process.

The two process options of a feed into the PFR and a higher flow rate of $Q_R = 20$ L/h can be combined and the results analyzed. Similar to the earlier simulation in Section 5.4.1, a higher flow rate leads to smaller residence times and substrate gradient, which can be seen in Appendix, Figure A.20, as well as the creation of more particles with unbalanced internal properties due to these environmental conditions. Therefore, the distributions of protein content considering all particles of the bioreactor system in Figure 5.18A are narrower and nosier without a clear distinction into subpopulations. The same applies to the metabolite content which shows a smaller range of values in Figure 5.18B and the local distributions overlap in Figure 5.18E. With the presence of more unbalanced particles, the growth rate of the particles have a larger range and even exhibits negative growth rates in Figure 5.18C. This is also the main difference to a population in a bioreactor system with a lower flow rate and even particles exposed to a higher substrate environment upon entering compartment 2 might grow with negative rate, which is depicted in Figure 5.18F. Nevertheless, high values of growth rate are only present in particles residing in a high substrate environment as it was the case for a flow rate of $Q_R = 1$ L/h.

The smaller substrate gradient has a significant effect on the time course of the variance of internal properties. Figure 5.18G shows a comparison of the variance for the different flow rates with a feed into the PFR. While the variance for lower flow rate follows the increase

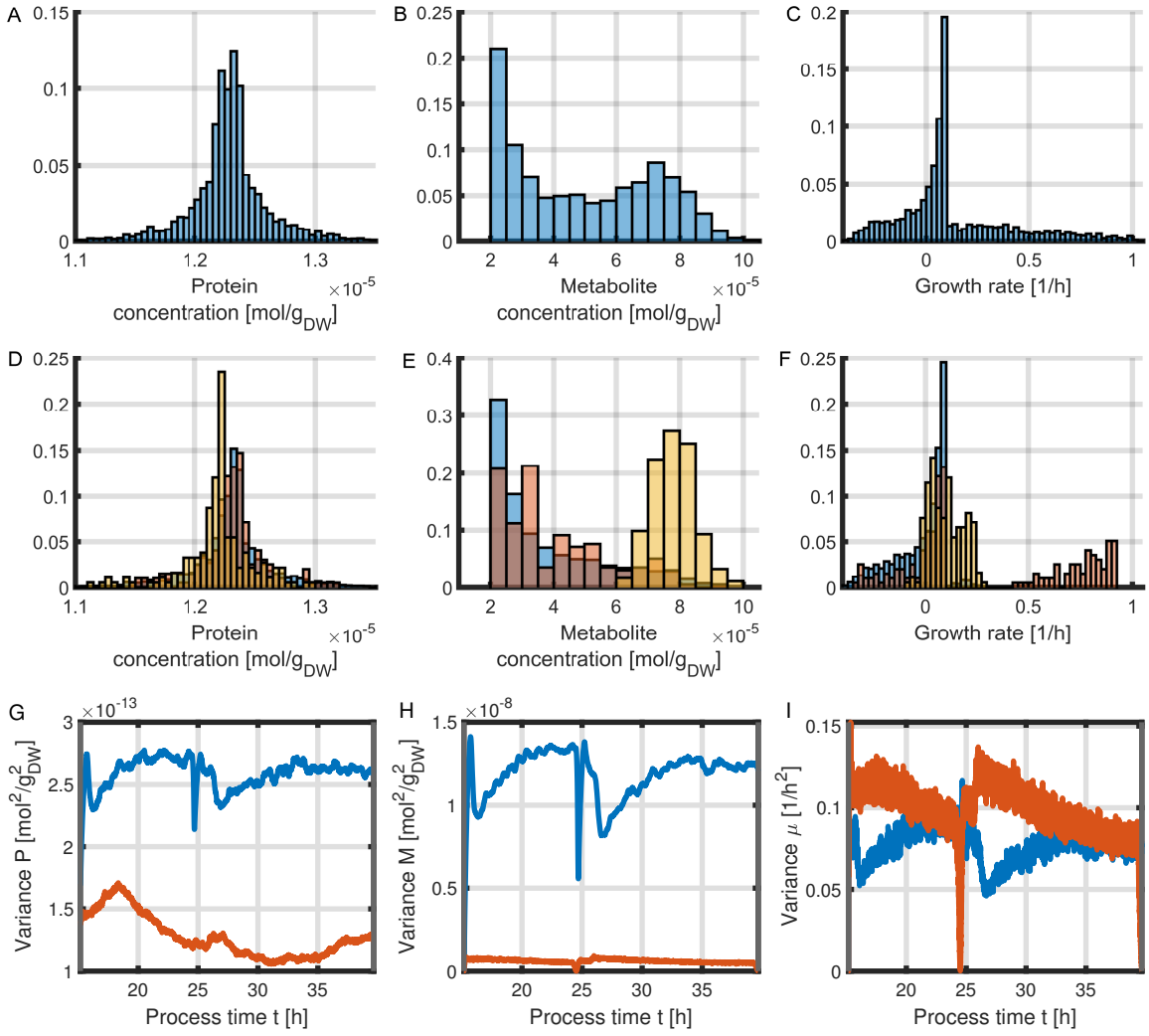

Figure 5.18: Comparison of the distributions of protein concentration (A,D), metabolite concentration (B,E) and growth rate (C,F) at $t = 38$ h with a feed into the PFR and flow rate of $Q_R = 20$ L/h. Normalized distribution for all particles over the whole bioreactor system (A, B, C) and for different compartments (D,E F; compartment 1 STR (blue), compartment 2 PFR inlet (orange) and compartment 11 PFR outlet (yellow)). Time course of the variance of the protein concentration (G), metabolite concentration (H) and growth rate (I).

of heterogeneity in the environment, the variance for higher flow rates does not follow this trend and is a result of all effects superimposed. These effects being intrinsic noise due to cell division, physical limitations of the substrate transport to the cell surface and the substrate gradient present in the environment. Interestingly, the variance of metabolite concentration in Figure 5.18H as well as the variance for the growth rate in Figure 5.18I resemble the variances in the case of a feed into the STR, but with the same high flow rate of $Q_R = 20$ L/h. This emphasizes the relation between the residence time and relaxation times of the internal properties as an important factor on population heterogeneity.

5.5 Short summary and discussion

Previous works approached the investigation of population heterogeneity with the focus on the effect of external noise [79, 118, 97, 51]. The environment of a large-scale bioreactor is recreated through the employment of a scaled-down bioreactor system, consisting of a well-mixed STR and a PFR in bypass, which is a common choice in order to study environmental changes [115, 167, 57]. The split into two environments enables a formation of substrate gradients during a fed-batch process with a feed inlet at the STR or the PFR. Optionally, pulse experiments can be conducted, but these scenarios were not investigated in the scope of this work. The dynamics of the environment is approximated by compartimentalization of the bioreactor system instead of using CFD to simulate the hydrodynamics of the bioreactor system. It has been shown that CM approaches provide a good compromise between computational effort and an accurate depiction of the hydrodynamics inside the bioreactor system [28, 112, 68].

This work provides improvements as it deals with both external sources, induced by substrate gradients in the environment, and intrinsic noise, due to redistribution of cellular content taking place at cell division. The biological phase is considered as a population of particles, each carrying a specific biomass and representing a group of biological cells with the same internal properties. The addition of the biomass carried by a virtual particle as internal property instead of a virtual particle representing a fixed number of biological cells serves the purpose of keeping the number of virtual particles constant over the course of the simulation. Otherwise, the number of virtual particles would increase exponentially, as the population is in balanced growth during the biomass production phase, which creates a bigger challenge when implementing the model.

The dynamical model of biological processes in the cells is based on a mass conservative coarse-grained approach from Chapter 3, where the growth rate is a result of the intracellular network rather than being described through an empirical function of external concentrations. An essential feature of the model is the substrate uptake rate calculation which rely upon both the environmental supply and the cell physiological state and, therefore, we move away from employing a black-box kinetic model towards a mechanistic model. The investigation of the time scales of the dynamical model, already obtained in Chapter 3 in Section 3.6.3 revealed two different time scales in the cell response when the cells are out of equilibrium (uptake rate, internal composition and growth rate are not balanced), and only one time scale when cells at equilibrium respond to external fluctuations. Without the inclusion of intrinsic noise due to cell division, the internal properties of the particles clearly reflect their local environment and the population is clearly divided into two subpopulations: a subpopulation of growing particles present in the STR and non-growing particles present in the starvation zone of the PFR. The only deciding factor for the formation of distinct subpopulation is the relation between the mean residence time in the compartments and the time scale of cellular response to external fluctuations. This finding shows the necessity to include intrinsic noise, which result in randomly distributed cells regarding their cell state, especially their protein content, when simulating a population in an changing environment in order to maintain different time scales in cellular response and decouple the cellular state from the local concentrations in the liquid phase. Through redistribution of cell properties, the effects of cell divisions is implemented without changing the number of particles as well. The new cell properties are chosen according to a Gaussian distribution, centered on the mother cell properties. Special care must be taken during the sampling of particles whose internal cell properties are redistributed. It has been shown that the choice of sampling weights affects the shape of the protein concentration as well as the growth rate dis-

tribution. The standard deviation of the Gaussian distribution on the other hand, only widens or narrows the protein concentration and growth rate distribution, while leaving weighted mean values unchanged. Other types of partitioning functions can be investigated and confronted to experimental results which are however difficult to obtain.

The time scales of mixing compared to the time scales of cellular response is an important factor on the extent of population heterogeneity, as mentioned earlier. For intense mixing, a broad distribution of growth rates for similar cellular compositions is observed, which is induced by very small substrate gradients. This is due to the fact that the majority of the population consist of unbalanced cells and their cellular content is not optimally adapted to their local environment leading to a diminished growth rate. Simultaneously, this also allows the development of fast growing particles, which is possible for particles reentering the STR. These results emphasize the necessity to approach population dynamics with segregated models, particularly when facing dynamic environments.

The scenarios of starvation and exposure to high substrate concentration through the simulation of a scaled-down bioreactor system have been analyzed separately, but the combination of the two can also be investigated in the same framework. The simulation of exposure to high substrate concentrations is accomplished through a feed into the inlet of the PFR resulting in higher substrate gradients. The assumption of increased population heterogeneity is confirmed by the increase of variance in the growth rate and protein concentration distribution. In this scenario, the effects of external noise dominates over the effects of intrinsic noise. With increasing flow rate of the liquid phase, the effects of intrinsic noise are more prominent again. Thus, we can conclude that the main factors influencing the extent of population heterogeneity are: the magnitude of the substrate gradient, the mean residence time in the compartments compared to the time scale of the internal properties to return to equilibrium in presence of intrinsic noise due to cell division.

Due to the numerical nature of the proposed framework, problems in maintaining positivity in the liquid phase can occur. This problem may arise during a time step in a compartment with very low substrate availability. Here, particles present in this compartment register the same concentration during the time step and the procedure of the consumption of substrate happens simultaneously leading to a possible higher sum of consumed substrate than available substrate. Therefore, the consumption of substrate of a particle has to be slowed down significantly in environments of low substrate. Here, we used a untypically large time scale for micro-mixing to maintain positivity. Smaller parameter values for this time scale can be used, but with higher computational effort through smaller time steps. Further improvements may be achieved by employing other strategies or functions to describe limiting factors on the substrate uptake.

The gradients applied in this work are purely on substrate level. Thus, it can be said that sufficient oxygen levels were assumed. In the experiments of the reference process, oxygen availability has proven to be very influential on the productivity of the process. In order to validate the results with fluorescent data linked to specific cellular traits, which can be obtained experimentally using multiple reporter strains, the model has to be extended to include other gradients such as oxygen. It can be a limiting factor, especially in the compartments of the PFR if the PFR is not aerated. In this case even in nutrient rich environments, substrate cannot be consumed by the particles at maximal rate. But due to the set-up of the framework of the Euler-Lagrange approach, it can easily extended through additional reaction rates in the coarse-grained description of the biological phase and an addition of the oxygen as second variable in the liquid phase. Here, one can posit the question if a third phase, the gas phase, is necessary to

be incorporated to depict the full dynamics of the scaled-down bioreactor system. Furthermore, this proposed framework can be applied to the full bioprocess including the L-phenylalanine production phase to study the effects of heterogeneity on the productivity.

6 Conclusion & Outlook

The main intention for this work was the development of a framework to model a segregated population in a bioprocess and subsequent investigation of emerging population heterogeneity induced by external and intrinsic noise. External noise occur especially in industrial-scale bioreactors due to mass transfer limitations and mixing insufficiencies leading to substrate gradients in the bioreactor the cells are exposed to. Intrinsic noise originates from the stochasticity of biochemical processes resulting in different expression levels of proteins, different sizes and structures. This work aimed to consider both the influence of substrate gradients and intrinsic noise due to cell division. The classical approach in biotechnology is to combine models describing the cellular environment in a typically well-mixed bioreactor with models describing all the reactions related to biomass. But these models average the values over the whole population, thus making it impossible to describe the processes in single cells. In general, there are two approaches that can quantify the degree of heterogeneity, which were both presented in the course of this work: a PBE in Chapter 4 and an Euler-Lagrange approach in Chapter 5. With special attention on biotechnological purposes, all simulations of each model set-up are performed using a L-phenylalanine production process with *E. coli* as reference process, where experimental data was obtained from [61] to compare the results with a special focus on the biomass production phase.

For both approaches, a model for the intracellular network of the cells was required. Since both model approaches are computational demanding on their own, the intracellular model had to be sufficiently simple, but still capture the most important cellular processes. For this purpose, the employment of a coarse-grained model was an appropriate choice, which has become popular to investigate the interdependence of the growth rate and the gene expression [106, 26, 163, 170]. A minimal coarse-grained model was developed in Chapter 3, which contains a structure in accordance to models of other works [143, 14] and an allocation of the proteome as in [8, 44, 142, 141]. The environment encountered by the cell population has an effect on the proteome allocation and gene expression [141] and this relation is a key property implemented in the coarse-grained model. Through the self-replicating nature of the model, the substrate uptake of the cell is regulated depending on the cells demand and state. In a nutrient rich environment, the substrate uptake rate is down-regulated due to a subsequent high pool of metabolites which leads to a higher ribosomal protein fraction and the synthesis of macromolecular cell components, such as proteins, is up-regulated. During times of starvation, the metabolites are exhausted and a higher protein fraction is allocated to transporter proteins in order to increase the uptake of substrate and conversion into metabolites. Furthermore, the growth rate of the cells is determined by the composition of the proteome and cell components and not by a empirical function dependent solely on the environment. This model was applied to the mentioned L-phenylalanine production process and was able to reproduce the dynamics of the process in line with the experimental data in Section 3.4.

The novelty of this work lies in the coupling of the coarse-grained model description of the intracellular network of the population based on strict mass conservation and the framework of

each a PBE and an Euler-Lagrange approach. While segregated model descriptions employing a PBE focused either only on the cell size or cell mass [54, 101], growth rate [108, 125] or only the intracellular components of the cells [145, 146, 147], the integration of the mass conservating coarse-grained model and the usage of cell mass and protein content as descriptor variables of the PBE enabled the investigation of microbial growth as depicted in e.g. Figure 4.5A under consideration of the cellular composition in Figure 4.4 and intrinsic noise due to cell division. Using a Finite Volume method, which was proven to be conserving the first moments, a fed-batch process was simulated and its results analyzed. The mean values of all involved variables did not differ between the segregated and unsegregated model, while differences emerge with the addition of the feed in Figure 4.3.

The proposed model has a few downsides. Due to the high computational effort of numerically solving a PBE, the number of variables had to be reduced to a total of two, whereas the original minimal coarse-grained model required a total of three variables: the metabolite and protein content as well as the cell mass. Additionally, with the inclusion of cell division the structure of the cell composition had to be adjusted as explained in Section 4.1. This complicates the interpretation and comparison of the reaction rates and intracellular components. The allocation into protein fractions T and R is still determined by a function of M given by Eq. (3.43) and (3.44), but with the changed interpretation of M, representing all cellular components except for T and R, the intracellular pool M in this model can not be directly linked to the metabolite pool M of the original model in Section 3.4.1. Therefore, the allocation function is not directly based on experimental data [139] and a new relation between M and the protein function had to be proposed. The new relation was set through an estimation using the range of P, T and R. Despite the modifications, the self-replicating nature of the model was kept for the modified version.

In case of the presented Euler-Lagrange model approach in Chapter 5, the equations of the intracellular network given by the coarse-grained model were integrated into the framework of a model with two phases describing the biological phase and liquid phase as proposed by Lapin et al. [81]. Thus, the approach inherits the advantages of the coarse-grained model being based on mass conservation and having the growth being a result of the network rather then being determined an empirical function. In line with the "Eulerian" approach, the biological phase was considered through a ensemble of virtual particles representing biological cells with the same cellular properties (intracellular content and location). The dynamics of the intracellular content (protein and metabolite concentration) of each particle are given through the intracellular network of Chapter 3. Therefore, each particle reacts according to the environment it perceives and the dynamics of the intracellular network and the liquid phase as well as the particles movement in the bioreactor can be characterized separately. This enables a more complex description of the intracellular network in contrast to the PBE approach, which has been done for a penicillin model containing multiple intracellular pools in [51] with a focus on the ability to describe the hydrodynamics with CM compared to a full CFD simulation. The movement of the particles followed a stochastic transport algorithm, as seen in [103, 27, 51], based on the residence time distribution. Due to the easier implementation of this framework, this approach was utilized to describe a scaled-down bioreactor system consisting of a STR and a PFR in bypass through a CM. The expansion of the PBE model to include the description of a bioreactor system can be done as well resulting in an Euler-Euler approach. Here, a PBE can be set up for each compartment of the bioreactor with an additional term for the influx and outflux of the population entering or leaving the compartment as done in [125]. But in doing so, the computational effort, which is already high, multiplies with the number of compartments

and, due to the more promising framework of the Euler-Lagrange approach, we refrained from implementing a Euler-Euler approach.

One origin of intrinsic noise lies in the process of cell division. The PBE is able to describe cell divisions, but a formulation for the cell division had to be assumed. In this work, we utilized a smooth function involving critical values following some previous works [98, 99, 129]. Thus, additional information about the critical values, where cell division at $m = 2\ m_c$ and $n = 2\ n_c$ is almost guaranteed, is required and generally difficult to obtain. In the simulations of this work, the critical values did not influence the results during the batch phase or in case of a constant nutrient rich environment, where the population resides in the upper left corner of the n-m-state space. In case, where the population moved away from the upper left corner, the critical values affected the results significantly. In the framework of the Euler-Lagrange approach, cell division is not modeled directly. In order to maintain a constant number of virtual particles for easier implementation, a virtual particle represents a total amount of biomass instead of a fixed number of biological cells. Furthermore, instead of modeling cell division itself, assuming division laws and finding a way to keep the number of virtual particles constant, the stochasticity arising from cell division was modeled as described in Section 5.1.3. Here, intrinsic noise due to cell division was manifested in the reshuffling of cell properties with the same frequency of cell divisions that happen during a time step according to the growth rates of the particles. The new properties of the particles were determined by a Gaussian distribution centered around the cell properties of the original particle. This sets this framework apart from previous works which only considered external noise [112, 51].

With intrinsic noise present, the cell state of the particles is decoupled from the local environment. Here, we introduced the concept of balanced and unbalanced cells. We used the term balanced cell if the intracellular composition of the particle is in steady state with the environment. It was shown in Section 5.3 and Section 5.4.1 that the degree of population heterogeneity is dependent on the time scale of mixing compared to the time scale of cellular response indicated by the relaxation time of the variables of the system. The cellular response happens on two different time scales depending on the type of perturbation. In Section 3.6.2, it was shown that the relaxation time for both metabolites and proteins is in the range of duzens of seconds in response to environmental changes for cells in balanced state. Thus, as expected, in absence of intrinsic noise in Section 5.2, the particles were in balanced state and able to quickly adapt to changes in the environment leading to distinct subpopulations reflecting their environment. The simulation of the bioreactor system exhibited a growing subpopulation in the compartment of the STR and a non-growing subpopulation in the PFR. The relaxation times of the metabolites and proteins to intrinsic noise in Section 3.6.3 were determined to happen on two different scales, where the metabolites converged within the same time as the relaxation time for external noise and the proteins required hours to converge as seen in Figure 3.12. This was expected as the proteins are of much higher molecular weight and its synthesis time is much longer than for metabolites. Due to the equal molecular weight in the modified model used for the PBE, no time scale analysis was performed for this case. With the addition of intrinsic noise on top of external noise in Section 5.3, the reshuffling of the internal properties due to cell division created unbalanced particles. This resulted in the decoupling of the cellular state in general from the environment the particle perceives with the protein content having a broad distribution due to the longer relaxation time and the metabolite content corresponding to the local substrate concentration in Figure 5.6. Thus, population heterogeneity is enhanced with the inclusion of intrinsic noise and these results further showed the necessity to describe the growth rate by other means than a function only dependent on the substrate concentration, which is in

agreement with other studies [1, 91]. Furthermore, it is consistent with the choice of the growth rate as an internal variable subject to a redistribution mechanisms at cell division in PBEs as seen in [108, 109, 110, 125].

With the established framework several scenarios were simulated and compared. Different flow rates through the bioreactor system, which means different mean residence times in the compartments, affected the degree of heterogeneity as mentioned. But most interestingly populations in a bioreactor with high flow rates, which result in a nearly homogeneous environment, are highly heterogeneous in regards to their growth rate distribution. This result further emphasizes the need to consider segregated models. Additionally, different feeding regimes were investigated resembling events of substrate excess.

The batch phase of the L-phenylalanine process which was simulated in this work as exemplary process, can be compared for the different model approaches. During this process phase the environment of the bioreactor system can be regarded as homogeneous with very small external noise due to stochasticity of the particle movement which can be neglected. Thus, only intrinsic noise is present. The mean values of the variables did not differ for PBE approach in Section 4.2 nor the Euler-Lagrange approach in Section 5.3 in comparison to the unsegregated simulation in Section 3.4.3. In both segregated approaches, the initial population was assumed to be normally distributed and therefore is formed of unbalanced cells. Because of the long relaxation time, the population remained normally distributed over the whole batch process phase in both approaches.

Due to the structure of the model, it can be extended to consider additional factors that affect the degree of heterogeneity and different bioreactor systems as well. In future works, the framework can be adapted to simulate the scenarios, which were experimentally performed by Neubauer et al. [114]. Here, fed-batch processes with the two different feeding profiles were investigated with the difference that the feed was constant. The PFR in this set-up was aerated and therefore no oxygen limitation was assumed. Other possibilities include investigations of large-scale bioreactors [51], different scaled-down bioreactor systems consisting of two STRs [122, 83] or short time responses to pulses of substrate in a chemostat culture [104, 161]. In the first case, the CM could be extended to a third dimension, where every compartment can have up to 6 neighbors. Thus, the matrix containing the flow field has to be adapted, but the transport of particles and description of the biological phase remain unchanged. As the hydrodynamics for large-scale bioreactors are more complex, insights of full CFD simulations as basis for the flow field matrix can be helpful [51, 28, 27].

In the current set-up of the bioprocess, the bypass is not aerated and oxygen limitations do occur. In the examplary L-phenylalanine production process, the low biomass to substrate yield and the absence of byproduct formation suggests a high respiration rate and this can heavily affect the productivity of the process. As mentioned in Section 5.5, we only considered one substrate gradient. In order to incorporate the full respiration process and the dependency of growth on oxygen availability, the coarse-grained model must be extended. A possibility is to incorporate a dependency of the substrate uptake rate on the oxygen in the liquid phase. The liquid phase present in the Euler-Lagrange model has to be extended to include the dissolved oxygen as well. Analogous to substrate availability, special care has to be taken such that oxygen levels do not fall below zero due to numerical errors. During energy production, the cells oxidize carbon-containing substrates to carbon dioxide. The oxygen uptake rate and carbon dioxide production rate can then be compared to experimental data and verified.

The established framework can be employed with the aim to optimize the product formation through a better understanding of the effects of population heterogeneity in industrial scale bioreactors. Currently, the model description and simulations of the segregated population only captured the biomass production phase of the L-phenylalanine production process, while the product formation phase is already implemented in the coarse-grained model in Section 3.4.3. After incorporation of oxygen dependent growth, the simulated process with the Euler-Lagrange approach can be extended to include the last process phase as well. For this, the liquid phase is expanded with a third scalar containing the L-phenylalanine concentration.

While we presented a promising framework to investigate population heterogeneity, it should be verified experimentally. The results were only compared to the mean values of the biomass, substrate concentration and growth rate provided by experimental data. Through the application of multi reporter strains monitoring growth, general stress response, oxygen limitation and L-phenylalanine production, fluorescent and proteome data can be integrated into the model and compared to the distributions of the proteome fractions corresponding to property of interest. While the fraction R can be directly linked to the stain monitoring growth and the fraction F_P to the stain monitoring the product formation, oxygen availability perceived by the cell and general stress do not have a direct counterpart in the model. Nevertheless, the interplay between at least growth, product formation and global oxygen availability can be evaluated in this way. Proteomics data, on the other hand, can be used to refine the intracellular network model to consider regulatory processes that may vary strongly during experiments, since some proteins might be activated or deactivated under certain stress conditions. This can give a hint towards which evolutionary strategies cells employ to cope with a fluctuating environment. Generally, high consistency between the simulation and experimental data leads to better prediction accuracy.

A well-established model is the foundation for further optimization to achieve higher product yields. One possibility is to find an optimal feeding strategy depending on the predicted property distributions of the cell population. Another possibility to increase productivity lies in the search of an optimal feeding rate through process optimization of the non-segregated model. The current process is divided into three process phases with a total of three switching time points, at which the feeding rate changes. These time points as well as the substrate concentration of the feed can be evaluated and improved based on optimization studies.

To conclude, segregated model approaches to investigate population heterogeneity induced by external and intrinsic noise due to cell division were presented and analyzed. Due to the flexible structure, it can be applied to various microbial populations in different cultivation conditions.

List of Variables and Functions

Symbol	Description	Unit	1st page
A	L-tyrosine concentration	g/L	30
c	Vector containing all intracellular components C_i	mol/g_{DW}	4
C_i	Intracellular components C_i	mol/g_{DW}	4
d	Dimension of domain Ω_ξ	-	11
F	L-phenylalanine concentration	g/L	29
$\mathcal{F}$	Matrix containing flow field of the CM	L/h	17
$\mathcal{F}_{i,j}$	Exit flow rate from compartment i to j	L/h	17
F_P	Proteins for L-phenylalanine production	mol/g_{DW}	29
M	Metabolite concentration	mol/g_{DW}	24
m	Total biomass of a particle in the CM	g_{DW}	60
m_X	Biomass weight	g_{DW}	4
N	Number of particles in the CM	-	60
$N(\xi, x, t)$	Number density function of a PBE	-	11
N_C	Number of intracellular components	-	4
n_{C_i}	Molar mass of component C_i	mol	4
$N_{p,i}$	Number of particles in compartment i	-	21
N_r	Number of intracellular reactions	-	4
$\mathcal{N}$	Stoichiometric matrix of the intracellular network	-	4
n_s	Number of extracellular concentrations in the CM	-	17
$N_t(t)$	Total number of particles of a PBE	-	12
n_z	Number of compartments in the CM	-	17
O	Acetate concentration	g/L	31
P	Protein concentration	mol/g_{DW}	24
$p(\xi, \xi', x)$	Partition probability density in a PBE	-	12
Q	Residual protein concentration	mol/g_{DW}	27
q_{in}	Feeding rate	L/h	31
Q_R	Flow rate of circulation in the bioreactor system	L/h	60
R	Ribosomal proteins	mol/g_{DW}	6
r	Reaction rate vector of the intracellular network	mol/g_{DW}h	4
$r_\xi(\xi, x)$	Rate of change of the internal properties ξ	-	11
$r_x(\xi, x)$	Rate of change of the external variables x	-	11
r_C	Rate of respiration	mol/g_{DW}h	29
r_F	Rate of L-phenylalanine production	mol/g_{DW}h	29
r_O	Rate of overflow metabolism	mol/g_{DW}h	24
r_P	Protein synthesis rate	mol/g_{DW}h	24
r_T	Substrate uptake rate	mol/g_{DW}h	24
r_U	Synthesis rate of residual biomass	mol/g_{DW}h	29

Symbol	Description	Unit	1st page
S	Substrate concentration	g/L	24
S_{in}	Feed concentration	g/L	5
$\mathcal{S}$	External concentrations	g/L	5
T	Transporter proteins	mol/g_{DW}	6
t	Time	h	3
U	Residual biomass concentration	mol/g_{DW}	29
V_{PFR}	Bioreactor volume of the PFR	L	60
V_R	Bioreactor volume	L	5
V_{STR}	Bioreactor volume of the STR	L	60
W	Matrix containing mass fractions of cell components	-	25
w	Vector containing the molecular weights of intracellular components	g/mol	4
x	External variables / location of a particle	-	11
$\Gamma(\xi, x)$	Division rate in the PBE	-	12
μ	Specific growth rate	1/h	3
ξ	Internal property of a particle	-	11
τ_r	Mean residence time	h	19
τ_u	Characteristic time for substrate uptake	h	35
τ	Relaxation time	h	37
Φ	Matrix containing transfer rates of external concentrations $\mathcal{S}$	mol/g_{DW}h	18
ϕ_S^b	Internal biological uptake capacity	mol/g_{DW}h	35
ϕ	Production or consumption rate of external concentrations $\mathcal{S}$	mol/g_{DW}h	13
Ω_ξ	Domain of the internal property ξ	-	11
Ω_x	Domain of the external variables	-	11

List of Abbreviations

E. coli	*Escherichia coli*
CFD	Computational fluid dynamics
CM	Compartment model
ODE	Ordinary differential equation
PBE	Population balance equation
PFR	Plug-flow bioreactor
STR	Stirred-tank bioreactor

Bibliography

[1] E. M. Abulesz and G. Lyberatos. Periodic operation of a continuous culture of baker's yeast. *Biotechnol. Bioeng.*, 34(6):741–749, 1989.

[2] M. Ackermann. A functional perspective on phenotypic heterogeneity in microorganisms. *Nat. Rev. Microbiol.*, 13(8):497–508, 2015.

[3] J. Almquist, M. Cvijovic, V. Hatzimanikatis, J. Nielsen, and M. Jirstrand. Kinetic models in industrial biotechnology - improving cell factory performance. *Metab. Eng.*, 24:38–60, 2014.

[4] A. A. Alonso, I. Molina, and C. Theodoropoulos. Modeling bacterial population growth from stochastic single-cell dynamics. *Appl. Environ. Microbiol.*, 80(17):5241–5253, 2014.

[5] T. B. Alter, L. M. Blank, and B. E. Ebert. Proteome regulation patterns determine escherichia coli wild-type and mutant phenotypes. *mSystems*, 6(2), 2021.

[6] V. I. Arnold. *Ordinary differential equations*. Universitet. Springer, Berlin, 1st ed., 2nd printing edition, 2006.

[7] J. Baldyga and J. R. Bourne. *Turbulent mixing and chemical reactions*. John Wiley & Sons, Chichester, 1999.

[8] U. Barenholz, L. Keren, E. Segal, and R. Milo. A minimalistic resource allocation model to explain ubiquitous increase in protein expression with growth rate. *PloS One*, 11(4):e0153344, 2016.

[9] J. C. Barrett and J. S. Jheeta. Improving the accuracy of the moments method for solving the aerosol general dynamic equation. *J. Aerosol Sci.*, 27(8):1135–1142, 1996.

[10] M. Basan, S. Hui, H. Okano, Z. Zhang, Y. Shen, J. R. Williamson, and T. Hwa. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nat.*, 528(7580):99–104, 2015.

[11] G. I. Bell and E. C. Anderson. Cell growth and division. *Biophys. J.*, 7(4):329–351, 1967.

[12] A. Ben-Israel and T. N. E. Greville. *Generalized inverses: Theory and applications*, volume 15 of *CMS Books in Mathematics*. Springer, New York, 2nd edition, 2012.

[13] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. van Dien, and J. D. Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. *Nat. Chem. Biol.*, 5(8):593–599, 2009.

[14] F. Bertaux, J. von Kügelgen, S. Marguerat, and V. Shahrezaei. A bacterial size law revealed by a coarse-grained model of cell physiology. *PLoS Comput. Biol.*, 16(9):e1008245, 2020.

[15] F. Bezzo, S. Macchietto, and C. C. Pantelides. General hybrid multizonal/CFD approach for bioreactor modeling. *AIChE J.*, 49(8):2133–2148, 2003.

[16] S. Bhoi and D. Sarkar. Hybrid finite volume and monte carlo method for solving multi-dimensional population balance equations in crystallization processes. *Chem. Eng. Sci.*, 217:115511, 2020.

[17] D. Binder, T. Drepper, K.-E. Jaeger, F. Delvigne, W. Wiechert, D. Kohlheyer, and A. Grünberger. Homogenizing bacterial cell factories: Analysis and engineering of phenotypic heterogeneity. *Metab. Eng.*, 42:145–156, 2017.

[18] T. Bollenbach, S. Quan, R. Chait, and R. Kishony. Nonoptimal microbial response to antibiotics underlies suppressive drug interactions. *Cell*, 139(4):707–718, 2009.

[19] E. Bosdriesz, D. Molenaar, B. Teusink, and F. J. Bruggeman. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *The FEBS J.*, 282(10):2029–2044, 2015.

[20] S. Bove, T. Solberg, and B. H. Hjertager. A novel algorithm for solving population balance equations: the parallel parent and daughter classes. derivation, analysis and testing. *Chem. Eng. Sci.*, 60(5):1449–1464, 2005.

[21] D. Boyer, G. Tarjus, and P. Viot. Shattering transition in a multivariable fragmentation model. *Phys. Rev. E, Stat. physics, plasmas, fluids, related interdisciplinary topics*, 51(2):1043–1046, 1995.

[22] N. A. Cookson, S. W. Cookson, L. S. Tsimring, and J. Hasty. Cell cycle-dependent variations in protein concentration. *Nucleic Acids Res.*, 38(8):2676–2681, 2010.

[23] R. Courant, E. Isaacson, and M. Rees. On the solution of nonlinear hyperbolic differential equations by finite differences. *Commun. on Pure Appl. Math.*, 5(3):243–255, 1952.

[24] A. Das, S. Bhoi, D. Sarkar, and J. Kumar. Sonofragmentation of rectangular plate-like crystals: Bivariate population balance modeling and experimental validation. *Cryst. Growth & Des.*, 20(8):5424–5434, 2020.

[25] K. M. Davis and R. R. Isberg. Defining heterogeneity within bacterial populations via single cell approaches. *BioEssays : news reviews molecular, cellular developmental biology*, 38(8):782–790, 2016.

[26] H. de Jong, S. Casagranda, N. Giordano, E. Cinquemani, D. Ropers, J. Geiselmann, and J.-L. Gouzé. Mathematical modelling of microbes: metabolism, gene expression and growth. *J. Royal Soc. Interface*, 14(136), 2017.

[27] A. Delafosse, S. Calvo, M.-L. Collignon, F. Delvigne, M. Crine, and D. Toye. Euler-Lagrange approach to model heterogeneities in stirred tank bioreactors – comparison to experimental flow characterization and particle tracking. *Chem. Eng. Sci.*, 134:457–466, 2015.

[28] A. Delafosse, M.-L. Collignon, S. Calvo, F. Delvigne, M. Crine, P. Thonart, and D. Toye. CFD-based compartment model for description of mixing in bioreactors. *Chem. Eng. Sci.*, 106:76–85, 2014.

[29] F. Delvigne, J. Baert, H. Sassi, P. Fickers, A. Grünberger, and C. Dusny. Taking control over microbial populations: Current approaches for exploiting biological noise in bioprocesses. *Biotechnol. J.*, 12(7), 2017.

[30] F. Delvigne, M. Boxus, S. Ingels, and P. Thonart. Bioreactor mixing efficiency modulates the activity of a prpos::gfp reporter gene in e. coli. *Microb. Cell Factories*, 8(1):1–17, 2009.

[31] F. Delvigne, J. Destain, and P. Thonart. Toward a stochastic formulation of microbial growth in relation to bioreactor performances: case study of an *E. coli* fed-batch process. *Biotechnol. Prog.*, 22(4):1114–1124, 2006.

[32] F. Delvigne and P. Goffin. Microbial heterogeneity affects bioprocess robustness: dynamic single-cell analysis contributes to understanding of microbial populations. *Biotechnol. J.*, 9(1):61–72, 2014.

[33] O. Diekmann, H. J. A. M. Heijmans, and H. R. Thieme. On the stability of the cell size distribution. *J. Math. Biol.*, 19(2):227–248, 1984.

[34] D. T. Doan, M. D. Hoang, A.-L. Heins, and A. Kremling. Applications of coarse-grained models in metabolic engineering. *Front. Mol. Biosci.*, 9, 2022.

[35] D. Dubnau and R. Losick. Bistability in bacteria. *Mol. Microbiol.*, 61(3):564–572, 2006.

[36] R. Dürr and A. Bück. Approximate moment methods for population balance equations in particulate and bioengineering processes. *Process.*, 8(4):414, 2020.

[37] R. Dürr, T. Müller, and A. Kienle. Efficient dqmom for multivariate population balance equations and application to virus replication in cell cultures. *IFAC-PapersOnLine*, 48(1):29–34, 2015.

[38] S. O. Enfors, M. Jahic, A. Rozkov, B. Xu, M. Hecker, B. Jürgen, E. Krüger, T. Schweder, G. Hamer, D. O'Beirne, N. Noisommit-Rizzi, M. Reuss, L. Boone, C. Hewitt, C. McFarlane, A. Nienow, T. Kovacs, C. Trägårdh, L. Fuchs, J. Revstedt, P. C. Friberg, B. Hjertager, G. Blomsten, H. Skogman, S. Hjort, F. Hoeks, H. Y. Lin, P. Neubauer, R. van der Lans, K. Luyben, P. Vrabel, and A. Manelius. Physiological responses to mixing in large scale bioreactors. *J. Biotechnol.*, 85(2):175–185, 2001.

[39] M. H. Ernst and G. Szamel. Fragmentation kinetics. *J. Phys. A: Math. Gen.*, 26(22):6085–6091, 1993.

[40] A. G. Franchini and T. Egli. Global gene expression in *Escherichia* coli k-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. *Microbiol.*, 152(Pt 7):2111–2127, 2006.

[41] A. G. Fredrickson and N. V. Mantzaris. A new set of population balance equations for microbial and cell cultures. *Chem. Eng. Sci.*, 57(12):2265–2278, 2002.

[42] A. G. Fredrickson and H. M. Tsuchiya. Continuous propagation of microorganisms. *AIChE J.*, 9(4):459–468, 1963.

[43] J. R. Garcia, H. J. Cha, G. Rao, M. R. Marten, and W. E. Bentley. Microbial *nar*-GFP cell sensors reveal oxygen limitations in highly agitated and aerated laboratory-scale fermentors. *Microb. Cell Factories*, 8(1):1–7, 2009.

[44] N. Giordano, F. Mairet, J.-L. Gouzé, J. Geiselmann, and H. de Jong. Dynamical allocation of cellular resources as an optimal control problem: Novel insights into microbial growth strategies. *PLoS Comput. Biol.*, 12(3):e1004802, 2016.

[45] R. González-Cabaleiro, A. M. Mitchell, W. Smith, A. Wipat, and I. D. Ofiţeru. Heterogeneity in pure microbial systems: Experimental measurements and modeling. *Front. Microbiol.*, 8:1813, 2017.

[46] C. González-Figueredo, R. Alejandro Flores-Estrella, and O. A. Rojas-Rejón. Fermentation: Metabolism, kinetic models, and bioprocessing. In N. Shiomi, editor, *Current Topics in Biochemical Engineering*. IntechOpen, [Erscheinungsort nicht ermittelbar], 2019.

[47] R. Grima, P. Thomas, and A. V. Straube. How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations? *The J. Chem. Phys.*, 135(8):084103, 2011.

[48] A. J. Grimbergen, J. Siebring, A. Solopova, and O. P. Kuipers. Microbial bet-hedging: the power of being different. *Curr. Opin. Microbiol.*, 25:67–72, 2015.

[49] S. Han, F. Delvigne, A. Brognaux, G. E. Charbon, and S. J. Sørensen. Design of growth-dependent biosensors based on destabilized GFP for the detection of physiological behavior of *Escherichia coli* in heterogeneous bioreactors. *Biotechnol. Prog.*, 29(2):553–563, 2013.

[50] C. Haringa, W. Tang, A. T. Deshmukh, J. Xia, M. Reuss, J. J. Heijnen, R. F. Mudde, and H. J. Noorman. Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines. *Eng. Life Sci.*, 16(7):652–663, 2016.

[51] C. Haringa, W. Tang, and H. J. Noorman. Stochastic parcel tracking in an Euler-Lagrange compartment model for fast simulation of fermentation processes. *Biotechnol. Bioeng.*, 2022.

[52] C. Haringa, W. Tang, G. Wang, A. T. Deshmukh, W. A. van Winden, J. Chu, W. M. van Gulik, J. J. Heijnen, R. F. Mudde, and H. J. Noorman. Computational fluid dynamics simulation of an industrial p. chrysogenum fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization. *Chem. Eng. Sci.*, 175:12–24, 2018.

[53] C. Hatzis, F. Srienc, and A. G. Fredrickson. Multistaged corpuscular models of microbial growth: Monte Carlo simulations. *Biosyst.*, 36(1):19–35, 1995.

[54] H. Heijmans. On the stable size distribution of populations reproducing by fission into two unequal parts. *Math. Biosci.*, 72(1):19–50, 1984.

[55] R. Heinrich and S. Schuster. *The Regulation of Cellular Systems*. Springer US, Boston, MA, 1996.

[56] A.-L. Heins, T. Johanson, S. Han, L. Lundin, M. Carlquist, K. V. Gernaey, S. J. Sørensen, and A. Eliasson Lantz. Quantitative flow cytometry to understand population heterogeneity in response to changes in substrate availability in *Escherichia coli* and *Saccharomyces cerevisiae* chemostats. *Front. Bioeng. Biotechnol.*, 7:187, 2019.

[57] A.-L. Heins and D. Weuster-Botz. Population heterogeneity in microbial bioprocesses: origin, analysis, mechanisms, and future perspectives. *Bioprocess Biosyst. Eng.*, 41(7):889–916, 2018.

[58] M. A. Henson. Dynamic modeling of microbial cell populations. *Curr. Opin. Biotechnol.*, 14(5):460–467, 2003.

[59] P. J. Hill and K. M. Ng. New discretization procedure for the breakage equation. *AIChE J.*, 41(5):1204–1216, 1995.

[60] M. D. Hoang, D. T. Doan, M. Schmidt, H. Kranz, A. Kremling, and A.-L. Heins. Application of an *Escherichia coli* triple reporter strain for at-line monitoring of single-cell physiology during L-phenylalanine production. *Eng. Life Sci.*, 23(1):e2100162, 2023.

[61] M. D. Hoang, I. Polte, L. Frantzmann, N. von den Eichen, A.-L. Heins, and D. Weuster-Botz. Investigation of population heterogeneity during l-phenylalanine production with an escherichia coli quadruple reporter strain in a novel two-compartment bioreactor combining a stirred-tank bioreactor with a coiled flow inverter. in preperation, 2023.

[62] S. L. Holland, T. Reader, P. S. Dyer, and S. V. Avery. Phenotypic heterogeneity is a selected trait in natural yeast populations subject to environmental stress. *Environ. Microbiol.*, 16(6):1729–1740, 2014.

[63] S. K. Hortsch and A. Kremling. Characterization of noise in multistable genetic circuits reveals ways to modulate heterogeneity. *PloS One*, 13(3):e0194779, 2018.

[64] H. Hristov, R. Mann, V. Lossev, S. D. Vlaev, and P. Seichter. A 3-D analysis of gas-liquid mixing, mass transfer and bioreaction in a stirred bio-reactor. *Food Bioprod. Process.*, 79(4):232–241, 2001.

[65] D. Huh and J. Paulsson. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.*, 43(2):95–100, 2011.

[66] S. Hui, J. M. Silverman, S. S. Chen, D. W. Erickson, M. Basan, J. Wang, T. Hwa, and J. R. Williamson. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.*, 11(1):784, 2015.

[67] C. D. Immanuel. Population balance model for cellular processes in biological systems: Biochemical and biomedical applications. *IFAC Proc. Vol.*, 37(9):365–370, 2004.

[68] N. Jourdan, T. Neveux, O. Potier, M. Kanniche, J. Wicks, I. Nopens, U. Rehman, and Y. Le Moullec. Compartmental modelling in chemical engineering: A critical review. *Chem. Eng. Sci.*, 210:115196, 2019.

[69] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.*, 6(6):451–464, 2005.

[70] M. Kafri, E. Metzl-Raz, F. Jonas, and N. Barkai. Rethinking cell growth models. *FEMS Yeast Res.*, 16(7), 2016.

[71] M. Kagoshima and R. Mann. Development of a networks-of-zones fluid mixing model for an unbaffled stirred vessel used for precipitation. *Chem. Eng. Sci.*, 61(9):2852–2863, 2006.

[72] W. J. Kelly. Using computational fluid dynamics to characterize and improve bioreactor performance. *Biotechnol. Appl. Biochem.*, 49(Pt 4):225–238, 2008.

[73] D. J. Kiviet, P. Nghe, N. Walker, S. Boulineau, V. Sunderlikova, and S. J. Tans. Stochasticity of metabolism and growth at the single-cell level. *Nat.*, 514(7522):376–379, 2014.

[74] S. Klumpp, M. Scott, S. Pedersen, and T. Hwa. Molecular crowding limits translation and cell growth. *Proc. Natl. Acad. Sci. United States Am.*, 110(42):16754–16759, 2013.

[75] P. L. Krapivsky and E. Ben-Naim. Scaling and multiscaling in models of fragmentation. *Phys. review. E, Stat. physics, plasmas, fluids, related interdisciplinary topics*, 50(5):3502–3507, 1994.

[76] J. Kumar, M. Peglow, G. Warnecke, S. Heinrich, and L. Mörl. Improved accuracy and convergence of discretized population balance for aggregation: The cell average technique. *Chem. Eng. Sci.*, 61(10):3327–3342, 2006.

[77] S. Kumar and D. Ramkrishna. On the solution of population balance equations by discretization—I. a fixed pivot technique. *Chem. Eng. Sci.*, 51(8):1311–1332, 1996.

[78] S. Kumar and D. Ramkrishna. On the solution of population balance equations by discretization—II. a moving pivot technique. *Chem. Eng. Sci.*, 51(8):1333–1342, 1996.

[79] M. Kuschel, F. Siebler, and R. Takors. Lagrangian trajectories to predict the formation of population heterogeneity in large-scale bioreactors. *Bioeng.*, 4(2), 2017.

[80] E. Kussell and S. Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Sci.*, 309(5743):2075–2078, 2005.

[81] A. Lapin, D. Müller, and M. Reuss. Dynamic behavior of microbial populations in stirred bioreactors simulated with Euler−Lagrange methods: Traveling along the lifelines of single cells †. *Ind. & Eng. Chem. Res.*, 43(16):4647–4656, 2004.

[82] A. Lapin, J. Schmid, and M. Reuss. Modeling the dynamics of *E. coli* populations in the three-dimensional turbulent field of a stirred-tank bioreactor—a structured–segregated approach. *Chem. Eng. Sci.*, 61(14):4783–4797, 2006.

[83] A. R. Lara, L. Leal, N. Flores, G. Gosset, F. Bolívar, and O. T. Ramírez. Transcriptional and metabolic response of recombinant *Escherichia coli* to spatial dissolved oxygen tension gradients simulated in a scale-down system. *Biotechnol. Bioeng.*, 93(2):372–385, 2006.

[84] G. Larsson, M. Trnkvist, E. S. Wernersson, C. Trgrdh, H. Noorman, and S.-O. Enfors. Substrate gradients in bioreactors: origin and consequences. *Bioprocess Eng.*, 14(6):281–289, 1996.

[85] A. Lemoine, F. Delvigne, A. Bockisch, P. Neubauer, and S. Junne. Tools for the determination of population heterogeneity caused by inhomogeneous cultivation conditions. *J. Biotechnol.*, 251:84–93, 2017.

[86] R. Lencastre Fernandes, M. Carlquist, L. Lundin, A.-L. Heins, A. Dutta, S. J. Sørensen, A. D. Jensen, I. Nopens, A. E. Lantz, and K. V. Gernaey. Cell mass and cell cycle dynamics of an asynchronous budding yeast population: experimental observations, flow cytometry data analysis, and multi-scale modeling. *Biotechnol. Bioeng.*, 110(3):812–826, 2013.

[87] R. Lencastre Fernandes, M. Nierychlo, L. Lundin, A. E. Pedersen, P. E. Puentes Tellez, A. Dutta, M. Carlquist, A. Bolic, D. Schäpper, A. C. Brunetti, S. Helmark, A.-L. Heins, A. D. Jensen, I. Nopens, K. Rottwitt, N. Szita, J. D. van Elsas, P. H. Nielsen, J. Martinussen, S. J. Sørensen, A. E. Lantz, and K. V. Gernaey. Experimental methods and modeling techniques for description of cell population heterogeneity. *Biotechnol. Adv.*, 29(6):575–599, 2011.

[88] R. J. LeVeque. *Numerical Methods for Conservation Laws*. Lectures in Mathematics. ETH Zürich Ser. Springer International Publishing, Cham, 2nd edition, 1992.

[89] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2012.

[90] S. F. Levy. Cellular heterogeneity: Benefits besides bet-hedging. *Curr. Biol. : CB*, 26(9):R355–7, 2016.

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

[92] J.-J. Liou, F. Srienc, and A. G. Fredrickson. Solutions of population balance models based on a successive generations approach. *Chem. Eng. Sci.*, 52(9):1529–1540, 1997.

[93] Y.-H. Liu, J.-X. Bi, A.-P. Zeng, and J.-Q. Yuan. A population balance model describing the cell cycle dynamics of myeloma cell cultivation. *Biotechnol. Prog.*, 23(5):1198–1209, 2007.

[94] G. Madras and B. J. McCoy. Reversible crystal growth–dissolution and aggregation–breakage: numerical and moment solutions for population balance equations. *Powder Technol.*, 143-144:297–307, 2004.

[95] A. W. Mahoney and D. Ramkrishna. Efficient solution of population balance equations with discontinuities by finite elements. *Chem. Eng. Sci.*, 57(7):1107–1119, 2002.

[96] A. Maitra and K. A. Dill. Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proc. Natl. Acad. Sci. United States Am.*, 112(2):406–411, 2015.

[97] F. Maluta, M. Pigou, G. Montante, and J. Morchain. Modeling the effects of substrate fluctuations on the maintenance rate in bioreactors with a probabilistic approach. *Biochem. Eng. J.*, 157:107536, 2020.

[98] N. V. Mantzaris. Stochastic and deterministic simulations of heterogeneous cell population dynamics. *J. Theor. Biol.*, 241(3):690–706, 2006.

[99] N. V. Mantzaris. From single-cell genetic architecture to cell population dynamics: quantitatively decomposing the effects of different population heterogeneity sources for a genetic network with positive feedback architecture. *Biophys. J.*, 92(12):4271–4288, 2007.

[100] N. V. Mantzaris, P. Daoutidis, and F. Srienc. Numerical solution of multi-variable cell population balance models: I. finite difference methods. *Comput. & Chem. Eng.*, 25(11-12):1411–1440, 2001.

[101] N. V. Mantzaris, J.-J. Liou, P. Daoutidis, and F. Srienc. Numerical solution of a mass structured cell population balance model in an environment of changing substrate concentration. *J. Biotechnol.*, 71(1-3):157–174, 1999.

[102] L. Marucci, M. Barberis, J. Karr, O. Ray, P. R. Race, M. de Souza Andrade, C. Grierson, S. A. Hoffmann, S. Landon, E. Rech, J. Rees-Garbutt, R. Seabrook, W. Shaw, and C. Woods. Computer-aided whole-cell design: Taking a holistic approach by integrating synthetic with systems biology. *Front. Bioeng. Biotechnol.*, 8:942, 2020.

[103] C. L. Mayorga Espinoza. *Flow structure analysis and velocity field reconstruction using Reduced Order Method (POD and DMD : application to stirred tank and oscillating bubble plume.)*. Theses, INSA de Toulouse, Apr. 2022.

[104] C. Meyer and W. Beyeler. Control strategies of continuous bioprocesses based on biological activities. *Biotechnol. Bioeng.*, 26(8):916–925, 1984.

[105] P. Mhaskar, M. A. Hjortsø, and M. A. Henson. Cell population modeling and parameter estimation for continuous cultures of saccharomyces cerevisiae. *Biotechnol. Prog.*, 18(5):1010–1026, 2002.

[106] D. Molenaar, R. van Berlo, D. de Ridder, and B. Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol. Syst. Biol.*, 5:323, 2009.

[107] J. Morchain and C. Fonade. A structured model for the simulation of bioreactors under transient conditions. *AIChE J.*, 55(11):2973–2984, 2009.

[108] J. Morchain, J.-C. Gabelle, and A. Cockx. Coupling of biokinetic and population balance models to account for biological heterogeneity in bioreactors. *AIChE J.*, 59(2):369–379, 2013.

[109] J. Morchain, J.-C. Gabelle, and A. Cockx. A coupled population balance model and CFD approach for the simulation of mixing issues in lab-scale and industrial bioreactors. *AIChE J.*, 60(1):27–40, 2014.

[110] J. Morchain, M. Pigou, and N. Lebaz. A population balance model for bioreactors combining interdivision time distributions and micromixing concepts. *Biochem. Eng. J.*, 126:135–145, 2017.

[111] M. Mori, S. Schink, D. W. Erickson, U. Gerland, and T. Hwa. Quantifying the benefit of a proteome reserve in fluctuating environments. *Nat. Commun.*, 8(1):1225, 2017.

[112] G. Nadal-Rey, D. D. McClure, J. M. Kavanagh, B. Cassells, S. Cornelissen, D. F. Fletcher, and K. V. Gernaey. Development of dynamic compartment models for industrial aerobic fed-batch fermentation processes. *Chem. Eng. J.*, 420:130402, 2021.

[113] G. Nadal-Rey, D. D. McClure, J. M. Kavanagh, S. Cornelissen, D. F. Fletcher, and K. V. Gernaey. Understanding gradients in industrial bioreactors. *Biotechnol. Adv.*, 46:107660, 2021.

[114] P. Neubauer, L. Häggström, and S. O. Enfors. Influence of substrate oscillations on acetate formation and growth yield in *Escherichia coli* glucose limited fed-batch cultivations. *Biotechnol. Bioeng.*, 47(2):139–146, 1995.

[115] P. Neubauer and S. Junne. Scale-down simulators for metabolic analysis of large-scale bioprocesses. *Curr. Opin. Biotechnol.*, 21(1):114–121, 2010.

[116] D. Nevozhay, R. M. Adams, E. van Itallie, M. R. Bennett, and G. Balázsi. Mapping the environmental fitness landscape of a synthetic gene circuit. *PLoS Comput. Biol.*, 8(4):e1002480, 2012.

[117] M. Nicmanis and M. J. Hounslow. A finite element analysis of the steady state population balance equation for particulate systems: Aggregation and growth. *Comput. & Chem. Eng.*, 20:S261–S266, 1996.

[118] A. Nieß, M. Löffler, J. D. Simen, and R. Takors. Repetitive short-term stimuli imposed in poor mixing zones induce long-term adaptation of *E. coli* cultures in large-scale bioreactors: Experimental evidence and mathematical model. *Front. Microbiol.*, 8:1195, 2017.

[119] J. D. Orth, I. Thiele, and B. Ø. Palsson. What is flux balance analysis? *Nat. Biotechnol.*, 28(3):245–248, 2010.

[120] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. Regulation of noise in the expression of a single gene. *Nat. Genet.*, 31(1):69–73, 2002.

[121] P. P. Pandey and S. Jain. Analytic derivation of bacterial growth laws from a simple model of intracellular chemical dynamics. *Theory Biosci. = Theor. den Biowissenschaften*, 135(3):121–130, 2016.

[122] M. Papagianni. Methodologies for scale-down of microbial bioprocesses. *J. Microb. & Biochem. Technol.*, s5, 2015.

[123] M. Y. Pavlov and M. Ehrenberg. Optimal control of gene expression for fast proteome adaptation to environmental change. *Proc. Natl. Acad. Sci. United States Am.*, 110(51):20527–20532, 2013.

[124] R. Penrose. A generalized inverse for matrices. *Math. Proc. Camb. Philos. Soc.*, 51(3):406–413, 1955.

[125] M. Pigou and J. Morchain. Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models. *Chem. Eng. Sci.*, 126:267–282, 2015.

[126] M. A. Pinto and C. D. Immanuel. Model of population heterogeneity and metabolic regulation in bioreactors. In *Proceedings of the 2005 American Control Conference*, pages 931–936, Evanston, IL and Piscataway, N.J, 2005. American Automatic Control Council.

[127] M. Planitz. 3. inconsistent systems of linear equations. *The Math. Gazette*, 63(425):181–185, 1979.

[128] S. Qamar, M. P. Elsner, I. A. Angelov, G. Warnecke, and A. Seidel-Morgenstern. A comparative study of high resolution schemes for solving population balances in crystallization. *Comput. & Chem. Eng.*, 30(6-7):1119–1131, 2006.

[129] V. Quedeville, H. Ouazaite, B. Polizzi, R. O. Fox, P. Villedieu, P. Fede, F. Létisse, and J. Morchain. A two-dimensional population balance model for cell growth including multiple uptake systems. *Chem. Eng. Res. Des.*, 132:966–981, 2018.

[130] C. Quirós, M. Herrero, L. A. García, and M. Díaz. Application of flow cytometry to segregated kinetic modeling based on the physiological states of microorganisms. *Appl. Environ. Microbiol.*, 73(12):3993–4000, 2007.

[131] D. Ramkrishna. Solution of population balance equations. *Chem. Eng. Sci.*, 26(7):1134–1136, 1971.

[132] D. Ramkrishna. *Population balances: Theory and applications to particulate systems in engineering*. Academic Press, San Diego, CA, 2000.

[133] D. Ramkrishna and M. R. Singh. Population balance modeling: Current status and future prospects. *Annu. Rev. Chem. Biomol. Eng.*, 5(1):123–146, 2014.

[134] A. S. Rathore, C. Sharma, and A. Persad. Use of computational fluid dynamics as a tool for establishing process design space for mixing in a bioreactor. *Biotechnol. Prog.*, 28(2):382–391, 2012.

[135] M. Reuss, S. Schmalzriedt, and M. Jenne. Application of computational fluiddynamics (cfd) to modeling stirred tank bioreactors. In K. Schügerl and K.-H. Bellgardt, editors, *Bioreaction Engineering*, pages 207–246. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.

[136] S. Rigopoulos and A. G. Jones. Finite-element scheme for solution of the dynamic population balance equation. *AIChE J.*, 49(5):1127–1139, 2003.

[137] J. Saha, N. Das, J. Kumar, and A. Bück. Numerical solutions for multidimensional fragmentation problems using finite volume methods. *Kinetic & Relat. Model.*, 12(1):79–103, 2019.

[138] S. Schmalzriedt, M. Jenne, K. Mauch, and M. Reuss. Integration of physiology and fluid dynamics. *Adv. Biochem. Eng.*, 80:19–68, 2003.

[139] A. Schmidt, K. Kochanowski, S. Vedelaar, E. Ahrné, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold, and M. Heinemann. The quantitative and condition-dependent escherichia coli proteome. *Nat. Biotechnol.*, 34(1):104–110, 2016.

[140] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Sci.*, 330(6007):1099–1102, 2010.

[141] M. Scott and T. Hwa. Shaping bacterial gene expression by physiological and proteome allocation constraints. *Nat. Rev. Microbiol.*, 2022.

[142] M. Scott, S. Klumpp, E. M. Mateescu, and T. Hwa. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol. Syst. Biol.*, 10:747, 2014.

[143] P. Sharma, P. P. Pandey, and S. Jain. Modeling the cost and benefit of proteome regulation in a growing bacterial cell. *Phys. Biol.*, 15(4):046005, 2018.

[144] Y. Shen, T. Fatemeh, L. Tang, and Z. Cai. Quantitative metabolic network profiling of escherichia coli: An overview of analytical methods for measurement of intracellular metabolites. *TrAC Trends Anal. Chem.*, 75:141–150, 2016.

[145] C.-C. Shu, A. Chatterjee, G. Dunny, W.-S. Hu, and D. Ramkrishna. Bistability versus bimodal distributions in gene regulatory processes from population balance. *PLoS Comput. Biol.*, 7(8):e1002140, 2011.

[146] C.-C. Shu, A. Chatterjee, W.-S. Hu, and D. Ramkrishna. Modeling of gene regulatory processes by population-mediated signaling: New applications of population balances. *Chem. Eng. Sci.*, 70:188–199, 2012.

[147] C.-C. Shu, A. Chatterjee, W.-S. Hu, and D. Ramkrishna. Role of intracellular stochasticity in biofilm growth. insights from population balance modeling. *PloS One*, 8(11):e79196, 2013.

[148] O. K. Silander, N. Nikolic, A. Zaslaver, A. Bren, I. Kikoin, U. Alon, and M. Ackermann. Correction: A genome-wide analysis of promoter-mediated phenotypic noise in escherichia coli. *PLoS Genet.*, 8(5), 2012.

[149] J. D. Simen, M. Löffler, G. Jäger, K. Schäferhoff, A. Freund, J. Matthes, J. Müller, and R. Takors. Transcriptional response of *Escherichia coli* to ammonia and glucose fluctuations. *Microb. Biotechnol.*, 10(4):858–872, 2017.

[150] A. M. Simons. Modes of response to environmental change and the elusive empirical evidence for bet hedging. *Proc. Biol. Sci.*, 278(1712):1601–1609, 2011.

[151] M. Singh, J. Kumar, A. Bück, and E. Tsotsas. An improved and efficient finite volume scheme for bivariate aggregation population balance equation. *J. Comput. Appl. Math.*, 308:83–97, 2016.

[152] P. Singh and M. K. Hassan. Kinetics of multidimensional fragmentation. *Phys. Rev. E, Stat. physics, plasmas, fluids, related interdisciplinary topics*, 53(4):3134–3144, 1996.

[153] H.-S. Song, W. Cannon, A. Beliaev, and A. Konopka. Mathematical modeling of microbial community dynamics: A methodological review. *Process.*, 2(4):711–752, 2014.

[154] R. Spann, K. V. Gernaey, and G. Sin. A compartment model for risk-based monitoring of lactic acid bacteria cultivations. *Biochem. Eng. J.*, 151:107293, 2019.

[155] G. A. Sprenger. From scratch to value: engineering *Escherichia coli* wild type cells to the production of l-phenylalanine and other fine chemicals derived from chorismate. *Appl. Microbiol. Biotechnol.*, 75(4):739–749, 2007.

[156] M. Stamatakis. Cell population balance, ensemble and continuum modeling frameworks: Conditional equivalence and hybrid approaches. *Chem. Eng. Sci.*, 65(2):1008–1015, 2010.

[157] M. Stratford, H. Steels, G. Nebe-von Caron, S. V. Avery, M. Novodvorska, and D. B. Archer. Population heterogeneity and dynamics in starter culture and lag phase adaptation of the spoilage yeast *Zygosaccharomyces bailii* to weak acid preservatives. *Int. J. Food Microbiol.*, 181:40–47, 2014.

[158] A. D. Tadmor and T. Tlusty. A coarse-grained biophysical model of e. coli and its application to perturbation of the rrna operon copy number. *PLoS Comput. Biol.*, 4(4):e1000038, 2008.

[159] H. Takahashi, T. Oshima, J. L. Hobman, N. Doherty, S. R. Clayton, M. Iqbal, P. J. Hill, T. Tobe, N. Ogasawara, S. Kanaya, and D. J. Stekel. The dynamic balance of import and export of zinc in escherichia coli suggests a heterogeneous population response to stress. *J. Royal Soc. Interface*, 12(106), 2015.

[160] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Sci.*, 329(5991):533–538, 2010.

[161] D. W. Tempest, O. M. Neijssel, and M. T. de Mattos. Regulation of carbon substrate metabolism in bacteria growing in chemostat culture. In *Environmental Regulation of Microbial Metabolism*, pages 53–69. Elsevier, 1985.

[162] A. S. Tomlin, L. Whitehouse, R. Lowe, and M. J. Pilling. Low-dimensional manifolds in tropospheric chemical systems. *Faraday Discuss.*, (120):125–46; discussion 197–213, 2001.

[163] B. D. Towbin, Y. Korem, A. Bren, S. Doron, R. Sorek, and U. Alon. Optimality and sub-optimality in a bacterial growth law. *Nat. Commun.*, 8:14123, 2017.

[164] M. Vanni. Approximate population balance equations for aggregation-breakage processes. *J. Colloid Interface Sci.*, 221(2):143–160, 2000.

[165] P. Vrábel, R. G. van der Lans, F. N. van der Schot, K. C. Luyben, B. Xu, and S.-O. Enfors. Cma: integration of fluid dynamics and microbial kinetics in modelling of large-scale fermentations. *Chem. Eng. J.*, 84(3):463–474, 2001.

[166] S. Waldherr, D. A. Oyarzún, and A. Bockmayr. Dynamic optimization of metabolic networks coupled with gene expression.

[167] G. Wang, W. Tang, J. Xia, J. Chu, H. Noorman, and W. M. van Gulik. Integration of microbial kinetics and fluid dynamics toward model-driven scale-up of industrial bioprocesses. *Eng. Life Sci.*, 15(1):20–29, 2015.

[168] G. Warren. Membrane partitioning during cell division. *Annu. Rev. Biochem.*, 62:323–348, 1993.

[169] M. Weiner, J. Tröndle, C. Albermann, G. A. Sprenger, and D. Weuster-Botz. Carbon storage in recombinant *Escherichia coli* during growth on glycerol and lactic acid. *Biotechnol. Bioeng.*, 111(12):2508–2519, 2014.

[170] A. Y. Weiße, D. A. Oyarzún, V. Danos, and P. S. Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci. United States Am.*, 112(9):E1038–47, 2015.

[171] S. Wielgoss, J. E. Barrick, O. Tenaillon, M. J. Wiser, W. J. Dittmar, S. Cruveiller, B. Chane-Woon-Ming, C. Médigue, R. E. Lenski, and D. Schneider. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc. Natl. Acad. Sci. United States Am.*, 110(1):222–227, 2013.

[172] D. M. Wolf, L. Fontaine-Bodin, I. Bischofs, G. Price, J. Keasling, and A. P. Arkin. Memory in microbes: quantifying history-dependent behavior in a bacterium. *PloS One*, 3(2):e1700, 2008.

[173] J. Zahradnik, R. Mann, M. Fialova, D. Vlaev, S. D. Vlaev, V. Lossev, and P. Seichter. A networks-of-zones analysis of mixing and mass transfer in three industrial bioreactors. *Chem. Eng. Sci.*, 56(2):485–492, 2001.

[174] A. Zaslaver, S. Kaplan, A. Bren, A. Jinich, A. Mayo, E. Dekel, U. Alon, and S. Itzkovitz. Invariant distribution of promoter activities in *Escherichia coli*. *PLoS Comput. Biol.*, 5(10):e1000545, 2009.

[175] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu. *The Finite Element Method: its Basis and Fundamentals*. Butterworth-Heinemann, 7th edition, 2013.

A Appendix

A.1 Supporting information for Chapter 3

A.1.1 Kernel of the cell composition matrix

In order to calculate the kernel of W, we analyze the detailed formulation of W

$$W \;=\; \mathrm{Id} \;-\; \underline{c}\,\underline{w}^T \;=\; \mathrm{Id} \;-\; \begin{pmatrix} c_1 w_1 & \cdots & c_1 w_n \\ \vdots & \ddots & \vdots \\ c_n w_1 & \cdots & c_n w_n \end{pmatrix}. \tag{A.1}$$

Let N_C be the dimension of W, corresponding to the number of intracellular components. The determinant of W can be easily calculated up to dimension 3 and for larger dimensions obtained from the Leibniz formula, leading to

$$\det(W) \;=\; 1 \;-\; \sum_i w_i c_i \tag{A.2}$$

which is zero due to strict mass conservation according to Eq. (2.16). One can show furthermore, that the kernel of W is one-dimensional, as the values of W have a positive sign on the diagonal and negative sign otherwise

$$\begin{pmatrix} + & - & \cdots & - \\ - & + & \cdots & - \\ \vdots & & \ddots & \vdots \\ - & \cdots & - & + \end{pmatrix}. \tag{A.3}$$

The rank of the matrix W is given by the maximum number of linearly independent column or row vectors of the matrix. As the N_C vectors of W all have exactly one positive entry at a different position of the vector, one needs N_C vectors in order to find a linear combination to form zero and the maximum number of linearly independent vectors is $N_C - 1$. This argumentation holds for both the set of column and row vectors. The system of equations for linear dependent vectors can be written as

$$0 \;=\; \lambda_i \;-\; c_i \sum_k \lambda_k w_k \tag{A.4}$$

If we choose $\lambda_i = s c_i$ with $s \in \mathbb{R}$, then the set of equations hold as the latter term forms

$$\sum_k s\; c_k w_k \;=\; s \sum_k c_k w_k \;=\; s. \tag{A.5}$$

Thus, the kernel of W is

$$c_{w,0} \;=\; s\;\underline{c}, \qquad s \in \mathbb{R}. \tag{A.6}$$

A.1.2 Flux analysis of the minimal model

In the example of the minimal model, $\mathcal{N}^+$ is given by

$$\mathcal{N}^+ = \begin{pmatrix} \frac{\alpha}{\alpha^2+1} & \frac{\gamma\alpha}{\beta+\beta\alpha^2} \\ 0 & \frac{1}{\gamma} \\ -\frac{1}{\alpha^2+1} & -\frac{\gamma}{\beta+\beta\alpha^2} \end{pmatrix}. \tag{A.7}$$

Therefore, we obtain

$$Id - \mathcal{N}^+\mathcal{N} = \begin{pmatrix} \frac{1}{\alpha^2+1} & 0 & \frac{\alpha}{\alpha^2+1} \\ 0 & 0 & 0 \\ \frac{\alpha}{\alpha^2+1} & 0 & \frac{\alpha^2}{\alpha^2+1} \end{pmatrix}. \tag{A.8}$$

With this formulation, we can see that $\text{rank}(\text{Id} - N^+N) = 1$, leading to a one-dimensional solution space for r. The image of $\text{Id} - \mathcal{N}^+\mathcal{N}$ is given by

$$\text{span}\left(\begin{pmatrix} \frac{1}{\alpha^2+1} \\ 0 \\ \frac{\alpha}{\alpha^2+1} \end{pmatrix}\right). \tag{A.9}$$

A.1.3 Kinetic parameters of different reaction rate sets for proteome allocation

The following kinetic parameters were used for the study of proteome allocation for 5 different model variations

	k_T	k_P	k_O
Linear kinetics	1000	3.5×10^4	3×10^6
Michaelis-Menten kinetics for M in r_P	1000	3.5	1.5×10^5
Michaelis-Menten kinetics for M in r_P and r_O	1000	8	440
Michaelis-Menten kinetics for T in r_T	4×10^{-3}	8	1.2×10^{-3}
Michaelis-Menten kinetics for M and T	3.2×10^{-3}	6×10^4	2

where the units of the kinetic parameters depend on the formulation of the reaction rate. Additionally, let $K_P = 0.5 \times 10^{-4}$ mol/g_{DW} be the Michaelis-Menten constant for M in r_P, $K_O = 2 \times 10^{-6}$ mol/g_{DW} the Michaelis-Menten constant for M in r_O and $K_T = 1 \times 10^{-6}$ mol/g_{DW} the Michaelis-Menten constant for T.

A.1.4 Parameter values for the simulation of a bioprocess

Table A.1: Overview of parameters values used for the simulation of a production process

Parameter	Description of parameter	Value	Unit
V_0	Initial reactor volume	1.35	L
X_0	Initial biomass concentration	0.014	g_{DW}/L
S_0	Initial substrate concentration	4.08	g/L
M_0	Initial metabolite concentration	$2.4 \cdot 10^{-4}$	mol/g_{DW}
P_0	Initial protein concentration	$1.19 \cdot 10^{-5}$	mol/g_{DW}
U_0	Initial residual biomass concentration	$0.87 \cdot 10^{-5}$	g_{DW}/l
F_0	Initial L-phenylalanine concentration	0.09	g/l
O_0	Initial acetate concentration	0	g/l
$F_{\max}$	Maximal fraction of L-phenylalanine proteins	0.05	-
ψ_P	Protein length	212	-
w_S	Molecular weight of substrate	92.09	g/mol
w_M	Molecular weight of metabolites	217	g_{DW}/mol
w_P	Molecular weight of proteins	$w_M \cdot \psi_P$	g_{DW}/mol
w_U	Molecular weight of residual biomass	$w_M \cdot \psi_P$	g_{DW}/mol
w_F	Molecular weight of L-phenylalanine	165.19	g/mol
k_T	Kinetic constant for transport rate	$3.75 \cdot 10^3$	1/h
k_P	Kinetic constant for protein synthesis rate	2	1/h
k_O	Kinetic constant for overflow metabolism rate	$1.5 \cdot 10^4$	g/molh
k_U	Kinetic constant for residual biomass synthesis rate	1.5	1/h
k_F	Kinetic constant for L-phenylalanine production rate	$1.8 \cdot 10^7$	g/molh
$k_{C,b}$	Kinetic constant for respiration rate (biomass production phase)	$1.2 \cdot 10^7$	g/molh
$k_{C,p}$	Kinetic constant for respiration rate (production phase)	$2 \cdot 10^7$	g/molh
K_T	Michealis-Menten constant for transport rate	0.05	g/l
K_P	Michealis-Menten constant for protein synthesis rate	$0.5 \cdot 10^{-4}$	mol/g_{DW}
K_U	Michealis-Menten constant for residual biomass synthesis rate	$0.5 \cdot 10^{-4}$	mol/g_{DW}
α	Stoichiometric coefficient for transport rate	1	-
β	Stoichiometric coefficient for protein synthesis rate	1	-
γ	Stoichiometric coefficient for protein synthesis rate	212	-
ϵ	Stoichiometric coefficient for residual biomass synthesis rate	1	-
δ	Stoichiometric coefficient for residual biomass synthesis rate	212	-
n_S	Stoichiometric coefficient for substrate rate	2	-
n_F	Stoichiometric coefficient for L-phenylalanine production rate	2	-
a_T	Constant of linear function relating T to M	-47.23	-
b_T	Constant of linear function relating T to M	0.223	-
a_R	Constant of linear function relating R to M	91.44	-
b_R	Constant of linear function relating R to M	0.131	-
t_ψ	Time constant in function ψ	5	h
T_{max}	Maximal concentration of T	2.76×10^{-6}	g_{DW}

A.1.5 Comparison of the substrate uptake rates

Between the two different models we cannot observe a big difference, except for a slight accumulation of substrate S if we include the effect of micro-mixing with $\tau_{mix} = 17 \times 10^{-3}$ h and

T_{max} is estimated with the assumption $0.57 = w_P\ P$ and the maximal fraction T_l

$$T_{max} = \frac{0.57}{w_P}\ T_l \tag{A.10}$$

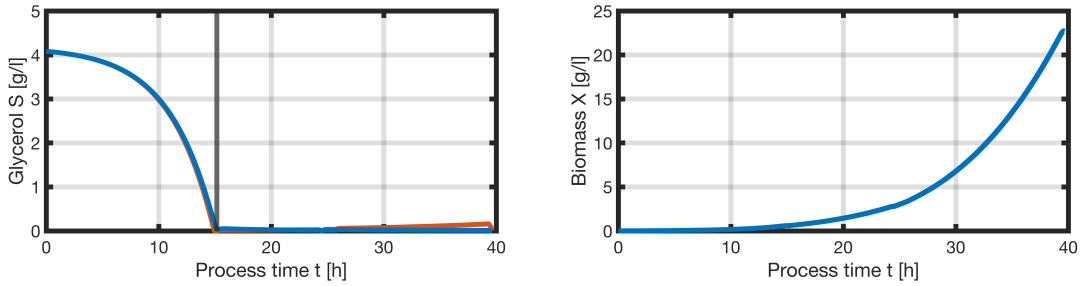

Figure A.1: Comparison of the two sets of simulated quantities for the substrate uptake rate with Michaelis-Menten kinetics (blue) and with micro-mixing (orange). Time course of the following concentrations: glycerol S and biomass X

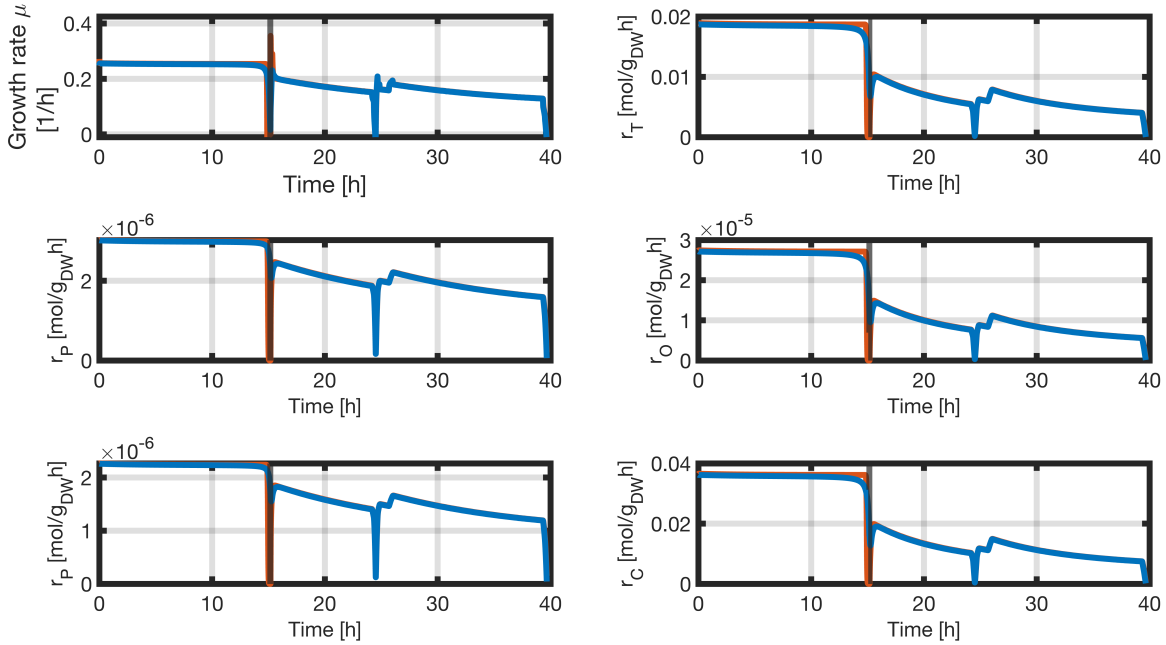

Figure A.2: Time course of the simulated specific growth rate μ, substrate transport rate r_T, protein synthesis rate r_P, respiration rate r_C for the substrate uptake rate with Michaelis-Menten kinetics (blue) and with micro-mixing (orange).

A.2 Supporting information for Chapter 4

Influence of cell division rate function

During the batch phase the differences between the protein concentration distributions are very small as seen in Figure A.3. While higher critical values m_c and n_c lead to a slighty narrower distribution, lower values lead to a broader distribution.

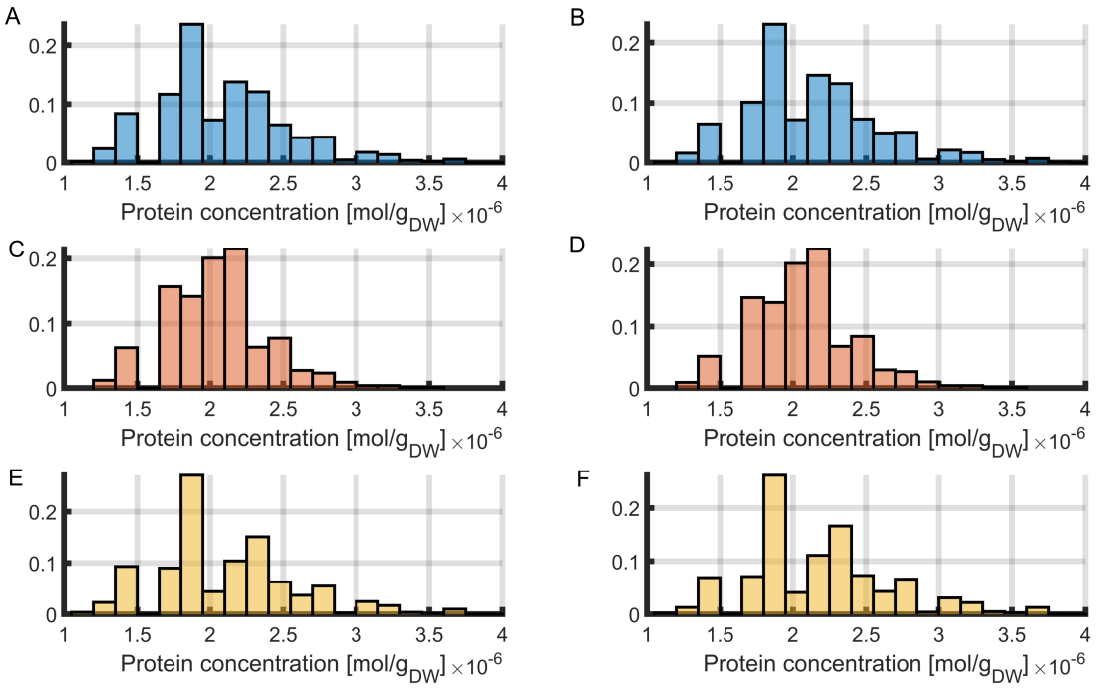

Figure A.3: Comparison of the protein distribution for division rate functions with different critical values m_c and n_c during the batch phase: base case with $m_c = 500$ fg and $n_c = 1.3 \times 10^{-3}$ fmol (A,B), higher critical values $m_c = 800$ fg and $n_c = 1.7 \times 10^{-3}$ fmol (C,D) and lower critical values $m_c = 400$ fg and $n_c = 1.3 \times 10^{-3}$ fmol (E,F). Protein distribution at time point $t = 7$ h (left column) and $t = 14$ h (right column).

In a constant environment, the cell population grows exponentially and exhibit internal properties, which correspond to values in the upper-left quarter of the state space. Thus, the cell population does not exhibit internal property values at the boundaries of the state space. We compare two cases: a base case with $m_c = 500$ fg and $n_c = 1.3 \times 10^{-3}$ fmol (A,B) and a case with higher critical values $m_c = 700$ fg and $n_c = 1.82 \times 10^{-3}$ fmol. Note, that the critical values for m_c and n_c are increased by the same factor, but running the model with different ratios between the two values does not show big differences. A look at the protein concentration distributions for two different time points in Figure A.4 reveals negligible differences between the two cases with an only slitghtly smaller variance for higher critical values.

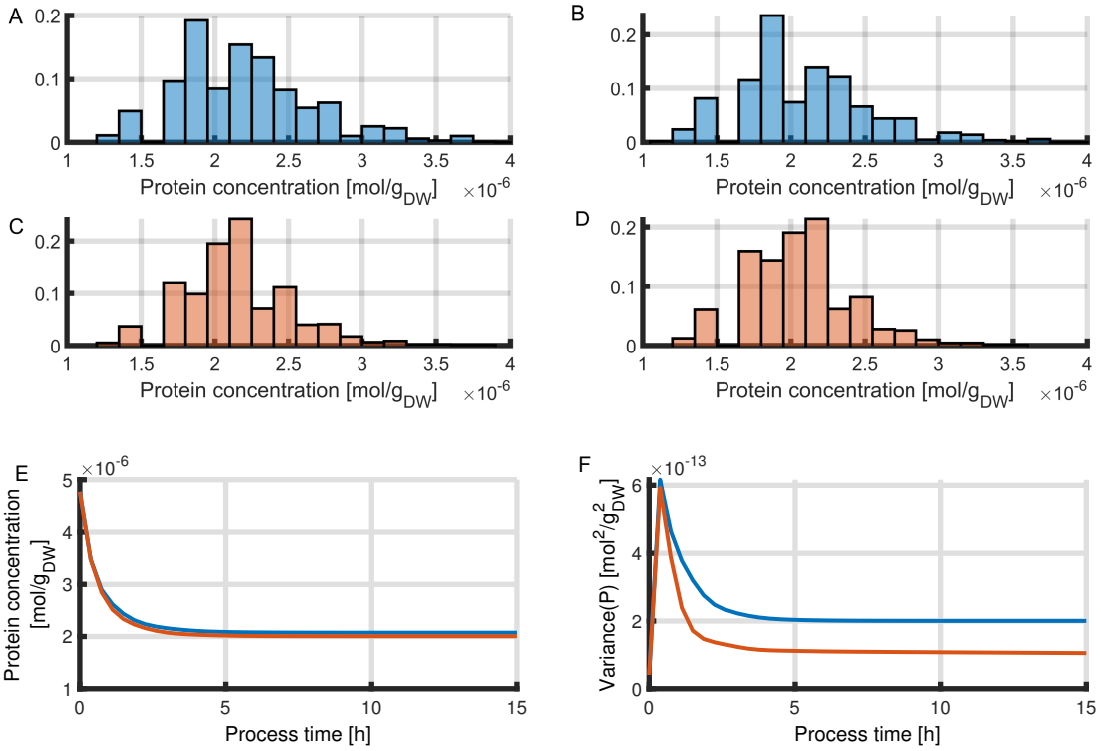

Figure A.4: Comparison of the protein distribution for division rate functions with different critical values m_c and n_c in a constant environment: base case with $m_c = 500$ fg and $n_c = 1.3 \times 10^{-3}$ fmol (A,B) and higher critical values $m_c = 700$ fg and $n_c = 1.82 \times 10^{-3}$ fmol (C,D). Protein distribution at time point $t = 7$ h (A,C) and $t = 14$ h (B,D). Mean protein concentration (EC) and variance of the protein concentration (F).

A.3 Supporting information for Chapter 5

A.3.1 Number of cell divisions

For the inclusion of intrinsic noise, we calculate the number of cell divisions. The mass of all cells after time step Δt is given by

$$\sum_i \mu_i m_i \Delta t + \sum_i m_i \tag{A.11}$$

Thus the number of cells after time step Δt can be calculated as

$$N \cdot \frac{\sum_i \mu_i \; m_i \; \Delta t + \sum_i m_i}{\sum_i m_i} = N \cdot \left(\frac{\sum_i \mu_i \; m_i \; \Delta t}{\sum_i m_i} + 1 \right) = N \cdot 2^k \tag{A.12}$$

Therefore, we obtain

$$\#\text{cell divisions} = N \sum_i 2^{i-1} = N \cdot (2^k - 1) = N \cdot \left(\frac{\sum_i \mu_i \; m_i \; \Delta t}{\sum_i m_i} \right) \tag{A.13}$$

A.3.2 Parameter values of the Euler-Lagrange simulation

In addition to the parameters used for the bioprocess in a well-mixed bioreactor from Appendix, Section A.1.4, following parameters are used in the Euler-Lagrange simulation.

Table A.2: Overview of parameters values used for the simulations

Parameter	Description of parameter	Value	Unit
$V_{STR,0}$	Initial reactor volume of the stirred tank reactor	0.9	l
V_{PFR}	Reactor volume of the plug flow reactor	0.45	l
b_f	Back-flow	0	l/h
σ_{M_0}	Initial metabolite concentration standard deriviation	$0.1 \cdot 10^{-4}$	mol/g_{DW}
σ_{P_0}	Initial protein concentration standard deriviation	$0.15 \cdot 10^{-5}$	mol/g_{DW}
τ_{mix}	Micro-mixing time	17×10^{-3}	h

A.3.3 Additional information about the contribution of cell division on population heterogeneity

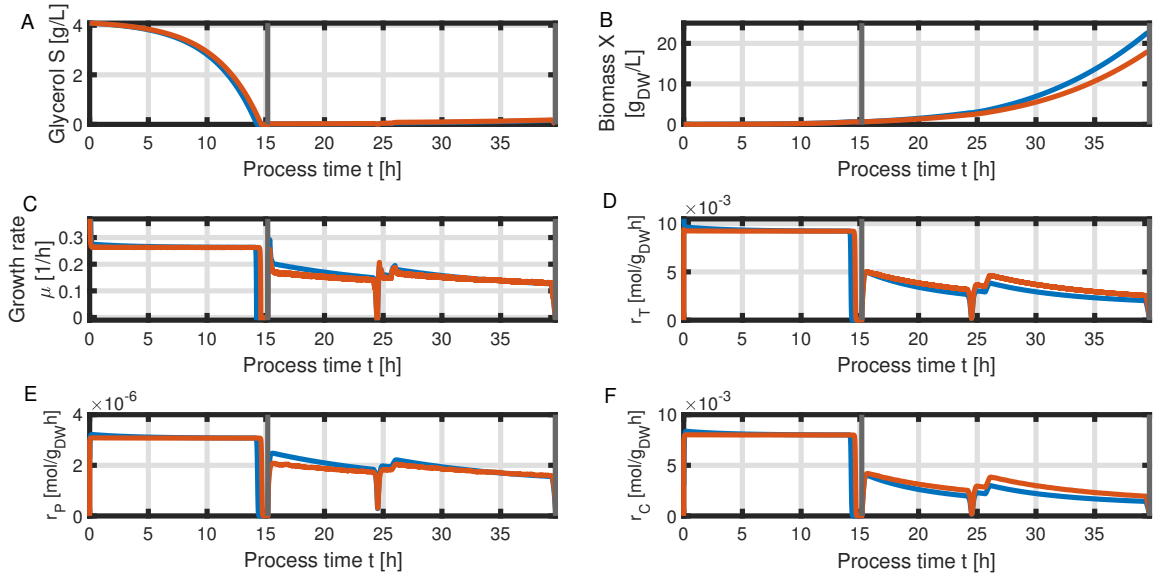

Figure A.5: Comparison of the substrate concentration S (A), biomass concentration X (B), growth rate μ (C), substrate transport rate r_T (D), protein synthesis rate r_P (E) and respiration rate r_C (F) in a STR (blue) and in a two-compartment bioreactor (orange) with intrinsic noise, where the quantities are mean values over all compartments and particles. The separation of the batch and fed-batch phase is indicated by the vertical black line.

A comparison between the simulation of a well-mixed bioreactor and the simulation of the bioreactor system with the inclusion of external noise and intrinsic noise in Figure A.5 shows the same relation as the comparison of the former with the simulation of the bioreactor system with only external noise in Figure 5.2.

The substrate and biomass concentration shows little variation with differences between the compartments occurring in the fed-batch phase. For the substrate concentration S a slight

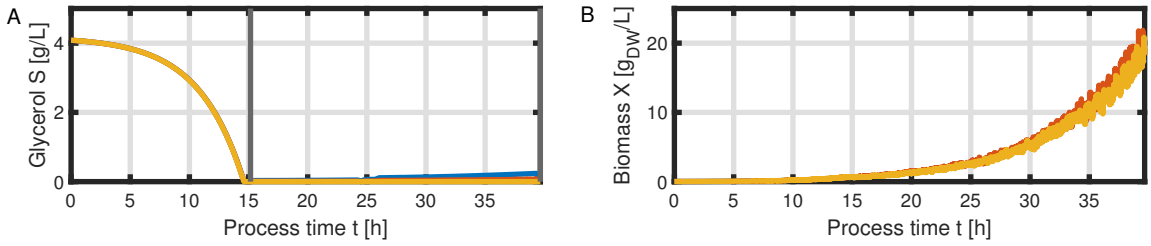

Figure A.6: Time course of substrate (A) and biomass concentration (B) for different compartments: compartment 1 STR (blue), compartment 2 PFR inlet (orange) and compartment 11 PFR outlet (yellow).

accumulation is observable in compartment 1 and a decrease of substrate concentration over the compartments with a concentration of zero in compartment 11 (Figure A.6).

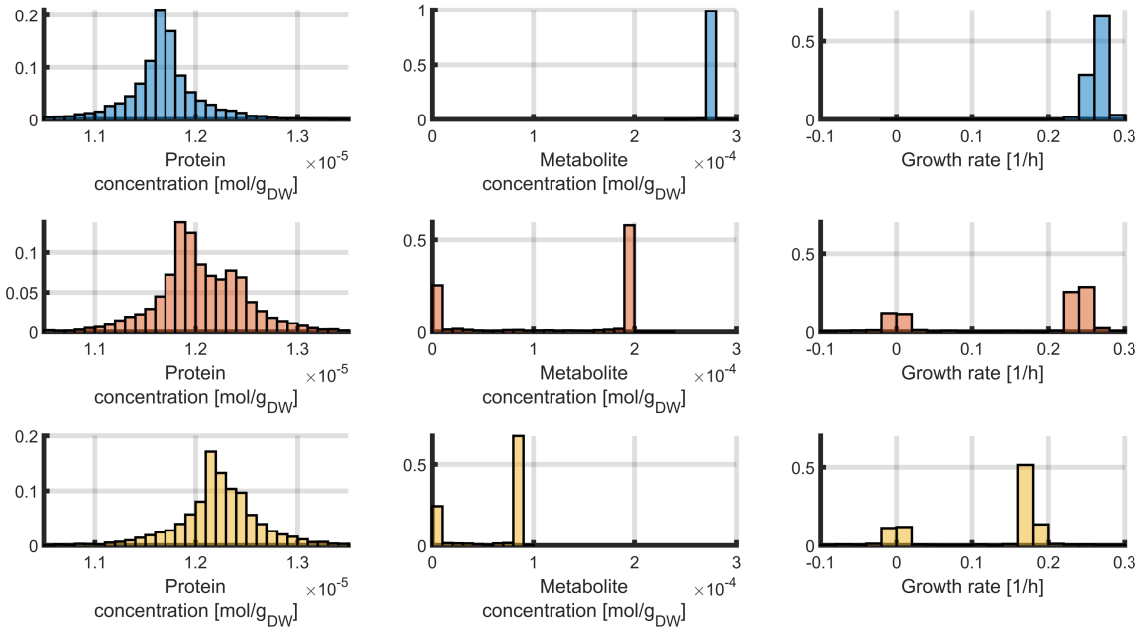

Figure A.7: Distributions of protein concentration (left column), metabolite concentration (middle column) and growth rate (right column) at different time points $t = 14$ h (top row), $t = 16$ h (middle row) and $t = 38$ h (bottom row) with intrinsic noise normalized to 1.

Figure A.7 shows the distribution of the internal properties considering all particles of the bioreactor system for three time points of the process, which are: $t = 14$ h (top row), $t = 16$ h (middle row) and $t = 38$ h (bottom row). The first row depicts the distributions of the protein concentration. With the inclusion of intrinsic noise, the protein distribution is much wider compared to the simulation only considering external noise, as described in Section 5.3, while the distributions for the metabolite concentration are not affected by the addition of internal noise. The distribution for the growth rate is split into two distinct subpopulations, which are slightly wider compared to the case with only external noise. Beside the difference in the shape of the distributions, the behavior over the time course mirrors the previous simulation, with a increased population heterogeneity at the beginning of the fed-batch and a followed decrease.

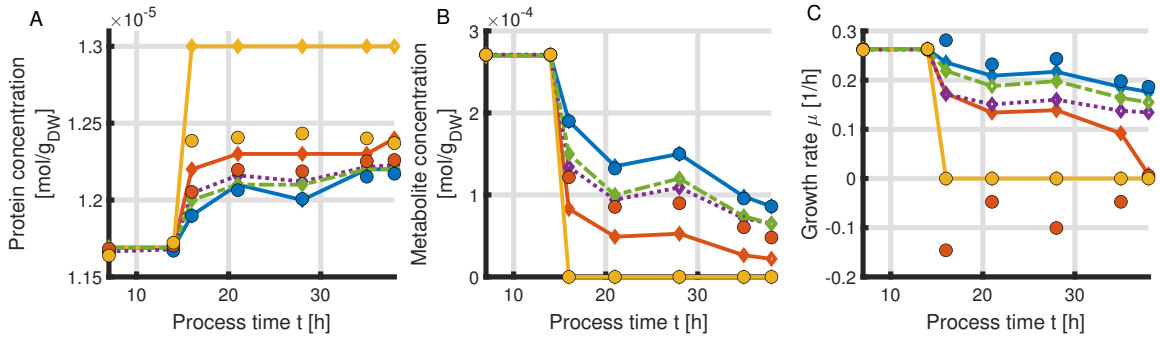

Figure A.8: Comparison of the local average values for compartments 1 (blue), 2 (orange) and 11 (yellow), indicated by circles, steady state values corresponding to the local environment (solid lines), global average (dotted purple line) and steady state corresponding to the global average environment (green dashed line) of internal properties for different time points of the process for the protein concentration (A), metabolite concentration (B) and growth rate (C).

Figure A.8 shows the local average values for the three different compartments, compared to the steady state values corresponding to the local environment, as well as the global average and steady state corresponding to the global average environment. Again, the particles of compartment 1 exhibit the same average protein concentration as the steady state corresponding to the local environment, while the average protein concentrations considering particles in compartment 2 or 11 lie below their respective steady state value. Interestingly, the global average is much closer to the values of compartment 1 and the global steady state. The values for the metabolite concentration as well as the growth rate mirror the behavior of the same variables under the consideration of external noise alone.

For the purpose of analyzing the influence of sampling weights considering the biomass of a particle, the relation between the total biomass a virtual particle carries and its internal properties is analyzed. This helps to understand the influence of intrinsic noise on the cell state of the cells represented by the virtual particles when considering the total biomass as sampling weights. Figure A.9A-C shows scatter plots of the total biomass and its internal properties as well as the particle's growth rate μ for all particles in the bioreactor system. No specific relation can be found for the protein concentration P and the total biomass of the particle in Figure A.9A

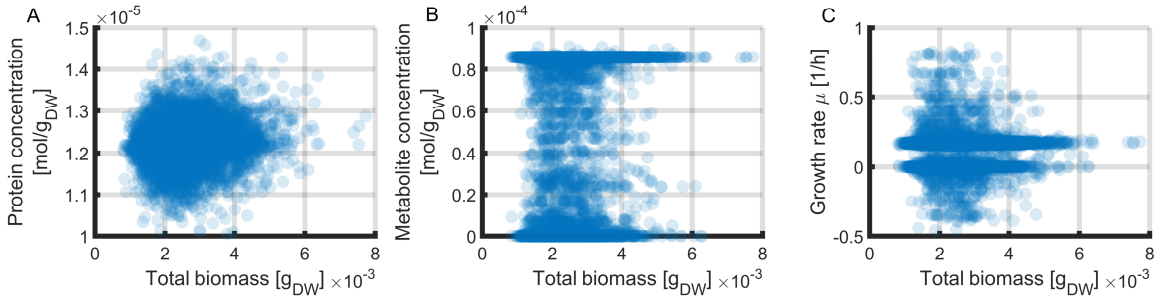

Figure A.9: Relation between total biomass of a virtual particle and its other internal properties: protein concentration (A), metabolite concentration (B) and growth rate (C) at $t = 38$ h.

as the particles are normally distributed around the average protein concentration and average total biomass. This is not surprising as both the dynamics of the protein concentration and the total biomass of a particle are slow processes and therefore are Gaussian shaped. Figure A.9B show two lines forming with metabolite concentration values of $M = 8.6 \times 10^{-5}$ mol/g_{DW} and $M = 3 \times 10^{-10}$ mol/g_{DW} and the rest of the particles contain values in between those two. Comparing the range of total biomass values the particles carry, we can observe that particles with higher total biomass are more likely to also exhibit higher metabolite concentrations, while particles with lower metabolite concentration tend to carry less biomass. This is reasonable, as particles with higher metabolite concentration are faster growing cells, thus are able to increase their biomass faster, which is confirmed in Figure A.9C.

Figure A.10 shows the protein concentration and growth rate distribution considering all particles of the bioreactor system for sampling for cell division with equal weights (blue) and weights considering the total biomass of the particles (orange) at two different time points $t = 26.5$ h (upper) and $t = 38$ h (lower row). A comparison of the protein distributions in Figure A.10A reveals a slightly wider distribution and lower peak for the subpopulation with lower protein concentration and a narrower distribution and higher peak for the subpopulation with higher protein concentration in case of weights considering particles' growth rate compared to equal weights. The time point is shortly after the switch to the second feed solution. After hours at $t = 38$ h in Figure A.10C, the population is more accustomed to the environment and the subpopulation with lower concentration values shows a higher peak at the mode. This outcome is in agreement with our prediction that particles in the STR are favored leading to lower variation on the right side of the protein concentration as those particles are less likely to be randomized. The average growth rate μ does not differ between both cases in Figure A.10B. Thus, the bias

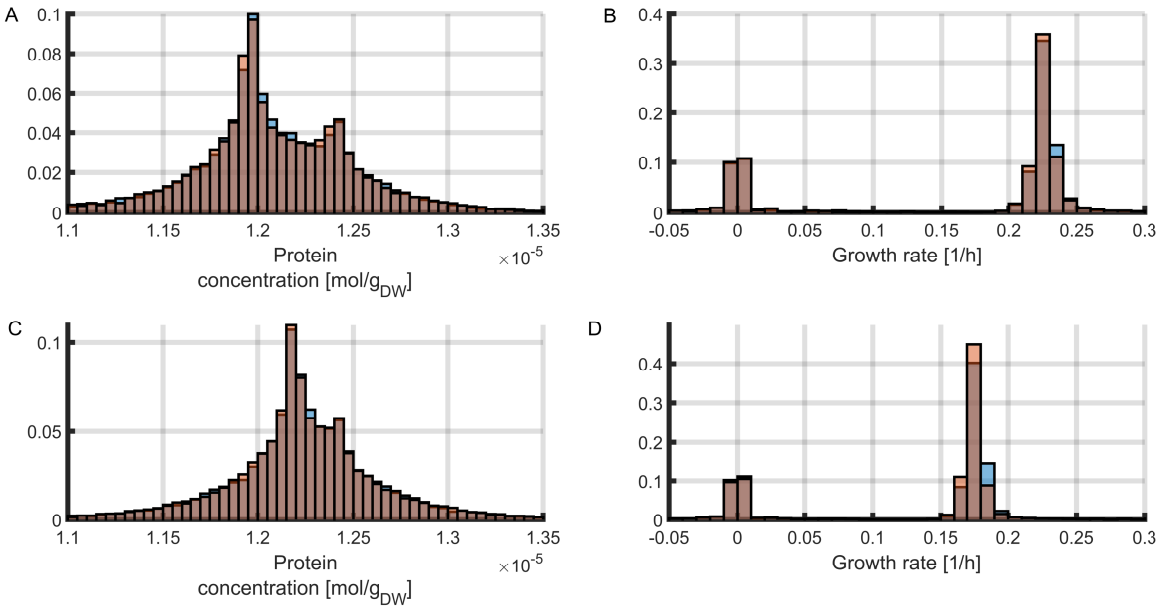

Figure A.10: Comparison of the protein concentration (A,C) and growth rate (B,D) distributions at $t = 26.5$ h (upper row) and $t = 38$ h (lower row) over the whole bioreactor normalized to 1 for sampling with equal weights (blue) and weights considering the biomass of the particles (orange) during the fed-batch phase.. Each case is the average of 5 simulations.

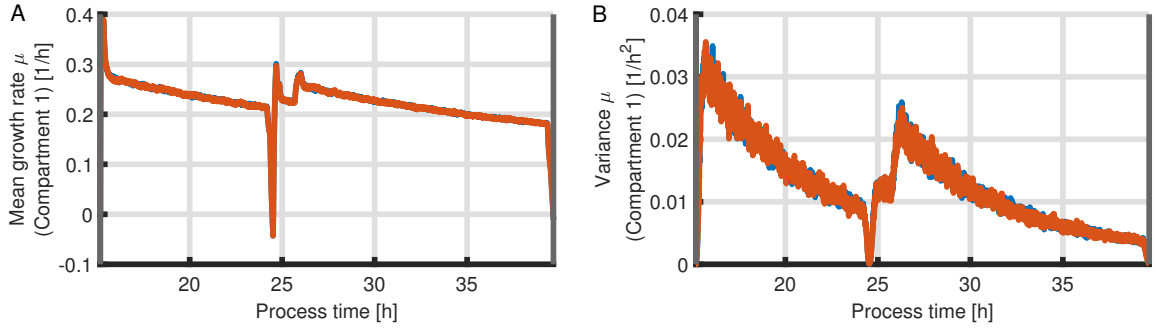

Figure A.11: Comparison of the mean growth rate (A) and variance of the growth rate (B) in compartment 1 for equal sampling weights (blue) and sampling weights considering the particle biomass (orange). Each case is the average of 5 simulations.

towards particles with higher total biomass does not affect the mean growth rate significantly. The growth rate distributions reveal slight differences especially for the growing particles as seen in Figure A.10D. The growth rate distribution in case of sampling weights considering the particle mass reveals less cells exhibiting higher growth rates as the mode of the subpopulation of growing cells and a higher peak at the mode growth rate. This relation is apparent at both time points. Therefore, we calculate the mean and variance considering particles of compartment 1. Figure A.11 shows the mean growth rate and variance of the growth rate of the particles residing

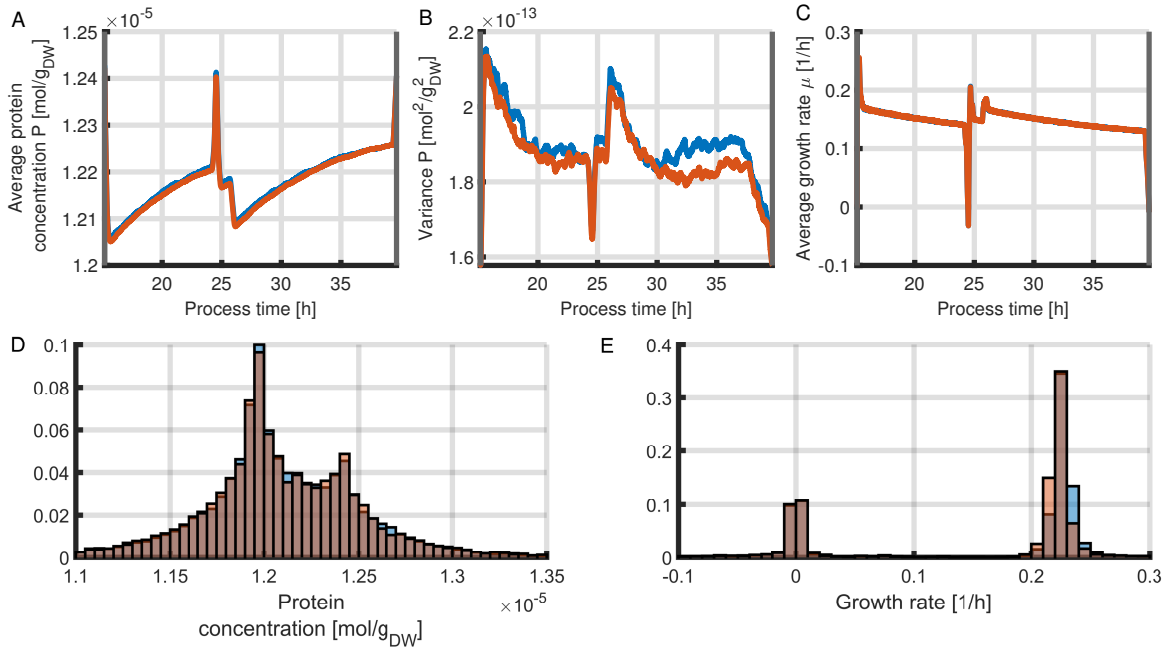

Figure A.12: Comparison of the protein concentration (A) and growth rate (B) distributions at t = 26.5 h over the whole bioreactor normalized to 1 for sampling with equal weights (blue) and weights considering the growth rate of the particles (orange) during the fed-batch phase. Each case is the average of 5 simulations.

in compartment 1 for both cases. Both quantities do not differ between the two sampling weight sets.

For sampling weights considering the growth rate of a particle, a close look at the protein concentration distribution reveals the same behavior as in the previous case and can be seen in the Figure A.12A. Investigating the growth rate distribution at $t = 26.5$ h in Figure A.12B shows differences especially for the subpopulation with higher growth rate, resembling the last case. With a stronger bias towards particles residing in compartment 1, the difference enhances, but the mean growth rate remains unaffected as mentioned earlier.

In case of sampling weights considering both the particle mass and growth rate, the mean growth rate as well as the variance of the growth rate considering particles in compartment 1 are comparable in both cases as seen in Figure A.13

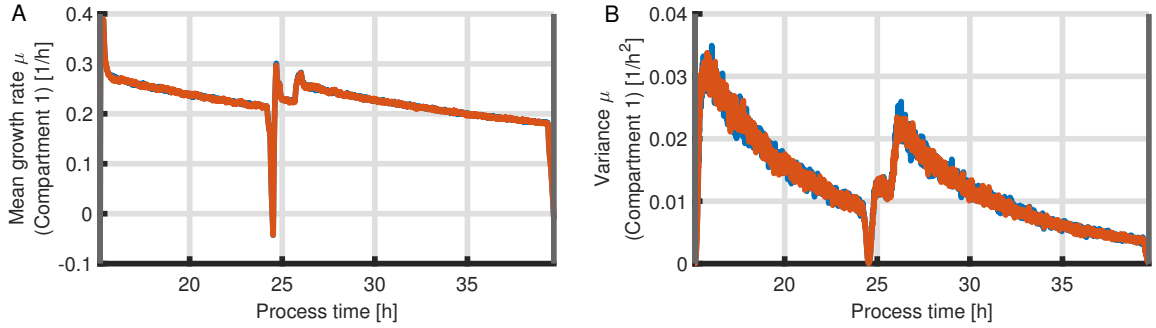

Figure A.13: Comparison of the mean growth rate (A) and variance of the growth rate (B) in compartment 1 for equal sampling weights (blue) and sampling weights considering the particle biomass and growth rate (orange). Each case is the average of 5 simulations.

Figure A.14 depicts the distributions of the internal properties at the end of the fed-batch $t = 38$ h for particles residing in the three different compartments using a normal distribution with a doubled (upper row) and halved standard (lower row) deviation for shuffling of the properties. Analogous to considering all particles of the bioreactor system, doubling of the standard deviation leads to wider protein concentration distributions in the compartments, while halving the standard deviation narrows the distributions. With larger variety in the compartments, the distributions of the compartments overlap more. The distributions of the metabolite concentration of the particles remain unchanged, as the dynamics for the metabolites are fast compared to the proteins. Thus, the different standard deviations do not affect the distributions in Figure A.14B and E. The differences in growth rate in Figure A.14C and F are small with no significant deviations in the distributions of compartment 2 and 11. But the range of growth rates exhibited by the subpopulation residing in comparment 1 is wider for a larger standard deviation.

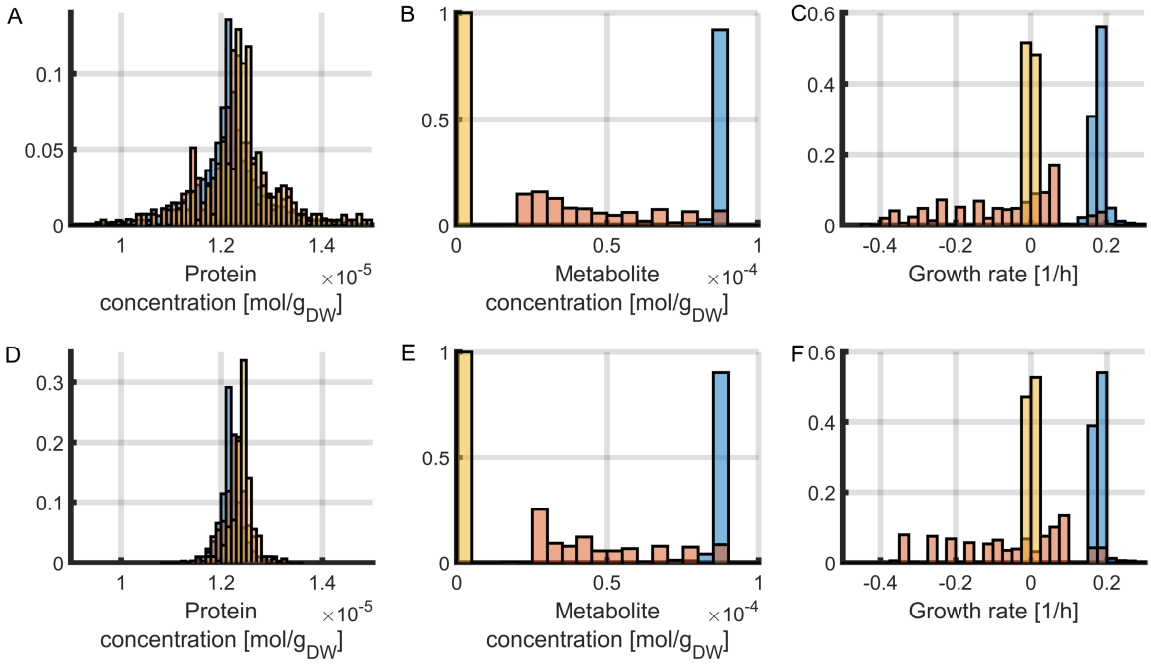

Figure A.14: Normalized distributions of protein concentrations (A,D), metabolite concentration (B,E) and growth rate (C,F) at $t = 38$ h considering particles in different compartments: compartment 1 STR (blue), compartment 2 PFR inlet (orange) and compartment 11 PFR outlet (yellow). The upper row shows the distributions for simulations using a normal distribution with $\sigma_M = 0.2\ M_0$ and $\sigma_P = 0.1\ P_0$ for reshuffling of the properties, the lower row uses $\sigma_M = 0.05\ M_0$ and $\sigma_P = 0.025\ P_0$.

A.3.4 Simulation of a bioreactor system with a very high flow rate

The simulation of the bioreactor system can also be performed with a very high flow rate of $Q_R = 120$ L/h. The simulated mean quantities such as substrate concentration, biomass concentration and internal reaction rates of the particles are depicted in Figure A.15. The simulation of both populations, homogeneous and segregated, are nearly identical.

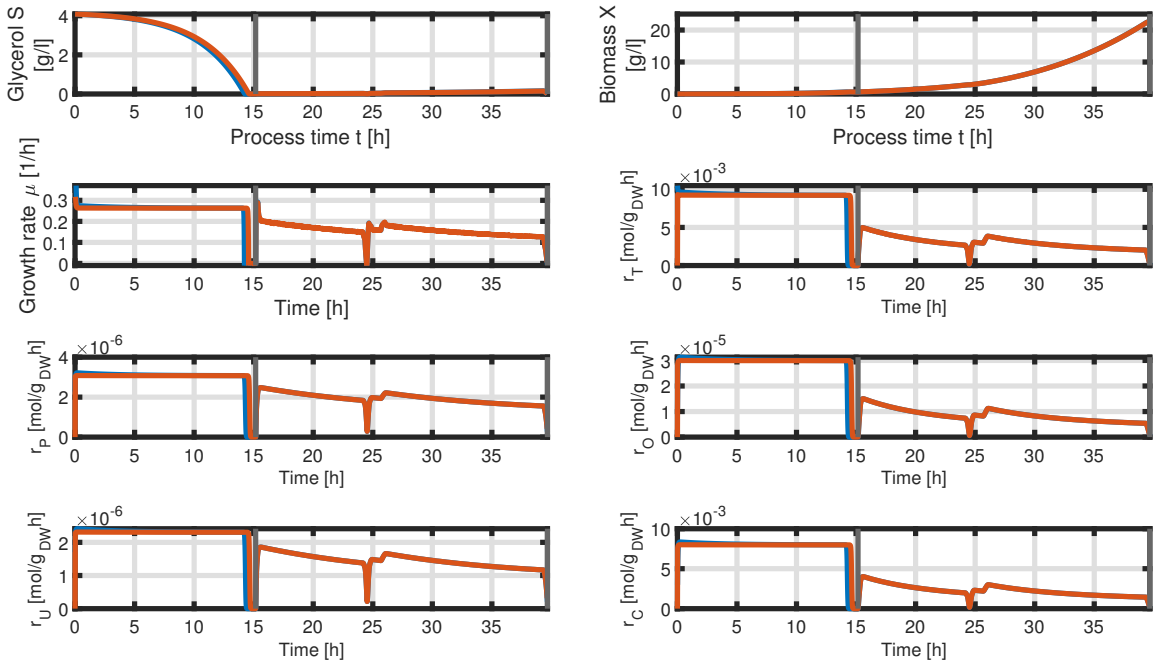

Figure A.15: Comparison of the substrate concentration S (A), biomass concentration X (B), growth rate μ (C), substrate transport rate r_T (D), protein synthesis rate r_P (E), over ow metabolism rate r_O (F), residual biomass synthesis rate r_U (G) and respiration rate r_C (H) in a STR (blue) and in a two-compartment bioreactor with $Q_R = 120$ L/h (orange), where the quantities are mean values over all compartments and particles. The seperation of the batch and fed-batch phase is indicated by the vertical black line.

With the high flow rate, the distribution of protein and metabolite concentration over all particles are very narrow and no subpopulations are visible in Figure A.16A and B. The distribution of the growth rate in Figure A.16C on the other hand still shows a split population into growing and non-growing particles. Distinguishing after the compartments reveal no differences between them. Thus, the two subpopulations in non-growing and growing particles can not be accounted to a specific environment. The environment is not split into two distinct environments and the substrate concentrations barely change between the compartments explaining the same internal property distributions. The different growth rates originate from the fact that, nevertheless, even very small changes in the environment lead to the formation of non-balanced particles and, therefore, non-growing particles.

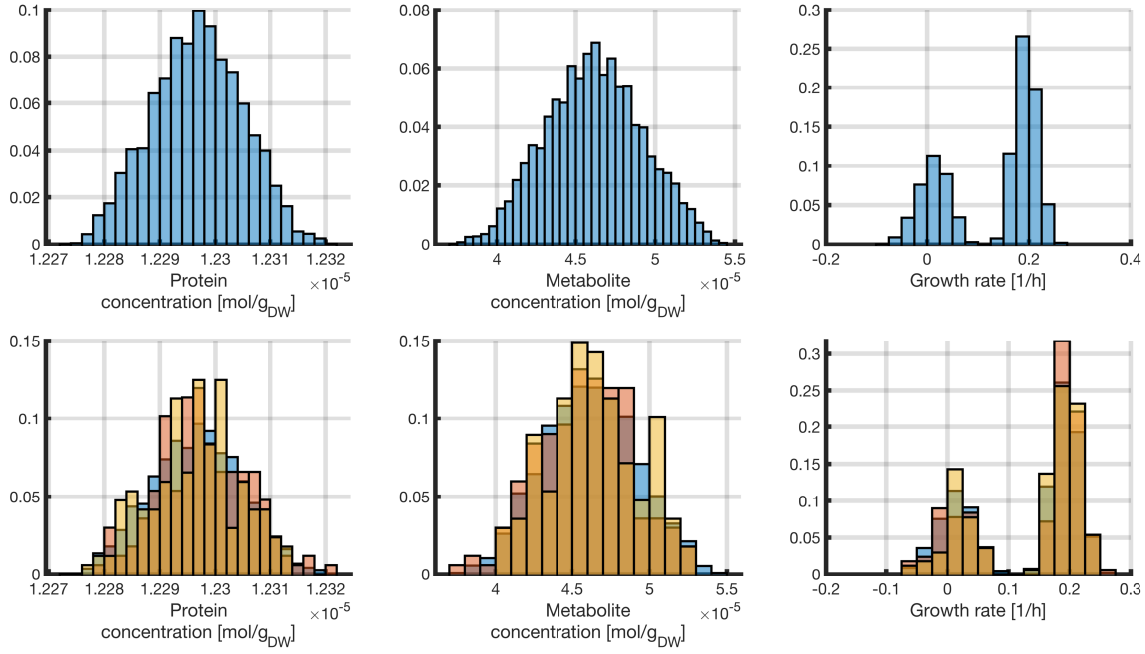

Figure A.16: Comparison of the distributions of protein concentration (left), metabolite concentration (middle) and growth rate (right) at $t = 38$ h with a flow rate of $Q_R = 120$ L/h normalized to 1 for all particles over the whole bioreactor system (upper row) and for different compartments (bottom row): Compartment 1 (blue), compartment 2 (orange) and compartment 11 (yellow).

A.3.5 Simulation for a bioreactor system with maximal variance for growth

Figure A.17 shows the internal properties for a simulation of the bioreactor system with a flow rate of $Q_R = 7.5$ L/h, which has been shown to lead to a maximal variance of growth rate. The protein distribution in compartment 1 and compartment 2 do not differ a lot and shifts slightly in compartment 11 in Figure A.17D. For the metabolite concentration, the range spans from nearly zero to values up to $M = 8 \times 10^{-5}$ mol/g_{DW} with particles exhibiting values in the whole range. This leads to a large variety of growth rates in Figure A.17C taking all values between $\mu_{min} = -0.3$ 1/h and $\mu_{max} = 0.5$ 1/h. In contrast to the cases in Figure 5.13F, the distribution is not split into two distinct subpopulations. This suggests that all particles except of the subpopulation present in compartment 1 are in transition and non-balanced.

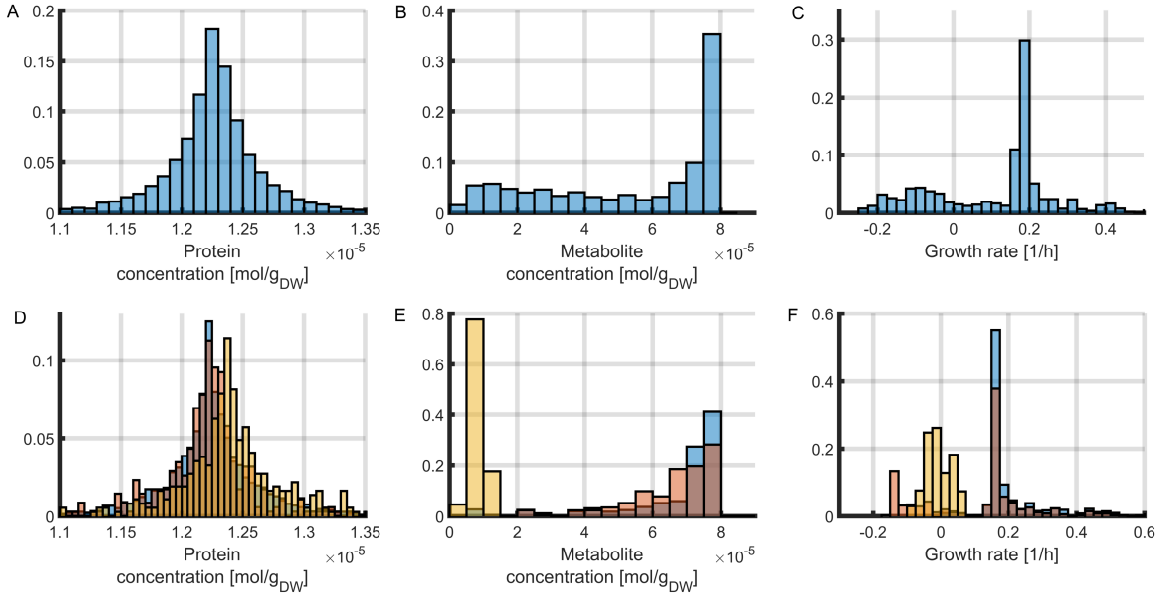

Figure A.17: Comparison of the distributions of protein concentration (left), metabolite concentration (middle) and growth rate (right) at $t = 38$ h with a flow rate of $Q_R = 120$ L/h normalized to 1 for all particles over the whole bioreactor system (upper row) and for different compartments (bottom row): Compartment 1 (blue), compartment 2 (orange) and compartment 11 (yellow).

A.3.6 Two-compartment bioreactor with feed into the PFR

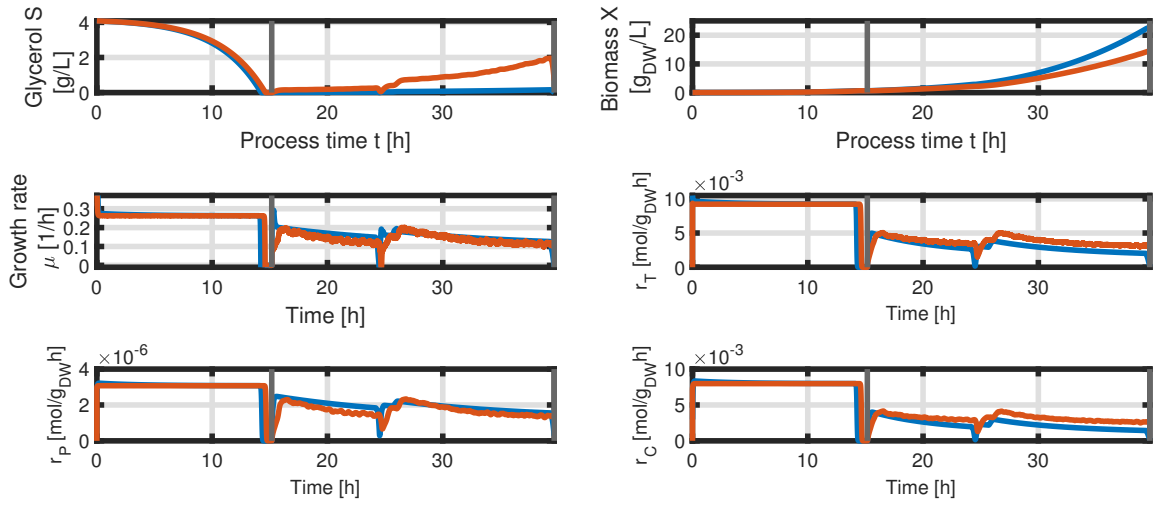

Figure A.18: Comparison of the substrate concentration S, biomass concentration X, growth rate μ, substrate transport rate r_T, protein synthesis rate r_P and respiration rate r_C in a STR (blue) and in a bioreactor system with feed into the PFR (orange), where the quantities are mean values over all compartments and particles. The seperation of the batch and fed-batch phase is indicated by the vertical black line.

Figure A.18 shows the weighted mean variables such as substrate concentration, biomass and internal reaction rates for the two feeding regimes in question: feed into the STR (blue) and feed into the PFR (orange). With addition of the feed into the PFR, substrate is accumulated in the PFR in Figure A.18A since particles encounter a high substrate environment which is to rich for them to consume, while they starve in the STR. This leads to a reduced biomass to substrate yield in Figure A.18B. The substrate concentration in the different compartments is depicted in Figure A.19, showing a high substrate concentration in compartment 2 and exhibiting the lowest substrate concentration in compartment 1.

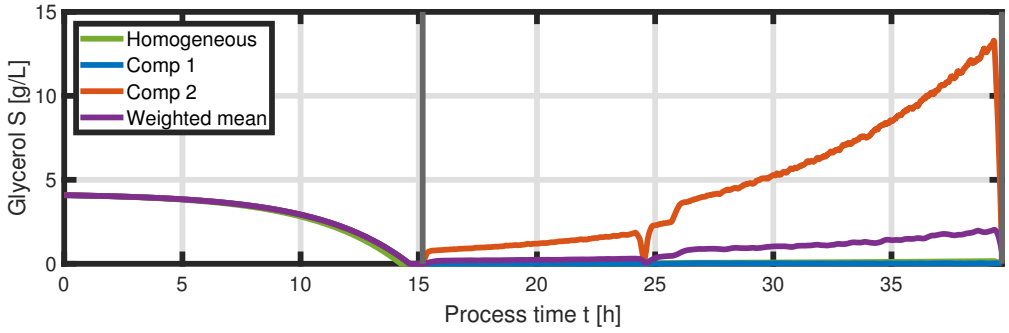

Figure A.19: Comparison of the substrate concentration in the well-mixed bioreactor (green), weighted mean (purple) and in different compartments: Compartment 1 (blue) and compartment 2 (orange).

With an increase of flow rate to $Q_R = 20$ L/h, the substrate gradient decreases, as seen in Figure A.20, and the biomass to substrate yield is almost the same as in a well-mixed envrionment.

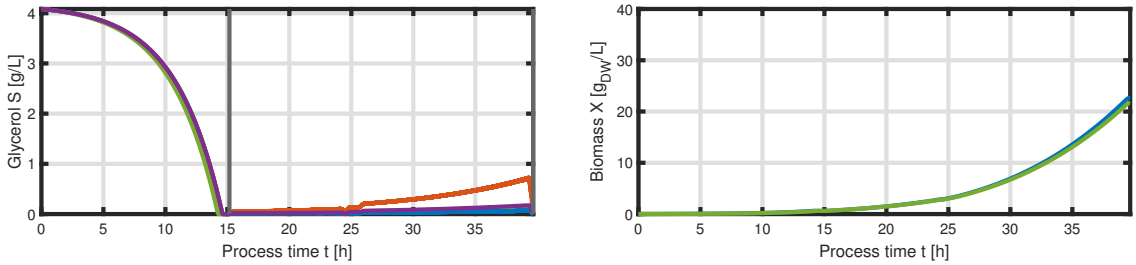

Figure A.20: Comparison of the substrate concentration in the well-mixed bioreactor (green), weighted mean (purple) and in different compartments: Compartment 1 (blue) and compartment 2 (orange).