Modeling, optimization, and predictive control for metabolic cybergenetics

Dissertation
zur Erlangung des akademischen Grades

Doktoringenieur
(Dr.-Ing.)

von Sebastián Espinel Ríos
geb. am 06.11.1993 in Pitalito, Huila, Kolumbien

genehmigt durch die Fakultät für Verfahrens- und Systemtechnik der
Otto-von-Guericke-Universität Magdeburg

Gutachter:

Dr.-Ing. Steffen Klamt

Dr.-Ing. Rolf Findeisen

Dr.-Ing. Steffen Waldherr

eingereicht am: 4. Oktober 2023
Promotionskolloquium am: 19. Dezember 2023

Abstract

This thesis outlines a framework for metabolic cybergenetics that employs computational methods to control gene expression of metabolism-relevant enzymes via external signals. This enables dynamic metabolic engineering through the modulation of intracellular metabolic fluxes. The framework systematically integrates concepts from synthetic biology, metabolic engineering, (machine-learning-supported) dynamic modeling, model-based optimization, predictive control, and estimation. The focus is on batch and fed-batch processes, although the framework can be extended to continuous processes. Two modeling approaches are considered: constraint-based dynamic modeling and (Gaussian-process-supported) quasi-unstructured/unsegregated kinetic modeling. The models are used for model-based optimization, i.e., to determine optimal inputs for maximizing production, including, e.g., cybergenetic inputs, feed rates, and initial concentrations. Repeatedly solving model-based optimization problems -model predictive control- can address uncertainties such as model-plant mismatch and disturbances. Model-based control requires information about the current system states. Real-time process monitoring of relevant states such as biomass components can be achieved with soft sensors, e.g., based on full information estimation. The applicability of the framework is outlined considering a case study dealing with enforced adenosine triphosphate (ATP) turnover for enhanced product yield or productivity, focusing on the anaerobic lactate fermentation by *Escherichia coli*. The ATP turnover is manipulated by modulating the expression of the ATPase (F_1-subunit), an enzyme catalyzing the hydrolysis of ATP, using an optogenetic approach. That is, light is used as a control input to fine-tune ATPase expression. Experimental validation involves open-loop control in batch, employing quasi-unstructured/unsegregated kinetic modeling. The presented framework allows full exploitation of all available input degrees of freedom while counteracting disturbances and uncertain model information. It opens the door to advanced biotechnological applications involving dynamic metabolic control; furthermore, its model-based nature can enable cost-effective process development, robust operation, and flexibility in biotechnology.

Acronyms

3PG	3-Phospho-D-glycerate
AA	Amino acid
AcCoA	Acetyl-CoA
ACE	Acetate
ADP	Adenosine diphosphate
AKG	Alpha-ketoglutarate
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
CcaR	Chromatic acclimation regulator
CcaS	Chromatic acclimation sensor
CO_2	Carbon dioxide
CoA	Coenzyme A
deFBA	Dynamic enzyme-cost flux balance analysis
DHAP	Dihydroxyacetone phosphate
ETH	Ethanol
E_i	Enzyme i
F6P	Fructose 6-phosphate
FIE	Full information estimation
FOR	Formate
FUM	Fumarate
G3P	Glyceraldehyde 3-phosphate
G6P	Glucose 6-phosphate
GLC	Glucose
IPTG	Isopropyl β-d-1-thiogalactopyranoside
MAL	Malate
mRNA	Messenger ribonucleic acid
MPC	Model predictive control
NAD	Nicotinamide adenine dinucleotide
NADH	NAD-reduced
LAC	Lactate
OAA	Oxaloacetic acid
OLO	Open-loop optimization
PEP	Phosphoenolpyruvate
PID	Proportional-integral-derivative
PYR	Pyruvate
Q	Quota
R	Ribosome
RL	Reinforcement learning
SE_FIE	Standard error of the estimate
SUCC	Succinate

Contents

1 Introduction and motivation

This introductory chapter begins by providing an overview of biotechnology, including the fundamental principles and emerging opportunities in the field. It then explores the *toolbox* of bioprocess optimization, highlighting the various degrees of freedom that can be exploited to maximize production efficiency. Furthermore, the chapter discusses open challenges in bioprocessing that serve as motivation for the concepts that will be developed throughout this thesis.

According to the Organization for Economic Co-operation and Development, biotechnology is defined as "the application of science and technology to living organisms, as well as parts, products, and models thereof, to alter living or non-living materials for the production of knowledge, goods, and services" [1]. Biotechnology offers sustainable and bio-based alternatives to non-renewable fossil-based chemicals, materials, and fuels. It also finds applications in the production of food (supplements), pharmaceuticals, and the bioremediation of environmental pollutants. As such, biotechnology can facilitate the development of the future biobased and circular economy, contributing to the achievement of the 2030 United Nations Sustainable Development Goals [2–4].

In this thesis, the scope of biotechnology is limited to microbial cell factories. As its name suggests, microbial cell factories are microorganisms engineered to produce valuable products, from renewable resources such as sugars or carbon dioxide [5]. Cells possess complex metabolic networks composed of interconnected metabolic pathways. These pathways encompass a series of biochemical reactions wherein biomolecules undergo conversion into energy and intermediate molecules (*catabolism*), which are utilized by the cells for the synthesis of building blocks for growth and maintenance purposes (*anabolism*). The reacting species or intermediate molecules in metabolic pathways are called *metabolites*. Enzymes, encoded by genes, catalyze most of the biochemical reactions within the cell. The *dynamic* interactions among the various components of metabolic networks explain the dynamic nature of microbial fermentations [6]. Henceforth, the term "fermentation" is used to refer to submersed microbial cultivations.

Industrial cell-based bioprocesses usually take place within sterile bioreactors; vessels comprising a gas phase and a liquid phase where microorganisms are suspended. Bioreactors are typically equipped with appropriate instrumentation and control systems to maintain an optimal environment (e.g., pH, temperature, and mixing) for cell growth and synthesis of products of interest [7]. Bioprocesses are commonly classified based on their operational mode, such as batch, fed-batch, or (semi)continuous [8]. In batch, nutrients are supplied at the start of

the process. In fed-batch, nutrients are continuously or intermittently fed during the process operation. In (semi)continuous systems, nutrients are fed (semi)continuously at the same rate as liquid culture is removed. In this thesis, the focus is on (fed-)batch setups which are one of the most prevalent operational modes in the biotechnology industry due to ease of implementation [9–11].

Compared to chemical processes, biotechnological processes generally operate under milder conditions (e.g., in terms of pH, temperature, and pressure) and in many cases exhibit higher selectivity toward specific products, resulting in reduced generation of harmful waste by-products [12, 13]. The potential of microbial biotechnology has been realized in numerous commercial processes. Some notable examples include the industrial production of ethanol, isobutanol, 1,3-propanediol, 1,4-butanediol, farnesene, lactic acid, cannabinoids, poly-β-hydroxybutyrate, and glutamate (for a more comprehensive list, refer to [14–17]). The global market size of microbial fermentation technology was estimated to be US$ 28.23 billion in 2021, with an expected compound annual growth rate of 7.8 % until 2030 [18]. These economic figures emphasize the relevance of (microbial) biotechnology in the global economy and society. However, it is important to note that, in several cases, bioprocesses still struggle to achieve the same level of economic competitiveness as (petro)chemical technologies, which hinders the development of a stronger bioeconomy [17].

Competitive bioprocesses must meet minimum requirements in terms of product yield, titer, and volumetric productivity [13, 19]. It is ideal for a bioprocess to achieve a high product yield, which represents the amount of product formed per unit of substrate consumed, contributing to economic viability. A high product titer is desirable to reduce the intensity and cost of downstream processing (product recovery, concentration, and purification). Furthermore, a high volumetric productivity rate, the product formed per unit of time per unit of culture volume, is key in minimizing capital costs as it reduces equipment size and manufacturing area requirements [13, 19]. In addition to meeting these criteria, the competitiveness of a bioprocess is influenced by factors such as process stability, reliability of operation, and robustness. These factors encompass, e.g., the genetic stability of cells and the ability to handle dynamic process disturbances and batch-to-batch process variations [13, 20, 21].

The naturally arising question is how bioprocesses can be optimized to maximize production efficiency. By focusing first on the process level, one can start by determining the best mode of operation for the bioreactor, such as batch, fed-batch, or continuous [22, 23]. Cultivation conditions, including pH, temperature, mixing, and oxygen supply, also play a crucial role in optimizing bioprocesses [24, 25]. Not less important, a well-designed composition of the growth medium can enhance process performance [26, 27]. Optimization of initial concentrations and dynamic feeding rates further contributes to improved process efficiency [28, 29]. While these strategies can influence *overall* metabolic functions, they alone cannot provide targeted interventions to fine-tune specific metabolic fluxes, unless the cell metabolism is engineered *a priori* to respond to specific process signals.

Metabolic engineering, i.e., rewiring metabolic networks, opens an area of opportunity for optimizing bioprocesses at the cellular level, the *heart* of the process. With this, one can obtain cells that overproduce (non-)native metabolites and proteins, and that are able to use non-conventional carbon sources as substrates [5, 30]. However, metabolic engineering often involves a trade-off between increasing product yield and decreasing biomass yield. This trade-off arises from the fact that maximizing the flux through production pathways diverts resources from growth, in turn resulting in a decline in volumetric productivity [31]. Therefore, the inherent trade-off between product yield and volumetric productivity must be carefully considered during bioprocess design due to its impact on the operational and capital costs of the plant [13].

Traditional metabolic engineering is characterized by static approaches, whereby enzymes involved in engineered metabolic pathways are constitutively expressed, i.e., at a constant rate [32–34]. Consequently, doing so limits the ability to externally fine-tune metabolic fluxes. For example, the product yield is *a priori* determined by the *fixed* metabolic flux distribution, and cannot be externally manipulated to *balance* product and growth trade-offs during the operation of the process. Moreover, cells are often engineered for a specific environment, thus adapting the cell metabolism to operate optimally under different conditions is difficult as there is no direct way to intervene metabolism in real-time. Overall, static metabolic engineering lacks the flexibility and adaptability required to effectively respond to changing conditions and demands.

To circumvent the challenges of static metabolic engineering, one can instead *dynamically* express metabolism-relevant proteins such as enzymes during the bioprocess operation. That is, as opposed to expressing metabolic pathways constitutively, one could opt for an inducible gene expression approach exploiting either native genetic *switches* (e.g., switching metabolism from aerobic to anaerobic conditions) or employing dedicated tunable gene expression systems [34–37]. This enables a degree of freedom for the dynamic manipulation of metabolic fluxes in cells. This idea could be exploited, e.g., for managing temporal growth and production trade-offs for improved yield and productivity in bioprocesses. Dynamic metabolic engineering can also alleviate the resource burden associated with expressing heterologous pathways or overexpressing native pathways because the cells would synthesize the target enzymes *only* when induced.

Dynamic metabolic engineering has been considered in the bioprocess engineering literature, resulting, e.g., in two-stage and three-stage fermentations, typically separating growth and induction/production phases [31, 34, 38–41]. A well-known application example of dynamic metabolic control is the decoupling of growth and production in recombinant protein synthesis by *Pichia pastoris* to maximize volumetric productivity, where methanol is used to induce the expression of target proteins following a growth phase [42]. An analogous strategy is often applied in recombinant protein production by *Escherichia coli* using isopropyl β-d-1-thiogalactopyranoside (IPTG) as an inducer [43]. Another example of dynamic metabolic control for enhancing volumetric productivity is the use of temperature to induce production after a growth phase in itaconic acid synthesis by *E. coli* [44].

The problem with using chemical inducers or process conditions such as temperature for dynamic gene expression is that fine-tuning these inputs is challenging due to irreversibility, mass/heat transfer limitations, and impact on untargeted cellular processes. Furthermore, chemicals such as IPTG or methanol can be costly and toxic, undesirable characteristics for large-scale implementation [43, 45]. Therefore, of increasing interest in the field of dynamic gene expression is optogenetics, which involves the use of light to modulate gene expression [46]. Optogenetics enables the precise control of fluxes along metabolic pathways by modulating enzyme expression [41, 47, 48]. It can also allow direct manipulation of cell growth via, e.g., the regulation of the expression of (anti)toxin proteins [49] and antibiotic resistance proteins [50]. In general, light as a control input provides tunability, reversibility, low toxicity, high orthogonality, and cost-effectiveness [51].

Despite the potential of dynamic metabolic engineering, there are significant challenges that need to be addressed toward facilitating its implementation. For example, what is the *optimal* dynamic metabolic flux distribution in the cell for maximizing the process efficiency? How can one *actuate* on specific intracellular metabolic fluxes to reach desired metabolic flux distributions over time? How can one optimally *control* cell metabolism such that it *adapts* online to disturbances and batch-to-batch variability? How can one estimate the metabolic state of a cell at a specific time point? These questions are addressed with the concept of metabolic cybergenetic systems, which is introduced in the next chapter.

2 Proposed metabolic cybergenetic framework and considered case study in a nutshell

In this chapter, the concept of metabolic cybergenetics is presented. Its capacity to facilitate the implementation of advanced biotechnological systems in the context of dynamic metabolic engineering is outlined. First, the following topics are introduced: model-based optimization and control, suitable dynamic modeling approaches of metabolism, as well as the pertinence of efficient bioprocess monitoring. The chapter finishes by introducing the core case study considered throughout this thesis: the dynamic manipulation of the adenosine triphosphate (ATP) turnover for dynamic metabolic engineering.

2.1 Model-based optimization and predictive control

Model-based optimization can be used to determine optimal dynamic inputs and initial conditions for maximizing production efficiency. It allows for the maximization of a specific cost function, such as product yield or productivity, while considering the (non-linear) dynamics of the process and system constraints. Previous studies have demonstrated the effectiveness of dynamic optimization in identifying optimal inputs for bioprocesses [52–56]. Note that if the model-based optimization is computed only once based on the current initial conditions of the plant and not reevaluated, it corresponds to an *open-loop* optimization [57]. That is, the inputs are applied without feedback, which may cause the system to perform suboptimally due to model-plant mismatch and disturbances.

A well-known feedback control strategy for bioprocesses is proportional-integral-derivative (PID) control [50, 58–60]. PID controllers compute the process input by applying proportional, integral, and derivative corrective actions based on the error between a reference and a measured state variable. PID controllers are, however, limited to maintaining set-points and do not explicitly maximize an (economic) objective function. Additionally, PID controllers are restricted to linear (or linearized) systems and cannot handle constraints.

A more advanced feedback control scheme is model predictive control (MPC) [37, 61–64] which can help to compensate for uncertainties, maximizing production efficiency while maintaining consistent process performance and product quality. In MPC, the model-based optimization problem is iteratively re-solved using the measured or estimated state of the plant at specific sampling instances. This re-initialization of the optimization problem with the updated state of the plant provides online feedback.

A promising data-driven feedback control approach for bioprocesses is reinforcement learning (RL) [65–67]. It uses machine learning to learn a control *policy* (input sequence) taken by an *agent* (controller) and applied to an *environment* (process) such that a *reward* (objective) function is maximized. Nevertheless, it is not straightforward to include system knowledge and constraints in RL and it often requires intense *exploration* before a sufficiently good policy is learned, resulting in high costs and long development times.

Due to the previous considerations, the focus of this thesis is on model-based dynamic optimization and predictive control approaches.

2.2 Dynamic modeling of metabolism

Modeling cell metabolism, necessary for dynamic optimization and predictive control of bioprocesses, poses challenges due to the intricate interplay of multiple mechanisms happening inside cells. These mechanisms encompass, e.g., transcription, translation, metabolic reactions, and allosteric regulation. The complexity arising from these mechanisms can give rise to non-identifiability problems during the construction of models due, e.g., to lack of suitable experimental data [68]. To overcome these obstacles, metabolic models often combine multiple cellular mechanisms into simplified equations and parameters [69]. However, these parameters are typically tailored to specific cultivation environments, and any deviations in conditions may compromise the accuracy and validity of the models. Furthermore, the inherent stochastic behavior of cells resulting from genetic instability and variable gene expression [70] is frequently neglected in metabolic models.

This thesis considers two modeling approaches of metabolism: constraint-based dynamic modeling and (machine-learning-supported) unsegregated/quasi-unstructured kinetic modeling.

Constraint-based models (cf. e.g. [37, 61, 71–76]) integrate mass balances of metabolism based on the stoichiometric matrix of a genome-scale metabolic network or a reduced version of it. These models are often underdetermined due to the presence of more unknowns (e.g., reaction rates) than equations. To address this issue, an optimization problem with a biologically meaningful objective function, such as maximizing growth, is typically formulated. Additional constraints based on, e.g., thermodynamics, flux capacity, and regulation, are imposed to narrow down the solution space. The first metabolic constraint-based models assumed steady-state conditions for simplicity, however, such strategies only serve to identify metabolic flux distributions but fail to capture transient metabolic phenomena. To enable the prediction of bioprocess dynamics, *dynamic* constraint-based models were introduced (cf. e.g. [72, 75, 76]). They consider time-varying species such as extracellular concentrations, while assuming pseudo-steady-state conditions of intracellular metabolites. In general, dynamic constraint-based models are solved by discretizing the differential equations, whereby the flux distribution is updated at each time step.

In the unsegregated/quasi-unstructured kinetic modeling approach, intracellular mechanisms are lumped up, focusing primarily on modeling the dynamics of extracellular species and

externally inducible intracellular components[1]. In this case, the model is *quasi*-unstructured because not all of the intracellular components are lumped up. The externally regulated intracellular components are still modeled. The cell population is treated as a homogeneous single component. The main objective of this modeling strategy is to reduce model complexity while capturing the essential dynamics of the process, although this simplification can come at the expense of model predictability and generalization properties [69, 77].

Machine-learning methods such as neural networks and Gaussian processes can be employed to augment prior knowledge-based models toward addressing model uncertainties, including oversimplifications and incorrect assumptions [78, 79]. Several studies have demonstrated the possibly improved predictability and generalization capabilities of machine-learning-supported models, commonly referred to as *hybrid* models (cf. e.g. [80–83]).

2.3 Bioprocess monitoring

Online monitoring of critical process parameters and key performance indicators associated with the critical quality attributes of a product is essential for effective process control. For example, model-based control approaches require process knowledge to provide feedback. Various methods are available for real-time monitoring of *extracellular* or general process variables such as pH, temperature, osmolarity, gas partial pressure, cell density, as well as various substrates and products. These methods encompass near-infrared spectroscopy, dielectric spectroscopy, fluorescence spectroscopy, as well as off-gas spectrometry, integrated high-performance liquid chromatography, and nanofluidic devices. For a comprehensive overview of state-of-the-art monitoring platforms for extracellular process variables, refer to [84–86].

In contrast, the real-time monitoring of *intracellular* variables, such as cell composition and metabolite concentrations, is still not established. This gap can be addressed by utilizing genetically encoded biosensors based on fluorescent proteins, enabling *in vivo* real-time estimation of intracellular states [87–90]. Nevertheless, the implementation and use of these biosensors is not straightforward. First, synthesizing the biosensors' machinery can impose a resource burden on the cell. Secondly, if multiple fluorescence-based biosensors are employed, interpreting the read-out becomes challenging due to potential overlapping in the output signal spectra. It is worth noting that proteomics and metabolomics can in principle be used to measure intracellular proteins and metabolites [91–93]. However, these methods are generally time-consuming, costly, and cannot be performed in a real-time manner.

In this thesis, *soft sensors* or state estimators are used to estimate unmeasured states of the system when hardware sensors are not available. Soft sensors employ a mathematical model of the process to infer unmeasured variables [86, 94–96]. Generally, soft sensors utilize input data from hardware sensors and actuators, and their output provides an *indirect* measurement of the state of the system, thereby facilitating process monitoring and feedback control.

[1] "Externally inducible" means that the gene expression of these components can be modulated using external signals.

2.4 Metabolic cybergenetics: a platform for dynamic metabolic engineering

The concept of metabolic cybergenetics, which forms the core of this thesis, is now formally introduced. The term "cybergenetics" refers to dynamic modulation of gene expression using computer-aided feedback control. In such scenarios, external corrective actions are calculated by an external controller[2] to keep or adjust the gene expression at desired levels [98, 99]. Currently, the utilization of cybergenetics in the field of biotechnology has primarily focused on, e.g., optogenetically regulating the expression of fluorescent proteins (*reporters*) and growth-regulatory proteins, such as antibiotic-resistance-conferring proteins and enzymes involved in the synthesis of essential amino acids [50, 60].

A notable advancement in the field of cybergenetics, holding substantial promise, is the realization of *metabolic* cybergenetic systems. Here, the term "metabolic" is added to highlight the fact that cybergenetics is extended to include dynamic metabolic engineering applications [51]. This encompasses scenarios where metabolic fluxes can be exploited as dynamic degrees of freedom for optimizing production efficiency. To do so, one can dynamically modulate the gene expression of key enzymes catalyzing metabolic fluxes of interest. The amount of these enzymes in the cell would constrain the achievable target intracellular fluxes, enabling one to *steer* metabolism toward desired directions.

To fully realize the potential of metabolic cybergenetic systems, the utilization of *model-based* optimization and predictive control methods is proposed. Thereby, both cybergenetic inputs and traditional process inputs, such as feed rates, are incorporated as optimization degrees of freedom. The optimization may involve maximizing volumetric productivity, achieving specific target product yields, or attaining desired product ratios, all while addressing system uncertainties through feedback control.

In summary, the metabolic cybergenetic framework proposed in this thesis comprises four key components, as depicted in Figure 2.1. First, it involves a cybergenetic input capable of dynamically inducing gene expression, such as light in optogenetics. Secondly, a manipulatable substrate feed stream is incorporated, particularly relevant for fed-batch systems, opening a further degree of freedom. Thirdly, the framework considers online (bio)sensors and soft sensors to enable real-time monitoring and estimation of the states of the process. Lastly, model-based dynamic optimization is employed in a closed-loop and fully automated manner. Feedback control is achieved via MPC by resolving the dynamic optimization problem with the updated state of the plant.

2.5 Case study

To highlight the applicability of metabolic cybergenetics, the dynamic control of ATP turnover for maximizing production in fermentations is considered throughout this thesis.

[2] As opposed to in-cell feedback control or self-regulation where the controller, typically encoded by genetic circuits, is located inside the cell [97].

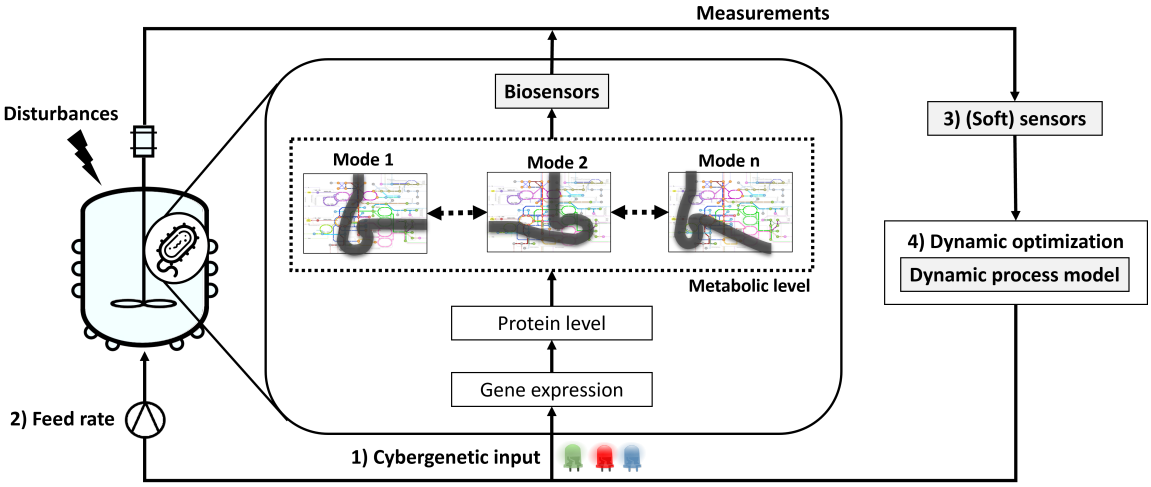


Figure 2.1: Components of the proposed metabolic cybergenetic systems, taking a fed-batch process as an example. The approach involves regulating the expression of key metabolism-related proteins through inducible gene expression systems. This enables dynamic transitions between various metabolic *modes* by manipulating external inputs, such as light in optogenetics. Model-based optimization is utilized to determine the optimal inputs of the system. The process is monitored using (bio)sensors and soft sensors. Feedback control is achieved by iteratively solving the optimization problem with the updated state of the plant.

2.5.1 Dynamic enforced ATP turnover

ATP, a crucial energy source in cells, regulates several if not all cellular processes [100]. Enforced ATP turnover has emerged as a promising strategy to enhance product yield and substrate uptake, particularly when the product pathway is linked to net ATP formation (Fig. 2.2) [40, 100–106]. This approach involves introducing an ATP drain or *wasting*[3] into the cell´s metabolism. This can be achieved by introducing ATP futile cycles or expressing an enzyme capable of directly hydrolyzing ATP. Either way, the ATP drain causes the cell to increase the product flux to compensate for the ATP loss, resulting in higher product yields and specific productivities. However, this also entails a drop in biomass yield and can lead to reduced volumetric productivity rates since biomass is the catalyst of the process [40, 102–107].

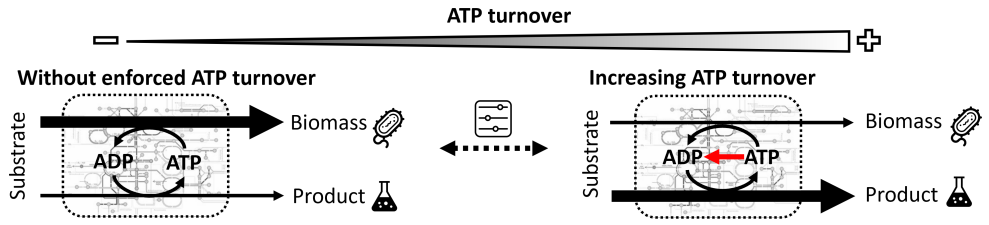


Figure 2.2: Impact of enforced ATP turnover on the cell´s metabolic flux distribution: no ATP turnover (left) against high ATP turnover (right). The red arrow indicates enforced ATP drain via ATP hydrolysis. ADP: adenosine diphosphate.

[3]Thus, the concept of "enforced ATP turnover" can be also found as "enforced ATP wasting" in the literature.

With this in mind, the fine-tuning of ATP turnover through online optimization and control represents a significant opportunity to enhance flexibility and explore new possibilities for biotechnological production. For example, one could achieve user-defined trade-offs between product yield and volumetric productivity in bioprocesses. This concept aligns with the principles of dynamic metabolic engineering, as different levels of ATP turnover would lead to distinct metabolic *modes* or temporal distributions of metabolic fluxes. Enforced ATP turnover has been investigated in both one-stage and two-stage fermentations [39, 102–107]. In two-stage processes, cells are initially grown without enforced ATP turnover, and then a production phase is initiated where (growth-arrested) cells undergo enforced ATP turnover. In contrast, one-stage processes involve continuous enforced ATP turnover throughout the entire fermentation process.

The current modeling approaches for simulating enforced ATP turnover [39] often fall short of the requirements[4] for model-based optimization and control. Therefore, further enhancements are necessary to improve the accuracy of these models. An exception is the recent development of a kinetic model of the energy and central carbon metabolism in *E. coli* that captures the effect of enforced ATP turnover in the cell´s metabolism [108]. Note that these types of kinetic models can provide great insight into metabolism, however, they are often large, highly non-linear, and parameterized, involving the kinetic description of metabolic fluxes as functions of dynamic states such as intracellular metabolites, enzymes, etc. Therefore, these models can be computationally challenging to handle in the context of model-based optimization and predictive control.

2.5.2 Anaerobic lactate fermentation by *E. coli*

As an application example of dynamic enforced ATP turnover, the anaerobic lactate synthesis by *E. coli* is considered throughout this thesis. Lactate is a platform chemical with applications in the food, beverage, cosmetic, pharmaceutical, and chemical industries [109]. Specifically, *E. coli* KBM10111 [102] is used as a basis for the case study, lacking genes for aldehyde-alcohol dehydrogenase, acetate kinase, and phosphate acetyltransferase. Consequently, the pathway from glucose to lactate becomes essential for regenerating the redox cofactors required in glycolysis, coupling lactate formation with ATP synthesis. This renders lactate fermentation on this strain suitable for the enforced ATP turnover strategy [102, 107]. More details on the specific metabolic pathway can be found in Section 4.4. Overall, *E. coli* KBM10111 exhibits favorable characteristics in terms of lactate production and minimal-to-no formation of by-products.

In the subsequent chapters, the question of how to efficiently manage enforced ATP turnover in a structured, adaptable, and flexible manner for optimizing bioprocesses will be addressed. The framework of metabolic cybergenetics will be built and generalized based on this case study.

[4] These include, e.g., good predictability, capturing dynamic transitions of metabolism, inducible gene expression, resource allocation, etc.

3 Structure and contributions

Initially, this thesis considers simplified assumptions and specific cases, gradually incorporating additional complexity. This iterative process leads to the establishment of a comprehensive metabolic cybergenetic framework capable of addressing a broader range of scenarios. The results are mostly based on simulations, however, the last chapter experimentally validates some of the outlined concepts, opening the door to more advanced experimental implementations.

Chapter 4 investigates through simulations the potential of dynamically modulating ATP turnover to maximize production in batch processes. The focus is on a simplified scenario where manipulation of ATP turnover is possible via direct modulation of an ATP-hydrolyzing intracellular flux, treating it as a *virtual* degree of freedom. Enforced ATP turnover is formulated as an optimal control problem based on constraint-based dynamic modeling. This approach aims to achieve an optimal balance between different objectives, such as product yield and volumetric productivity. MPC is used to address uncertainties such as model-plant mismatch and disturbances. The case study in this chapter focuses on anaerobic lactate fermentation by *E. coli* in batch.

Chapter 5 presents a strategy for dynamically manipulating ATP turnover by employing an optogenetic gene expression system that regulates the F_1-subunit of the ATPase. The latter catalyzes a reaction where ATP is hydrolyzed. This aims to shift the degree of freedom of the optimization from the *virtual* ATP-hydrolyzing flux in Chapter 4 to a tunable light input. To incorporate the dynamics of the optogenetic actuator, an expansion of the constraint-based dynamic model proposed in Chapter 4 is performed. An optimal control problem is formulated to determine the light inputs that maximize production. Control simulations are conducted in an open-loop manner, assuming no model-plant mismatch. Additionally, it is assumed that the cells in the bioreactor receive homogeneous light for ATPase induction. The case study in this chapter also involves anaerobic lactate fermentation by *E. coli* in batch.

Chapter 6 generalizes and expands the methods outlined in Chapters 4-5 to develop a comprehensive framework for metabolic cybergenetics. It presents a constraint-based dynamic model that can accommodate diverse metabolic systems with external regulation of the expression of metabolism-relevant proteins. The constraint-based dynamic model covers fed-batch systems and considers average gradients of cybergenetic inputs in bioreactors. The chapter explores optimal control of cybergenetic inputs and conventional process inputs such as feed rates in open-loop. It also evaluates the effectiveness of MPC as a feedback control strategy for metabolic cybergenetics. The incorporation of a soft sensor to monitor the intracellular

biomass composition is outlined, needed to facilitate MPC implementations. The case study in this chapter focuses on anaerobic lactate fermentation by *E. coli* in fed-batch.

Finally, as proof of concept of external dynamic control of metabolism, Chapter 7 experimentally validates a (Gaussian-process-supported) model-based optimization strategy for optogenetic ATPase modulation. Specifically, the focus is on anaerobic lactate fermentation by *E. coli* in batch with open-loop optogenetic control. An alternative quasi-unstructured/unsegregated kinetic model is employed. This simplified modeling approach facilitates model parameterization and streamlines the model-based optimization routine, offering experimental and computational advantages over the constraint-based dynamic modeling presented in the previous chapters.

Some results and parts of this thesis have either been published or are intended for future publication, and will not be explicitly cited within this thesis. Chapters 4-7 are mainly based, respectively, on:

- Espinel-Ríos S., Bettenbrock K., Klamt S., Findeisen R. (2022). Maximizing batch fermentation efficiency by constrained model-based optimization and predictive control of adenosine triphosphate turnover. *AIChE J.*, **68**, 4, e17555.

- Espinel-Ríos S., Morabito B., Pohlodek J., Bettenbrock K., Klamt S., Findeisen R. (2022). Optimal control and dynamic modulation of the ATPase gene expression for enforced ATP wasting in batch fermentations. *IFAC-PapersOnLine*, **55**, 7, 174–180.

- Espinel-Ríos S., Morabito B., Pohlodek J., Bettenbrock K., Klamt S., Findeisen R. (2023). Toward a modeling, optimization, and predictive control framework for fed-batch metabolic cybergenetics. *Biotechnol Bioeng, early view, doi: 10.1002/bit.28575*.

- Espinel-Ríos S., Behrendt G., Morabito B., Bauer J., Pohlodek J., Schütze A., Bettenbrock K., Klamt S., Findeisen R. (TBA). Dynamic optimization of ATPase expression exploiting hybrid Gaussian-process-supported models and optogenetic modulation with experimental validation. *To be submitted*.

Other publications were generated during the doctoral research as part of collaborations and side projects, but they have not been included in this thesis. These articles include:

- Morabito B., Pohlodek J., Kranert L., Espinel-Ríos S., Findeisen R. (2022). Efficient and simple Gaussian process supported stochastic model predictive control for bioreactors using HILO-MPC. *IFAC-PapersOnLine*, **55**, 7, 922–927.

- Boecker S., Espinel-Ríos S., Bettenbrock K., Klamt S. (2022). Enabling anaerobic growth of *Escherichia coli* on glycerol in defined minimal medium using acetate as redox sink. *Metab Eng*, **73**, 50–57.

- Espinel-Ríos S., Huber N., Alcalá-Orozco E. A., Morabito B., Rexer T. F., Reichl U., Klamt S., Findeisen R. (2022). Cell-free biosynthesis meets dynamic optimization and control: a fed-batch framework. *IFAC-PapersOnLine*, **55**, 23, 92–97.

- Espinel-Ríos S., Morabito B., Bettenbrock K., Klamt S., Findeisen R. (2022). Soft sensor for monitoring dynamic changes in cell composition. *IFAC-PapersOnLine*, **55**, 23, 98–103.
- Espinel-Ríos S., Bettenbrock K., Klamt S., Avalos J. L., Findeisen R. (2023). Machine learning-supported cybergenetic modeling, optimization and control for synthetic microbial communities, *Comput Aided Chem Eng*, **52**, 2601–2606, ISBN: 978-0-443-15274-0.
- Espinel-Ríos S., Slaviero G., Bettenbrock K., Klamt S., Findeisen R. (2023). Monitoring intracellular metabolite concentrations by moving horizon estimation based on kinetic modeling. *IFAC-PapersOnLine*, **56**, 2, 4608–4613.
- Espinel-Ríos S., Kok R., Klamt S., Avalos J. L., Findeisen R., "Batch-to-batch optimization with model adaptation leveraging Gaussian processes: the case of optogenetically assisted microbial consortia", in *2023 23rd International Conference on Control, Automation and Systems (ICCAS)*, Yeosu, Korea, Republic of: IEEE, (2023), pp. 1292–1297.

4 Maximizing batch fermentation efficiency through virtual ATP turnover

The aim of this chapter is to assess the potential of dynamically modulating ATP turnover to maximize process efficiency. Section 4.1 presents a constraint-based dynamic modeling approach to capture the impact of enforced ATP turnover on fermentation dynamics. An optimal control problem is formulated in Section 4.2 to maximize production, treating ATP turnover, linked to an ATP-hydrolyzing reaction flux, as a *virtual* degree of freedom. Section 4.3 outlines a shrinking-horizon MPC formulation to address disturbances and model-plant mismatch. The effectiveness of the optimization and predictive control of ATP turnover is demonstrated through simulations in Section 4.4, using the anaerobic lactate fermentation by *E. coli* as a case study.

It is important to note that while this chapter assumes the direct fine-tuning of the ATP-hydrolyzing reaction flux, its practical implementation is not straightforward. Nevertheless, this simplification helps to first explore the theoretical benefits of dynamically manipulating ATP turnover before delving into practical considerations, which will be the focus of the next chapters. Moreover, the methods here outlined will serve as a basis for further generalizations in the context of metabolic cybergenetic systems.

4.1 Dynamic enzyme-cost flux balance analysis

Dynamic enzyme-cost flux balance analysis (deFBA) [37, 61, 75], a constraint-based dynamic modeling framework, is considered to capture the influence of enforced ATP turnover on cellular metabolism. deFBA enables the prediction of changes in intracellular composition resulting from temporal metabolic adaptations and takes into consideration the *cost* associated with the production of cellular components. In deFBA, the biomass is composed of enzymes, ribosomes, and quota compounds[1], represented by the vector $\boldsymbol{p} \in \mathbb{R}^{n_p}$, while the corresponding molecular weights of the biomass components are contained in the vector $\boldsymbol{b} \in \mathbb{R}^{n_p}$. Therefore, biomass dry weight $B \in \mathbb{R}$ in $\mathrm{g/L}$ can be computed as

$$B(t) = \boldsymbol{b}^{\mathsf{T}} \boldsymbol{p}(t), \tag{4.1}$$

where $t \in \mathbb{R}^{+}$ is the time.

[1]These elements are essential for cell structure, maintenance, and growth, although they are not directly involved in catalytic reactions. These elements encompass, e.g., non-catalytic proteins, DNA, lipids, carbohydrates, and other small molecules.

Remark 4.1: Notation.

Henceforth, the time dependency of the variables is omitted when clear from the context. Furthermore, throughout this thesis, bold fonts are utilized to represent vectors and matrices, while non-bold fonts are used for scalar variables and parameters.

The dynamics of extracellular metabolites $\boldsymbol{z} \in \mathbb{R}^{n_z}$ and intracellular components $\boldsymbol{p}$ can be described by

$$\left[\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}t}, \frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}\right]^{\mathsf{T}} = \boldsymbol{S}_{\mathrm{zp}}\boldsymbol{V}. \tag{4.2}$$

The reaction fluxes for exchange, metabolic, and biomass-production reactions are combined in the flux vector $\boldsymbol{V} \in \mathbb{R}^{n_V}$. The stoichiometric matrix $\boldsymbol{S}_{zp} \in \mathbb{R}^{(n_z+n_p)\times n_V}$ relates the species in $\boldsymbol{z}$ and $\boldsymbol{p}$ to the reaction fluxes.

Quasi-steady-state conditions are assumed for the intracellular metabolites $\boldsymbol{m} \in \mathbb{R}^{n_m}$, implying that the change of these metabolites is much faster relative to the extracellular metabolites and biomass components[2]. Therefore,

$$\frac{\mathrm{d}\boldsymbol{m}}{\mathrm{d}t} = \boldsymbol{S}_m\boldsymbol{V} = \boldsymbol{0}, \tag{4.3}$$

where $\boldsymbol{S}_m \in \mathbb{R}^{n_m\times n_V}$ represents the stoichiometric matrix relating the species in $\boldsymbol{m}$ to the reaction fluxes.

Assuming that enzymes operate under substrate saturation conditions, i.e., that they are close to their maximum reaction rates, the metabolic fluxes are constrained by the intracellular enzyme concentration

$$\sum_{j\in\mathrm{cat}(i)} \left|\frac{V_j}{k_{\mathrm{cat},j}}\right| \leq p_i, \quad \forall i \in [1, n_p], \tag{4.4}$$

where $\mathrm{cat}(i)$ represents the set of reactions catalyzed by enzyme p_i and $\boldsymbol{k}_{\mathrm{cat}} \in \mathbb{R}^{n_{\mathrm{cat}}}$ comprises the catalytic constants. $|\cdot|$ denotes the absolute value operator.

A lumped quota compound $p_Q \in \mathbb{R}$ corresponds, at minimum, to a fraction $\varphi_Q \in [0,1]$ of the biomass dry weight

$$\varphi_Q(\boldsymbol{b}^{\mathsf{T}}\boldsymbol{p}) \leq p_Q. \tag{4.5}$$

In addition, metabolic fluxes are subject to biologically feasible lower and upper bounds

$$\boldsymbol{V}_{\min} \leq \boldsymbol{V} \leq \boldsymbol{V}_{\max}. \tag{4.6}$$

Likewise, feasible bounds for the dynamic states are considered

$$\boldsymbol{p}_{\min} \leq \boldsymbol{p} \leq \boldsymbol{p}_{\max},\ \boldsymbol{z}_{\min} \leq \boldsymbol{z} \leq \boldsymbol{z}_{\max}. \tag{4.7}$$

[2]This can be a limitation of the model if the internal metabolites actually accumulate in the cell, influencing the resulting metabolic state.

The conditions of the system at the initial time t_0 are given by

$$\boldsymbol{p}(t_0) = \boldsymbol{p}_0,\, \boldsymbol{z}(t_0) = \boldsymbol{z}_0. \tag{4.8}$$

The system of equations (4.1)-(4.8) is frequently underdetermined due to, e.g., lack of mathematical descriptions of the reaction rates, which leads to more (unknown) fluxes than (known) equations. To tackle this challenge, it is commonly assumed that cells have evolved to optimize a specific biological cost function, denoted as

$$\max_{\boldsymbol{V}(\cdot)} \quad \int_{t_0}^{t_0+\Delta t_{\text{bio}}} F_V(\cdot)\,\mathrm{d}t, \tag{4.9a}$$

with, for example, the vector *function* of metabolic fluxes $\boldsymbol{V}(\cdot)$ as decision variable. $F_V(\cdot)$ represents the assumed objective function that the cell maximizes over the time horizon $[t_0, t_0 + \Delta t_{\text{bio}}]$, where Δt_{bio} is the length of the horizon. In deFBA models, it is common to choose the integral of the biomass as the cell´s objective function [37, 61, 75, 76, 110–112], which has demonstrated effectiveness under normal growth conditions but may not hold true during substrate starvation [110]. In this thesis, the integral of the biomass is assumed to be the cell´s objective function as only scenarios where substrate is available for conversion are considered.

Therefore, batch fermentation processes can be described by

$$\max_{\boldsymbol{V}(\cdot)} \quad \int_{t_0}^{t_0+\Delta t_{\text{bio}}} B\,\mathrm{d}t, \tag{4.10a}$$

$$\text{s.t.} \quad \text{Eqs.}\,(4.1) - (4.8). \tag{4.10b}$$

Solving the above dynamic optimization problem allows for simulating and predicting the cell's dynamic behavior.

4.2 Open-loop control of virtual ATP turnover

Using the previous model, one can set up a model-based optimization problem to find the ATP turnover values for maximizing production. The F_1-subunit of the ATPase[3], responsible for the hydrolysis of ATP into ADP, is considered as the enforced ATP turnover mechanism [40, 103–105, 107, 108, 113]. The *virtual* ATPase flux is denoted as $V_{\text{ATPase}} \in \mathbb{R}$.

Temporal changes in the product and biomass yields are expected to occur when the ATP turnover is dynamically manipulated. Therefore, the goal is to determine an optimal strategy for

[3] From now on, the F_1-subunit of the ATPase enzyme/gene will be referred to as the "ATPase enzyme/gene", unless unclear from the context.

manipulating the ATPase flux that maximizes the efficiency of the batch process, defined by the cost function $J(\cdot)$. This dynamic optimization problem is formulated as

$$\max_{V_{\mathrm{ATPase}}(\cdot)} \quad J(\cdot), \tag{4.11a}$$

$$\text{s.t.} \quad (4.10). \tag{4.11b}$$

The approach provides the flexibility to consider various optimization objectives, such as maximizing product-to-substrate yield, volumetric productivity, or a multi-objective cost function. The decision variable in the optimal control problem corresponds to the manipulated ATPase flux *function* $V_{\mathrm{ATPase}}(\cdot)$. In the deFBA model, the ATPase flux is treated as determined by the controller, hence the cell has no *choice* over it. It should be noted that the optimization in (4.11) represents a bilevel optimal control problem if the constrained model (4.10) is used, which involves both an inner and outer optimization problem. The numerical connotations of the bilevel optimization nature of the optimal control problem and the infinite dimensionality of the functions $V_{\mathrm{ATPase}}(\cdot)$ and $\boldsymbol{V}(\cdot)$ in the optimization problems are addressed in Section 4.4.2.

4.3 Shrinking-horizon model predictive control of virtual ATP turnover

Finding an optimal open-loop input is possible by solving problem (4.11). However, this does not allow counteracting disturbances or model uncertainties. As introduced in Section 2.1, MPC implements an optimal control problem where the open-loop input trajectory is only partially applied. As time progresses, the control action is periodically re-evaluated, taking into account the updated state of the plant (Fig. 4.1). This re-evaluation allows for incorporating online feedback [57, 114, 115]. As a result, the adverse effects of short-term unknown uncertainties such as model-plant mismatch and disturbances can be mitigated.

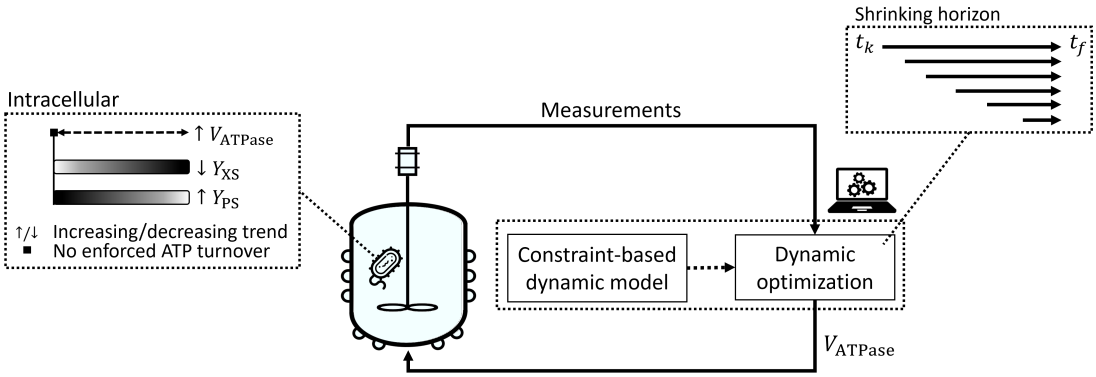


Figure 4.1: Shrinking-horizon MPC strategy with enforced ATP turnover via the ATPase flux as a *virtual* degree of freedom. The optimal control problem is solved, and the predicted ATPase flux is *applied* to the plant. State measurements are used to iteratively re-solve the control problem over a shrinking time horizon. Y_{PS}: product-on-substrate yield, Y_{XS}: biomass-on-substrate yield.

Without loss of generality, it is assumed that state measurements are available at equidistant intervals. The sampling times for measurements are denoted as t_k, where $t_k := kh_s$, $k \in \mathbb{N}_0$, and

h_s is a fixed sampling interval. Additionally, the controller is designed to predict the system's behavior up to a final time $t_f := Nh_s$, where $N \in \mathbb{N}$ represents the number of steps in the prediction horizon. This results in a shrinking prediction horizon spanning from t_k to t_f. The shrinking-horizon MPC of the virtual ATPase flux reads

$$\max_{V_{\text{ATPase}}(\cdot)} \quad J(\cdot), \tag{4.12a}$$

$$\text{s.t.} \quad \max_{\boldsymbol{V}(\cdot)} \int_{t_k}^{t_k+\Delta t_{\text{bio}}} B\,\mathrm{d}t, \tag{4.12b}$$

$$\text{s.t.} \quad \text{Eqs. (4.1)- (4.7)}, \tag{4.12c}$$

$$\left[\boldsymbol{p}(t_k)^\mathsf{T}, \boldsymbol{z}(t_k)^\mathsf{T}\right]^\mathsf{T} = \left[\tilde{\boldsymbol{p}}_{\boldsymbol{k}}^\mathsf{T}, \tilde{\boldsymbol{z}}_{\boldsymbol{k}}^\mathsf{T}\right]^\mathsf{T}, \tag{4.12d}$$

where $\tilde{\boldsymbol{p}}_{\boldsymbol{k}}$ and $\tilde{\boldsymbol{z}}_{\boldsymbol{k}}$ denote the measured values of $\boldsymbol{p}_{\boldsymbol{k}}$ and $\boldsymbol{z}_{\boldsymbol{k}}$ at t_k.

4.4 Anaerobic lactate fermentation by *E. coli* with virtual ATP turnover

The lactate synthesis by *E. coli* KBM10111[4], which is coupled with ATP formation under anaerobic conditions, is employed as a case study for the dynamic enforced ATP turnover strategy.

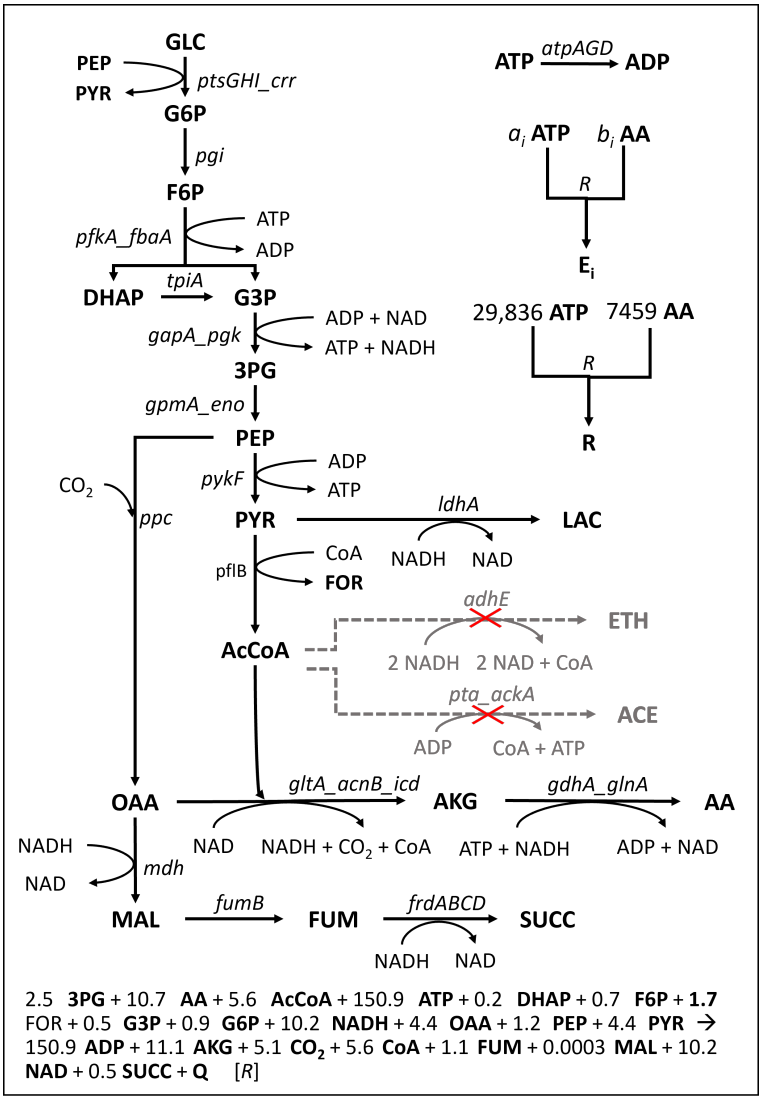
4.4.1 Resource allocation model

The derivation of the lactate fermentation model followed a similar protocol as suggested by [116]. The NetworkReducer algorithm [117] was employed to construct a reduced metabolic network based on the ECGS genome-scale model of *E. coli* [118], using literature values for the strains´ growth, glucose, and lactate extracellular rates [102]. This generated the network depicted in Fig. 4.2. A reaction was included to account for amino acid production, with glutamine selected as the reference amino acid. Furthermore, an ATP-hydrolyzing reaction catalyzed by the ATPase was incorporated. The catalytic constants for the reactions were determined using median values from the BRENDA [119] and SABIO-RK [120] databases[5].

To determine the quota reaction, it was assumed that approximately 67 % of the *E. coli´*s cell dry weight consists of quota elements, while the remaining 33 % is attributed to catalytic enzymes and ribosomes [121]. The catalytic constants for biomass-producing reactions were computed based on a translation rate of ribosomes (12 amino acids per second) [121]. The molecular weight values of gene products were obtained from UniProt [122]. Note that, unless otherwise stated, the vectors $\boldsymbol{p}$ and $\boldsymbol{z}$ in the deFBA model are given in mM, while the fluxes are expressed in mM/h. The derived resource allocation model can be seen in Table 4.1. Note that the ATPase flux corresponds to reaction no. 9 in the latter table.

[4]Throughout the rest of this chapter, this strain will be referred to as "*E. coli*", unless otherwise unclear from the context.

[5]Minor adjustments were made for specific enzymes (ptsGHI_crr, pfkA_fbaA, gapA_pgk, gpmA_eno, ppc, and gdhA_glnA) by selecting individual catalytic constant entries from the databases that better aligned with the experimental data.



3PG: 3-phospho-D-glycerate; AA: amino acid; AcCoA: acetyl-CoA; ACE: acetate; ADP: adenosine diphosphate; AKG: alpha-ketoglutarate; ATP: adenosine triphosphate; CO_2: carbon dioxide; CoA: coenzyme A; DHAP: dihydroxyacetone phosphate; ETH: ethanol; E_i: enzyme i; F6P: fructose 6-phosphate; FOR: formate; FUM: fumarate; G3P: glyceraldehyde 3-phosphate; G6P: glucose 6-phosphate; GLC: glucose; LAC: lactate; MAL: malate; NAD: nicotinamide adenine dinucleotide; NADH: NAD-reduced; OAA: oxaloacetic acid; PEP: phosphoenolpyruvate; PYR: pyruvate; Q: quota; R: ribosome; SUCC: succinate.

Figure 4.2: Resource allocation model for anaerobic lactate fermentation by *E. coli* KBM10111. The metabolism is anaerobic as no external electron acceptors are used. Genes of catalytic species are italicized. Lumped enzymes are denoted by underscore symbols. Blocked pathways are shown in gray.

Table 4.2: Metabolic reactions and associated parameters in the resource allocation model for the considered anaerobic lactate fermentation.

Metabolic part				**Biomass-producing part**[a]			
No.	**Reaction**	**Enzyme(s)**[b]	k_{cat} $[\text{min}^{-1}]$	**No.**	**Reaction**	b [g/mmol]	k_{cat} $[\text{min}^{-1}]$
1	GLC + PEP → G6P + PYR	ptsGHI_crr	12,600	17	2358 AA + 9432 ATP → 9432 ADP + ptsGHI_crr	255.85	0.31
2	G6P → F6P	pgi	84,600	18	1098 AA + 4392 ATP → 4392 ADP + pgi	123.06	0.66
3	ATP + F6P → ADP + DHAP + G3P	pfkA_fbaA	630	19	1998 AA + 7992 ATP → 7992 ADP + pfkA_fbaA	217.66	0.36
4	DHAP → G3P	tpiA	49,002	20	510 AA + 2040 ATP → 2040 ADP + tpiA	53.94	1.41
5	ADP + G3P + NAD → 3PG + ATP + NADH	gapA_pgk	3,120	21	1711 AA + 6844 ATP → 6844 ADP + gapA_pgk	183.25	0.42
6	3PG → PEP	gpmA_eno	1,500	22	1364 AA + 5456 ATP → 5456 ADP + gpmA_eno	148.42	0.53
7	ADP + PEP → ATP + PYR	pykF	7,920	23	1880 AA + 7520 ATP → 7520 ADP + pykF	202.92	0.38
8	NADH + PYR → LAC + NAD	ldhA	83,520	24	1316 AA + 5264 ATP → 5264 ADP + ldhA	146.14	0.55
9	ATP → ADP	atpAGD	612	25	3206 AA + 12824 ATP → 12824 ADP + atpAGD	348.22	0.22
10	CO_2 + PEP → OAA	ppc	32,400	26	3532 AA + 14128 ADP → 14128 ADP + ppc	396.25	0.2
11	CoA + PYR → AcCoA + FOR	pflB	714	27	1520 AA + 6080 ATP → 6080 ADP + pflB	170.71	0.47
12	AcCoA + NAD + OAA → AKG + CO_2 + CoA + NADH	gltA_acnB_icd	192	28	5124 AA + 20496 ATP → 20496 ADP + gltA_acnB_icd	566.6	0.14
13	AKG + ATP + NADH → AA + ADP + NAD	gdhA_glnA	3,000	29	8310 AA + 33240 ATP → 33240 ADP + gdhA_glnA	914.33	0.09
14	NADH + OAA → MAL + NAD	mdh	17,328	30	624 AA + 2496 ATP → 2496 ADP + mdh	64.67	1.15
15	MAL → FUM	fumB	444	31	1096 AA + 4384 ATP → 4384 ADP + fumB	120.21	0.66
16	FUM + NADH → NAD + SUCC	frdABCD	14,400	32	1096 AA + 4384 ATP → 4384 ADP + frdABCD	121.22	0.66

Table continuation.

Metabolic part				**Biomass-producing part**[a]			
No.	**Reaction**	**Enzyme(s)**[b]	k_{cat} $[\text{min}^{-1}]$	**No.**	**Reaction**	b [g/mmol]	k_{cat} $[\text{min}^{-1}]$
				33	7459 AA + 29836 ATP → 29836 ADP + R	2,500.00	0.1
				34	2.5 3PG + 10.7 AA + 5.6 AcCoA + 150.9 ATP + 0.2 DHAP + 0.7 F6P + 1.7 FOR + 0.5 G3P + 0.9 G6P + 10.2 NADH + 4.4 OAA + 1.2 PEP + 4.4 PYR → 150.9 ADP + 11.1 AKG + 5.1 CO_2 + 5.6 CoA + 1.1 FUM + 0.0003 MAL + 10.2 NAD + 0.5 SUCC + Q	1.00[c]	66

[a] It is assumed that ribosomes catalyze all biomass-producing reactions.
[b] Gene IDs used for naming the enzymes. Underscore symbols are used to indicate lumped enzymes.
[c] Reaction 34 given in terms of g/L of Q; therefore, $b_Q = 1$ for simplification purposes.
3PG: 3-phospho-D-glycerate; AA: amino acid; AcCoA: acetyl-CoA; ADP: adenosine diphosphate; AKG: alpha-ketoglutarate; ATP: adenosine triphosphate; CO_2: carbon dioxide; CoA: coenzyme A; DHAP: dihydroxyacetone phosphate; F6P: fructose 6-phosphate; FOR: formate; FUM: fumarate; G3P: glyceraldehyde 3-phosphate; G6P: glucose 6-phosphate; GLC: glucose; LAC: lactate; MAL: malate; NAD: nicotinamide adenine dinucleotide; NADH: NAD-reduced; OAA: oxaloacetic acid; PEP: phosphoenolpyruvate; PYR: pyruvate; Q: quota; R: ribosome; SUCC: succinate.
Species in $\boldsymbol{z}$: CO_2, FOR, GLC, LAC, SUCC.
Species in $\boldsymbol{p}$: all listed enzymes, Q, and R.
Species in $\boldsymbol{m}$: 3PG, AA, AcCoA, ADP, AKG, ATP, CoA, DHAP, F6P, FUM, G3P, G6P, MAL, NAD, NADH, OAA, PEP, and PYR.
$\varphi_Q = 0.67$.

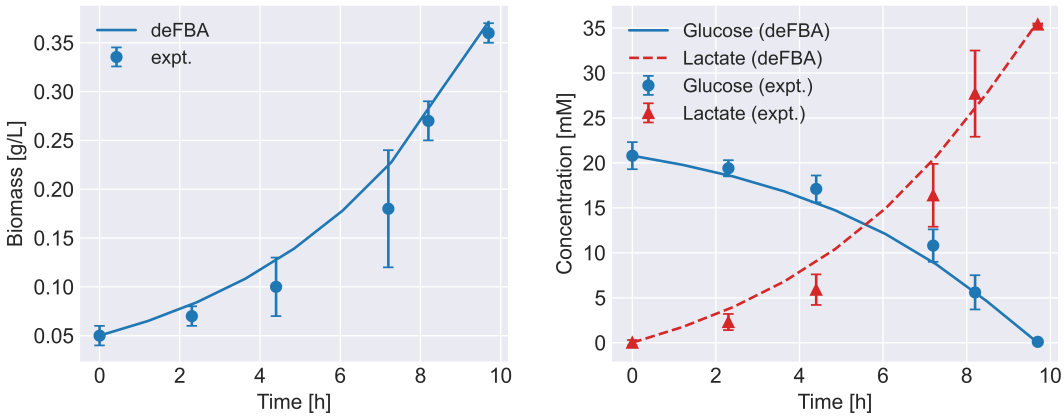


Figure 4.3: Comparison of deFBA model simulation with experimental data (expt.) [102] for biomass, glucose, and lactate concentrations. Scenario without enforced ATP turnover.

The derived deFBA model for lactate fermentation was validated by comparing its predictions with published experimental data [102] (Fig. 4.3). The validation focused on the scenario without enforced ATP turnover. To run the deFBA simulation, the vector $\boldsymbol{p}(t_0)$ was initialized as described in Remark 4.2. Overall, taking into account the standard deviations, the dynamic model showed good agreement with the experimental data.

Remark 4.2: Initial concentration of biomass components.

The estimation of $\boldsymbol{p}(t_0)$ was performed using resource balance analysis [37]. This approach relies on resource allocation theory [123], which assumes that cells aim to maximize growth while optimally allocating the biomass components. It considers steady-state conditions and uses the known initial biomass dry weight concentration for the estimation. Further details can be found in Appendix A.1. This method is consistently applied throughout this thesis whenever mentioned.

4.4.2 Open-loop optimization results

The objective function for the optimal control problem in (4.11) is chosen to maximize the final lactate concentration, denoted as $J(\cdot) = z_{\mathrm{LAC}}(t_f)$. The problem involved eight control actions. For simplicity, $t_0 + \Delta t_{\mathrm{bio}} = t_f$ was assumed to avoid covering time frames beyond t_f, likely to correspond to substrate starvation scenarios, for which the biomass integral objective function is deemed unreliable, as previously mentioned. Note that the metabolic flux values were bounded between 0 and 1000 mM/h. Also, the final time was not an optimization variable but rather *a priori* fixed in the optimization problems. Refer to Remark 4.3 for more information on the numerical solution of the optimization problems.

Remark 4.3: Numerical solution of the optimization problems.

Optimization problems pose a challenge when they are formulated with decision variables as *functions* (infinite-dimensional) and in a bilevel manner. Practical solutions can be obtained by approximating these problems using finite-dimensional approaches, see e.g. [57, 115]. In this thesis, piece-wise constant inputs are assumed and orthogonal collocation with Lagrange interpolation polynomials is employed to discretize the ordinary differential equations [75]. Furthermore, bilevel optimizations are transformed into single-level optimizations with complementarity constraints by applying the Karush-Kuhn-Tucker conditions to the lower-level optimization problem, following an optimistic approach [124–126]. Throughout this thesis, optimizations are implemented in Python (https://www.python.org/) using the CasADi library [127] and the IPOPT solver [128], unless indicated otherwise.

Metabolic trade-offs resulting from enforced ATP turnover

Initially, the focus was on analyzing the influence of fermentation time on the volumetric productivity in batch processes with enforced ATP turnover. As mentioned in Section 2.5, the enforced ATP turnover mechanism involves a trade-off between product yield and volumetric productivity. If longer process times are allowed in the optimal control problem compared to the system without enforced ATP turnover, it is expected that the enforced ATP turnover mechanism will be *activated* to increase the product yield and consequently the final lactate concentration[6].

The optimal control problem of ATP turnover was solved in open-loop using different batch times to explore a range of trade-offs between average fermentation metrics[7] such as product-on-substrate yield ($Y_{\mathrm{PS,batch}}$), biomass-on-substrate yield ($Y_{\mathrm{XS,batch}}$), and product volumetric productivity ($r_{\mathrm{P,batch}}$). This analysis resulted in multiple sets of $Y_{\mathrm{PS,batch}}$, $Y_{\mathrm{XS,batch}}$, and $r_{\mathrm{P,batch}}$ for selected values of t_f (see Figure 4.4). All scenarios presented in Figure 4.4 correspond to fermentations with 100 % substrate consumption efficiencies; furthermore, *no model-plant mismatch was assumed at this stage*. In the open-loop optimizations, resource balance analysis (cf. Remark 4.2) was utilized to estimate the initial $\boldsymbol{p}(0)$ vector based on the given initial biomass concentration.

The scenario without dynamic enforced ATP turnover, depicted in green in Fig. 4.4, showed the highest volumetric productivity, as anticipated. Longer final batch times led to distinct ATP turnover trajectories, as will be illustrated in later examples, resulting in increased batch product yields at the expense of biomass yield and product volumetric productivity. With longer final batch times, the predicted product-on-substrate yield approached the maximum theoretical value of 2 mol lactate per mol glucose, indicating a gradual redirection of substrate flux toward lactate synthesis. In essence, a longer allowed final batch time reduces the dependence on biomass

[6] Provided the same initial conditions and that all substrate available in the fermentation is fully depleted by the end of the batch.

[7] These metrics were calculated for the entire batch duration.

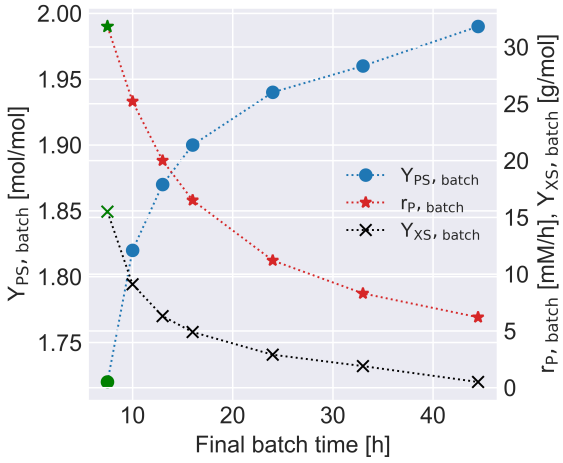


Figure 4.4: Impact of final batch time on yields ($Y_{\mathrm{PS,batch}}$ and $Y_{\mathrm{XS,batch}}$) and volumetric productivity ($r_{\mathrm{P,batch}}$). Plot based on open-loop optimizations without model-plant mismatch and given 100 % substrate consumption. Green: scenario without enforced ATP turnover. Relevant initial conditions for all scenarios: $B(0) = 0.59\,\mathrm{g/L}$, $z_{\mathrm{GLC}}(0) = 139\,\mathrm{mM}$, $z_{\mathrm{LAC}}(0) = 0\,\mathrm{mM}$.

concentration to maximize final lactate titer, as lower fermentation rates are compensated by higher product yields. These findings illustrate that the ATPase flux, as an *intracellular* degree of freedom, enables different trade-offs between product yield and volumetric productivity, potentially bringing more flexibility and adaptability to the operation of bioprocesses.

Fermentation dynamics under enforced ATP turnover

To exemplify the fermentation dynamics under the effect of enforced ATP turnover, the scenario for a batch time of 16 h (cf. Fig. 4.4) is considered. The corresponding fermentation metrics for the 16 h-batch are $Y_{\mathrm{PS,batch}} = 1.90\,\mathrm{mol/mol}$, $r_{\mathrm{P,batch}} = 16.5\,\mathrm{mM/h}$, and $Y_{\mathrm{XS,batch}} = 4.9\,\mathrm{g/mol}$. Compared to the scenario without enforced ATP turnover, this represents an 11 % increase in the batch product yield at the expense of a 48 % drop in the product volumetric productivity. As shown in Fig. 4.5, the optimizer predicted gradual increments of ATPase flux over time. Therefore, biomass showed a higher growth rate at the beginning compared to the later stages of the process. On the other hand, the product concentration exceeded the maximum achievable with the scenario without enforced ATP turnover, for the same initial conditions. In summary, a good balance was achieved by the optimizer in improving product yield against reducing biomass yield, ultimately maximizing the final lactate concentration in the specified batch time.

The optimization results indicated a gradual increase in enforced ATP turnover, rather than a distinct phase of fully OFF-ATP turnover (*growth-dominant*) followed by fully ON-ATP turnover (*production-dominant*). In order to compare the performance of the optimization against a steep step-like (OFF-to-ON) ATPase-flux trajectory, the optimal control problem was modified accordingly. However, it was consistently found that such a scenario was not feasible using the deFBA model. This limitation arises from the cell's inability to instantly reallocate resources to achieve the necessary ATPase enzyme level for the fully ON ATP-turnover phase. In

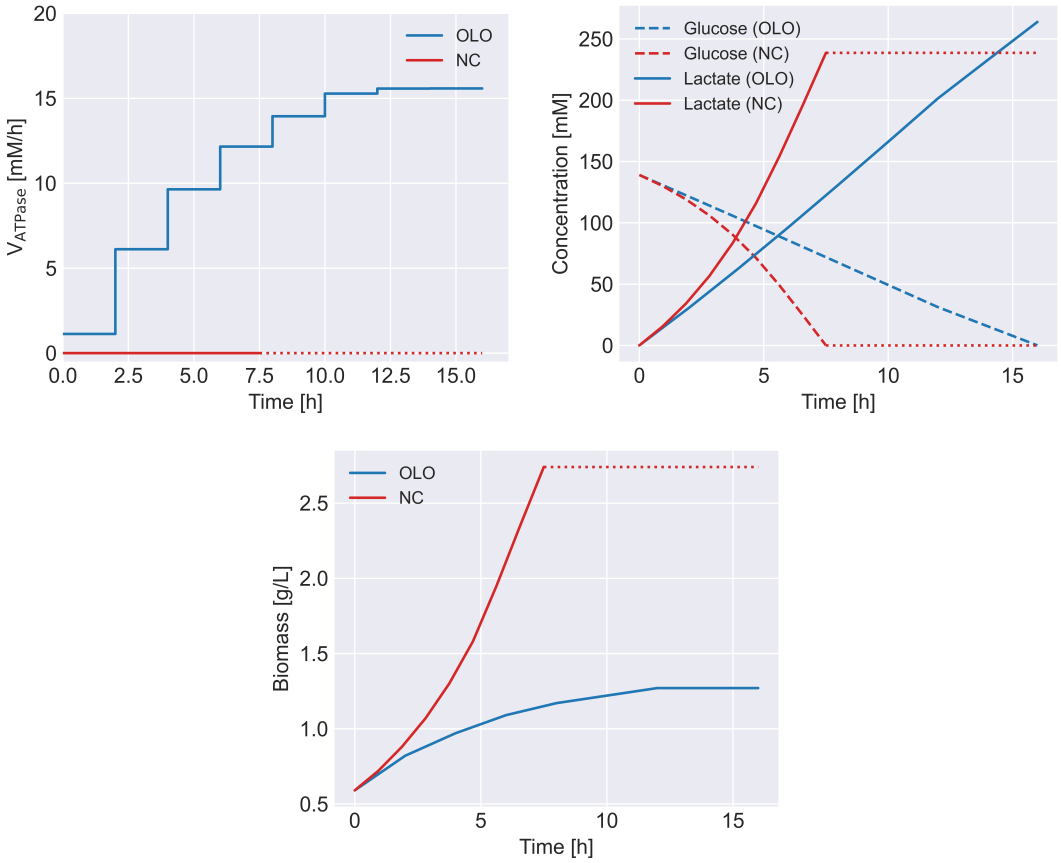


Figure 4.5: Open-loop optimization (OLO) results for a final batch time of 16 h compared to the case without ATP turnover (NC). Initial conditions: $B(0) = 0.59\,\mathrm{g/L}$, $z_{\mathrm{GLC}}(0) = 139\,\mathrm{mM}$, $z_{\mathrm{LAC}}(0) = 0\,\mathrm{mM}$. No model-plant mismatch.

contrast, the predicted gradual increments of ATPase flux were feasible, as the cell can gradually produce and accumulate the necessary ATPase without compromising the resources needed for maintaining proper growth and metabolic functions. In other words, from a mathematical perspective, the upper bounds of the fluxes in the deFBA model are constrained by the product of enzyme concentrations and catalytic constants (Eq. (4.4)). Consequently, based on the deFBA simulation results, one cannot abruptly impose a high ATPase flux immediately after a fully OFF ATP-turnover stage because there would not be sufficient ATPase.

The dynamic evolution of intracellular enzyme concentrations for the above-mentioned open-loop optimization with enforced ATP turnover is shown in Fig. 4.6. A gradual buildup of the ATPase, necessary to enable the increasing ATPase fluxes as predicted by the optimizer, is observed. The enzyme distribution within cells subject to enforced ATP turnover exhibits a different profile compared to the scenario without enforced ATP turnover. This difference arises from the dynamic reallocation of cellular resources toward ATPase synthesis, thereby impacting the intracellular levels of other enzymes. For example, enzymes such as pfkA_fbaA, gpmA_eno, and gapA_pgk experienced significant reductions as the ATPase concentration in the cell increased.

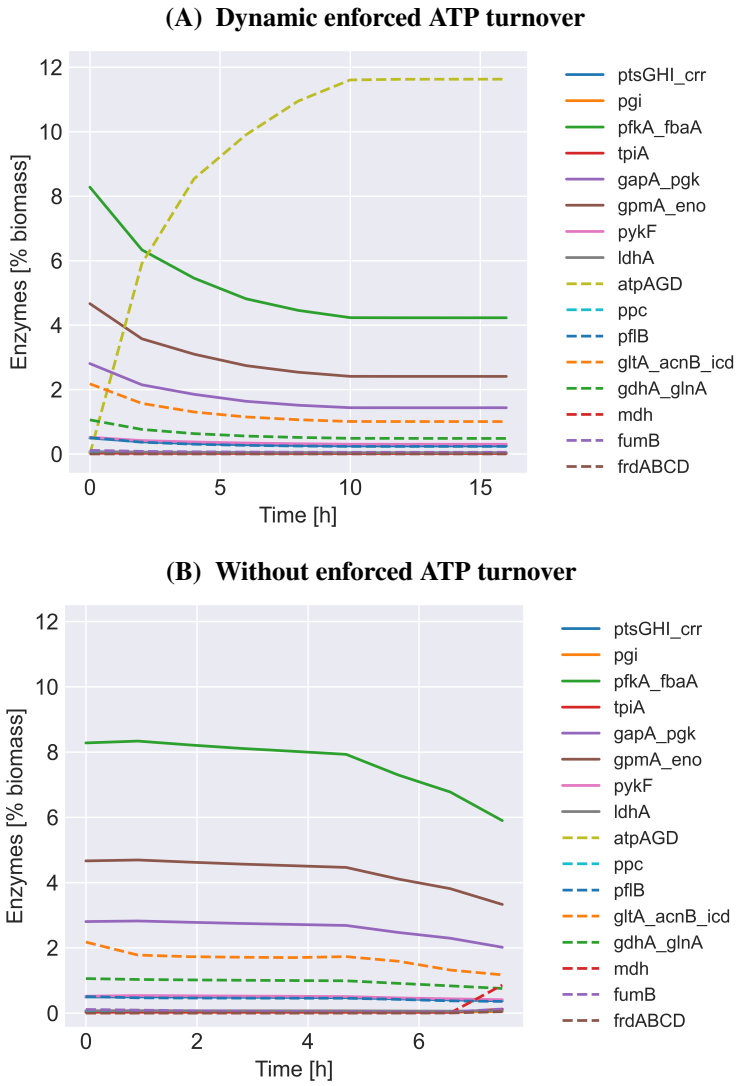


Figure 4.6: Temporal changes in enzyme concentrations for the scenarios shown in Fig. 4.5.

4.4.3 Closed-loop control results

The potential of shrinking-horizon MPC to handle model-plant mismatch is now assessed. For all MPC simulations in this section, a final batch time of 13 h is used[8]. The MPC simulations employed resource balance analysis to obtain $\boldsymbol{p}(t_0)$ (cf. Remark 4.2). Two MPC scenarios are considered:

- **MPC scenario 1** corresponds to the *nominal* case, where there is no model-plant mismatch. In addition, it is assumed that all the system states can be measured. While this scenario may seem trivial and very optimistic, it serves as an initial test of the MPC implementation.

- **MPC scenario 2** examines model-plant mismatch. To introduce this mismatch, the k_{cat} values for reactions 3, 5, 6, 12, and 13 (cf. Table 4.1) were reduced by a scaling factor

[8] Based on the open-loop optimal control analysis in Fig. 4.4, without model-plant mismatch, this scenario corresponds to $Y_{\text{PS,batch}} = 1.87\,\text{mol/mol}$, $r_{\text{P,batch}} = 20\,\text{mM/h}$, and $Y_{\text{XS,batch}} = 6.3\,\text{g/mol}$.

of 0.75. Additionally, φ_Q was increased to 0.69. The controller utilizes the nominal parameters listed in Table 4.1, while the plant operates with the modified parameters. It is assumed that all states can be measured.

As anticipated, for MPC scenario 1 (Fig. 4.7), the optimal ATPase-flux trajectory obtained from the open-loop optimization aligned with the MPC result. This can be attributed to Bellman's principle of optimality [129] as a shrinking-horizon scheme was employed in the MPC, i.e., end-pieces of optimal trajectories in the nominal case are optimal.

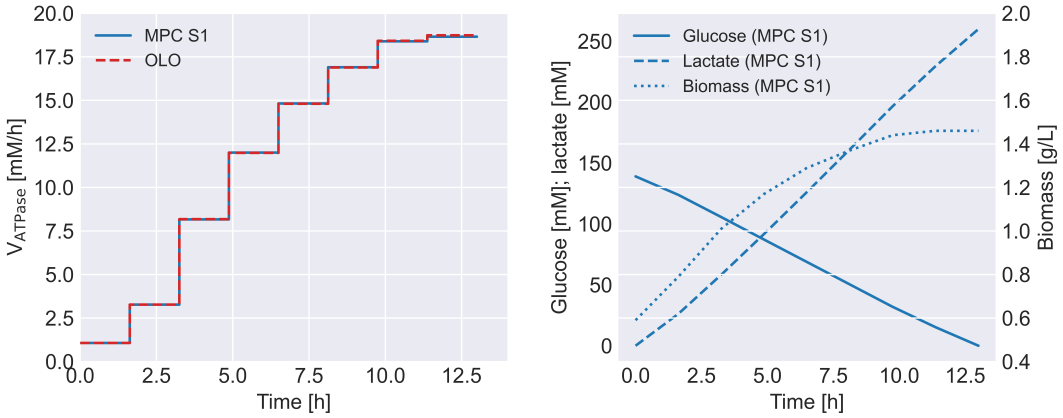


Figure 4.7: Results for MPC scenario 1 (S1) against the open-loop system (OLO). Only the concentrations for MPC S1 of glucose, lactate, and biomass are shown because the trends for the OLO system were identical. Initial conditions: $B(0) = 0.59\,\mathrm{g/L}$, $z_{\mathrm{GLC}}(0) = 139\,\mathrm{mM}$, $z_{\mathrm{LAC}}(0) = 0\,\mathrm{mM}$. Batch time: $t_f = 13\,\mathrm{h}$. No model-plant mismatch.

Figure 4.8 illustrates the performance of MPC scenario 2. MPC successfully accounted for the reduced growth rate by delivering a *corrected* ATPase-flux trajectory, which is less intense compared to the open-loop system. As previously explained, higher ATPase fluxes (linked to higher ATP turnovers) correspond to increased product yields and decreased biomass yields. Consequently, there was greater biomass accumulation toward the middle-to-end phase of the batch in MPC scenario 2 due to the less intense ATP turnover. Additionally, MPC exhibited complete consumption of glucose in the medium, in contrast to the open-loop system with model-plant mismatch, which only achieved 90 % glucose consumption.

4.5 Summary

An optimal control strategy for maximizing the efficiency of batch anaerobic lactate fermentation by *E. coli* levering an ATP-hydrolyzing flux as a virtual degree of freedom was presented. It combines constraint-based modeling with repeated solutions of an optimal control problem -shrinking horizon MPC-. The example demonstrated the potential of enforced ATP turnover for obtaining trade-offs between product yield and volumetric productivity in a flexible way. The fermentation dynamics were described using constraint-based dynamic modeling. MPC showed potential for feedback control of enforced ATP turnover in the presence of model uncertainty.

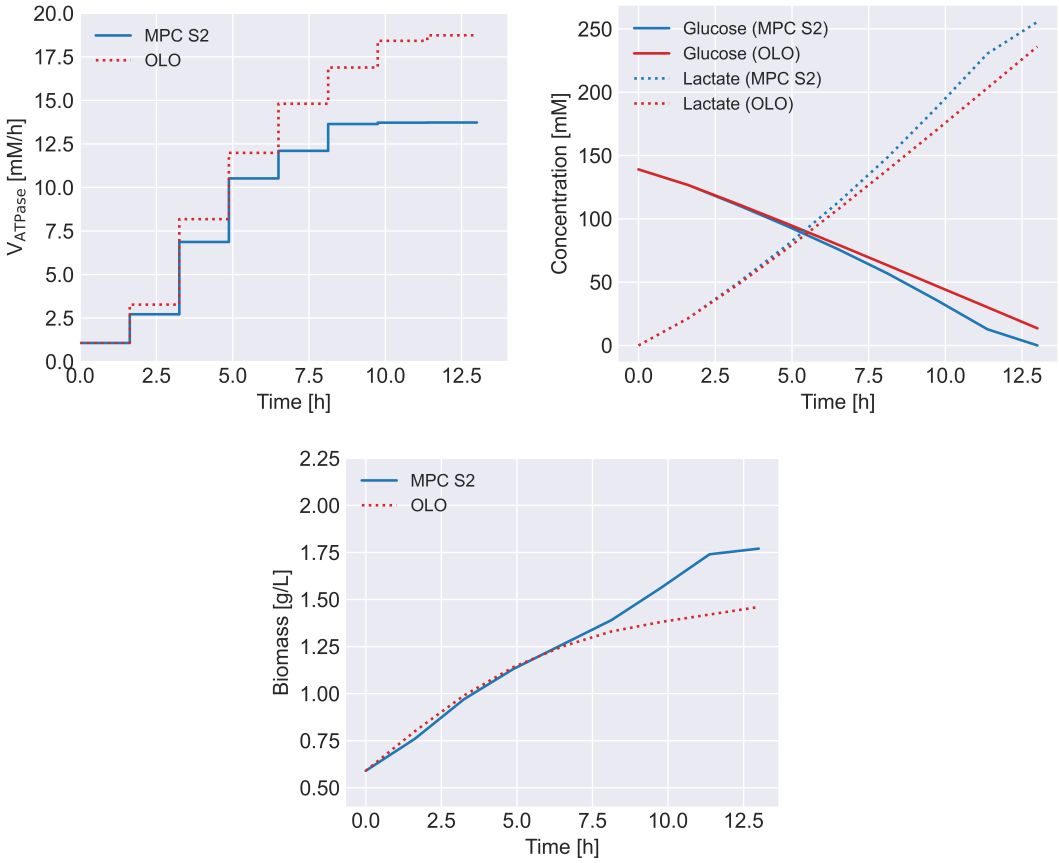


Figure 4.8: Results for MPC scenario 2 (S2) against the open-loop system (OLO). Initial conditions: $B(0) = 0.59\,\mathrm{g/L}$; $z_{\mathrm{GLC}}(0) = 139\,\mathrm{mM}$; $z_{\mathrm{LAC}}(0) = 0\,\mathrm{mM}$. Batch time: $t_f = 13\,\mathrm{h}$. Model-plant mismatch considered.

Real-life applications of the presented control strategy face several challenges that will be addressed in the subsequent chapters. First, ATP turnover, connected to the intracellular ATPase flux, was treated as a virtual degree of freedom. However, direct manipulation of the ATPase flux is not straightforward. Here, for simplicity, it was assumed that the cell could accommodate the required ATPase concentration to enable the predicted ATPase fluxes while satisfying the considered resource allocation constraints of the deFBA model. For practical implementation, however, it is essential to find practical approaches for dynamically manipulating intracellular fluxes such as the ATPase flux, e.g., using inducible gene expression systems to modulate the intracellular amount of the catalytic enzymes.

Secondly, in the MPC simulations, it was assumed that all states can be measured online. While this assumption may hold for biomass and extracellular species, measuring intracellular biomass components poses technical difficulties. Therefore, the development of efficient strategies, e.g., soft sensors, for monitoring these intracellular variables is necessary. Another point is how to extend the approach to other bioreactor modes of operation, such as fed-batch, to increase the degrees of freedom for optimization. Furthermore, a remaining question is how to extend and generalize these methods for metabolic cybergenetics.

5 Dynamic modulation of the gene expression of the ATPase in batch

In the previous chapter, optimal control problems were formulated based on constraint-based dynamic modeling to maximize batch fermentation efficiency through enforced ATP turnover. To do so, an ATP-hydrolyzing flux was considered as a *virtual* degree of freedom. This allowed for achieving different trade-offs between product yield and volumetric productivity. However, one of the challenges for the practical realization of such an approach is how to effectively fine-tune the ATP-hydrolyzing flux in cells. Section 5.1 of this chapter outlines a solution strategy involving the use of an optogenetic system that regulates the expression of the ATPase, responsible for the hydrolysis of ATP into ADP. With optogenetics, one can in principle fine-tune the intracellular ATPase concentration with light, and thereby the level of ATP turnover.

Light serves as a favorable control input for dynamic gene expression due to its tunability, reversibility, low toxicity, high orthogonality, and cost-effectiveness compared to alternative inducers such as chemical agents [51]. This motivated optogenetics for modulating the ATPase expression.

Finding the optimal light inputs is, however, not straightforward. To enable *model-based* optimization of such a system, the constraint-based dynamic model introduced in the previous chapter is extended in Section 5.2 to incorporate the dynamics of the optogenetic actuator. Then, in Section 5.3, an open-loop optimization problem is formulated to determine optimal light inputs that maximize production. Consistent with the previous chapter, the anaerobic lactate fermentation by *E. coli* remains as the case study. The case-specific extended constraint-based dynamic model introduced in this chapter will serve as a basis for developing a generalized cybergenetic modeling framework in the next chapter.

5.1 Proposed optogenetic system for ATPase expression

It is considered that the optogenetic regulation of ATPase (F_1-subunit) utilizes the CcaS/CcaR (chromatic acclimation sensor/regulator) system, derived from cyanobacteria [60, 130, 131]. Upon green light exposure, CcaS undergoes autophosphorylation and subsequently phosphorylates CcaR. Phosphorylated CcaR dimerizes and acts as a transcriptional activation factor. In contrast, red light induces CcaS dephosphorylation, resulting in repression of gene expression. The overall approach is outlined in Fig. 5.1.

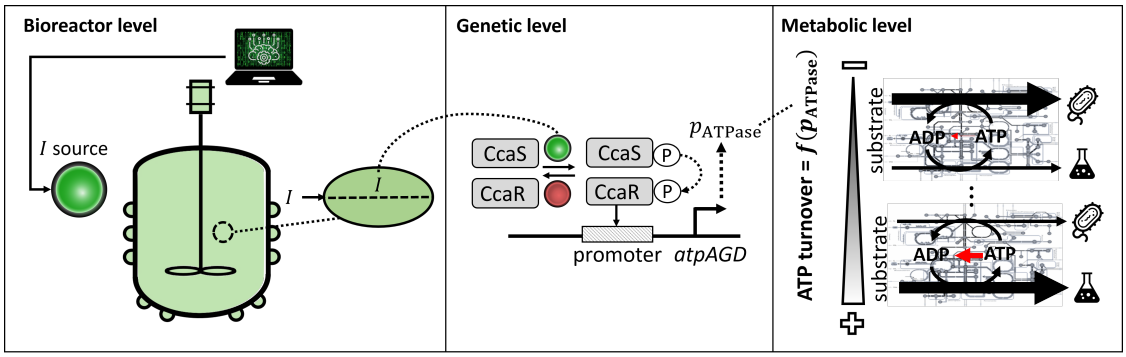


Figure 5.1: Illustration of the Ccas/CcaR optogenetic system for modulating ATPase expression (gene *atpAGD*) in batch and its impact on cellular metabolism. The higher the light intensity, the higher the ATPase expression, the higher the ATPase concentration p_{ATPase}, and the higher the ATP turnover. Homogeneous light penetration through the bioreactor is assumed. I: green light input.

5.2 Extended constraint-based dynamic model with optogenetic ATPase expression

The extended constraint-based dynamic model with optogenetic control of ATPase expression in batch processes is outlined in this section. The model introduced in the previous chapter (cf. Section 4.1) is used as a basis for the extended model[1].

The biomass components $\boldsymbol{p} \in \mathbb{R}^{n_p}$ are redefined as

$$\boldsymbol{p} := \left[p_{\mathrm{ATPase}}, \boldsymbol{p}_{\mathrm{unr}}^{\mathsf{T}}\right]^{\mathsf{T}}, \tag{5.1}$$

where $p_{\mathrm{ATPase}} \in \mathbb{R}$ denotes the concentration of the *regulated* ATPase, and $\boldsymbol{p}_{\mathrm{unr}} \in \mathbb{R}^{n_{p\mathrm{unr}}}$ represents a vector encompassing the remaining biomass components, such as *unregulated* enzymes, ribosomes, and quota elements. In this work, the term "unregulated cell components" refers to components whose expression is not externally modulated, as opposed to "regulated" ones whose expression can be modulated by external signals. Therefore, these terms do not refer to native regulation mechanisms within the cell, which will always take place.

The dynamics of ATPase read

$$\frac{\mathrm{d}p_{\mathrm{ATPase}}}{\mathrm{d}t} = F_{\mathrm{ATPase}}(B,I) - D_{\mathrm{ATPase}}(p_{\mathrm{ATPase}}), \tag{5.2}$$

where $F_{\mathrm{ATPase}} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $D_{\mathrm{ATPase}} : \mathbb{R} \to \mathbb{R}$ represent functions of production and degradation of ATPase, respectively. $B \in \mathbb{R}$ is the biomass dry weight and $I \in \mathbb{R}$ corresponds to the green light input. In this chapter, it is assumed that all cells receive the same light input, thus the induction is homogeneous. Furthermore, it is assumed that the system is under a constant red light background, high enough to repress induction by other light sources/wavelengths and low enough to enable induction by green light.

[1] Throughout this thesis, variables that have been introduced in earlier chapters will be re-introduced in each new chapter for the sake of completeness and self-containment.

The dose-response function for the CcaS/CcaR optogenetic system has been previously modeled following a Hill function [130]. Normalized per biomass concentration, it is assumed that the ATPase response function $\eta_{\text{ATPase}} : \mathbb{R} \to \mathbb{R}$, linked to the light dose, can be described by

$$\eta_{\text{ATPase}}(I) = \alpha + \beta \frac{I^{\delta}}{K^{\delta} + I^{\delta}}, \tag{5.3}$$

where $\alpha \in \mathbb{R}$ is an input-independent basal rate of ATPase production (accounting, e.g., for promoter *leakage*), $\beta \in \mathbb{R}$ is an input-dependent maximum rate of ATPase production, $K \in \mathbb{R}$ is a saturation constant, and $\delta \in \mathbb{R}$ is the Hill coefficient that determines how steep the function is.

Therefore,

$$F_{\text{ATPase}}(B, I) = B\eta_{\text{ATPase}}(I). \tag{5.4}$$

The degradation of ATPase is modeled as the product of a constant ATPase degradation rate $d_{\text{ATPase}} \in \mathbb{R}$ and the ATPase concentration

$$D(p_{\text{ATPase}}) = d_{\text{ATPase}} p_{\text{ATPase}}. \tag{5.5}$$

Note that the model implicitly considers the dilution of ATPase due to cell growth. As cellular components are modeled separately, the production of other components (such as enzymes, ribosomes, and quota elements) *dilutes* the proportion of ATPase within the cell.

For simplicity, it is assumed that the ATPase operates under substrate saturation conditions, approaching its maximum reaction rate. Therefore, the ATPase flux $V_{\text{ATPase}} \in \mathbb{R}$ is *equal to* the product of the ATPase concentration and its corresponding catalytic constant $k_{\text{cat,ATPase}} \in \mathbb{R}$

$$\left| \frac{V_{\text{ATPase}}}{k_{\text{cat,ATPase}}} \right| = p_{\text{ATPase}}. \tag{5.6}$$

Note that $V_{\text{ATPase}} \in \boldsymbol{V}$, where $\boldsymbol{V} \in \mathbb{R}^{n_V}$ comprises the fluxes.

In contrast, the reaction rates catalyzed by $\boldsymbol{p}_{\text{unr}}$ are *less than or equal to* the product of the corresponding enzyme concentration and the catalytic constant

$$\sum_{j \in \text{cat}_{\text{unr}}} \left| \frac{V_j}{k_{\text{unr},j}} \right| \leq p_{\text{unr}_i}, \quad \forall i \in [1, n_{p_{\text{unr}}}], \tag{5.7}$$

where $\text{cat}_{\text{unr}}(i)$ represents the set of reactions catalyzed by p_{unr_i}.

The underlying difference between Eqs. (5.6) and (5.7) is, while the former assumes that the ATPase flux can be set externally by optogenetically adjusting the ATPase level, the latter assumes that the unregulated fluxes are determined by the cell. In other words, the flux in Eq. (5.6) is *set*, hence an equality constraint. In contrast, fluxes in Eq. (5.7) are *optimized* by the cell following resource allocation constraints, hence an inequality constraint.

The metabolic cost of producing the ATPase via the CcaS/CcaR optogenetic system is accounted for by the constraint

$$V_{p_{\text{ATPase}}} - \frac{\mathrm{d}p_{\text{ATPase}}}{\mathrm{d}t} = 0, \tag{5.8}$$

where $V_{p_{\text{ATPase}}} \in \mathbb{R}$ is the flux through the ATPase-producing reaction (see reaction no. 25 in Table 4.1).

Furthermore, the dynamics of extracellular metabolites $\boldsymbol{z} \in \mathbb{R}^{n_z}$ and unregulated intracellular components $\boldsymbol{p}_{\text{unr}}$ can be described by

$$\left[\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}t}, \frac{\mathrm{d}\boldsymbol{p}_{\text{unr}}}{\mathrm{d}t}\right]^{\mathsf{T}} = \boldsymbol{S}_{z p_{\text{unr}}} \boldsymbol{V}. \tag{5.9}$$

The stoichiometric matrix $\boldsymbol{S}_{z p_{\text{unr}}} \in \mathbb{R}^{(n_z + n_{p_{\text{unr}}}) \times n_V}$ relates the species in $\boldsymbol{z}$ and $\boldsymbol{p}_{\text{unr}}$ to the reaction fluxes.

Note that the other constraints introduced in the previous chapter, namely Eqs. (4.1), (4.3), and (4.5)-(4.8), remain unchanged.

Integrating the constraints into the model formulation outlined in Section 4.1, leads to the following extended optimization problem

$$\max_{\boldsymbol{V}(\cdot)} \quad \int_{t_0}^{t_0+\Delta t_{\text{bio}}} B \,\mathrm{d}t, \tag{5.10a}$$

$$\text{s.t.} \quad \text{Eqs.}\,(4.1), (4.3), (4.5)-(4.8), (5.1)-(5.9), \tag{5.10b}$$

whose solution allows simulating and predicting the batch fermentation dynamics with optogenetic regulation of the ATPase.

5.3 Optimal modulation of the ATPase expression

An optimal control problem is formulated to determine the optimal green light input *function* $I(\cdot)$ that maximizes batch efficiency

$$\max_{I(\cdot)} \quad J(\cdot) \tag{5.11a}$$

$$\text{s.t.} \quad (5.10), \tag{5.11b}$$

where $J(\cdot)$ captures the objective function of the optimization.

Note that the key difference between the optimization problems in (4.11) and (5.11) is, in the former the degree of freedom or decision variable is the ATPase flux (an intracellular, *virtual* variable), while in the latter it is light (a tunable process parameter linked to an intracellular genetic actuator). This makes the approach more suitable for practical implementations, following of course a proper characterization of the optogenetic and metabolic systems.

5.4 Optogenetic open-loop control of ATPase in anaerobic lactate fermentation by *E. coli*

The objective function is defined as the maximization of the lactate concentration at the end of the batch, $J(\cdot) = z_{\mathrm{LAC}}(t_f)$. Consistent with the previous chapter, different batch times in the optimization are expected to result in different ATP turnover policies and, consequently, in varying average fermentation metrics. In other words, the choice of batch time imposes limitations on the maximum achievable lactate titer. Furthermore, any enhancement in product yield should maintain adequate biomass formation and efficient substrate utilization to ultimately maximize the final lactate titer.

The optimal control problem described in (5.11) is solved for three different batch durations: 13, 15, and 19 h. These specific batch times were chosen to investigate their impact on the fermentation metrics, following a similar motivation as outlined in Section 4.4.2. As a reference, a scenario without light induction was also considered, where the control input remained zero at all time points. To do so, the bilevel optimization problem in (5.11) was discretized and converted into a single-level non-linear program with complementary constraints as explained in Remark 4.3. Note that the initial value for the intracellular biomass composition in the simulations was estimated using resource balance analysis (cf. Remark 4.2).

5.4.1 Model parameters

The deFBA model for batch lactate fermentation, as derived in Chapter 4, consists of 16 metabolic reactions and 18 biomass-producing reactions, including the synthesis of individual enzymes, resulting in a total of 34 fluxes. The model includes 5 external metabolites, namely glucose, lactate, formate, succinate, and carbon dioxide, along with 18 internal metabolites and 18 cell components. The model parameters include 34 catalytic constants and 18 molecular weights, as shown in Table 4.1. The additional parameters regarding the CcaS/CcaR optogenetic system are listed in Table 5.1.

Table 5.1: Parameters of the CcaS/CcaR optogenetic system for the simulations.

Item	Value	Unit	Ref./Note
δ	2.490	1	[130]
K	0.138	$\mathrm{W/m^2}$	[130]
α	$2 \cdot 10^{-6}$	mmol/g/h	Note 1
β	$1 \cdot 10^{-4}$	mmol/g/h	Note 1
d_{ATPase}	$6.3 \cdot 10^{-2}$	1/h	Note 2

Note 1. Biologically sound parameter values derived from feasible deFBA simulations.
Note 2. Estimated as $d_{\mathrm{ATPase}} = \frac{\ln(2)}{t_{0.5}}$, where $t_{0.5}$ is the ATPase protein half-life time [132].

5.4.2 Open-loop optimization results

The open-loop optimization results for the different batch times are presented in Figs. 5.2-5.5. No model-plant mismatch is considered. Various fermentation metrics were calculated to compare the performance across the different scenarios (Table 5.2). These metrics include the average product-to-substrate yield ($Y_{\mathrm{PS,batch}}$), the average biomass-to-substrate yield ($Y_{\mathrm{XS,batch}}$), and the product volumetric productivity ($r_{\mathrm{P,batch}}$).

Table 5.2: Average fermentation metrics in lactate fermentation with optogenetic regulation of the ATPase in batch, maximizing the titer of lactate at t_f.

t_f [h]	$Y_{\mathrm{PS,batch}}$ $\left[\frac{\mathrm{mol}}{\mathrm{mol}}\right]$	$Y_{\mathrm{XS,batch}}$ $\left[\frac{\mathrm{g}}{\mathrm{mol}}\right]$	$r_{\mathrm{P,batch}}$ $\left[\frac{\mathrm{mmol}}{\mathrm{L\cdot h}}\right]$	$z_{\mathrm{LAC}}(t_f)$ mM
7.5*	1.72	15.3	31.8	238.8
13	1.87	7.2	20.0	259.8
15	1.90	5.6	17.6	263.6
19	1.92	4.5	14.0	266.8

* Without induction of the ATPase expression.

Consistent with the previous chapter (cf. Section 4.4.2), longer batch times in the optimal control problem, compared to the system without induction of the ATPase, are expected to *activate* the enforced ATP turnover mechanism. The optimal optogenetic modulation of the ATPase indeed enabled an improvement in product yield for the scenarios with batch times longer than that of the fermentation with uninduced ATPase[2] (cf. Table 5.2). This enhancement came at the expense of reduced biomass yield as a result of the enforced ATP turnover, leading to lower volumetric productivity.

Specifically, the product yield increased by 8.7, 10.3, and 11.7 % for batch times of 13, 15, and 19 h, respectively, compared to the case without induction of the ATPase (7.5 h-batch time). A longer batch fermentation time resulted in a higher product yield enhancement, as the enforced ATP turnover could be applied for a longer duration while allowing for lower volumetric productivity. That is, the reduced fermentation rates were compensated by the higher product yields.

It can be argued that running a 7.5-h process twice, without ATPase induction, could potentially render a higher overall production compared to running a 19-hour process once with ATPase induction. While this point may be valid, drawing a definitive conclusion is not straightforward due to additional factors that need to be considered. Setting up a batch fermentation is labor-intensive and often involves a significant amount of unproductive time or downtime [133]. Between batches, various tasks such as equipment cleaning, bioreactor sterilization, culture inoculation, etc., are required. Furthermore, a bioprocess encompasses not only the fermentation itself but also upstream and downstream unit operations. Therefore, selecting a specific trade-off between yield and volumetric productivity should be part of an

[2] Assuming the same initial conditions and depletion of all available substrate.

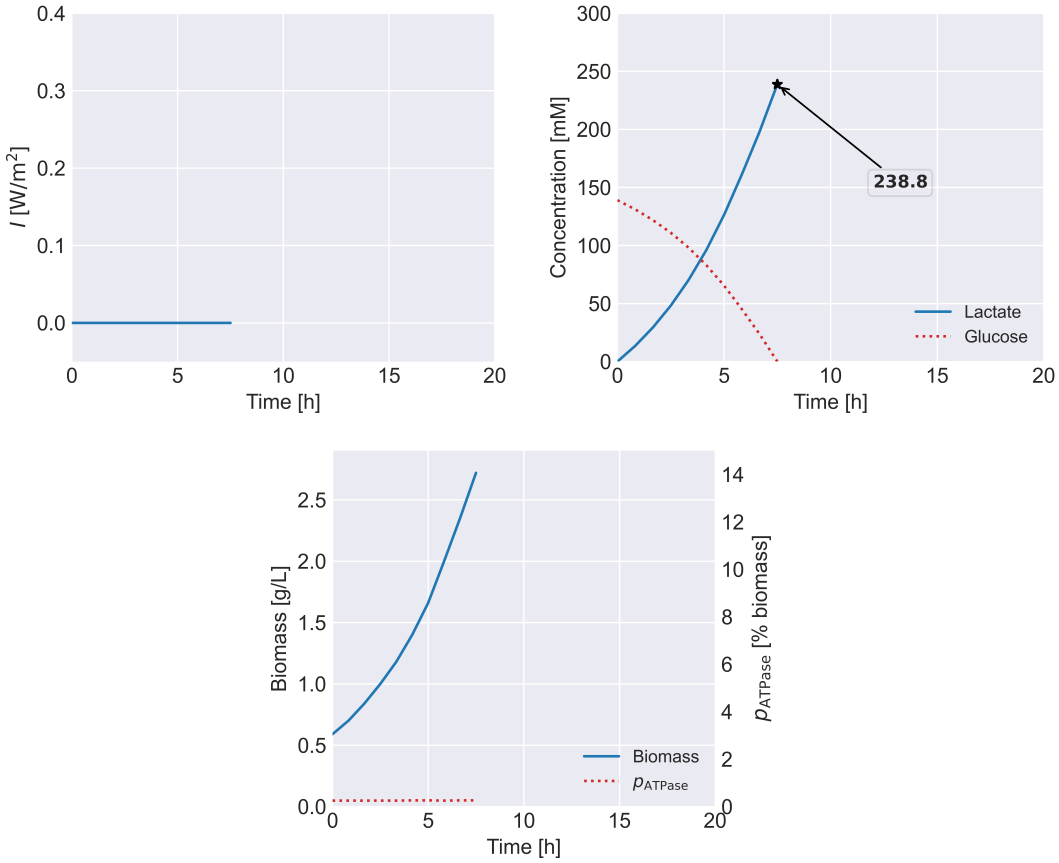


Figure 5.2: Batch lactate fermentation without induction of ATPase expression, with a batch time of 7.5 h. Initial conditions: $B(0) = 0.59\,\mathrm{g/L}$, $z_{\mathrm{GLC}}(0) = 139\,\mathrm{mM}$, $z_{\mathrm{LAC}}(0) = 0\,\mathrm{mM}$. No model-plant mismatch. This scenario is used for comparison purposes.

overall plant optimization that takes into account techno-economic aspects, as well as potential environmental impact. For example, if the substrate is costly, high yields are to be preferred.

The gene expression of the ATPase reached approximately 10-11 % of the biomass dry weight in the studied scenarios. In general, overexpression of proteins can adversely impact cell growth and productivity due to resource burden [134]. Therefore, the observed decrease in growth rate with increasing ATPase expression can be attributed to both the enforced ATP turnover mechanism and the potential burden associated with the cost of protein synthesis. In fact, a higher level of ATPase induction can result in reduced availability of amino acids and energy co-factors (cf. e.g. Eq. (5.8)), which are necessary for synthesizing other enzymes and cell components. The optimizer takes this potential resource burden into consideration when determining the optimal light input trajectories since resource allocation is an inherent aspect of the proposed constraint-based dynamic model.

Notably, the light trajectories predicted by the open-loop controller corresponded to a gradual accumulation of the intracellular ATPase. These profiles resemble the ATPase concentration trends predicted for the optimal scenarios in the previous chapter when considering the ATPase flux as a *virtual* degree of freedom (cf. e.g. Fig. 4.6-A).

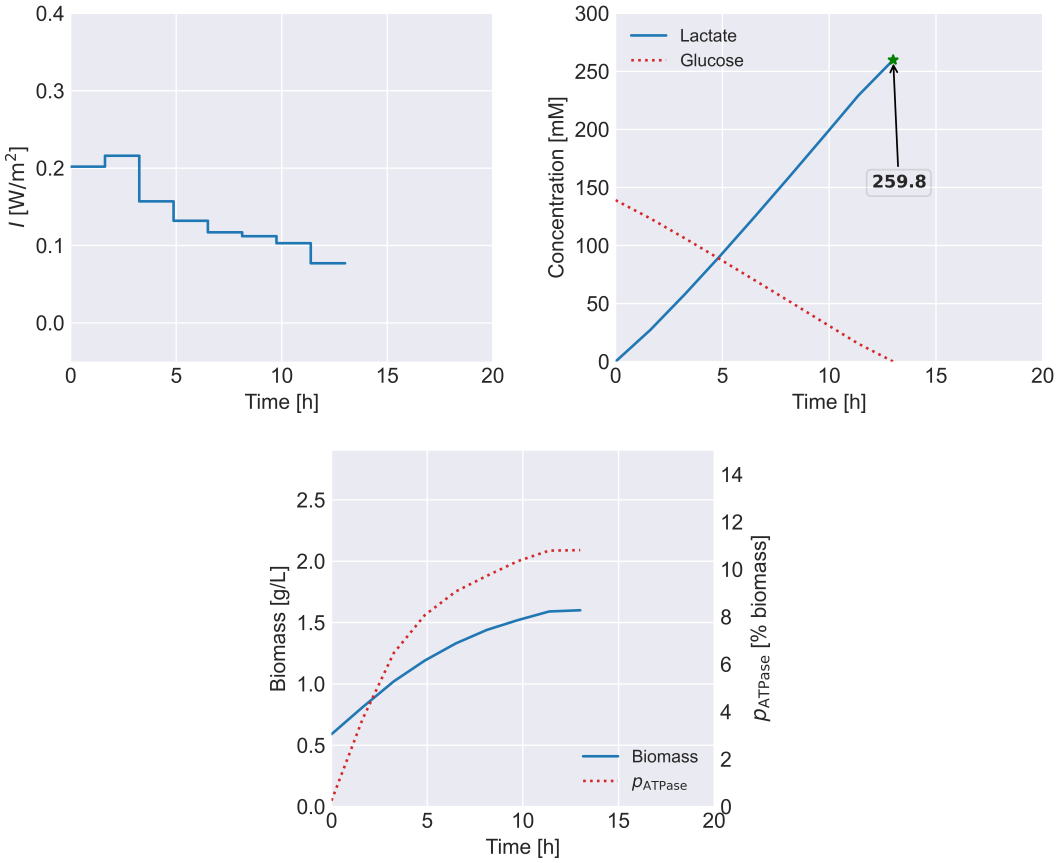


Figure 5.3: Optimization results for batch lactate fermentation with optogenetic modulation of ATPase for a batch time of 13 h. Initial conditions: $B(0) = 0.59\,\mathrm{g/L}$, $z_{\mathrm{GLC}}(0) = 139\,\mathrm{mM}$, $z_{\mathrm{LAC}}(0) = 0\,\mathrm{mM}$. No model-plant mismatch.

5.5 Summary

This chapter presented a model-based optimization strategy for modulating ATP turnover in cells through the optogenetic regulation of ATPase in batch processes. The deFBA model introduced in Chapter 4 was extended to incorporate the dynamics of the CcaS/CcaR optogenetic gene expression system, which was assumed to regulate ATPase expression. By solving an appropriate optimal control problem, using light as a dynamic degree of freedom, it was possible to enhance the product yield of the fermentation process at the expense of volumetric productivity.

Various trade-offs between product yield and volumetric productivity were explored. The ability to optogenetically modulate ATP turnover offers a practical way to adjust the strain's performance from one batch to another, enabling the plant to readily adapt to changes. Unlike traditional static metabolic engineering, which requires the design of new microorganisms or strains for each desired performance metric, the proposed approach exploits the same production strain.

Note that in this chapter the focus was on open-loop optimization without considering model-plant mismatch. However, open-loop optimization is often affected by uncertainties such as model-plant mismatch and disturbances. Therefore, the subsequent chapter investigates the

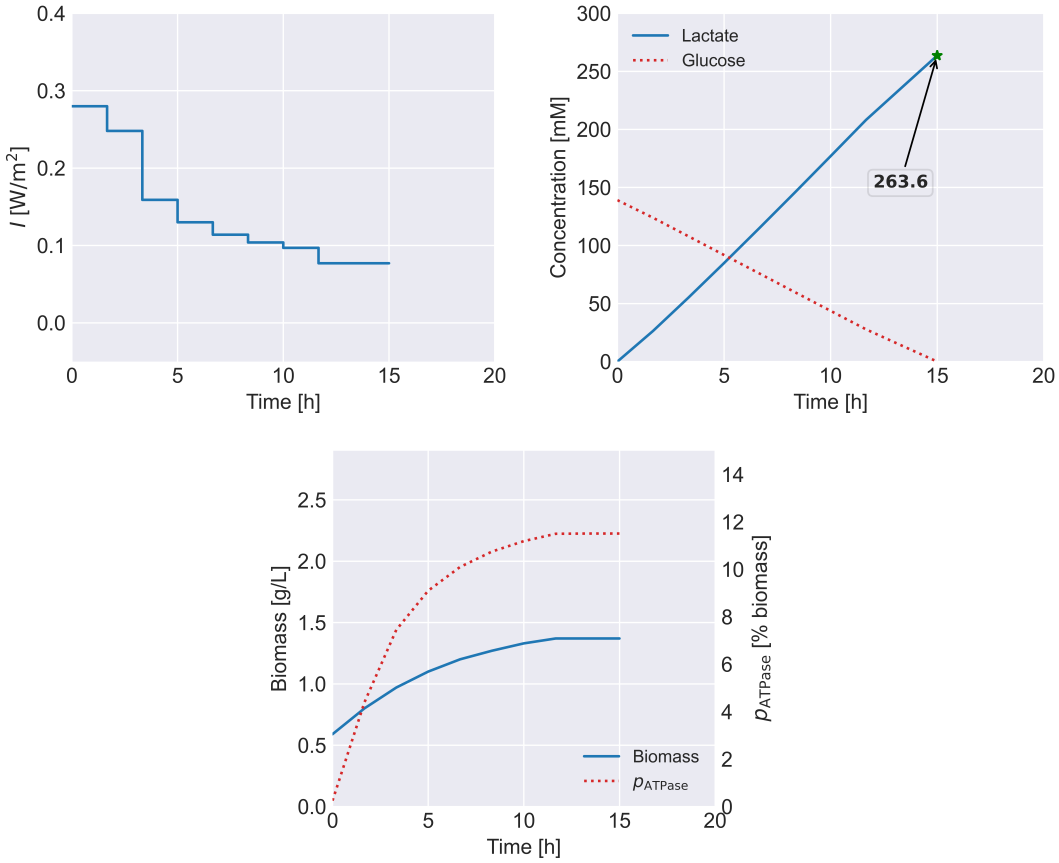


Figure 5.4: Optimization results for batch lactate fermentation with optogenetic modulation of ATPase for a batch time of 15 h. Initial conditions: $B(0) = 0.59\,\mathrm{g/L}$, $z_{\mathrm{GLC}}(0) = 139\,\mathrm{mM}$, $z_{\mathrm{LAC}}(0) = 0\,\mathrm{mM}$. No model-plant mismatch.

application of MPC to address these uncertainties. Additionally, the next chapter explores the utilization of state estimators to facilitate process monitoring in the context of feedback control. Fed-batch systems will also be considered, which provide an additional degree of freedom for process optimization.

Furthermore, the implementation of optogenetics in large-scale bioreactors remains a challenge. The extended constraint-based dynamic model outlined in this chapter assumed uniform light intensity and homogeneous induction across all cells. While this assumption may hold true for small-scale bioreactors, it becomes questionable for larger bioreactors where various types of gradients can emerge. For example, high cell densities in bioreactors may lead to issues concerning light penetration, potentially resulting in light-limited fermentation performance. This aspect is also addressed in the subsequent chapter.

Finally, the next chapter offers a generalized framework for (fed-batch) metabolic cybergenetics. It provides a modeling, optimization, and predictive control framework for dynamic metabolic engineering applications, potentially spanning beyond the specific example of optogenetically-assisted enforced ATP turnover.

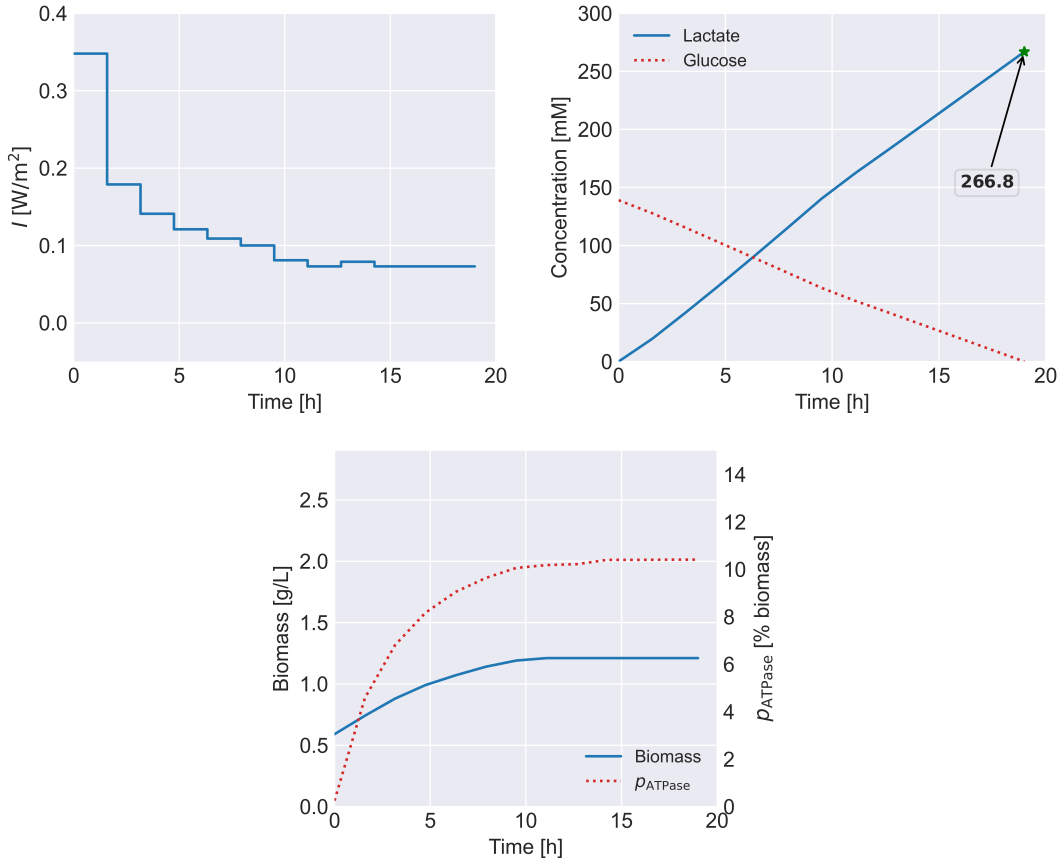


Figure 5.5: Optimization results for batch lactate fermentation with optogenetic modulation of ATPase for a batch time of 19 h. Initial conditions: $B(0) = 0.59\,\mathrm{g/L}$, $z_{\mathrm{GLC}}(0) = 139\,\mathrm{mM}$, $z_{\mathrm{LAC}}(0) = 0\,\mathrm{mM}$. No model-plant mismatch.

6 Generalized framework for metabolic cybergenetics in fed-batch: modeling, optimization, control, and estimation

In Chapters 4-5, the models and control formulations focused on the specific case study of enforced ATP turnover. Even though the latter will continue to be the core application example in this thesis, it raises the question of how one can formulate a *generalized* framework that can be employed in other dynamic metabolic engineering applications and considering other fermentation modes of operations such as fed-batch.

In this chapter, a generalized framework for modeling, optimization, and predictive control of metabolic cybergenetics is presented. It exploits the concept of dynamically manipulating intracellular metabolic fluxes to optimize bioprocesses. This is enabled by fine-tuning the gene expression of metabolism-relevant proteins by external inputs. Refer to Section 2.4 for more details; Fig. 2.1 offers an overview of the considered metabolic cybergenetic components.

Without loss of generality, this chapter focuses primarily on fed-batch processes, which offer inherent advantages compared to pure batch setups. Fed-batch processes involve the introduction of a concentrated feed (typically substrate) into the bioreactor, thereby extending the production phase, and turning the feed into an additional degree of freedom for process optimization. In general, fed-batch fermentations can offer higher productivity and the generation of more concentrated product streams [135, 136].

This chapter is organized as follows. In Section 6.1, a generalized constraint-based dynamic cybergenetic modeling approach is presented, which integrates metabolism, resource allocation, and inducible gene expression. The model considers average cybergenetic input gradients (uneven distribution) inside the bioreactor, relevant for large-scale systems. The derived model serves as the foundation for model-based optimization, predictive control, and estimation of metabolic cybergenetic systems, as discussed in Sections 6.2 and 6.3.

The framework is assessed considering the optogenetic modulation of ATPase in anaerobic lactate fermentation by *E. coli* (Section 6.4). In contrast to Chapter 5, here a *fed-batch* regime with *non-homogeneous* light penetration is examined, considering both open-loop optimization *and* MPC. Furthermore, the use of a *soft sensor* for reconstructing the biomass composition to facilitate predictive control is outlined.

6.1 Modeling of metabolic cybergenetic systems in fed-batch

Consistent with the constraint-based dynamic modeling approach presented in the preceding chapters, it is assumed that cells consist of metabolic enzymes, ribosomes, and quota elements. These biomass components are represented by the molar vector $\boldsymbol{p} \in \mathbb{R}^{n_p}$. Therefore, the total biomass $B_T \in \mathbb{R}$ in fed-batch systems can be calculated as

$$B_T = v_L(\boldsymbol{b}^T \boldsymbol{p}). \tag{6.1}$$

Here, $v_L \in \mathbb{R}$ refers to the bioreactor volume and $\boldsymbol{b} \in \mathbb{R}^{n_p}$ represents a vector comprising appropriate molecular weights of $\boldsymbol{p}$.

The vector $\boldsymbol{p}$ is redefined to combine the concentrations of regulated proteins $\boldsymbol{p}_{\text{reg}} \in \mathbb{R}^{n_{p\text{reg}}}$ and the concentrations of the remaining unregulated biomass components $\boldsymbol{p}_{\text{unr}} \in \mathbb{R}^{n_{p\text{unr}}}$. Thus, $\boldsymbol{p} := \left[\boldsymbol{p}_{\text{reg}}^{\mathsf{T}}\ \boldsymbol{p}_{\text{unr}}^{\mathsf{T}}\right]^{\mathsf{T}}$. Remark that in this work the term "regulated proteins" refers to the fact that the expression of these proteins can be modulated by *external* signals through suitable genetic systems, such as light-inducible gene expression systems [137, 138]. Similarly, as opposed to "regulated cell components", the expression of "unregulated cell components" is not modulated by *external* signals. As mentioned before, these terms do not refer to the native internal cell regulation which will always take place.

Regarding the externally inducible gene expression systems, a clear distinction is made between the inputs manipulated by the controller, denoted as $\boldsymbol{u_s} \in \mathbb{R}^{n_u}$ (such as light in optogenetics), and the values perceived by the cells within the bioreactor, denoted as $\boldsymbol{u_c} \in \mathbb{R}^{n_u}$. This differentiation is crucial because it is possible for $\boldsymbol{u_s}$ to differ from $\boldsymbol{u_c}$ depending on factors such as the characteristics of the input and the bioreactor. This discrepancy becomes particularly relevant in large-scale setups where conditions tend to be less homogeneous or when the input values received by the cells are influenced by factors such as cell density.

In Section 6.4, $\boldsymbol{u_c}$ will be derived to consider average light penetration gradients in cases where light is used to induce gene expression. For the sake of generality, the input perceived by the cells is determined by a function $\boldsymbol{f}_u : \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_{\theta_u}} \to \mathbb{R}^{n_u}$, which maps the input at the source $\boldsymbol{u_s}$ to an *average* input $\bar{\boldsymbol{u}}_c$

$$\bar{\boldsymbol{u}}_c = \boldsymbol{f}_u(\boldsymbol{u}_s, \boldsymbol{x}, \boldsymbol{\theta}_u). \tag{6.2}$$

Here, $\boldsymbol{x} \in \mathbb{R}^{n_x}$ represents all the model states, $\boldsymbol{\theta}_u \in \mathbb{R}^{n_{\theta_u}}$ comprises parameters of $\boldsymbol{f}_u(\cdot)$, and $\bar{\boldsymbol{u}}_c$ denotes the average value of $\boldsymbol{u}_c$ in the bioreactor under well-mixed conditions. The introduction of $\bar{\boldsymbol{u}}_c$ simplifies the model by considering only changes over time while still accounting for average input gradients.

The change in the amount of regulated proteins can be described as

$$\frac{\mathrm{d}\left(v_L \boldsymbol{p}_{\text{reg}}\right)}{\mathrm{d}t} = \boldsymbol{F}_{\text{reg}}(B, \bar{\boldsymbol{u}}_c) - \boldsymbol{D}_{\text{reg}}(\boldsymbol{p}_{\text{reg}}), \tag{6.3}$$

where $\boldsymbol{F}_{\text{reg}} : \mathbb{R} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_{p\text{reg}}}$ and $\boldsymbol{D}_{\text{reg}} : \mathbb{R}^{n_{p\text{reg}}} \to \mathbb{R}^{n_{p\text{reg}}}$ represent protein production and degradation functions, respectively. The proposed modeling framework also considers *dilution* due to cellular growth, wherein the production of other biomass components gradually reduces the proportion of regulated proteins within the cell.

Transcription factors have the ability to transition between active and inactive states in response to specific signals. When in the active state, these transcription factors can bind to the promoter region of regulated genes, thereby exerting control over the transcription process by activating or repressing it. Ribosomes translate messenger ribonucleic acid (mRNA) molecules, which are generated through transcription, into proteins. In most bacteria such as *E. coli*, transcription and translation processes are tightly coupled [139, 140]. To capture lumped transcription and translation phenomena, the use of dose-response functions $\boldsymbol{\eta} : \mathbb{R}^{n_u} \to \mathbb{R}^{n_{p\text{reg}}}$ is proposed. Therefore, protein production is expressed as

$$\boldsymbol{F}_{\text{reg}}(B, \bar{\boldsymbol{u}}_c) = B\boldsymbol{\eta}(\bar{\boldsymbol{u}}_c). \tag{6.4}$$

The next step is to link the dynamics of externally regulated proteins captured by Eqs. (6.1)-(6.4) to the broader metabolism and resource allocation phenomena. To do so, deFBA-related constraints (cf. Sections 4.1 and 5.2), modified to account for fed-batch systems, are incorporated as follows.

The change in the amount of extracellular metabolites, such as substrates and products, is modeled as

$$\frac{\mathrm{d}(v_L \boldsymbol{z})}{\mathrm{d}t} = F_{\text{in}} \boldsymbol{z}_{\text{in}} + v_L(\boldsymbol{S}_{\boldsymbol{z}} \boldsymbol{V}) - \boldsymbol{D}_{\boldsymbol{z}}(\boldsymbol{z}). \tag{6.5}$$

In this equation, $\boldsymbol{z} \in \mathbb{R}^{n_z}$ denotes the molar vector of extracellular metabolites, $\boldsymbol{S}_{\boldsymbol{z}} \in \mathbb{R}^{n_z \times n_V}$ is the stoichiometric matrix of $\boldsymbol{z}$, and $\boldsymbol{V} \in \mathbb{R}^{n_V}$ represents the fluxes in molar amount per time. The function $\boldsymbol{D}_z : \mathbb{R}^{n_z} \to \mathbb{R}^{n_z}$ describes the degradation of $\boldsymbol{z}$, which is not considered in the models presented in previous chapters. $F_{\text{in}} \in \mathbb{R}$ represents the feed rate and $\boldsymbol{z}_{\text{in}} \in \mathbb{R}^{n_z}$ consists of the feed concentrations of $\boldsymbol{z}$. It is assumed in the model that the feed contains only substrates, which can be different for other scenarios.

The change in the amount of unregulated biomass components is described by

$$\frac{\mathrm{d}(v_L \boldsymbol{p}_{\text{unr}})}{\mathrm{d}t} = v_L(\boldsymbol{S}_{p_{\text{unr}}} \boldsymbol{V}) - \boldsymbol{D}_{\text{unr}}(\boldsymbol{p}_{\text{unr}}). \tag{6.6}$$

Here, $\boldsymbol{S}_{p_{\text{unr}}} \in \mathbb{R}^{n_{p\text{unr}} \times n_V}$ represents the stoichiometric matrix of $\boldsymbol{p}_{\text{unr}}$, and $\boldsymbol{D}_{\text{unr}} : \mathbb{R}^{n_{p\text{unr}}} \to \mathbb{R}^{n_{p\text{unr}}}$ captures the degradation of $\boldsymbol{p}_{\text{unr}}$. Note that the latter term is not considered in the models presented in previous chapters.

Including biomass-producing reactions in the network (cf. e.g. Table 4.1) enables capturing the resource *cost* of synthesizing biomass components. Therefore, a constraint is added to link the expression of regulated proteins to a resource cost

$$\boldsymbol{V}_{p_{\text{reg}}} - \frac{\mathrm{d}\boldsymbol{p}_{\text{reg}}}{\mathrm{d}t} = \boldsymbol{0}, \tag{6.7}$$

where $\boldsymbol{V}_{p_{\text{reg}}} \in \mathbb{R}^{n_{p_{\text{reg}}}}$ represents the fluxes of externally regulated protein-producing reactions.

Quasi-steady-state dynamics are considered for the intracellular metabolites

$$\frac{\mathrm{d}(v_L \boldsymbol{m})}{\mathrm{d}t} = v_L(\boldsymbol{S}_{\boldsymbol{m}} V) = \boldsymbol{0}. \tag{6.8}$$

Here, $\boldsymbol{m} \in \mathbb{R}^{n_m}$ represents the molar vector of intracellular metabolites, while $\boldsymbol{S}_{\boldsymbol{m}} \in \mathbb{R}^{n_m \times n_V}$ denotes the stoichiometric matrix of $\boldsymbol{m}$.

The metabolic fluxes of reactions associated with enzymes in $\boldsymbol{p}_{\text{reg}}$ are constrained by the following equality

$$\sum_{j \in \text{cat}_{\text{reg}}(i)} \left| \frac{V_j}{k_{\text{reg},j}} \right| = p_{\text{reg}_i}, \quad \forall i \in [1, n_{p_{\text{reg}}}], \tag{6.9}$$

where $\text{cat}_{\text{reg}}(i)$ represents the set of reactions catalyzed by the enzyme p_{reg_i}.

The constraints introduced in previous chapters, Eqs. (4.5)-(4.6), and (5.7), remain unchanged.

Note that Eq. (5.7) imposes an upper bound on the metabolic fluxes (inequality constraint) by considering the product of the enzyme concentrations and the corresponding catalytic constants. In contrast, Eq. (6.9) is formulated as an equality constraint. As discussed in the previous chapter, this distinction arises from the assumption that external control can be exerted over these (regulated) fluxes by modulating the amount of the associated catalytic proteins. On the other hand, the (unregulated) fluxes in Eq. (5.7), whose expression of the associated catalytic proteins cannot be adjusted externally, are *optimized* by the cell, e.g., toward maximizing growth.

The liquid volume of the bioreactor changes over time as a result of the substrate feed rate F_{in}

$$\frac{\mathrm{d}v_L}{\mathrm{d}t} = F_{\text{in}}. \tag{6.10}$$

Bounds are taken into account for the dynamic states

$$\boldsymbol{p}_{\min} \leq \boldsymbol{p} \leq \boldsymbol{p}_{\max}, \, \boldsymbol{z}_{\min} \leq \boldsymbol{z} \leq \boldsymbol{z}_{\max}, \, v_{L_{\min}} \leq v_L \leq v_{L_{\max}}. \tag{6.11}$$

The initial conditions of the system at time t_0 are expressed as

$$\boldsymbol{p}(t_0) = \boldsymbol{p}_0, \, \boldsymbol{z}(t_0) = \boldsymbol{z}_{\boldsymbol{0}}, \, v_L(t_0) = v_{L0}. \tag{6.12}$$

In summary, the constraint-based dynamic model for fed-batch metabolic cybergenetic systems can be expressed as the following optimization problem

$$\max_{\boldsymbol{V}(\cdot)} \quad \int_{t_0}^{t_0+\Delta t_{\text{bio}}} B_T \, \mathrm{d}t, \tag{6.13a}$$

$$\text{s.t.} \quad \text{Eqs.}\,(4.5)-(4.6),(5.7),(6.1)-(6.12), \tag{6.13b}$$

where it is assumed that the cell maximizes the integral of the biomass, and $\boldsymbol{V}(\cdot)$ is a *function* of the resulting metabolic flux distribution. By solving this dynamic optimization problem, it

becomes possible to simulate and predict the behavior of the cell in a fed-batch regime subject to cybergenetic inputs. Note that the model formulation is also applicable to batch systems by setting Eq. (6.10) equal to zero. In the case of continuous processes, an additional dilution term associated with a flow rate leaving the bioreactor can be incorporated.

6.2 Optimal control of metabolic cybergenetic systems

To optimize the performance of the cell metabolism described by an objective function $J(\cdot)$, the constraint-based dynamic model derived in the previous section is employed. The dynamic process inputs are collected in $\boldsymbol{u_p}$, hence $\boldsymbol{u_p} := [\boldsymbol{u_s}(\cdot)^\mathsf{T}, F_\text{in}(\cdot)]^\mathsf{T}$. The model parameters are represented by the vector $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$. Remark that the dynamic states are denoted by $\boldsymbol{x}$, hence $\boldsymbol{x} := [\boldsymbol{p}_\text{reg}^\mathsf{T}, \boldsymbol{z}^\mathsf{T}, \boldsymbol{p}_\text{unr}^\mathsf{T}, v_L]^\mathsf{T}$. The optimal control problem to determine the optimal inputs for the plant becomes

$$\max_{\boldsymbol{u_p}, \boldsymbol{x_0}} \quad J(\cdot), \tag{6.14a}$$

$$\text{s.t.} \quad (6.13), \tag{6.14b}$$

$$\mathbf{0} \leq \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u_p}, \boldsymbol{\theta}), \tag{6.14c}$$

where Eq. (6.14c) captures additional system constraints, encompassing aspects such as physical limitations, safety considerations, or economic-related matters. Note that the decision variables of the optimization comprise in principle $\boldsymbol{u_p}$ (*dynamic* degrees of freedom) as well as the initial conditions of the plant $\boldsymbol{x_0}$ (*static* degrees of freedom).

$J(\cdot)$ can be defined in various ways. It may involve maximizing production, maintaining a desired set-point, or tracking a reference trajectory, among other possibilities. In the case of batch mode operation, where the process is not fed, $\boldsymbol{u_s}(\cdot)$ can be selected as the only dynamic degree of freedom. Note that the optimal control problem stated in (6.14) represents an open-loop optimization, as the optimal inputs are applied to the system without feedback.

6.3 Shrinking-horizon model predictive control of metabolic cybergenetic systems

In the context of fed-batch processes, shrinking-horizon MPC is employed to address the impact of uncertainties associated with open-loop control, including model-plant mismatch and disturbances.

Similar to the MPC formulation presented in Chapter 4, the optimal control problem is repeatedly solved at predefined sampling times. At these time instances, the system states are either measured directly or estimated using an observer or soft sensor. This feedback mechanism enables the controller to utilize the current state information and take corrective control actions. Let t_k represent the sampling times when measurements are obtained. Without loss of generality, equidistant sampling times are assumed, denoted as $t_k := kh_s$, where $k \in \mathbb{N}_0$ and h_s denotes the fixed sampling interval. Additionally, it is assumed that the controller predicts the system

behavior up to the final time $t_f := Nh_s$, where $N \in \mathbb{N}$ represents the number of steps in the prediction horizon. Consequently, the prediction horizon *shrinks* at each sampling time. The formulation of the shrinking-horizon MPC at time t_k can be expressed as [57, 115]

$$\max_{\boldsymbol{u_p}, \boldsymbol{x_0}} \quad J(\cdot), \tag{6.15a}$$

$$\text{s.t.} \quad \max_{\boldsymbol{V}(\cdot)} \int_{t_k}^{t_k+\Delta t_{\text{bio}}} B_T \, \mathrm{d}t, \tag{6.15b}$$

$$\text{s.t.} \quad \text{Eqs. (4.5)-(4.6), (5.7), (6.1)-(6.11),} \tag{6.15c}$$

$$\boldsymbol{x}(t_k) = \tilde{\boldsymbol{x}}_{\boldsymbol{k}}, \tag{6.15d}$$

$$\text{Eq. (6.14c),} \tag{6.15e}$$

where $t \in [t_k, t_f]$. Moreover, $\tilde{\boldsymbol{x}}_{\boldsymbol{k}}$ denotes the measured value of $\boldsymbol{x}$ at t_k.

The culture volume can be easily monitored by tracking the applied feed rate, and real-time sensors for biomass and extracellular metabolite concentrations are available [84–86]. However, sensors for (intracellular) biomass components are generally unavailable, hindering the implementation of MPC as outlined in (6.15).

6.3.1 Full information estimation: estimating unmeasured cell components

To tackle the challenge of real-time monitoring of cell composition, resource balance analysis (cf. Remark 4.2) was previously proposed [37]. Despite its relative simplicity, resource balance analysis does not always provide a reliable *quantitative* estimation of the *dynamic* cell composition [37]. This is explained by the fact that it considers quasi-steady-state conditions, which is a very optimistic assumption given the dynamic nature of metabolism[1]. Furthermore, it uses only the current measurements for the estimation, neglecting historical process measurements.

The use of a full information estimation algorithm [115] is outlined as an alternative to resource balance analysis [37] for inferring *dynamic* changes in the cell composition. The estimation strategy offers in principle better properties in terms of optimality as it is based on the dynamic model of metabolism and can incorporate constraints, as well as historical and current measurements. Note that a direct comparison between resource balance analysis and full information estimation for inferring dynamic changes in cell composition has been previously performed in the context of glycerol fermentation by *E. coli* with varying oxygen uptake rate [141]. In the latter case, the full information estimator showed better estimation performance than resource balance analysis, which motivated its incorporation in the proposed metabolic cybergenetic framework.

Let $(\cdot)_i$ represent a general optimization variable calculated at time t_i. The dynamic Eqs. (6.3), (6.5), (6.6), and (6.10) are combined into a vector function $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u_p}, \boldsymbol{\theta})$. Similar to MPC, equidistant sampling times are assumed for the estimator, although non-equidistant

[1] Especially if the cell´s metabolism is dynamically excited with external inputs, which is the underlying idea of metabolic cybergenetics.

sampling times are also possible. At time t_k, the full information estimator solves the following optimization problem

$$\min_{\boldsymbol{x}_0,\boldsymbol{\theta},\boldsymbol{\omega}} \quad \left\| \begin{bmatrix} \boldsymbol{x_0} \\ \boldsymbol{\theta} \end{bmatrix} - \begin{bmatrix} \hat{\boldsymbol{x}}_0 \\ \hat{\boldsymbol{\theta}} \end{bmatrix} \right\|_{\mathbf{P}}^2 + \sum_{i=0}^{k} \|\boldsymbol{y}(t_i) - \tilde{\boldsymbol{y}}_{\boldsymbol{i}}\|_{\mathbf{R}}^2 + \|\boldsymbol{\omega_i}\|_{\mathbf{Q}}^2 \tag{6.16a}$$

$$\text{s.t.} \quad \max_{\boldsymbol{V}(\cdot)} \int_{t_i}^{t_i+\Delta t_{\text{bio}}} B_T \, \mathrm{d}t \tag{6.16b}$$

$$\text{s.t.} \quad \boldsymbol{x}(t_i + h_s) = \boldsymbol{x}(t_i) + \int_{t_i}^{t_i+h_s} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u_p}, \boldsymbol{\theta}) \, \mathrm{d}t + \boldsymbol{\omega_i}, \tag{6.16c}$$

$$\mathbf{0} \leq \boldsymbol{c}(\boldsymbol{x}(t_i), \boldsymbol{u_p}(t_i), \boldsymbol{\theta}), \tag{6.16d}$$

$$\boldsymbol{y}(t_i) = \boldsymbol{h}(\boldsymbol{x}(t_i), \boldsymbol{u_p}(t_i), \boldsymbol{\theta}), \tag{6.16e}$$

$$\mathbf{0} \leq \boldsymbol{g}(\boldsymbol{x}(t_i), \boldsymbol{u_p}(t_i), \boldsymbol{\theta}), \tag{6.16f}$$

where $\|\boldsymbol{a}\|_{\mathbf{A}}^2 := \boldsymbol{a}^\mathsf{T}\mathbf{A}\boldsymbol{a}$ and $i \in [0, ..., k]$. The matrices $\mathbf{P}$, $\mathbf{R}$, and $\mathbf{Q}$ are weighting matrices with appropriate dimensions. The function $\boldsymbol{c} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u+1} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_c}$ represents the model constraints, and $\boldsymbol{h} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u+1} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_y}$ comprises the equations of measurements $\boldsymbol{y} \in \mathbb{R}^{n_y}$. Note that $k+1$ denotes the number of samples collected up to time t_k. The measurements start from t_0, thus at t_k there are $k+1$ measurements available. The full information estimator considers all the past measurements, whereas a *moving horizon estimator* refers to using only the measurements within a specific time window [115, 142]. Depending on the specific needs and computational limitations, one could opt for a moving horizon estimation.

The optimization variables of the full information estimator are the initial state $\boldsymbol{x_0}$, the parameter $\boldsymbol{\theta}$, and the state noise $\boldsymbol{\omega} := [\boldsymbol{\omega_0}^\mathsf{T}, ..., \boldsymbol{\omega}_k^\mathsf{T}]^\mathsf{T}$. The solution of the full information estimation problem is denoted by $(\cdot)^*$, while $(\hat{\cdot})$ represents the prior information of a variable[2]. The states at t_k, reconstructed using $\boldsymbol{x}_0^*$, $\boldsymbol{\omega}^*$ and $\boldsymbol{\theta}^*$ (cf. Eq. (6.16c)), can be used in the MPC to enable online feedback control. Furthermore, the objective function in (6.16a) consists of three terms. The first term represents the discrepancy between the prior information of states and model parameters and the actual estimated values at the beginning of the estimation horizon, i.e., the *arrival cost*. The second term quantifies the difference between the predicted measurements and the actual measurements $\tilde{\boldsymbol{y}}_{\boldsymbol{i}}$ for every $i \in [0, ..., k]$. The final term accounts for the influence of the *state noise*. In order to capture potential model uncertainty, the state noise $\boldsymbol{\omega_i}$ is incorporated into the model at each sampling time (cf. Eq. (6.16c)). The inclusion of state noise is convenient when addressing model-plant mismatch arising from parameter or structural model uncertainties.

The choice of weights $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{R}$ should correspond to the relative importance of the different terms. For example, in scenarios characterized by low measurement noise and high model uncertainty, the weight $\mathbf{R}$ should be *larger* than $\mathbf{Q}$ to reflect a greater reliance on the measurements compared to the model. Similarly, if the prior values are considered more trustworthy than the measurements, the matrix $\mathbf{P}$ should be *larger* than $\mathbf{R}$.

[2]This could be the *best* measurement or *best* guess available for that specific variable.

6.4 Anaerobic lactate fermentation by *E. coli* with optogenetic control of ATPase expression

As a case study, the anaerobic lactate fermentation by *E. coli* in fed-batch is used. It is assumed that the ATPase expression in the cell is regulated through the optogenetic Ccas/CcaR system to enable dynamic enforced ATP turnover. The concentration of the ATPase is denoted as $p_{\text{ATPase}} \in \mathbb{R}$. A single regulated protein is considered, hence $\boldsymbol{p_{\text{reg}}} := p_{\text{ATPase}}$. From now on, $\boldsymbol{u_s} := I_s$ and $\bar{\boldsymbol{u}}_{\boldsymbol{c}} := \bar{I}_c$, where I_s denotes the manipulated green light intensity exerted by the controller, and $\bar{I}_c$ represents the average value of green light perceived by the cells within the bioreactor. There are two dynamic degrees of freedom, the cybergenetic input (green light) and the substrate (glucose) feed, hence $\boldsymbol{u_p} := [I_s(\cdot), F_{\text{in}}(\cdot)]^{\mathsf{T}}$. The initial conditions of the plant are not optimized. A scheme of the considered setup and control strategy is presented in Fig. 6.1.

Additionally, a flat-panel photobioreactor configuration is considered, which comprises two flat surfaces interconnected by a narrow gap [143]. One side of the flat-panel bioreactor is illuminated by a green light source. This particular geometry maximizes the illuminated area in relation to the culture volume, rendering it particularly suitable for optogenetic applications.

6.4.1 Model and process considerations

Similar to Eq. (5.3), the dose-response function for ATPase expression is assumed to follow a Hill function [130]:

$$\eta_{\text{ATPase}}(\bar{I}_c) = \alpha + \beta \frac{\bar{I}_c^{\delta}}{K^{\delta} + \bar{I}_c^{\delta}}, \tag{6.17}$$

however, the difference is that here the *average* light input inside the bioreactor is considered. $\alpha \in \mathbb{R}$ represents an input-independent basal rate of production, which could be attributed to factors such as promoter leakage or basal gene expression. $\beta \in \mathbb{R}$ denotes an input-dependent maximum rate of production, while $K \in \mathbb{R}$ is a saturation constant. The exponent $\delta \in \mathbb{R}$ corresponds to the Hill coefficient.

To make the study more realistic, it is assumed that light penetration within the bioreactor is not uniform due to the interference of cells along the light beam. Let $l \in \mathbb{R}$ represent the distance between the two plates of the flat-panel bioreactor (see Fig. 6.1-Bioreactor level as a reference). By performing a balance over an infinitesimally small distance dl, considering the perpendicular incidence of light onto the illuminated flat surface, and assuming a well-mixed culture, integration from l_0 to l yields

$$I_c(l,B) = I_c(l_0,B)e^{-a_\lambda B(l-l_0)}, \tag{6.18}$$

where $I_c(l_0,B) = I_s$, and $a_\lambda \in \mathbb{R}$ denotes a lumped biomass-specific constant that accounts for the effects of light scattering and absorption. The derivation of the above equation resembles

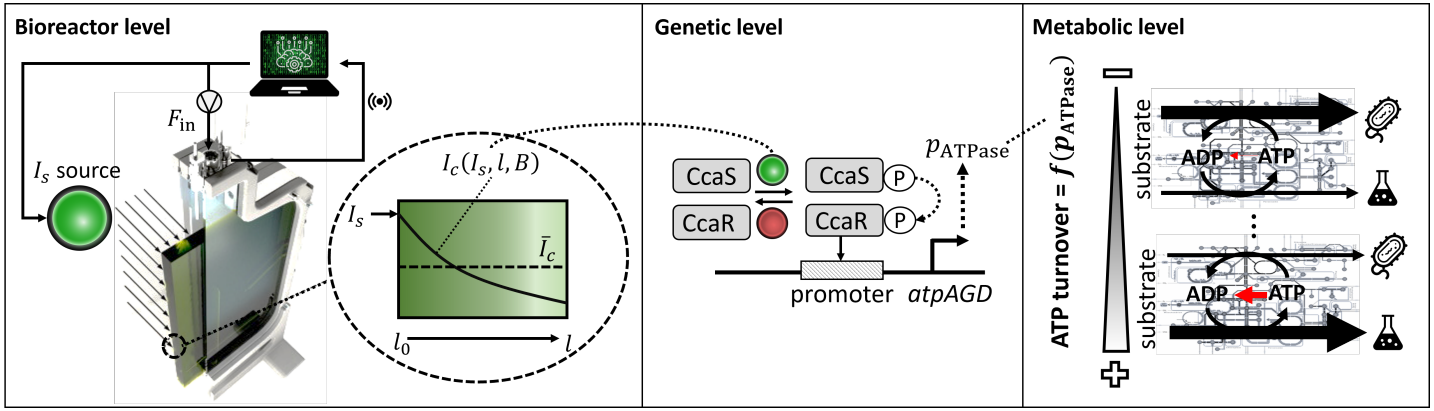


Figure 6.1: Scheme of the CcaS/CcaR optogenetic system for regulating the expression of the ATPase in fed-batch. The left side showcases the configuration of the flat-panel (photo)bioreactor alongside the control scheme, with the average light input within the culture. On the right side, the impact of varying ATPase concentration on metabolism is depicted. Bioreactor image adapted from [145].

that of the Lambert-Beer law [144]. The average light $\bar{I}_c$ can be determined by calculating the mean integral of $I_c(l,B)$ from $l_0 = 0$ to l

$$\bar{I}_c(B) = \frac{I_s}{a_\lambda B l}\left(1 - e^{-a_\lambda B l}\right). \tag{6.19}$$

Other relevant considerations for the case study are as follows. It is assumed that the degradation terms in Eqs. (6.5)-(6.6) are negligible. However, for the ATPase, the degradation term is given by

$$D_{\text{ATPase}} = d_{\text{ATPase}} p_{\text{ATPase}} v_L, \tag{6.20}$$

where $d_{\text{ATPase}} \in \mathbb{R}$ represents a constant rate of degradation of the ATPase.

The dynamics of lactate fermentation are based on the deFBA model outlined in the previous chapters. See Table 4.1 and Fig. 4.2 for a summary of the considered resource allocation model. In addition, the model parameters for the CcaS/CcaR system, the flat-panel bioreactor, and the initial conditions of the fed-batch process are presented in Table 6.1.

The chosen cost function in the optimization problem aimed to maximize the lactate concentration after a final time t_f of 30 hours $J = z_{\text{LAC}}(t_f)$ and 12 control actions $N = 12$. Remark constraints on the inputs were considered, specifically $I_s = [0,1]\,\text{W/m}^2$ and $F_{\text{in}} = [0,1]\,\text{L/h}$. An additional constraint was imposed to ensure complete consumption of glucose (the feeding substrate) at the end of the process, $z_{\text{GLC}}(t_f) = 0$. Lastly, the volume of the bioreactor should not exceed the maximum working volume capacity $v_{L_{\max}}$, thus the constraint $v_L \leq v_{L_{\max}}$ was included. Refer to Remark 4.3 for more information on the numerical solution of the optimization problems.

6.4.2 Open-loop optimization results

Fig. 6.2 presents the results of the open-loop optimization under the assumption of *no model-plant mismatch*. Four scenarios are considered:

- **Scenario 1 (S1)** represents a high-strength inducible CcaS/CcaR system, with the highest β value.
- **Scenario 2 (S2)** represents a medium-strength inducible CcaS/CcaR system, with a medium β value.
- **Scenario 3 (S3)** represents a low-strength inducible CcaS/CcaR system, with the lowest β value.
- **No induction (NI)** corresponds to the scenario without an inducible ATPase, where the CcaS/CcaR system is not present and thus no enforced ATP turnover occurs.

In the case without enforced ATP turnover (NI)[3], the final lactate concentration was 1434.3 mM. In contrast, in scenario S1 it reached 1572.3 mM (↑ 10 %), in scenario S2

[3]Note that in the NI scenario the only degree of freedom was the substrate feed rate.

Table 6.1: Relevant parameters and initial conditions of the nominal model.

Item	Value	Unit	Ref./Note
δ	2.490	1	[130]
K	0.138	$\mathrm{W/m^2}$	[130]
α	$2 \cdot 10^{-6}$	mmol/g/h	Note 1
β, S1	$1 \cdot 10^{-4}$	mmol/g/h	Note 1
β, S2	$2.5 \cdot 10^{-5}$	mmol/g/h	Note 1
β, S3	$1 \cdot 10^{-5}$	mmol/g/h	Note 1
d_{ATPase}	$6.3 \cdot 10^{-2}$	1/h	Note 2
l	0.022	m	Note 3
$v_{L_{\max}}$	45	L	Note 3
a_λ	$1 \cdot 10^{-2}$	$\mathrm{m^2/g}$	Note 4
$z_{\mathrm{GLC,in}}$	2220	mM	—
$x_{\mathrm{GLC}}(0)$	139	mM	—
$x_{\mathrm{LAC}}(0)$	0	mM	—
$x_{\mathrm{CO_2}}(0)$	0	mM	—
$x_{\mathrm{FOR}}(0)$	0	mM	—
$x_{\mathrm{SUCC}}(0)$	0	mM	—
$B(0)$	0.59	g/L	—
$\boldsymbol{p}(0)$	Note 5	mM	—
$v_L(0)$	30	L	—

Note 1. Biologically sound parameter values derived from feasible deFBA simulations for different induction strength scenarios ($\mathrm{S_i}$). **Note 2**. Estimated as $d_{\mathrm{ATPase}} = \frac{\ln(2)}{t_{0.5}}$, where $t_{0.5}$ is the ATPase protein half-life time [132]. **Note 3**. Based on a pilot-scale flat-panel photobioreactor [146]. **Note 4**. Assumed biologically sound order of magnitude. Estimated to be approximately $1/30$ of typical parameter values for microalgae [147]. **Note 5**. Estimated from $B(0)$ using resource balance analysis (cf. Remark 4.2).

1538.5 mM (↑ 7 %), and in scenario S3 1498.4 mM (↑ 4 %). In all fermentations, the maximum bioreactor volume was reached and all glucose was depleted. This indicates that the same net amount of glucose was fed and consumed in all scenarios. Therefore, the relative improvements in product titer observed in the aforementioned scenarios corresponded, proportionally, to average enhancements in product yield *and* volumetric productivity. Chapters 4 and 5 discussed the trade-off between increased product yield and decreased volumetric productivity as a result of enforced ATP turnover during one-stage batch fermentations. This chapter demonstrates that, in a fed-batch system, it is possible to simultaneously enhance both the product yield through the ATP turnover mechanism *and* the volumetric productivity by implementing a substrate feed. However, if one simulates a batch process, e.g., taking as a reference the same light regime as in scenario S1, starting with the maximal volume and the same total amount of sugar as added in

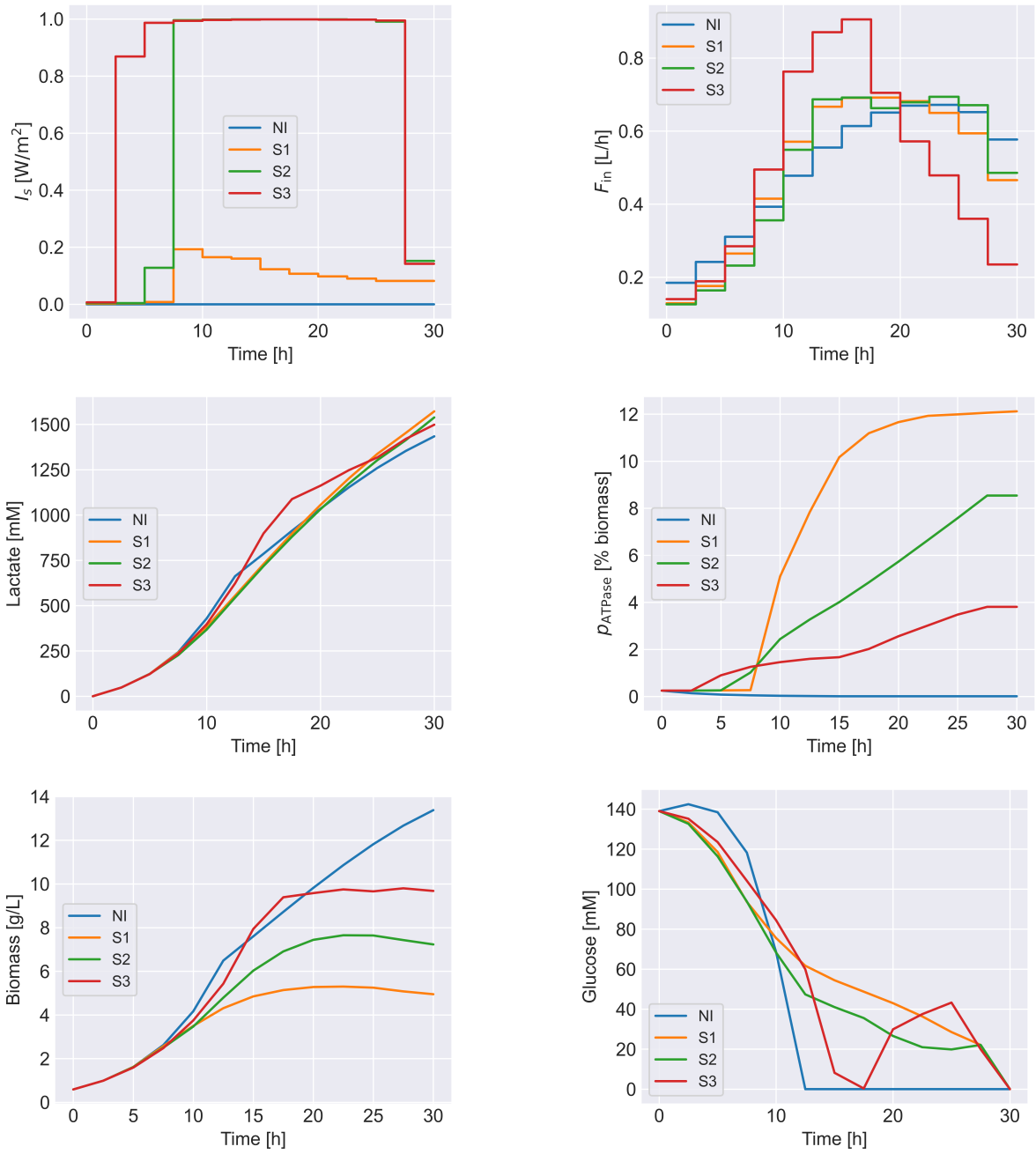


Figure 6.2: Open-loop fed-batch simulations *without* model-plant mismatch. The simulations considered the following scenarios based on different induction strengths of the CcaS/CcaR system: NI (no induction), S1 (high-strength induction), S2 (medium-strength induction), and S3 (low-strength induction).

the fed-batch counterpart (not shown), an equivalent volumetric productivity is observed. Following the same logic, for the uninduced scenario, the adapted batch version finishes earlier than the corresponding fed-batch NI. Therefore, the improved volumetric productiviy in the induced fed-batch cases against the uninduced fed-batch in Fig. 6.2 is mainly explained by the imposed upper bound on the feed rate. Of course, in reality, starting the process at very high substrate concentrations may be impractical due to viscosity issues and substrate inhibition, hence the fed-batch system would still be relevant.

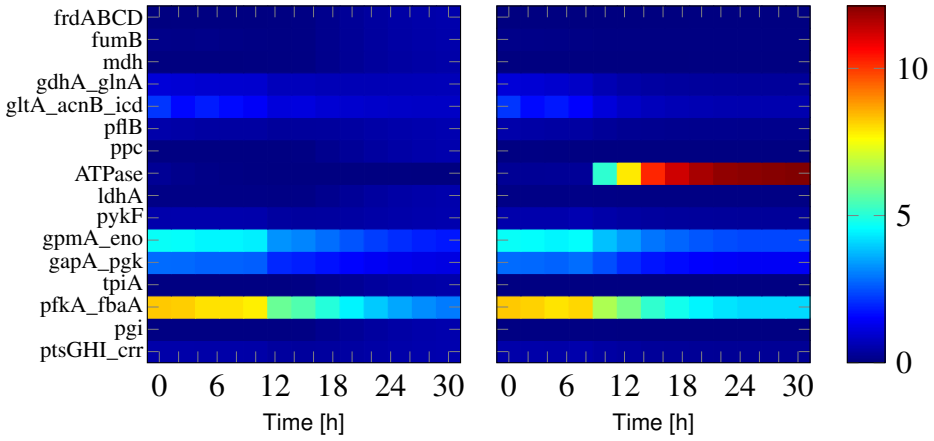


Figure 6.3: Heat map of the relative expression of enzymes in the open-loop scenarios NI (left) and S1 (right) in Fig. 6.2. The color bar on the right corner indicates the scale as a percentage of biomass dry weight.

In scenario S1, the maximum ATPase concentration reached 12.1 % of the cell dry weight, whereas in S2 and S3, it was 8.5 % and 3.8 %, respectively. As expected, this observation confirms that the induction strength of the CcaS/CcaR system directly influences the achievable expression of the ATPase. In addition, in comparison to scenario NI, scenarios S1, S2, and S3 exhibited 63, 46, and 28 % lower final biomass concentrations, respectively. This reduction can be attributed to the combined impact of lower biomass yields resulting from ATP turnover and the potential resource burden associated with ATPase induction. Despite the lower biomass growth rates in the scenarios with ATPase induction, the increased ATP turnover rates successfully improved the final lactate titer. In each of the induction scenarios, the feed rate exhibited an initial gradual increase, followed by a continuous decrease around the midpoint of the fermentation to avoid exceeding the maximum bioreactor volume capacity.

Due to the induced ATPase expression, the cell must adjust the remaining biomass components to optimize growth, which is assumed as the cell's objective function. An advantage of the outlined model-based optimization approach is the incorporation of resource allocation constraints. As an example, the dynamic profile of enzymes during the open-loop fermentation (S1) in Fig. 6.2 is illustrated with a heat map in Fig. 6.3. The induction of the ATPase results in a reconfiguration of the unregulated or *native* enzyme distribution, consistent with the observations in Section 4.4.2-Fig. 4.6. For instance, when comparing the profiles of enzymes frdABCD, fumB, mdh, ppc, ldhA, tpiA, and pgi, they have lower concentrations in the ATPase induction cases, once the ATPase starts to accumulate, compared to the NI fermentation.

6.4.3 Closed-loop control results

Open-loop control fails to account for model uncertainties, unforeseen disturbances, and process changes (e.g., change in bioreactor parts, change in inoculum preparation, etc.). Therefore, shrinking-horizon MPC is considered for addressing system uncertainty. For simplicity, the closed-loop control simulations focused on the high-strength inducible CcaS/CcaR system.

To introduce model-plant mismatch, the catalytic constants of enzymes pfkA_fbaA, gpmA_eno, gapA_pgk, gltA_acnB_icd, and gdhA_glnA were scaled down by a factor of 0.98. Additionally, the parameters δ and d_{ATPase} were scaled down by 0.97 and 0.98, respectively. The *modified* model was employed for plant simulations, while the controller operated based on the *nominal* model. Two cases are considered:

- **Model predictive control-scenario 1 (MPC 1)** assumes that all states can be measured online without any measurement noise.
- **Model predictive control-scenario 2 (MPC 2)** assumes that online measurements of the ATPase[4], biomass dry weight, and extracellular metabolites are available. Gaussian white noise (with a standard deviation of 1 % with respect to the exact plant value) is added to the *measurements*. The cell composition is estimated using full information estimation as described in Section 6.3.1[5]. The reconstructed cell composition and the extracellular online measurements are then passed to the MPC.

The results of the MPC simulations are shown in Fig. 6.4. Additionally, the open-loop scenario (*with model-plant mismatch*) is included as a reference case. In the open-loop controller, a final lactate concentration of 1449.5 mM was achieved, with 64.3 mM of net unconsumed glucose. Initially, the applied light input resulted in an ATPase concentration of 12.6% of the cell dry weight, followed by a slight decrease to 11.9 %.

In MPC 1, the adjustment of the light input led to a higher fraction of ATPase in the cell compared to the open-loop fermentation[6]. This, combined with the adjusted feed rates, resulted in a final lactate titer of 1567.8 mM. This represents an improvement of 8 % compared to the open-loop optimization. Unlike the open-loop optimization, in MPC 1, there was no unconsumed glucose at the end of the process. The results of MPC 1 should be interpreted with caution; they are *too optimistic* as they assume full state measurement without any measurement noise.

MPC 2 provides a more realistic scenario as it involves state estimation and considers measurement noise. The estimation results of biomass components at different sampling times are shown in Fig. 6.5, along with the calculated standard error (SE_FIE) of the estimates. The latter is computed as

$$\text{SE\_FIE} = \sqrt{\frac{\sum_{j=0}^{N-1}(p_{i,j} - p_{i,j}^*)^2}{n_T}}, \forall i \in [1, n_{p_i}], \tag{6.21}$$

where n_T is the total number of estimates for p_i. The values of the estimated biomass components are denoted as p_i^*.

[4]The ATPase could be measured, e.g., using a fluorescence-based biosensor [148].

[5]For simplicity, the used full information estimation only considers the second term of the objective function in Eq. (6.16a). It disregards state noise and assumes constant model parameters, hence eliminating the need for parameter estimation. The identity matrix is chosen for matrix **R**.

[6]Remark that this open-loop fermentation refers to the one with model-plant mismatch. Do not confuse with the open-loop optimizations in Fig. 6.2 where no model-plant mismatch was considered.

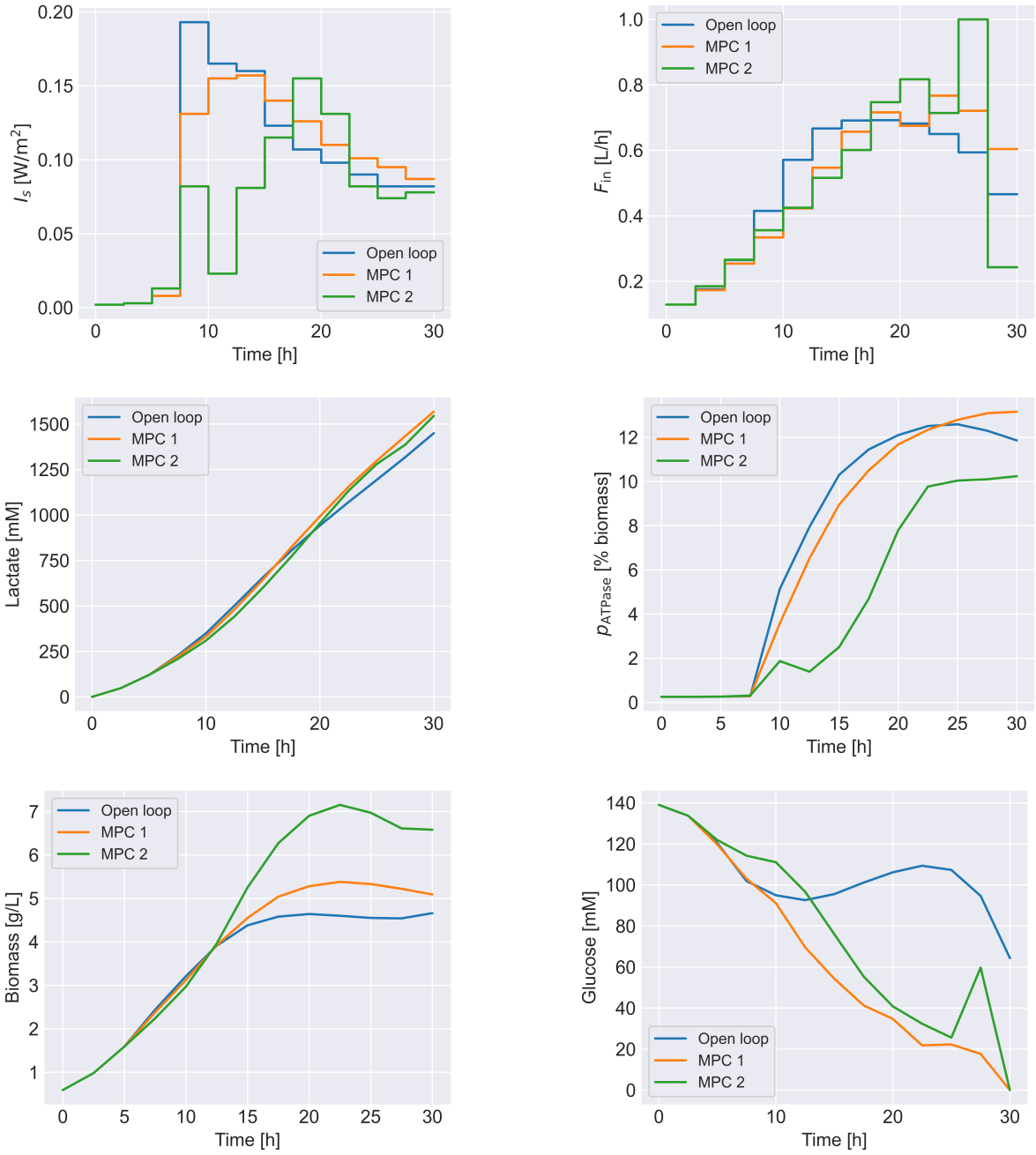


Figure 6.4: Closed-loop fed-batch simulations with model uncertainty for the high-strength inducible CcaS/CcaR system. MPC 1: full state measurement without measurement noise. MPC 2: noisy measurements of p_{ATPase}, B, and $\boldsymbol{z}$; full information estimator employed to infer $\boldsymbol{p}$. The figure includes the open-loop case.

In general, the full information estimator tracked well the dynamic changes in biomass composition. It is important to note that the estimation improved as the fermentation process advanced. For example, during the initial one-third of the process, the estimations for enzymes such as pfkA_fbaA, gltA_acnB_icd, and gdhA_glnA were less accurate, but later on they improved. This progressive improvement can be attributed to the growing estimation horizon and data available, i.e., with more sampling points, more process data can be used within the

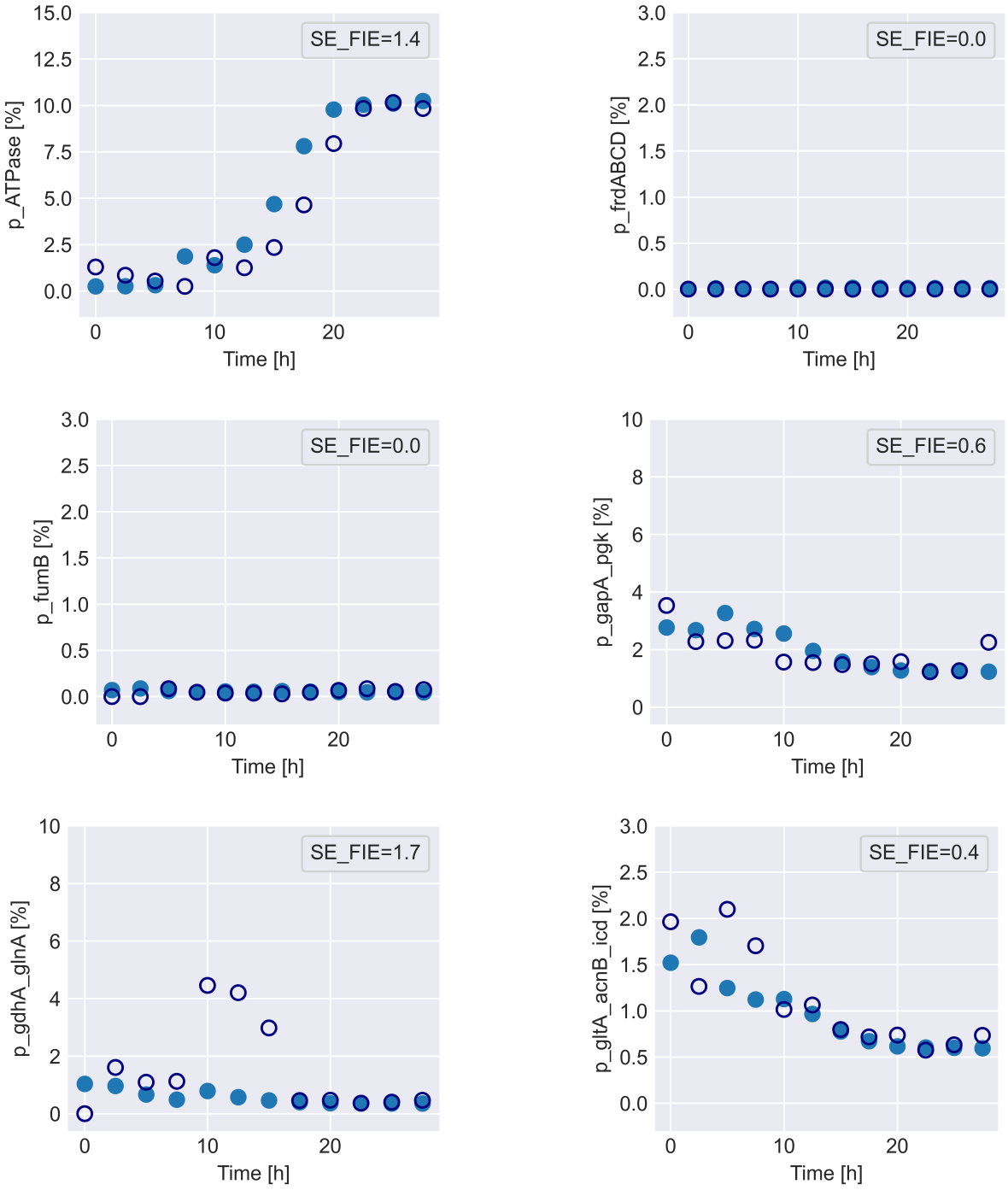


Figure 6.5: Online estimation of cell components, expressed as a percentage of cell dry weight. Species are shown in alphabetical order. Filled circles represent the exact values, while empty circles represent the estimated states. The standard error of the estimate (SE_FIE) is provided to indicate the accuracy of the estimation.

estimation. MPC 2 achieved an ATPase concentration of approximately 10.2 % of the dry cell weight, resulting in increased biomass accumulation compared to MPC 1 and the open-loop optimization. MPC 2 adjusted the feed rates to ensure complete consumption of glucose by the

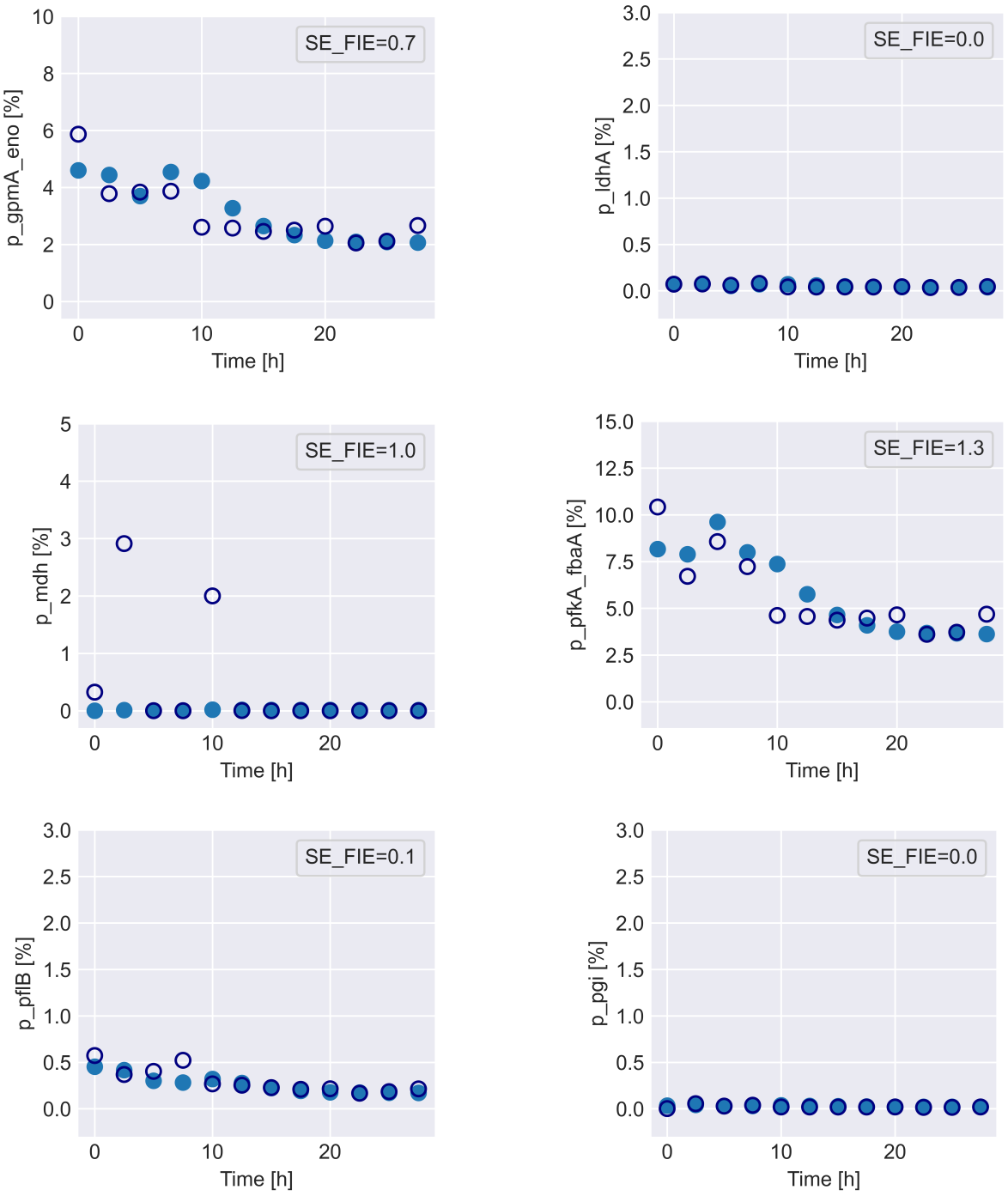


Figure continuation.

end of the process. As a result, this scenario reached a final lactate concentration of 1544.4 mM, approaching the value obtained in MPC 1.

6.5 Summary

In this chapter, cybergenetics was fused with model-based optimization and predictive control for dynamic metabolic engineering applications. This led to a generalized (fed-batch) metabolic

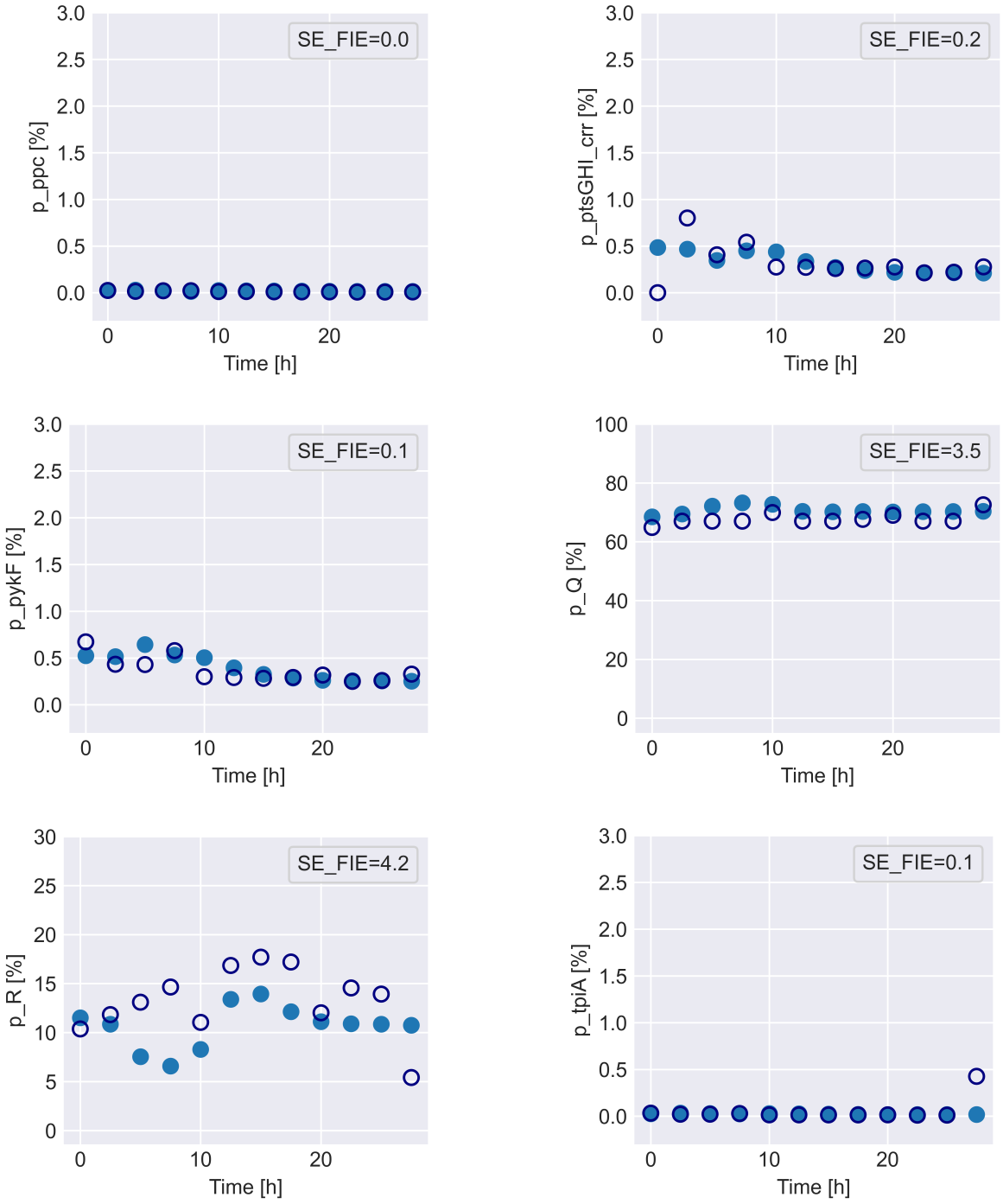


Figure continuation.

cybergenetic framework, where dynamic metabolic control is realized by fine-tuning the gene expression of a selected metabolism-relevant protein. To do so, a constraint-based dynamic modeling framework was developed that integrated the dynamics of metabolic reactions, resource allocation, and external adjustment of gene expression. The constraint-based dynamic model was used to formulate optimal control problems both in open-loop and closed-loop (via MPC).

In addition, a soft sensor based on full information estimation was outlined to reconstruct the biomass composition, necessary to enable predictive control.

The framework was evaluated by focusing on enforced ATP turnover through optogenetic modulation of the ATPase in fed-batch anaerobic lactate fermentation by *E. coli*. In contrast to previous chapters, here non-homogeneous light penetration in the bioreactor and fed-batch operation was considered. Optimal control of the light input and the substrate feed improved in simulations the process performance in terms of product titer *and* volumetric productivity. Furthermore, introducing feedback through MPC, optionally coupled with full information estimation to reconstruct biomass components, managed to mitigate uncertainties to a satisfactory extent.

Motivated by the concept of metabolic cybergenetics and the results of this chapter, the primary goal of the following chapter is to devise an experimental strategy for implementing optogenetic control of metabolism. This is done by utilizing enforced ATP turnover as a demonstrative example, limited to open-loop control due to technical considerations. Additionally, the next chapter addresses the question of whether and how the modeling and optimization framework could be simplified for practical applications.

7 Experimental validation of optogenetic open-loop control of ATPase in batch

While the modeling approach described in Chapters 5-6 provides valuable insights into metabolism and resource allocation, its complexity and computational demands may pose challenges for experimental implementation. Obtaining measurements of the dynamic states required to parameterize and validate such models can be technically difficult due to, e.g., lack of sensors and unavailability of analytical technologies. Although the use of soft sensors (as discussed in [37, 141] and in Section 6.3.1) is a possibility, their applicability depends on the number of measurable states and the availability of validated mathematical models[1].

Since the constraint-based dynamic models in Chapters 5-6 are optimization problems on their own, model-based optimization using such models is of a bilevel nature. Assumptions need to be made regarding the relationship between the upper- and lower-level optimization problems (cf. [124–126]). So far, an optimistic approach has been adopted to solve these problems, where the bilevel optimization is reformulated into a single-level optimization by substituting the lower-level problem with its corresponding Karush-Kuhn-Tucker conditions (cf. Remark 4.3). This transformation renders non-convex mathematical programs with complementarity constraints, which are generally difficult to solve due to the non-linearity of these constraints. Furthermore, the inclusion of Lagrange multipliers or dual variables, derived from the Karush-Kuhn-Tucker conditions, increases the size of the optimization problem. Overall, this led to numerical challenges in the bilevel optimizations presented in the previous chapters, requiring in many cases iteratively reruning the optimizations with different initial guesses, different linear solvers in IPOPT, e.g., MA27, MA57, MA77, MA86, and MA97 (https://licences.stfc.ac.uk/product/coin-hsl), and increasing convergence tolerances until an acceptable solution was found.

In this chapter, a simplified modeling framework is proposed for fermentations with optogenetic control of the ATPase, although the approach can also be applied to other processes involving external control of gene expression of metabolism-relevant proteins. This is achieved by modeling only the most relevant extracellular states and the intracellular components under external control. This leads to a system of differential equations as opposed to constraint-based dynamic models which are formulated as optimization problems. To compensate for the approximations made, machine learning is used to obtain a hybrid model. In situations where there is significant model uncertainty, the use of Gaussian processes [149] is outlined to learn the

[1] Recall that the model used in the previous chapters, in particular the part regarding the dynamics of the optogenetic actuator, was built based on literature values and biologically sound assumptions.

errors in the known dynamic equations. A single-level optimization problem can be formulated using the simplified dynamic model, avoiding the complexity of bilevel optimizations.

A scheme of the experimental control strategy is presented in Fig. 7.1. Due to technical considerations[2] and ease of experimental implementation, the focus is on batch processes and open-loop control. This serves, nevertheless, as a proof of concept of external modulation of metabolism-relevant proteins for bioprocess optimization.

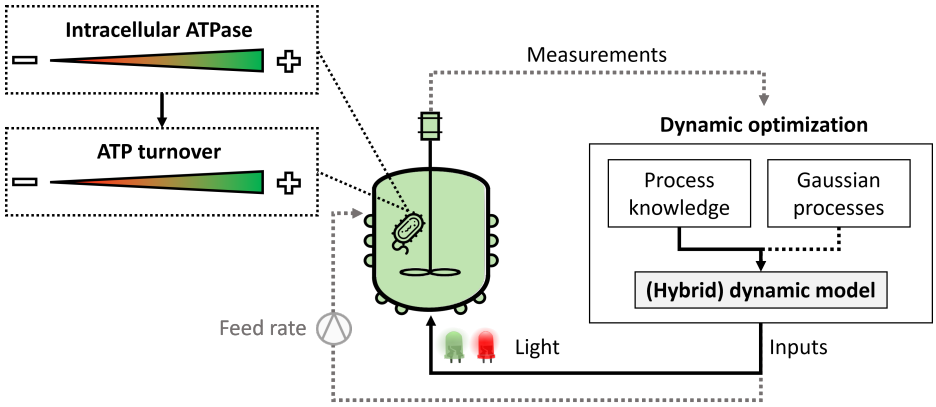


Figure 7.1: Scheme of the proposed control strategy for experimental implementation. Control can be performed in an open-loop or closed-loop manner. In this chapter, open-loop control is considered. Inputs can include intracellular and extracellular degrees of freedom (e.g., light intensity for ATPase induction and substrate feed in fed-batch fermentations). In this chapter, the focus is on batch processes. The model used for dynamic optimization can be knowledge-based or a hybrid model incorporating machine learning components such as Gaussian processes.

This chapter is organized as follows. Section 7.1 provides an overview of the genetically engineered biological system and experimental setup. The corresponding modeling framework and dynamic optimization problem are outlined in Sections 7.2 and 7.3, respectively. The anaerobic lactate fermentation by *E. coli* with optogenetic control of the ATPase is considered, which is the same biological system discussed in Chapters 5-6. The experimental results validating the proposed model and open-loop optimization are presented in Section 7.4.

7.1 Genetically engineered strain and experimental setup

To recapitulate, an *E. coli* strain with deletions of the ethanol and acetate production pathways [102] is considered as a basis for the experimental validation (cf. Fig.7.2-A). Under anaerobic conditions, lactate synthesis in this engineered microorganism becomes the main fermentation pathway for achieving redox balance, making it suitable for enforced ATP turnover.

The two-component optogenetic system CcaS/CcaR is utilized to establish control over the expression of ATPase F_1-subunit (*atpAGD*) in *E. coli* (cf. Section 5.1). For experimental validation, *E. coli* was engineered by inserting an *atpAGD* expression cassette into the chromosome, where all genes are regulated by the optimized *cpcG2* promoter [150]. Additionally, plasmid

[2]For example, lack of real-time (soft) sensors and automated actuators in the laboratory.

pPLPCB(S) [151] was introduced to enable the production of phycocyanobilin, a cofactor essential for photo-sensing, and plasmid pGB-MPI-23 was used to express the CcaS/CcaR proteins. In the resulting *E. coli* strain sGB015, ATPase expression can be regulated using light (cf. Fig. 7.2-B). Refer to Appendix A.2 for a detailed description of the genetic engineering procedure.

Anaerobic fermentation experiments with *E. coli* sGB015 were carried out at 37 °C in a Certomat BS-1 incubator (B. Braun Biotech International) (cf. Fig. 7.2-C,D). Light-emitting diode (LED) arrays (Osram OSLON SSL LED green λ_{peak} = 521 nm; Osram OSLON SSL LED red λ_{peak} = 660 nm) were employed to generate red and green light outputs. The intensity of the emitted light was set by adjusting the supplied current. The corresponding photon flux density was determined using an ULM-500 Universal Light Meter (Heinz Walz GmbH) and expressed in μmol m^{-2} s^{-1}. Each experiment was performed with three biological replicates. Refer to Appendix A.3 for a more detailed description of the fermentation procedure and analytical measurements.

7.2 Hybrid quasi-unstructured/unsegregated modeling of a fermentation with inducible ATPase

The modeling approach for fermentations with optogenetic regulation of the ATPase is outlined in this section. The notation is kept general for clarity of presentation. The specific model assumptions for the lactate fermentation are presented in Section 7.4.1.

7.2.1 Model formulation

The dynamics of the biomass $B \in \mathbb{R}$, *rate-limiting* external substrates $\boldsymbol{s} \in \mathbb{R}^{n_s}$, *rate-limiting* (by)products and products of interest $\boldsymbol{\rho} \in \mathbb{R}^{n_\rho}$, as well as the intracellular ATPase $E \in \mathbb{R}$ are considered in the model. B, $\boldsymbol{s}$, and $\boldsymbol{\rho}$ are expressed in mass per culture volume, and E is expressed in mass of ATPase per biomass. The biomass is treated as a homogeneous population of cells, making the model *unsegregated*. Except for the dynamics of the intracellular ATPase, the model is *quasi-unstructured* as the remaining intracellular components are neglected. The dynamic input $u_l \in \mathbb{R}$ represents the green light photon flux density under a constant red light background. For simplicity, homogeneous light penetration in the bioreactor is assumed.

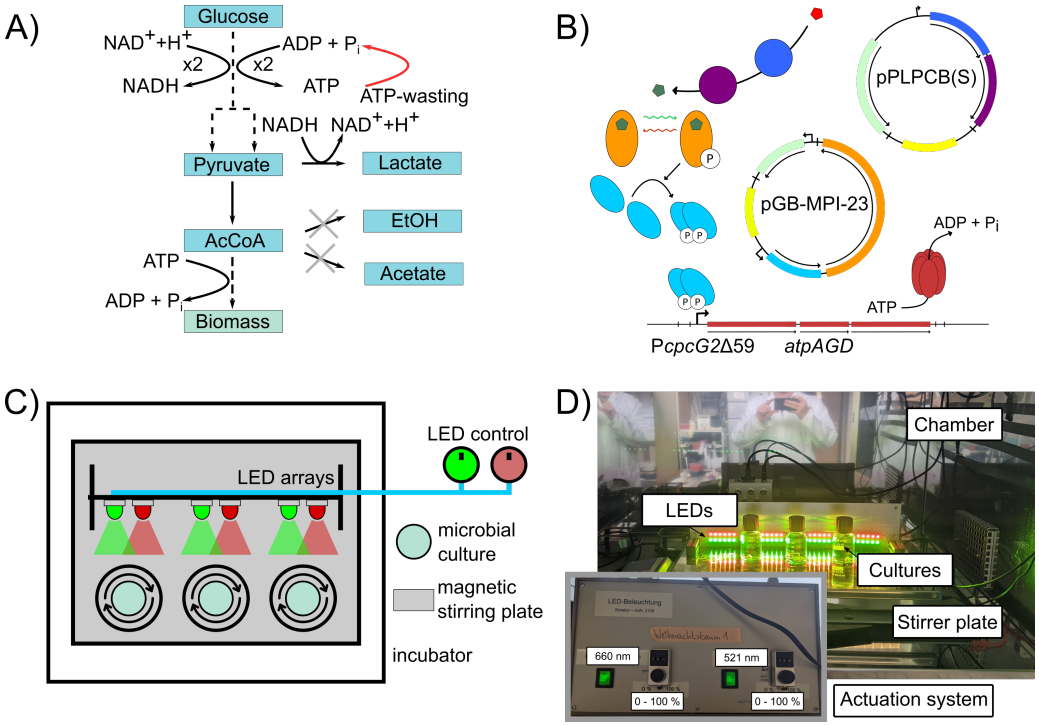


Figure 7.2: A) Simplified core metabolism of *E. coli* sGB015 with enforced ATP turnover. B) Light-inducible ATPase in *E. coli* sGB015 managed by three heterologous genetic elements: pPLPCB(S), pGB-MPI-23, and the chromosomal insertion of PcpcG2Δ59-*atpAGD*-rrnBT1. Genes (italicized) and their related proteins (non-italicized) are color-coded. pPLPCB(S) expresses *ho1* (dark blue) and *pcyA* (purple), enabling the conversion of heme (red pentagon) into phycocyanobilin (green pentagon), the chromophore necessary for CcaS to detect light. The expression of *ccaS* (orange) and *ccaR* (pale blue) is facilitated by pGB-MPI-23. CcaS autophosphorylates upon green light-induced conformational change, leading to CcaR phosphorylation, dimerization, and functioning as a transcription factor for PcpcG2Δ59. Promoters are shown as arrows, open reading frames as arrows adjacent to their respective genes, terminators as black perpendicular lines, origins of replication in yellow, and antibiotic resistances in pale green. C) Fermentation setup with green and red light actuation system based on LEDs. D) Photograph of the actual setup in C).

The process dynamics can be described as follows

$$\frac{\mathrm{d}B}{\mathrm{d}t} = \boldsymbol{S}_B \boldsymbol{r}(\boldsymbol{x}, u_l, \boldsymbol{\psi}) + \boldsymbol{Q}_B \boldsymbol{w}(\boldsymbol{x}, u_l, \boldsymbol{\tau}), \tag{7.1a}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \boldsymbol{S}_E \boldsymbol{r}(\boldsymbol{x}, u_l, \boldsymbol{\psi}) + \boldsymbol{Q}_E \boldsymbol{w}(\boldsymbol{x}, u_l, \boldsymbol{\tau}), \tag{7.1b}$$

$$\frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}t} = \boldsymbol{S}_s \boldsymbol{r}(\boldsymbol{x}, u_l, \boldsymbol{\psi}) + \boldsymbol{Q}_s \boldsymbol{w}(\boldsymbol{x}, u_l, \boldsymbol{\tau}), \tag{7.1c}$$

$$\frac{\mathrm{d}\boldsymbol{\rho}}{\mathrm{d}t} = \boldsymbol{S}_p \boldsymbol{r}(\boldsymbol{x}, u_l, \boldsymbol{\psi}) + \boldsymbol{Q}_\rho \boldsymbol{w}(\boldsymbol{x}, u_l, \boldsymbol{\tau}), \tag{7.1d}$$

$$\boldsymbol{x} := [B, E, \boldsymbol{s}^\mathsf{T}, \boldsymbol{\rho}^\mathsf{T}]^\mathsf{T}, \tag{7.1e}$$

$$\boldsymbol{x}(t_0) = \boldsymbol{x_0}. \tag{7.1f}$$

The function $\boldsymbol{r} : \mathbb{R}^{n_x} \times \mathbb{R} \times \mathbb{R}^{n_\psi} \to \mathbb{R}^{n_r}$ represents the reaction rates of the process, encompassing production, consumption, degradation, and dilution rates[3]. The parameters of the reaction rates are comprised in $\boldsymbol{\psi} \in \mathbb{R}^{n_\psi}$. $\boldsymbol{S_B} \in \mathbb{R}^{1\times n_r}$, $\boldsymbol{S_E} \in \mathbb{R}^{1\times n_r}$, $\boldsymbol{S_s} \in \mathbb{R}^{n_s\times n_r}$, $\boldsymbol{S_\rho} \in \mathbb{R}^{n_\rho\times n_r}$ map the coefficients of the reaction rates to the system´s differential equations. The previous elements constitute the *knowledge-based* portion of the model.

Additionally, the dynamic equations in (7.1a)-(7.1d) account for model uncertainty due to, e.g., oversimplified or incorrect assumptions. This model *error* is defined by the vector-valued function $\boldsymbol{w} : \mathbb{R}^{n_x} \times \mathbb{R} \times \mathbb{R}^{n_\tau} \to \mathbb{R}^{n_w}$ and captures the dynamics not described by the knowledge-based part of the equations. The vector $\boldsymbol{\tau} \in \mathbb{R}^{n_\tau}$ represents the parameters of $\boldsymbol{w}$. $\boldsymbol{Q_B} \in \mathbb{R}^{1\times n_w}$, $\boldsymbol{Q_E} \in \mathbb{R}^{1\times n_w}$, $\boldsymbol{Q_s} \in \mathbb{R}^{n_s\times n_w}$, and $\boldsymbol{Q_\rho} \in \mathbb{R}^{n_\rho\times n_w}$ map the functions describing the model error to the differential equations.

When the knowledge-based component of the dynamic equations describes the real system to a satisfactory extent, one could simply neglect $\boldsymbol{w}$ from the model. However, when there is a significant mismatch between the knowledge-based part of the model and the actual system, a hybrid modeling approach can be adopted (cf. [80–83] for examples of hybrid models). This approach involves augmenting the knowledge-based part of the model with a data-driven/machine-learning component. It is important to note that experiments in biotechnology often involve substantial costs and time requirements, resulting in limited availability of training data sets for machine learning. In this chapter, Gaussian processes [149] are considered for *learning* $\boldsymbol{w}$ as they often offer good approximation quality even with small training data sets.

7.2.2 Gaussian process regression

Gaussian processes offer a probabilistic distribution over functions. They provide predictions along with a measure of the prediction uncertainty. In this section, a brief description of Gaussian processes is presented. For more detailed information, the reader is referred to [64, 149, 152, 153].

Consider a Gaussian process regressor where $\kappa \in \mathbb{R}$ represents the label (regression output). Additionally, the features (regression inputs) are defined as $\boldsymbol{v} \in \mathbb{R}^{n_v}$. The aim of the Gaussian process regressor is to model a function $\zeta : \mathbb{R}^{n_v} \to \mathbb{R}$ from noisy observations κ of $\zeta(\boldsymbol{v})$, expressed as

$$\kappa = \zeta(\boldsymbol{v}) + \varepsilon, \tag{7.2}$$

where ε represents Gaussian-distributed (measurement) noise with zero mean and variance σ_n^2, hence $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$.

$\boldsymbol{V_d} \in \mathbb{R}^{n_v\times n_d}$ collects the supplied training inputs and $\boldsymbol{L} \in \mathbb{R}^{1\times n_d}$ represents the training outputs, where n_d corresponds to the number of training data sets. It is assumed that the labels follow a normal distribution

$$\zeta(\boldsymbol{v}) \sim \mathcal{N}(m(\boldsymbol{v}), k(\boldsymbol{v}, \boldsymbol{v})). \tag{7.3}$$

[3] For example, in case of fed-batch and continuous systems.

Here, the mean *function* is $m : \mathbb{R}^{n_v} \to \mathbb{R}$ and the kernel or covariance *function* is $k : \mathbb{R}^{n_v} \times \mathbb{R}^{n_v} \to \mathbb{R}$.

Gaussian processes utilize a *prior distribution* of functions, defined by a prior mean function and a prior covariance function. The choice of the kernel function determines the form of the prior distribution, and it is typically selected to be infinitely differentiable, smooth, and continuous. Let $\boldsymbol{v_i} \in \mathbb{R}^{n_v \times 1}$ and $\boldsymbol{v_j} \in \mathbb{R}^{n_v \times 1}$ be two arbitrary input vectors. In this chapter, the squared-exponential kernel function is employed

$$k(\boldsymbol{v_i}, \boldsymbol{v_j}|\boldsymbol{\tau}) = \sigma^2 \exp\left(\frac{-(\boldsymbol{v_i} - \boldsymbol{v_j})^\mathsf{T}(\boldsymbol{v_i} - \boldsymbol{v_j})}{2d^2}\right), \tag{7.4}$$

where $\sigma^2 \in \mathbb{R}$ represents the signal variance and $d \in \mathbb{R}$ corresponds to the length-scale.

The covariance *matrix* $\boldsymbol{K} \in \mathbb{R}^{n_d \times n_d}$ is obtained from the chosen kernel function and the supplied training data

$$\boldsymbol{K} = \begin{bmatrix} k(\boldsymbol{v_1}, \boldsymbol{v_1}) & \cdots & k(\boldsymbol{v_1}, \boldsymbol{v}_{n_d}) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{v}_{n_d}, \boldsymbol{v}_1) & \cdots & k(\boldsymbol{v}_{n_d} \boldsymbol{v}_{n_d}) \end{bmatrix}. \tag{7.5}$$

Note that the elements of the covariance matrix characterize the *neighborhood* or *similarity* between data points in the feature space via the kernel function.

The hyperparameters of the kernel function are optimized by maximizing the log marginal likelihood, thus $\boldsymbol{\tau}^* = \arg\max_{\boldsymbol{\tau}} \log p(\boldsymbol{L}|\boldsymbol{V_d}, \boldsymbol{\tau})$. The log marginal likelihood is given by

$$\begin{aligned} \log(p(\boldsymbol{L}|\boldsymbol{V_d}, \boldsymbol{\tau})) = &-\frac{1}{2}\boldsymbol{L}^\mathsf{T}(\boldsymbol{K} + \sigma_n^2 \boldsymbol{I_d})^{-1}\boldsymbol{L} - \frac{1}{2}\log\left(|\boldsymbol{K} + \sigma_n^2 \boldsymbol{I_d}|\right) \\ &- \frac{n_d}{2}\log(2\pi), \end{aligned} \tag{7.6}$$

where $\boldsymbol{\tau}$ represents the hyperparameters, defined as $\boldsymbol{\tau} := [\sigma^2, d, \sigma_n^2]$, and $\boldsymbol{I_d}$ denotes the appropriate-sized identity matrix.

The conditional posterior of the Gaussian process, given optimal hyperparameters, follows a normal distribution for a test input vector $\boldsymbol{v^*} \in \mathbb{R}^{n_v \times 1}$. This can be expressed as $p(\bar{\boldsymbol{\zeta}}(\boldsymbol{v}^*)|\boldsymbol{V_d}, \boldsymbol{L}) \sim \mathcal{N}(\bar{\boldsymbol{\zeta}}, \Sigma)$, where $\bar{\boldsymbol{\zeta}}$ represents the predictive mean and Σ denotes the predictive variance. The predictive mean is given by

$$\bar{\boldsymbol{\zeta}}(\boldsymbol{v}^*) = \tilde{\boldsymbol{k}}^\mathsf{T}(\boldsymbol{K} + \sigma_n^2 \boldsymbol{I_d})^{-1}\boldsymbol{L}, \tag{7.7}$$

and the predictive variance is given by

$$\Sigma(\boldsymbol{v}^*) = k(\boldsymbol{v}^*, \boldsymbol{v}^*) - \tilde{\boldsymbol{k}}^\mathsf{T}(\boldsymbol{K} + \sigma_n^2 \boldsymbol{I_d})^{-1}\tilde{\boldsymbol{k}}, \tag{7.8}$$

where $\tilde{\boldsymbol{k}} := [k(\boldsymbol{v_1}, \boldsymbol{v}^*), ..., k(\boldsymbol{v}_{n_d}, \boldsymbol{v}^*)]^\mathsf{T}$.

Here, $\boldsymbol{\zeta}(\boldsymbol{v})$ is defined as $\boldsymbol{\zeta}(\boldsymbol{v}) := w_i(\boldsymbol{v})$, where w_i represents the model error of a specific differential equation i. In the proposed approach, multiple Gaussian process regressors are

employed, corresponding to the number of model errors n_w. Therefore, the regressors are multi-input single-output Gaussian processes. The features used in the Gaussian processes can be, e.g., model states and inputs.

7.3 Gaussian-process-supported optimal control of optogenetic ATPase expression

To determine optimal light input trajectories for maximizing the efficiency of the process with optogenetic modulation of the ATPase, an optimal control is formulated

$$\max_{u_l(\cdot),\boldsymbol{x_0}} \quad J(\cdot), \tag{7.9a}$$

$$\text{s.t.} \quad \text{Eqs. (7.1a)} - \text{(7.1f)}, \tag{7.9b}$$

$$0 \leq \boldsymbol{g}(\boldsymbol{x}, u_l, \boldsymbol{\psi}, \boldsymbol{\tau}), \tag{7.9c}$$

where $J(\cdot)$ represents the cost function that quantifies the process efficiency, while $\boldsymbol{g} : \mathbb{R}^{n_x} \times \mathbb{R} \times \mathbb{R}^{n_\psi} \times \mathbb{R}^{n_\tau} \to \mathbb{R}^{n_g}$ denotes additional system constraints. The decision variables of the optimization problem encompass both static and dynamic degrees of freedom: the dynamic green light photon flux density $u_l(\cdot)$ and the initial state concentrations $\boldsymbol{x_0}$. Note that for the model only the mean of the Gaussian process is considered and the uncertainty, represented by the covariance information, is neglected. It could be, however, integrated, considering stochastic MPC formulations [64].

7.4 Experimental validation

This section outlines the experimental validation of the proposed modeling and open-loop control strategy for the anaerobic lactate fermentation by *E. coli* with optogenetic modulation of the ATPase in batch.

7.4.1 Modeling

The following dynamic states are considered: glucose ($s_G \in \mathbb{R}$), lactate ($\rho_L \in \mathbb{R}$), *E. coli*'s biomass ($B_c \in \mathbb{R}$), and intracellular ATPase ($E \in \mathbb{R}$). Thus, $\boldsymbol{s} := s_G$, $\boldsymbol{\rho} := \rho_L$, and $B := B_c$. This represents a significant reduction in the number of dynamic states compared to the constraint-based dynamic model outlined in the previous chapters. That is, the number of dynamic states is reduced from 23 (cf. Table 4.1) to only 4.

The proposed batch model is represented by the following differential equations

$$\frac{\mathrm{d}s_G}{\mathrm{d}t} = -q_G(s_G, E, \psi)B_c + w_G(s_G, B_c, \rho_L, E, u_l, \tau), \tag{7.10a}$$

$$\frac{\mathrm{d}B_c}{\mathrm{d}t} = \mu(s_G, E, \psi)B_c + w_c(s_G, B_c, \rho_L, E, u_l, \tau), \tag{7.10b}$$

$$\frac{\mathrm{d}\rho_L}{\mathrm{d}t} = q_L(s_G, E, \psi)B_c + w_L(s_G, B_c, \rho_L, E, u_l, \tau), \tag{7.10c}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = q_E(u_l, \psi) - d_E(E, \psi), \tag{7.10d}$$

$$s_G(t_0) = s_{G_0}, B_c(t_0) = B_{c_0}, \rho_L(t_0) = \rho_{L_0}, E(t_0) = E_0. \tag{7.10e}$$

In these equations, q_G, μ, q_L, q_E, and d_E are known kinetic *functions* with appropriate parameters, while w_G, w_c, and w_L represent Gaussian process regression functions that describe the model error. It should be noted that Eqs. (7.10a)-(7.10d) assume only two rate-limiting components in the kinetic functions, namely glucose and the light-inducible ATPase. Dilution or degradation effects are neglected in equations (7.10a)-(7.10c). Incorporating E into the model enables capturing potential time *delays* in extracellular uptake and production rates resulting from the lumped ATPase transcription and translation phenomena[4].

In terms of parameterizing the model, obtaining measurements of extracellular concentrations is relatively straightforward. However, quantifying the intracellular ATPase poses a challenge. Therefore, it is assumed for simplicity that the dynamics of the intracellular ATPase depend solely on its concentration and the light input, as shown in Eq. (7.10d). This variable is treated as a *virtual* variable[5] expressed in virtual units (VU) per gram of biomass, with no model uncertainty considered in Eq. (7.10d). Note that in the fermentation experiments the inoculum/preculture preparation follows a standardized protocol (cf. Appendix A.3). Specifically, the preculture is consistently grown under red light conditions, without ATPase induction. Consequently, for all fermentation experiments, the initial ATPase concentration was set arbitrarily to $E(t_0) = 0\,\mathrm{VU/g}$. The precise biological interpretation of Eq. (7.10d) and its associated parameters is neglected, as long as they effectively help to capture the dynamics of the extracellular species.

[4] Assuming that there is no additional delay associated with the light input and the start of the lumped transcription/translation process.

[5] An alternative approach, as demonstrated in Chapter 6, could involve estimating the ATPase concentration using soft sensors. However, this requires a validated mathematical model, which is currently unavailable.

The kinetic rates in the model are described by

$$q_G(s_G,E) = q_{G_{\max}} \left(\frac{s_G}{s_G+k_G}\right)\left(1+\frac{E^{n_1}}{E^{n_1}+k_{GV}^{n_1}}\right), \tag{7.11a}$$

$$\mu(s_G,E) = Y_{BG}\left(q_G(s_G,E)-m_G\right)\left(1-\frac{E^{n_2}}{E^{n_2}+k_{BV}^{n_2}}\right), \tag{7.11b}$$

$$q_L(s_G,E) = \left(Y_{LB}\mu(s_G,E)+m_L\right)\left(1+\frac{E^{n_3}}{E^{n_3}+k_{LV}^{n_3}}\right), \tag{7.11c}$$

$$q_E(u_l) = q_{E_0} + q_{E_{\max}} \frac{u_l^{n_4}}{u_l^{n_4}+k_u^{n_4}}. \tag{7.11d}$$

$$d_E(E) = k_d E. \tag{7.11e}$$

In the absence of enforced ATP turnover, i.e., $E(t)=0$, the specific substrate uptake rate (Eq. (7.11a)) follows conventional Monod-type kinetics [154], the specific growth rate (Eq. (7.11b)) is governed by Pirt's equation for substrate distribution [155], and the specific lactate production rate (Eq. (7.11c)) obeys the Leudeking-Piret's equation for catabolic products [156]. However, to account for changes in the specific rates due to enforced ATP turnover, i.e., for $E(t)\geq 0$, Hill-type *activation* terms [157] are incorporated into Eqs. (7.11a)-(7.11c), either increasing (+) or decreasing (-) the specific rates. The production of ATPase is *activated* by green light following the Hill function (Eq. (7.11d)) [157]. Lastly, an average lumped dilution/degradation rate of ATPase is assumed in Eq. (7.11e). The above kinetic functions have associated parameters[6] that need to be estimated. As in Chapter 5, a constant red light background is assumed, which is high enough to repress induction by other light sources/wavelengths and low enough to enable induction by green light. Therefore, for simplicity, the model only considers the effect of the green light input.

7.4.2 Model fitting and training

To validate the proposed modeling strategy, five fermentation experiments were conducted. In these experiments, different constant green light inputs were applied: 0, 175, 349, 524, and 873 μmol m^{-2} s^{-1}. The results, depicted by the blue bars in Fig. 7.3, show that the lactate-on-glucose yield for the *entire batch*[7] ($Y_{LG,batch}$) increased with higher light inputs. It reached approximately 1 g/g with the light input $u_l = 873$ μmol m^{-2} s^{-1}, which corresponds to the maximum theoretical yield. Furthermore, increasing light inputs led to decreasing *average* biomass-on-glucose yields ($Y_{BG,batch}$) and lactate volumetric productivities ($r_{L,batch}$). These results align with the simulations presented in the previous chapters.

A parameter estimation procedure was performed using COPASI with a particle swarm algorithm [158] based on the batch experiments presented in Fig. 7.4. The parameter estimation incorporated data from all the batch experiments simultaneously. It should be noted that these

[6] $\psi := [k_{BV}, k_G, k_{GV}, k_{LV}, m_G, m_L, n_1, n_2, n_3, q_{G_{\max}}, Y_{BG}, Y_{LB}, q_{E_0}, q_{E_{\max}}, n_4, k_u, k_d]^T$

[7] Calculated using only initial and final concentrations.

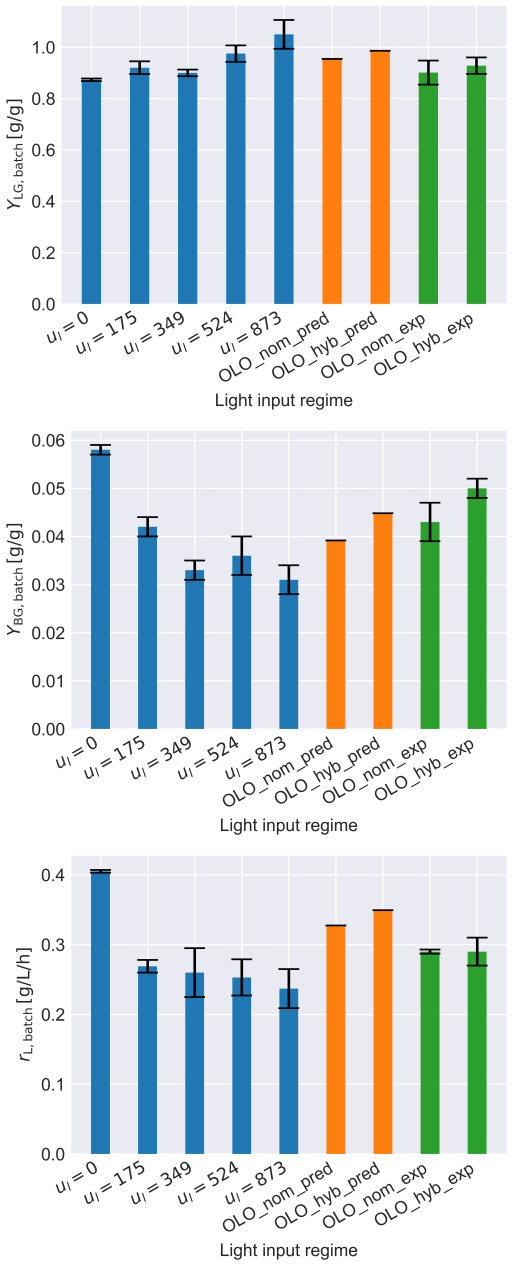


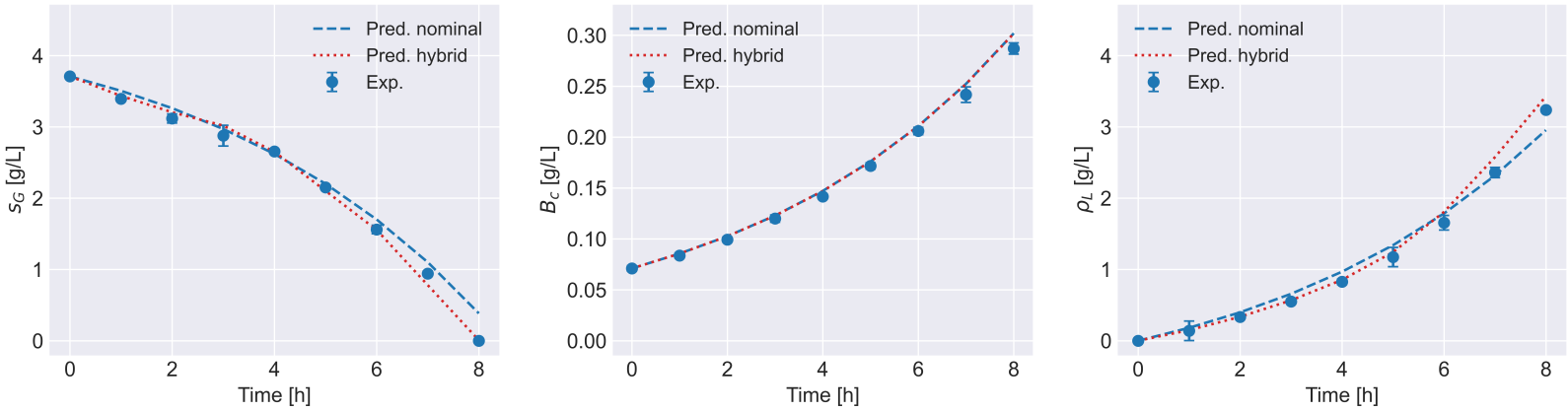
Figure 7.3: Average lactate-on-glucose yield ($Y_{\mathrm{LG,batch}}$, top), biomass-on-glucose yield ($Y_{\mathrm{BG,batch}}$, middle), and lactate volumetric productivity ($r_{\mathrm{L,batch}}$, bottom) calculated for: the modeling (wet-lab) experiments (blue bars); predicted open-loop optimization results using the nominal (OLO_nom_pred) and hybrid (OLO_hyb_pred) models (orange bars); and the experimental results for the open-loop optimizations using the nominal (OLO_nom_exp) and hybrid (OLO_hyb_exp) models (green bars). u_l is given in μmol m^{-2} s^{-1}.

parameters[8] were estimated based only on the knowledge-based component of the model. This

[8] Optimized parameters: $k_{\mathrm{BV}} = 2.605 \times 10^{-4}$ VU/g, $k_G = 5.340 \times 10^{-7}$ g/L, $k_{\mathrm{GV}} = 1.053 \times 10^{-6}$ VU/g, $k_{\mathrm{LV}} = 1.002 \times 10$ VU/g, $m_G = 1.232 \times 10^{-6}$ g/g/h, $m_L = 1.910$ g/g/h, $n_1 = 1.000 \times 10^{-2}$, $n_2 = 1.028 \times 10^{-1}$, $n_3 = 1.000 \times 10^{1}$, $q_{G_{\max}} = 1.731$ g/g/h, $Y_{\mathrm{BG}} = 1.083 \times 10^{-1}$ g/g, $Y_{\mathrm{LB}} = 2.204$ g/g, $q_{E_0} = 1.000 \times 10^{-6}$ VU/g/h, $q_{E_{\max}} = 1.000 \times 10$ VU/g/h, $n_4 = 4.718$, $k_u = 3.729 \times 10^{2}$ μmol/m^2/s, and $k_d = 0.988$ 1/h.

model, referred to as the *nominal* model throughout this chapter, exhibited a good fit to the experimental data for all the tested constant light inputs.

A) Batch with constant $u_l = 0\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$



B) Batch with constant $u_l = 175\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$

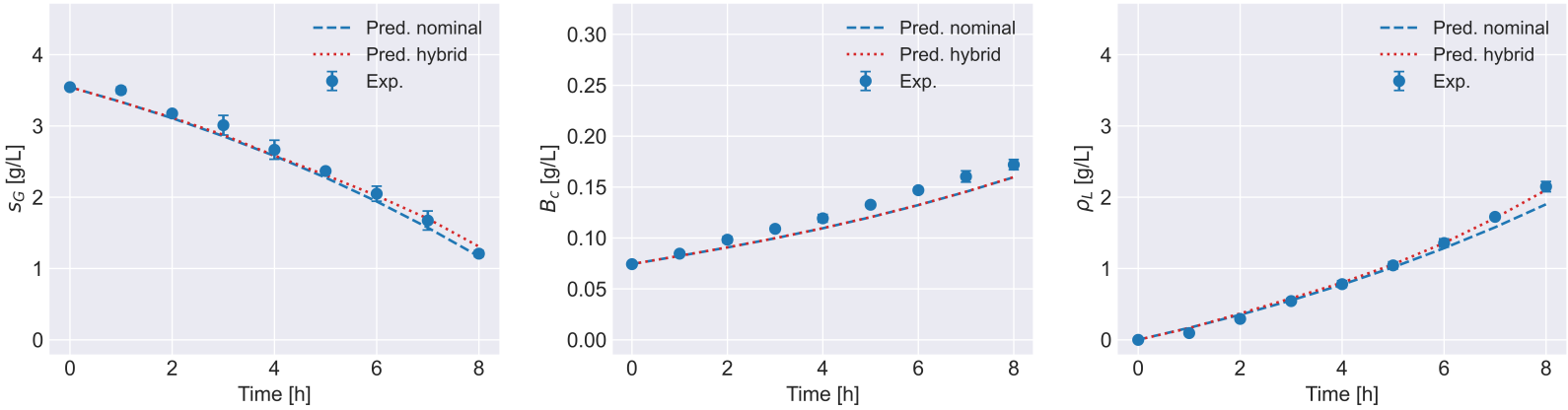
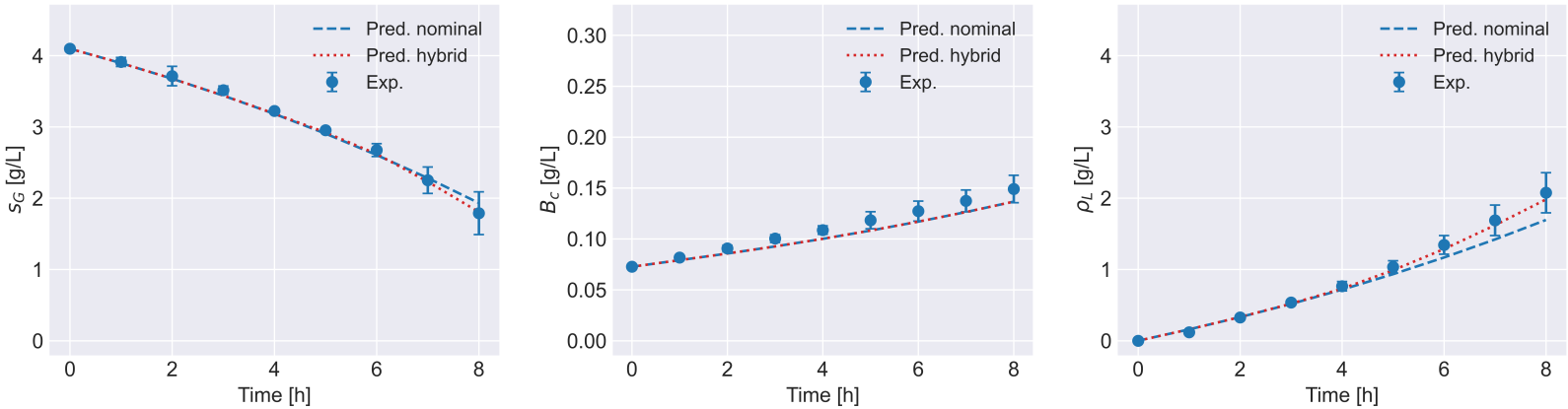


Figure 7.4: Model fitting of the nominal and hybrid models to experimental data under different constant light inputs.

C) Batch with constant $u_l = 349\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$



D) Batch with constant $u_l = 524\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$

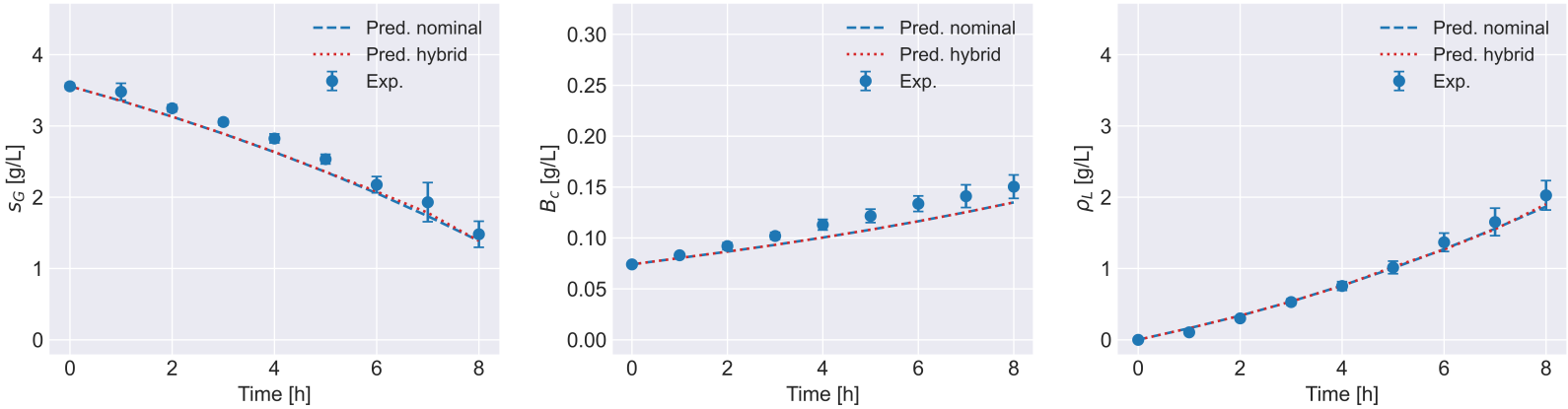


Figure continuation.

E) Batch with constant $u_l = 873\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$

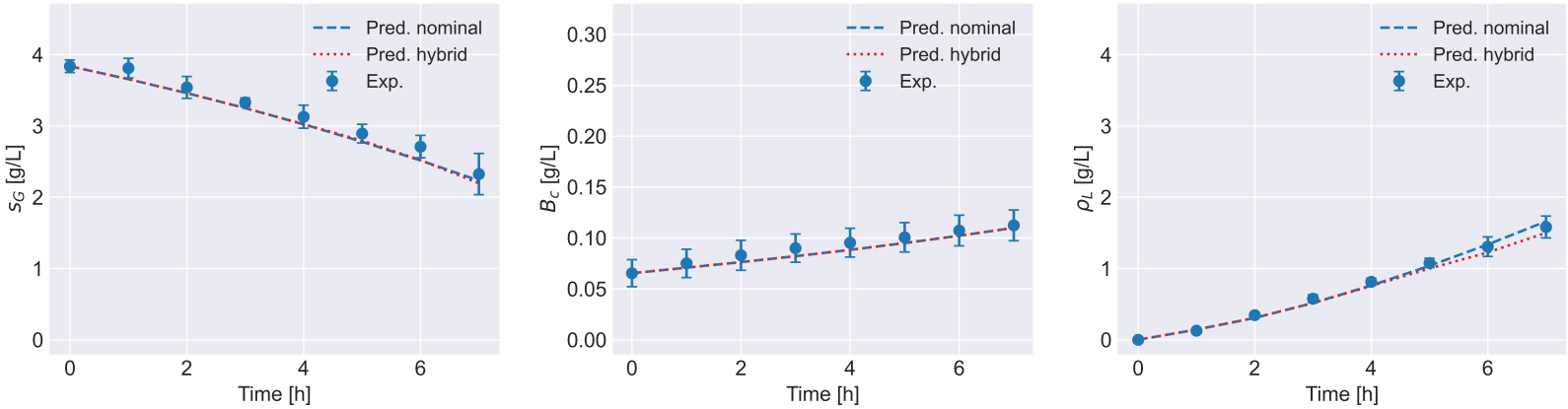


Figure continuation.

Despite the reasonable fitting achieved by the nominal model, Gaussian process regressors were implemented as a proof of concept to address the remaining model-plant mismatch. The goal was to learn the model error w_i in equations (7.10a)-(7.10c). The inputs for each Gaussian process regressor consisted of the dynamic input and the model states, while the output was the model-plant error. The model error w_i was estimated using the equation

$$w_i(t_k) = \frac{x_e(t_{k+1}) - x_e(t_k)}{t_{k+1} - t_k} - \frac{x_m(t_{k+1}) - x_m(t_k)}{t_{k+1} - t_k}, \tag{7.12}$$

where $x_e(t_k)$ and $x_e(t_{k+1})$ denote the experimental state values at sampling times t_k and t_{k+1}, respectively. Similarly, $x_m(t_k)$ and $x_m(t_{k+1})$ represent the predicted state values at times t_k and t_{k+1} from the knowledge-based part of the model. At each sampling time, $x_m(t_k) := x_e(t_k)$. Note that the model combining the knowledge-based part with the Gaussian process regressors is referred to as the *hybrid* model. The hybrid model provided a slightly better fit to the experimental data, particularly in the later stages of the fermentations (cf. Fig. 7.4).

The Python toolbox for machine-learning-supported optimal control, known as HILO-MPC [152], was utilized for training the Gaussian process regressors and solving the optimal control problems outlined in this chapter.

7.4.3 Open-loop optimization results

Optimal control problems as outlined in Section 7.3 were constrained by the derived nominal and hybrid models. The objective was to maximize the final lactate concentration within an 8 h time frame in a batch process with optogenetically regulated ATPase expression. The green light and the initial glucose concentration were considered as dynamic and static degrees of freedom, respectively. Piece-wise constant light inputs with a length of 1 h were employed for discretizing the system. The dynamic input was constrained to the range of values used for model fitting/training, as stated in Eq. (7.13c). It was also demanded that all glucose is to be depleted by the end of the fermentation, as stated in Eq. (7.13d). Additionally, the optimizer was constrained to achieve a user-defined batch lactate-on-glucose yield $\tilde{Y}_{\mathrm{LG,batch}}$, as stated in Eq. (7.13e). Lower and upper bounds for the initial glucose concentration were considered. These bounds were set to zero and $s_{G_{\max}} = 5$ g/L, respectively, as stated in Eq. (7.13f). The resulting optimization problem is formulated as

$$\begin{aligned}
\max_{u_l(\cdot),\, s_G(t_0)} \quad & \rho_L(t_f), && \text{(7.13a)}\\
\text{s.t.} \quad & \text{Eqs.(7.10a)} - \text{(7.10e)}, && \text{(7.13b)}\\
& 0 \leq u_l \leq 873, && \text{(7.13c)}\\
& s_G(t_f) = 0, && \text{(7.13d)}\\
& \frac{\rho_L(t_f) - \rho_L(t_0)}{s_G(t_0) - s_G(t_f)} = \tilde{Y}_{\mathrm{LG,batch}}, && \text{(7.13e)}\\
& 0 < s_G(t_0) \leq s_{G_{\max}}. && \text{(7.13f)}
\end{aligned}$$

The model error $\boldsymbol{w}$ in Eqs. (7.10a)-(7.10c) is neglected when employing the *nominal* model in the optimizations. However, if the *hybrid* model is utilized, the Gaussian-process-based model error $\boldsymbol{w}$ is incorporated into Eqs. (7.10a)-(7.10c). To demonstrate the *flexibility* of enforced ATP turnover for obtaining *user-defined* batch-to-batch yield improvements, two target values for $\tilde{Y}_{\mathrm{LG,batch}}$ were selected. For the optimization based on the nominal model, $\tilde{Y}_{\mathrm{LG,batch}}$ was set to 0.954 g/g, while for the optimization based on the hybrid model, it was set to 0.986 g/g. These values correspond to a 9 and 13 % increase in product yield, respectively, when compared to the scenario without ATPase induction (Fig. 7.3, $u_l = 0$).

Note that if the target $\tilde{Y}_{\mathrm{LG,batch}}$ is higher than what can be achieved by the cell without ATPase induction (cf. Fig. 7.3), it is expected that the optimizer will make use of the enforced ATP turnover mechanism to enhance the product yield. The results of the open-loop optimizations based on the nominal and hybrid models are shown in Fig. 7.5[9].

In the open-loop optimizations, the predicted light input followed a two-stage trajectory. The first phase involved no ATPase induction, i.e., with a light input of 0 $\mu\mathrm{mol\,m^{-2}\,s^{-1}}$, followed by a second phase with ATPase induction at the maximum light input of 873 $\mu\mathrm{mol\,m^{-2}\,s^{-1}}$. The key difference between the two scenarios is the switching time to the second phase, which was triggered after 3 hours for the optimization based on the nominal model and after 4 hours for the optimization based on the hybrid model.

It is worth noting that the predicted input, which followed an OFF-ON light trajectory, differs from the more gradual trajectories predicted when constraint-based dynamic models were considered (cf. Chapters 5-6). This discrepancy can be attributed to the fact that the aforementioned constraint-based dynamic models consider redox and energy aspects, as well as resource allocation phenomena, which are not accounted for in the current simplified modeling approach. Nonetheless, the bang-bang predicted inputs were deemed good *approximations* since they were in line with conventional two-phase fermentation approaches for optimizing bioprocesses [31, 38, 39].

The optimal initial glucose concentration determined by the optimizer based on the nominal model was 2.745 g/L, while it was 2.834 g/L for the optimizer based on the hybrid model. The predicted batch lactate-on-glucose yield for both fermentations met the demanded values (Fig. 7.3, orange bars). Compared to the scenario without ATPase induction (Fig. 7.3, $u_l = 0$), the increase in batch lactate-on-glucose yield was predicted to be at the expense of a 33 % and 23 % decrease in biomass-on-glucose yield, respectively. This corresponded to a predicted drop in volumetric productivity of 19 % and 14 %, respectively.

The experimental validation of the open-loop optimizations matched well the overall trends as predicted by the optimal control problems. However, there were some inconsistencies between the model predictions and the actual experimental results (cf. e.g. the lactate profile of the hybrid model-based optimization). Also, a slight discrepancy was observed between the target and experimental initial concentrations, which is understandable considering human error. This

[9] A direct comparison of these open-loop optimizations is, however, difficult due to the selection of different $\tilde{Y}_{\mathrm{LG,batch}}$ values for each scenario.

is more noticeable for the glucose initial concentration in the optimization based on the hybrid model.

The final lactate concentrations in the validation experiments were 2.324 ± 0.020 and $2.317 \pm 0.018\,\mathrm{g/L}$ for the optimizations based on the nominal and hybrid models, respectively. Similarly, the final lactate-on-glucose yields were 0.901 ± 0.047 and 0.928 ± 0.032 g/g, respectively. When comparing the latter yields to the scenario without ATPase induction ($u_l = 0$ constant), the *average* increase in lactate-on-glucose yield was 3 % and 6 %, respectively (cf. Fig. 7.3, green bars). These values were lower than the predicted/demanded $\tilde{Y}_{\mathrm{LG,batch}}$ targets. However, the reader should bear in mind that these were open-loop optimizations, i.e., without online corrective actions.

7.5 Summary

A model-based optimization strategy for open-loop optogenetic control of the ATPase to maximize production efficiency through enforced ATP turnover was proposed and experimentally implemented. The outlined approach involved a simplified modeling framework utilizing a quasi-unstructured/unsegregated kinetic model. Furthermore, the option of employing hybrid models combining knowledge-based and machine-learning components (such as Gaussian processes) to support model-based optimization was presented.

For experimental validation, *E. coli* was engineered to carry the CcaS/CcaR system to achieve optogenetic control of ATPase expression. By solving optimal control problems using knowledge-based and hybrid Gaussian-process-supported models, the concentration of lactate was maximized given predefined product yields while guaranteeing complete consumption of glucose. Since the optimization was implemented in an open-loop fashion, some model-plant mismatch was observed. To address potential uncertainties, feedback control schemes such as MPC coupled with state estimators, as discussed in Chapter 6, could be experimentally implemented in the future. The experimental implementation of MPC schemes was beyond the scope of this thesis.

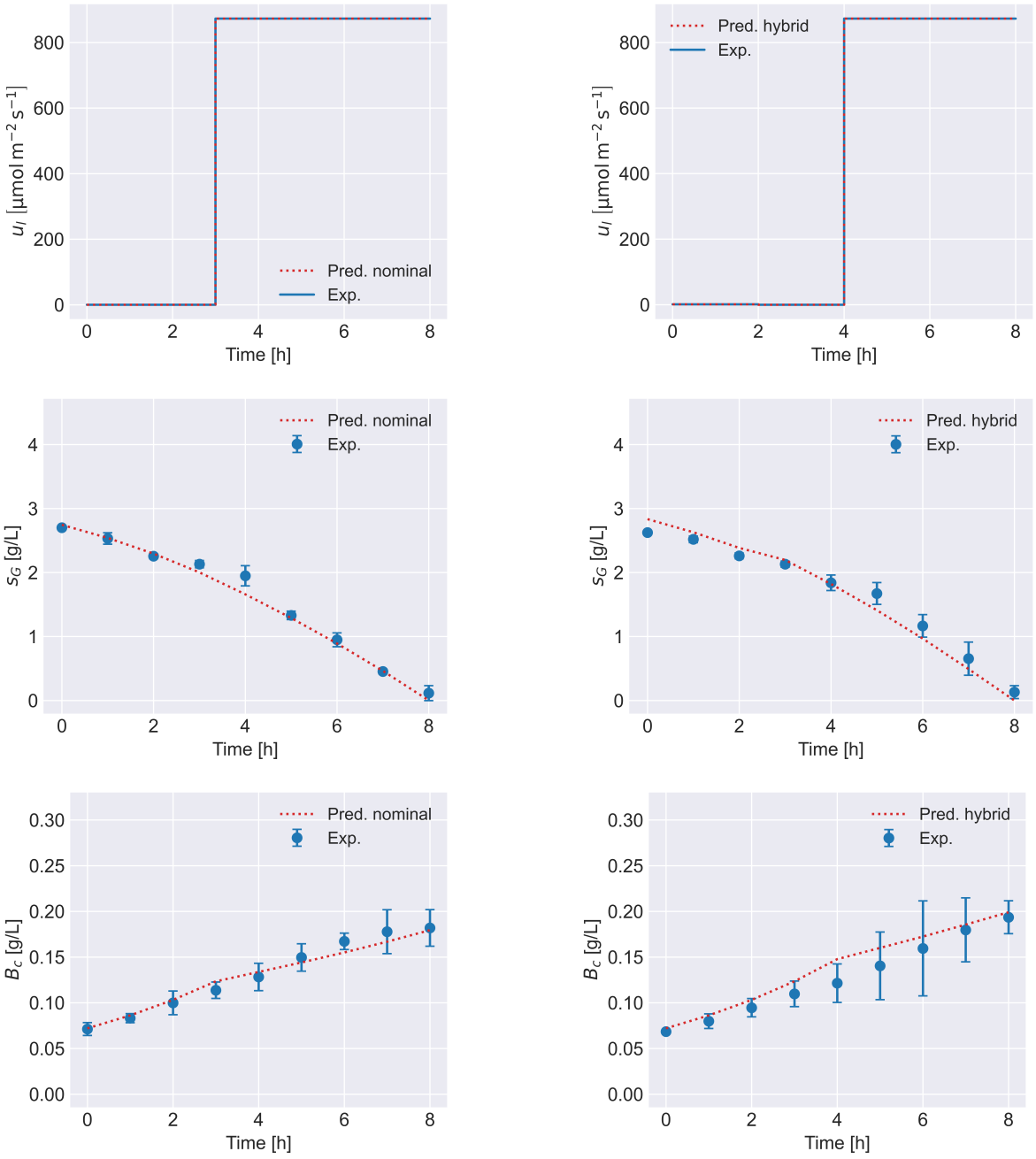


Figure 7.5: Comparison of open-loop optimization prediction and experimental implementation. The left side shows the optimization using the nominal model, while the right side shows the optimization using the hybrid model.

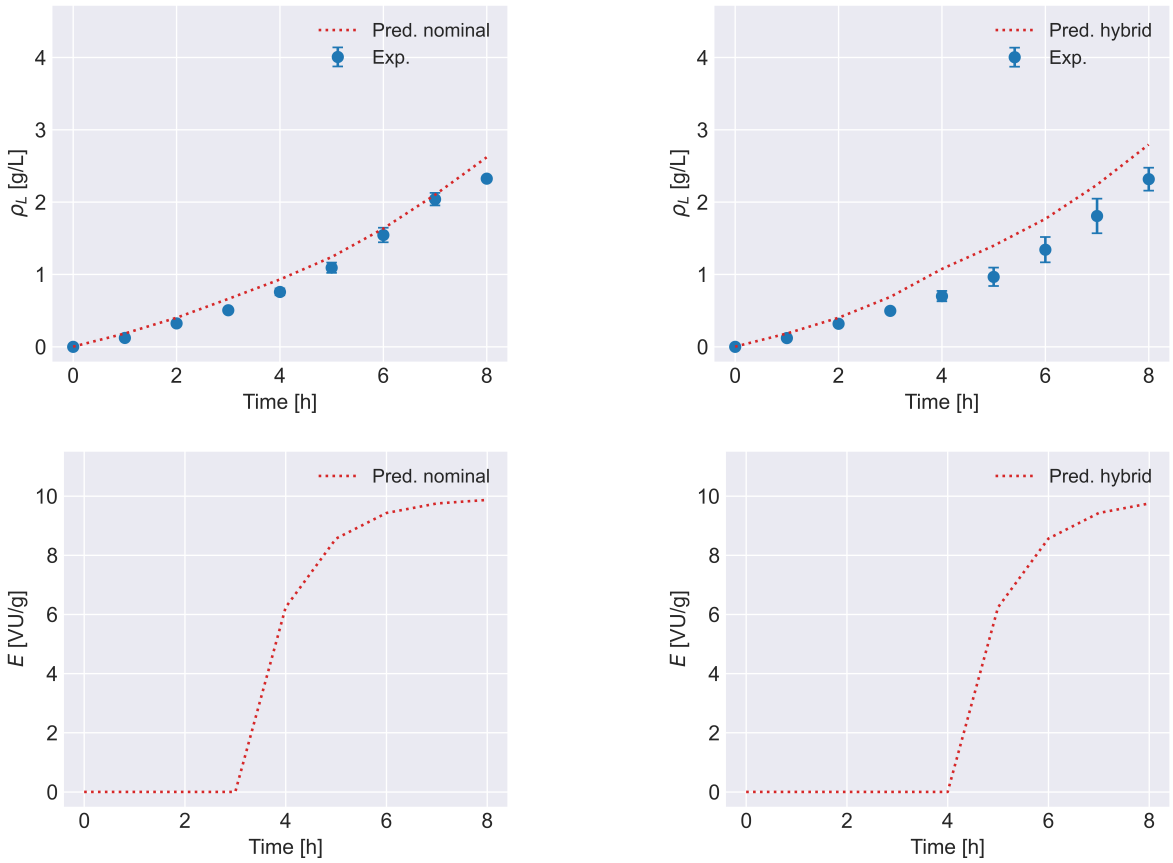


Figure continuation.

8 Conclusion and future perspectives

In this thesis, the concept of cybergenetics was considered toward enabling dynamic metabolic engineering, leading to the concept of *metabolic* cybergenetic systems. The main idea was to manipulate intracellular metabolic fluxes, a powerful degree of freedom for maximizing production in bioprocesses. This is enabled by modulating the gene expression of key metabolism-relevant enzymes via external inputs, such as light in the case of optogenetics. The presented metabolic cybergenetic framework followed a process systems engineering approach, integrating concepts such as synthetic biology, metabolic engineering, (machine-learning-supported) dynamic modeling, model-based optimization, predictive control, and estimation.

From the modeling side, two main approaches were considered: 1) constraint-based dynamic modeling, which integrates the dynamics of metabolic reactions, resource allocation, and externally inducible gene expression, and 2) simplified quasi-unstructured/unsegregated kinetic modeling, where only the most relevant external species are modeled, along with the gene expression of metabolism-relevant proteins triggered by external inputs. For the latter modeling strategy, Gaussian processes, a machine-learning method, was outlined as a way to compensate for potential model uncertainties arising from, e.g., oversimplifications and wrong model assumptions.

Model-based dynamic optimization was consistently employed throughout this thesis to obtain optimal system inputs, comprising both dynamic and static degrees of freedom. Dynamic inputs included *virtual* metabolic fluxes, light in the context of optogenetic modulation, and feed rates in the context of fed-batch fermentations. Static variables included initial conditions such as initial substrate concentrations. Applying model-based optimization in an open-loop manner does not take care of process uncertainties and disturbances, which often leads to suboptimal performance. Therefore, MPC was proposed to address uncertainties such as model-plant mismatch and disturbances, whereby the optimal control problem is resolved with the measured/estimated state of the cybergenetic system, thus incorporating state feedback.

Efficient and real-time process monitoring becomes essential to enable predictive control approaches. For example, the outlined constraint-based dynamic model captures the dynamics of the intracellular components (such as enzymes, ribosomes, and quota elements), thus monitoring biomass composition is required to reinitialize the MPC problem at each sampling instance. Even if one opts for the simplified quasi-unstructured/unsegregated kinetic modeling approach, the externally regulated metabolism-relevant proteins still need to be monitored. Therefore, a soft sensor based on full information estimation was proposed to infer the intracellular

components based on easier-to-measure variables such as biomass and extracellular metabolite concentrations. However, bear in mind that the efficiency of such soft sensors will depend on the states that can be measured, and whether they are sufficient for reconstructing the system.

To evaluate the potential of the proposed metabolic cybergenetic framework, dynamic enforced ATP turnover was selected as the core case study throughout this thesis. The idea was to exploit the concept of enforced ATP turnover for enhanced yield/productivity. Specifically, an engineering *E. coli* strain where the lactate-producing pathway is coupled to net ATP synthesis under anaerobic conditions was considered.

As a first step, enforced ATP turnover, linked to a *virtual* intracellular ATP-hydrolyzing flux was used as the manipulated variable. Optimal control problems using constraint-based dynamic modeling were formulated to find trade-offs between the enhancement of product yield and the decline in volumetric productivity in batch processes, resulting from the enforced ATP turnover mechanism. Predictive control simulations considering the virtual ATP-hydrolyzing flux were conducted, demonstrating the potential of MPC to address model-plant mismatch. An extended constraint-based dynamic model for anaerobic lactate fermentation was outlined, accounting for the dynamics of an optogenetic system regulating the expression of ATPase. The latter enzyme catalyzes the hydrolysis reaction of ATP into ADP, hence a direct mechanism for enforced ATP turnover. Open-loop optimizations using the extended constraint-based dynamic model were performed to maximize production with light as an external control input. Homogeneous light penetration in the bioreactor and batch operation was considered.

Motivated by the previous results, a *generalized* constraint-based dynamic modeling framework was presented to account for other potential dynamic metabolic engineering applications. The generalized model considered non-homogeneous light penetration (in the case of optogenetics) and accounted for fed-batch modes of operation. Model-based optimization and predictive control schemes were considered. A full information estimator was used as a soft sensor of the biomass components in the MPC simulations. Using the anaerobic lactate fermentation in fed-batch with optogenetic manipulation of the ATPase as an example of a metabolic cybergenetic system, the applicability of the modeling, optimization, predictive control, and estimation framework was demonstrated.

An *E. coli* strain was genetically engineered to enable *experimental* optogenetic modulation of the ATPase in anaerobic lactate fermentation. Furthermore, a quasi-unstructured/unsegregated kinetic model was derived and validated with experiments. Gaussian processes were considered to learn the uncertain parts of the model equations, although other machine-learning methods are also possible. Model-based optimizations were performed and experimentally validated. The experimental implementation only involved open-loop control, as a proof of concept. Future work could involve the experimental implementation of predictive control schemes, if necessary coupled with soft sensors. MPC could help to mitigate potential system uncertainties such as uncertain initial conditions, model-plant mismatch, and disturbances, that open-loop control cannot address.

Note that the initial biomass concentrations were not optimized in the optimization examples provided throughout this thesis. However, the inoculum concentration could be a valuable degree of freedom to explore in the future. For instance, in the context of dynamic ATP turnover, one could start the fermentation with a high cell density, thereby inducing ATPase expression at a high level right from the beginning of the process. This could potentially lead to an improvement in volumetric productivity.

The dynamic models presented in this thesis did not take into account the stability of gene expression systems. However, in reality, if the employed gene expression systems are unstable or burden the cell, the cells can mutate to inactivate/reverse/modify the engineered elements [159]. A recent example of modeling genetic stability in engineered cells is given by [160]. Therefore, evolutionary stability may be integrated into future metabolic cybergenetic models to improve long-term predictability, particularly relevant to continuous bioprocesses.

In the constraint-based dynamic modeling, metabolic fluxes were constrained only by the concentration of the catalytic enzymes and the corresponding catalytic constants. Furthermore, in the quasi-unstructured/unsegregated kinetic model, the rates were only a function of the regulated enzyme and limiting substrate concentrations. However, the reaction rates are in principle also a function of intracellular metabolites (cf. e.g. [108]), which the outlined constraint-based model assumes for simplicity to be in pseudo-steady-state conditions. In metabolic cybergenetics, one aims to dynamically manipulate metabolic fluxes, which may lead to changing dynamic profiles of intracellular metabolites. Therefore, in the future, one may consider incorporating the dynamics of key intracellular metabolites in the models to better describe the reaction rates. This would, however, increase the complexity of process monitoring. Recently, we proposed the use of moving horizon estimation based on kinetic modeling to reconstruct the dynamics of intracellular metabolites [161].

Another area of opportunity is the extension of the presented metabolic cybergenetic framework to synthetic microbial communities for biotechnological production. The advantage of microbial consortia is that complex metabolic pathways can be split into smaller sub-modules via *division of labor*, whereby each sub-module is catalyzed by a different microbial sub-population [162, 163]. With synthetic microbial communities, the challenge is how to optimally manipulate sub-population levels and avoid mono-culture formation. Recently, we addressed this challenge in [164], where optogenetics is considered to regulate sub-population growth via a toxin/antitoxin system. In simulations, model-based optimization and predictive control showed the potential to maximize production using microbial consortia.

One aspect that could hinder the applicability of the proposed framework is the absence of suitable real-time sensors, either hard or soft sensors, for bioprocess monitoring. This can severely limit closed-loop control implementations. Alternatively, in such instances, the feedback could be realized after the process is finished and the relevant dynamic states are measured offline in a batch-to-batch manner. This approach was recently explored in the context of synthetic microbial communities [165]. Machine-learning methods could be employed to learn the uncertain parts of the differential equations. Gaussian processes could serve this

purpose, with the additional advantage that they could allow the controller to be "aware" of areas of high uncertainty. Therefore, the objective function of the optimal control problem could incorporate an "exploitation" component, i.e., the economic cost function, and potentially also an "exploration" component based on the uncertainty of the Gaussian processes.

Throughout this thesis, the focus was on cell-based bioprocesses, where cells and their *machinery* catalyze substrate-to-product conversions. However, cell-based biotechnology has drawbacks. For instance, certain metabolites or products can be toxic, diminishing growth and production when toxic levels are reached. Additionally, intrinsic trade-offs exist in cell metabolism, such as biomass versus product synthesis [166, 167]. One solution to these issues is cell-free biosynthesis, where (*de novo*) metabolic pathways operate outside cells in more favorable conditions than those within the cell. This also minimizes competing pathways that interfere with product synthesis, potentially resulting in higher yields and productivity. Volumetric productivity in cell-free processes is influenced by enzyme concentrations, which can be exploited as an optimization degree of freedom [168]. In this context, a *synthetic* cell can feature externally regulated enzyme production machinery [169, 170], constructing synthetic pathways in real-time while dynamically adjusting enzyme concentrations if necessary. Thus, the cybergenetic methods outlined in this thesis can also be extended or adapted to cell-free production.

Personally, I believe that the outlined *metabolic* cybergenetic framework can contribute to the development of novel and advanced biotechnological applications as we can exploit the modulation of intracellular metabolic fluxes for optimizing bioprocesses. That is, we are now targeting the *heart* of the process. Moreover, the model-based feature of the presented framework, potentially augmented with machine-learning methods, can contribute to shortening and reducing the cost of process development and obtaining a more robust, consistent, and flexible operation. I look forward to what the future shall bring in the field of metabolic cybergenetics.

A Appendix

A.1 Resource balance analysis

The resource balance analysis method for estimating $\boldsymbol{p}(t_0) \in \mathbb{R}^{n_p}$ is formulated as [37]

$$\begin{aligned}
\max_{\mu, \boldsymbol{V}, \boldsymbol{p}} \quad & \mu && \text{(A.1a)}\\
\text{s.t.} \quad & \boldsymbol{S}_p \boldsymbol{V} - \mu \boldsymbol{p} = \boldsymbol{0}, && \text{(A.1b)}\\
& \tilde{B}(t_0) = \boldsymbol{b}^{\mathsf{T}} \boldsymbol{p}, && \text{(A.1c)}\\
& \boldsymbol{S_m} \boldsymbol{V} = \boldsymbol{0}, && \text{(A.1d)}\\
& \sum_{j \in \text{cat}_i} \left| \frac{V_j}{k_{\text{cat},j}} \right| \leq p_i, \quad \forall i \in [1, n_{p_i}] && \text{(A.1e)}\\
& \varphi_Q \boldsymbol{b}^{\mathsf{T}} \boldsymbol{p} \leq p_Q, && \text{(A.1f)}\\
& \boldsymbol{V}_{\min} \leq \boldsymbol{V} \leq \boldsymbol{V}_{\max}. && \text{(A.1g)}
\end{aligned}$$

This *static* optimization aims to maximize the growth rate $\mu \in \mathbb{R}$ while considering the initial dry weight measurement $\tilde{B}(t_0) \in \mathbb{R}$. It involves allocating $\boldsymbol{p} \in \mathbb{R}^{n_p}$ to achieve an optimal distribution of $\boldsymbol{V} \in \mathbb{R}^{n_V}$. $\boldsymbol{S}_p \in \mathbb{R}^{n_p \times n_V}$ is the stoichiometric matrix of $\boldsymbol{p}$. The rate of production of $\boldsymbol{p}$ is assumed to be equally diluted by cell growth.

A.2 Genetic engineering procedure

The strain KBM10111s (=MG1655 Δ*adhE* Δ*ackA-pta*) [102] was the starting point to create an *E. coli* strain that exhibits a light-inducible expression of *atpAGD* and produces lactate as the main fermentation product. Initially, pGB-MPI-035, a modified version of pSKA397 [60], along with pTNS3 [171] for Tn7-based insertion of *atpAGD* regulated by PcpcG2Δ59 [150], downstream of *glmS* [172], were recombined into KBM10111s. This yielded the strain sGB013 (=MG1655 Δ*adhE* Δ*ackA-pta* Tn7::*cat*-PcpcG2Δ59-*atpAGD*-rrnBT1). Subsequently, the chloramphenicol resistance cassette of sGB013 was discarded via FLP recombination with pCP20 [173]. This rendered the strain sGB014 (=MG1655 Δ*adhE* Δ*ackA-pta* Tn7::PcpcG2Δ59-*atpAGD*-rrnBT1).

The genes *ccaS*/*ccaR* for the optogenetic system, regulated through PccaR, were inserted in an *sfgfp*-deficient variant of pSKA413 [60] (pGB-MPI-23). Furthermore, the *Synechocystis* PCC6803 genes *ho1* and *pcyA*, which enable the conversion of heme into phycocyanobilin,

can be expressed from pPLPCB(S) [151]. Both plasmids were subsequently transferred into sGB014, resulting in the ultimate strain employed for all experimental procedures, denoted as *E. coli* sGB015 (=MG1655 Δ*adhE* Δ*ackA-pta* Tn7::PcpcG2Δ59-*atpAGD*-rrnBT1 pPLPCB(S) pGB-MPI-23).

A.3 Fermentation experiments and analytical measurements

Single colonies of *E. coli* sGB015 were used to inoculate 10 mL aerobic cultures using LB0 medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) at 37 °C in 100 mL shake flasks equipped with baffles, and agitated at 200 rpm. Subsequently, precultures were prepared using a standard defined medium [174] consisting of 4 g/L glucose, 34 mM NaH_2PO_4, 64 mM K_2HPO_4, 20 mM $(NH_4)_2SO_4$, 9.52 mM $NaHCO_3$, 1 μM $Fe(SO_4)_4$, 300 μM $MgSO_4$, 1 μM $ZnCl_2$, and 10 μM $CaCl_2$, supplemented with 150 μg/mL spectinomycin and 25 μg/mL chloramphenicol. These precultures were inoculated from the LB0 culture in 50 mL Schott flasks containing 25 mL culture volume and cultivated overnight under red light at 37 °C and 180 rpm.

The main fermentation experiments were initiated using fresh standard defined medium in 50 mL flasks with a culture volume of 50 mL and inoculated from the latter preculture to specified cell densities. The flasks were screwed tightly to prevent gas exchange. To take samples, the culture vessels were briefly transferred inside a Whitley A25 anaerobic workstation (Meintrup DWS Laborgeräte GmbH) providing an oxygen-free atmosphere composed of 80 % N_2, 10 % CO_2, and 10 % H_2.

Quantitative measurement of glucose was performed using the HK assay kit (Megazyme Ltd.). Lactate concentrations were determined through reversed-phase high-performance liquid chromatography using an Inertsil ODS-3 column (5 μm, RP-18 100A, 250 x 4.6 mm) (GL Sciences Inc.). The chromatographic system was operated at a flow rate of 1.0 mL/min, employing a running buffer of 0.1 M $NH_4H_2PO_4$ at pH 2.6, and a temperature of 40 °C. An injection volume of 10 μL was used, and detection was performed using an UV-DAD detector at 210 nm.

A.4 Online resources

The codes associated with this thesis can be found here: https://bitbucket.org/codes-to-share/phd_thesis/src/master/. Sequence files containing all the genetic elements employed in the experimental part of this thesis are accessible here: https://doi.org/10.17617/3.H5GT8I.

Bibliography

[1] Eurostat (2018). *Statistics Explained*, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Biotechnology (**accessed:** 2023/01/13).

[2] Timmis K. et al. (2017). The contribution of microbial biotechnology to sustainable development goals. *Microb Biotechnol*, **10**, 5, 984–987.

[3] Yang D., Cho J. S., Choi K. R., Kim H. U., Lee S. Y. (2017). Systems metabolic engineering as an enabling technology in accomplishing sustainable development goals. *Microb Biotechnol*, **10**, 5, 1254–1258.

[4] French K. E. (2019). Harnessing synthetic biology for sustainable development. *Nat Sustain*, **2**, 4, 250–252.

[5] Cho J. S., Kim G. B., Eun H., Moon C. W., Lee S. Y. (2022). Designing microbial cell factories for the production of chemicals. *JACS Au*, **2**, 8, 1781–1799.

[6] Dräger A., Planatscher H. (2013). Metabolic Networks, *Encyclopedia of Systems Biology*, 1249–1251, ISBN: 978-1-4419-9862-0 978-1-4419-9863-7.

[7] Meyer H.-P., Minas W., Schmidhalter D. (2016). Industrial-scale fermentation, *Industrial Biotechnology*, 1–53, ISBN: 978-3-527-80783-3 978-3-527-34181-8.

[8] Zhong J.-J. (2011). Bioreactor engineering, *Comprehensive Biotechnology*, 257–269, ISBN: 978-0-444-64047-5.

[9] Croughan M. S., Konstantinov K. B., Cooney C. (2015). The future of industrial bioprocessing: batch or continuous? *Biotechnol Bioeng*, **112**, 4, 648–651.

[10] Kopp J., Slouka C., Spadiut O., Herwig C. (2019). The rocky road from fed-batch to continuous processing with *E. coli*. *Front Bioeng Biotechnol*, **7**, 328.

[11] Kumar A., Udugama I. A., Gargalo C. L., Gernaey K. V. (2020). Why is batch processing still dominating the biologics landscape? Towards an integrated continuous bioprocessing alternative. *Processes*, **8**, 12, 1641.

[12] Kiss A. A., Grievink J., Rito-Palomares M. (2015). A systems engineering perspective on process integration in industrial biotechnology: process integration in industrial biotechnology. *J Chem Technol Biotechnol*, **90**, 3, 349–355.

[13] Woodley J. M. (2020). Towards the sustainable production of bulk-chemicals using biotechnology. *N Biotechnol*, **59**, 59–64.

[14] Sanford K., Chotani G., Danielson N., Zahn J. A. (2016). Scaling up of renewable chemicals. *Curr Opin Biotechnol*, **38**, 112–122.

[15] Carpine R., Olivieri G., Hellingwerf K. J., Pollio A., Marzocchella A. (2020). Industrial production of poly-β-hydroxybutyrate from CO_2: can cyanobacteria meet this challenge? *Processes*, **8**, 3, 323.

[16] Chandel A. K., Garlapati V. K., Jeevan Kumar S. P., Hans M., Singh A. K., Kumar S. (2020). The role of renewable chemicals and biofuels in building a bioeconomy. *Biofuel Bioprod Biorefin*, **14**, 4, 830–844.

[17] Nielsen J., Tillegreen C. B., Petranovic D. (2022). Innovation trends in industrial biotechnology. *Trends Biotechnol*, **40**, 10, 1160–1172.

[18] Custom Market Insights (2022). *Global microbial fermentation technology market 2022 – 2030*, https://www.custommarketinsights.com/report/microbial-fermentation-technology-market/ (**accessed:** 2023/01/16).

[19] Lynch M. D. (2021). The bioprocess TEA calculator: an online technoeconomic analysis tool to evaluate the commercial competitiveness of potential bioprocesses. *Metab Eng*, **65**, 42–51.

[20] Son H.-I., Weiss A., You L. (2021). Design patterns for engineering genetic stability. *Cur Opin Biomed Eng*, **19**, 100297.

[21] Olsson L., Rugbjerg P., Torello Pianale L., Trivellin C. (2022). Robustness: linking strain design to viable bioprocesses. *Trends Biotechnol*, **40**, 8, 918–931.

[22] Yang O., Prabhu S., Ierapetritou M. (2019). Comparison between batch and continuous monoclonal antibody production and economic analysis. *Ind Eng Chem Res*, **58**, 15, 5851–5863.

[23] Carlozzi P., Touloupakis E. (2021). Bioplastic production by feeding the marine *Rhodovulum sulfidophilum* DSM-1374 with four different carbon sources under batch, fed-batch and semi-continuous growth regimes. *N Biotechnol*, **62**, 10–17.

[24] Zhou X., Zhou X., Xu Y. (2017). Improvement of fermentation performance of *Gluconobacter oxydans* by combination of enhanced oxygen mass transfer in compressed-oxygen-supplied sealed system and cell-recycle technique. *Bioresour Technol*, **244**, 1137–1141.

[25] Liang X., Li C., Cao W., Cao W., Shen F., Wan Y. (2021). Fermentative production of fructo-oligosaccharides using *Aureobasidium pullulans*: effect of dissolved oxygen concentration and fermentation mode. *Molecules*, **26**, 13, 3867.

[26] Ham S., Yoon H., Park J.-M., Park Y. G. (2021). Optimization of fermentation medium for indole acetic acid production by *Pseudarthrobacter* sp. NIBRBAC000502770. *Appl Biochem Biotechnol*, **193**, 8, 2567–2579.

[27] Stylianou E., Pateraki C., Ladakis D., Vlysidis A., Koutinas A. (2021). Optimization of fermentation medium for succinic acid production using *Basfia succiniciproducens*. *Environ Technol*, **24**, 101914.

[28] Mozumder M. S. I., De Wever H., Volcke E. I., Garcia-Gonzalez L. (2014). A robust fed-batch feeding strategy independent of the carbon source for optimal polyhydroxybutyrate production. *Process Biochem*, **49**, 3, 365–373.

[29] Fei Q., O'Brien M., Nelson R., Chen X., Lowell A., Dowe N. (2016). Enhanced lipid production by *Rhodosporidium toruloides* using different fed-batch feeding strategies with lignocellulosic hydrolysate as the sole carbon source. *Biotechnol Biofuels*, **9**, 1, 130.

[30] Yuan S.-F., Alper H. S. (2019). Metabolic engineering of microbial cell factories for production of nutraceuticals. *Microb Cell Fact*, **18**, 1, 46.

[31] Lalwani M. A., Zhao E. M., Avalos J. L. (2018). Current and future modalities of dynamic control in metabolic engineering. *Curr Opin Biotechnol*, **52**, 56–65.

[32] Holtz W. J., Keasling J. D. (2010). Engineering static and dynamic control of synthetic pathways. *Cell*, **140**, 1, 19–23.

[33] Zhang F., Carothers J. M., Keasling J. D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat Biotechnol*, **30**, 4, 354–359.

[34] Brockman I. M., Prather K. L. (2015). Dynamic knockdown of *E. coli* central metabolism for redirecting fluxes of primary metabolites. *Metab Eng*, **28**, 104–113.

[35] Venayak N., Anesiadis N., Cluett W. R., Mahadevan R. (2015). Engineering metabolism through dynamic control. *Curr Opin Biotechnol*, **34**, 142–152.

[36] Hartline C. J., Schmitz A. C., Han Y., Zhang F. (2021). Dynamic control in metabolic engineering: theories, tools, and applications. *Metab Eng*, **63**, 126–140.

[37] Jabarivelisdeh B., Carius L., Findeisen R., Waldherr S. (2020). Adaptive predictive control of bioprocesses with constraint-based modeling and estimation. *Comput Chem Eng*, **135**, 106744.

[38] Burg J. M., Cooper C. B., Ye Z., Reed B. R., Moreb E. A., Lynch M. D. (2016). Large-scale bioprocess competitiveness: the potential of dynamic metabolic control in two-stage fermentations. *Curr Opin Chem Eng*, **14**, 121–136.

[39] Klamt S., Mahadevan R., Hädicke O. (2018). When do two-stage processes outperform one-stage processes? *Biotechnol J*, **13**, 2, 1700539.

[40] Boecker S., Harder B.-J., Kutscha R., Pflügl S., Klamt S. (2021). Increasing ATP turnover boosts productivity of 2,3-butanediol synthesis in *Escherichia coli*. *Microb Cell Factories*, **20**, 1, 63.

[41] Zhao E. M., Lalwani M. A., Chen J.-M., Orillac P., Toettcher J. E., Avalos J. L. (2021). Optogenetic amplification circuits for light-induced metabolic control. *ACS Synth Biol*, **10**, 5, 1143–1154.

[42] Liu W.-C., Inwood S., Gong T., Sharma A., Yu L.-Y., Zhu P. (2019). Fed-batch high-cell-density fermentation strategies for *Pichia pastoris* growth and production. *Crit Rev Biotechnol*, **39**, 2, 258–271.

[43] Khani M.-H., Bagheri M. (2020). Skimmed milk as an alternative for IPTG in induction of recombinant protein expression. *Protein Expr Purif*, **170**, 105593.

[44] Harder B.-J., Bettenbrock K., Klamt S. (2018). Temperature-dependent dynamic control of the TCA cycle increases volumetric productivity of itaconic acid production by *Escherichia coli*. *Biotechnol Bioeng*, **115**, 1, 156–164.

[45] Vogl T., Hartner F. S., Glieder A. (2013). New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. *Curr Opin Biotechnol*, **24**, 6, 1094–1101.

[46] Hoffman S. M., Tang A. Y., Avalos J. L. (2022). Optogenetics illuminates applications in microbial engineering. *Annu Rev Chem Biomol Eng*, **13**, 1, 373–403.

[47] Tandar S. T., Senoo S., Toya Y., Shimizu H. (2019). Optogenetic switch for controlling the central metabolic flux of *Escherichia coli*. *Metab Eng*, **55**, 68–75.

[48] Lalwani M. A. et al. (2021). Optogenetic control of the lac operon for bacterial chemical and protein production. *Nat Chem Biol*, **17**, 1, 71–79.

[49] Lalwani M. A., Kawabe H., Mays R. L., Hoffman S. M., Avalos J. L. (2021). Optogenetic control of microbial consortia populations for chemical production. *ACS Synth Biol*, **10**, 8, 2015–2029.

[50] Gutiérrez Mena J., Kumar S., Khammash M. (2022). Dynamic cybergenetic control of bacterial co-culture composition via optogenetic feedback. *Nat Commun*, **13**, 1, 4808.

[51] Carrasco-López C., García-Echauri S. A., Kichuk T., Avalos J. L. (2020). Optogenetics and biosensors set the stage for metabolic cybergenetics. *Curr Opin Biotechnol*, **65**, 296–309.

[52] Jabarivelisdeh B., Waldherr S. (2016). Improving bioprocess productivity using constraint-based models in a dynamic optimization scheme. *IFAC-PapersOnLine*, **49**, 26, 245–251.

[53] Jabarivelisdeh B., Findeisen R., Waldherr S. (2018). Model predictive control of a fed-batch bioreactor based on dynamic metabolic-genetic network models. *IFAC-PapersOnLine*, **51**, 19, 34–37.

[54] Ryu K. H., Kim B., Lee J. H. (2019). A model-based optimization of microalgal cultivation strategies for lipid production under photoautotrophic condition. *Comput Chem Eng*, **121**, 57–66.

[55] Nimmegeers P., Vallerio M., Telen D., Impe J., Logist F. (2018). Interactive multi-objective dynamic optimization of bioreactors under parametric uncertainty. *Chem Ing Tech*, cite.201800082.

[56] Rio-Chanona E. A. del, Ahmed N. R., Wagner J., Lu Y., Zhang D., Jing K. (2019). Comparison of physics-based and data-driven modelling techniques for dynamic optimisation of fed-batch bioprocesses. *Biotechnol Bioeng*, **116**, 11, 2971–2982.

[57] Findeisen R., Allgöwer F., "An introduction to nonlinear model predictive control", in *21st Benelux Meeting on Systems and Control*, 11, (2002), pp. 119–141.

[58] Khan O., Madhuranthakam C. M. R., Douglas P., Lau H., Sun J., Farrell P. (2018). Optimized PID controller for an industrial biological fermentation process. *J Process Control*, **71**, 75–89.

[59] Harcum S. W., Elliott K. S., Skelton B. A., Klaubert S. R., Dahodwala H., Lee K. H. (2022). PID controls: the forgotten bioprocess parameters. *Disc Chem Eng*, **2**, 1.

[60] Milias-Argeitis A., Rullan M., Aoki S. K., Buchmann P., Khammash M. (2016). Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nat Commun*, **7**, 1, 12546.

[61] Jabarivelisdeh B., Waldherr S. (2018). Optimization of bioprocess productivity based on metabolic-genetic network models with bilevel dynamic programming. *Biotechnol Bioeng*, **115**, 7, 1829–1841.

[62] Morabito B., Kienle A., Findeisen R., Carius L. (2019). Multi-mode model predictive control and estimation for uncertain biotechnological processes. *IFAC-PapersOnLine*, **52**, 1, 709–714.

[63] Morabito B., Pohlodek J., Matschek J., Savchenko A., Carius L., Findeisen R. (2021). Towards risk-aware machine learning supported model predictive control and open-loop optimization for repetitive processes. *IFAC-PapersOnLine*, **54**, 6, 321–328.

[64] Morabito B., Pohlodek J., Kranert L., Espinel-Ríos S., Findeisen R. (2022). Efficient and simple Gaussian process supported stochastic model predictive control for bioreactors using HILO-MPC. *IFAC-PapersOnLine*, **55**, 7, 922–927.

[65] Petsagkourakis P., Sandoval I. O., Bradford E., Zhang D., Del Rio-Chanona E. (2019). Reinforcement learning for batch-to-batch bioprocess optimisation, *Comput Aided Chem Eng*, **46**, 919–924, ISBN: 978-0-12-818634-3.

[66] Petsagkourakis P., Sandoval I., Bradford E., Zhang D., Del Rio-Chanona E. (2020). Reinforcement learning for batch bioprocess optimization. *Comput Chem Eng*, **133**, 106649.

[67] Treloar N. J., Fedorec A. J. H., Ingalls B., Barnes C. P. (2020). Deep reinforcement learning for the control of microbial co-cultures in bioreactors. *PLoS Comput Biol*, **16**, 4, e1007783.

[68] Vasilakou E. et al. (2016). Current state and challenges for dynamic metabolic modeling. *Curr Opin Microbiol*, **33**, 97–104.

[69] Zhang D., Del Rio-Chanona E. A., Petsagkourakis P., Wagner J. (2019). Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization. *Biotechnol Bioeng*, **116**, 11, 2919–2930.

[70] Peng M., Liang Z. (2020). Degeneration of industrial bacteria caused by genetic instability. *World J Microbiol Biotechnol*, **36**, 8, 119.

[71] Varma A., Palsson B. O. (1994). Metabolic flux balancing: basic concepts, scientific and practical use. *Nat Biotechnol*, **12**, 10, 994–998.

[72] Mahadevan R., Edwards J. S., Doyle F. J. (2002). Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J*, **83**, 3, 1331–1340.

[73] Mahadevan R., Schilling C. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng*, **5**, 4, 264–276.

[74] Machado D., Costa R. S., Ferreira E. C., Rocha I., Tidor B. (2012). Exploring the gap between dynamic and constraint-based models of metabolism. *Metab Eng*, **14**, 2, 112–119.

[75] Waldherr S., Oyarzún D. A., Bockmayr A. (2015). Dynamic optimization of metabolic networks coupled with gene expression. *J Theor Biol*, **365**, 469–485.

[76] Liu L., Bockmayr A. (2020). Regulatory dynamic enzyme-cost flux balance analysis: a unifying framework for constraint-based modeling. *J Theor Biol*, **501**, 110317.

[77] Kyriakopoulos S. et al. (2018). Kinetic modeling of mammalian cell culture bioprocessing: the quest to advance biomanufacturing. *Biotechnol J*, **13**, 3, 1700229.

[78] Sokolov M., Von Stosch M., Narayanan H., Feidl F., Butté A. (2021). Hybrid modeling — a key enabler towards realizing digital twins in biopharma? *Curr Opin Chem Eng*, **34**, 100715.

[79] Tsopanoglou A., Jiménez Del Val I. (2021). Moving towards an era of hybrid modelling: advantages and challenges of coupling mechanistic and data-driven models for upstream pharmaceutical bioprocesses. *Curr Opin Chem Eng*, **32**, 100691.

[80] Marques R. et al. (2017). Hybrid modeling of microbial exopolysaccharide (EPS) production: the case of *Enterobacter* A47. *J Biotechnol*, **246**, 61–70.

[81] Zhang D., Savage T. R., Cho B. A. (2020). Combining model structure identification and hybrid modelling for photo-production process predictive simulation and optimisation. *Biotechnol Bioeng*, **117**, 11, 3356–3367.

[82] Vega-Ramon F., Zhu X., Savage T. R., Petsagkourakis P., Jing K., Zhang D. (2021). Kinetic and hybrid modeling for yeast astaxanthin production under uncertainty. *Biotechnol Bioeng*, **118**, 12, 4854–4866.

[83] Rogers A. W., Song Z., Ramon F. V., Jing K., Zhang D. (2023). Investigating 'greyness' of hybrid model for bioprocess predictive modelling. *Biochem Eng J*, **190**, 108761.

[84] Fung Shek C., Betenbaugh M. (2021). Taking the pulse of bioprocesses: at-line and in-line monitoring of mammalian cell cultures. *Curr Opin Biotechnol*, **71**, 191–197.

[85] Reardon K. F. (2021). Practical monitoring technologies for cells and substrates in biomanufacturing. *Curr Opin Biotechnol*, **71**, 225–230.

[86] Reyes S. J., Durocher Y., Pham P. L., Henry O. (2022). Modern sensor tools and techniques for monitoring, controlling, and improving cell culture processes. *Processes*, **10**, 2, 189.

[87] Liu D., Evans T., Zhang F. (2015). Applications and advances of metabolite biosensors for metabolic engineering. *Metab Eng*, **31**, 35–43.

[88] Rogers J. K., Church G. M. (2016). Genetically encoded sensors enable real-time observation of metabolite production. *Proc Natl Acad Sci USA*, **113**, 9, 2388–2393.

[89] Mahr R., Frunzke J. (2016). Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl Microbiol Biotechnol*, **100**, 1, 79–90.

[90] Torello Pianale L., Rugbjerg P., Olsson L. (2022). Real-time monitoring of the yeast intracellular state during bioprocesses with a toolbox of biosensors. *Front Microbiol*, **12**, 802169.

[91] Li W. et al. (2018). Systematically integrated metabonomic-proteomic studies of *Escherichia coli* under ciprofloxacin stress. *J Proteomics*, **179**, 61–70.

[92] Pomraning K. R. et al. (2021). Integration of proteomics and metabolomics into the design, build, test, learn cycle to improve 3-hydroxypropionic acid production in *Aspergillus pseudoterreus*. *Front Bioeng Biotechnol*, **9**, 603832.

[93] Wang C.-Y., Lempp M., Farke N., Donati S., Glatter T., Link H. (2021). Metabolome and proteome analyses reveal transcriptional misregulation in glycolysis of engineered *E. coli*. *Nat Commun*, **12**, 1, 4929.
[94] Brunner V., Siegl M., Geier D., Becker T. (2021). Challenges in the development of soft sensors for bioprocesses: a critical review. *Front Bioeng Biotechnol*, **9**, 722202.
[95] Luo Y., Kurian V., Ogunnaike B. A. (2021). Bioprocess systems analysis, modeling, estimation, and control. *Curr Opin Chem Eng*, **33**, 100705.
[96] Rathore A. S., Nikita S., Jesubalan N. G. (2022). Digitization in bioprocessing: the role of soft sensors in monitoring and control of downstream processing for production of biotherapeutic products. *Biosens Bioelectron: X*, **12**, 100263.
[97] Del Vecchio D., Dy A. J., Qian Y. (2016). Control theory meets synthetic biology. *J R Soc Interface*, **13**, 120, 20160380.
[98] Hsiao V., Swaminathan A., Murray R. M. (2018). Control theory for synthetic biology: recent advances in system characterization, control design, and controller implementation for synthetic biology. *IEEE Control Syst*, **38**, 3, 32–62.
[99] Khammash M. H. (2022). Cybergenetics: theory and applications of genetic control systems. *Proc IEEE*, **110**, 5, 631–658.
[100] Man Z., Guo J., Zhang Y., Cai Z. (2020). Regulation of intracellular ATP supply and its application in industrial biotechnology. *Crit Rev Biotechnol*, **40**, 8, 1151–1162.
[101] Hädicke O., Klamt S. (2015). Manipulation of the ATP pool as a tool for metabolic engineering. *Biochem Soc Trans*, **43**, 6, 1140–1145.
[102] Hädicke O., Bettenbrock K., Klamt S. (2015). Enforced ATP futile cycling increases specific productivity and yield of anaerobic lactate production in *Escherichia coli*: ATP wasting to improve yield and productivity. *Biotechnol Bioeng*, **112**, 10, 2195–2199.
[103] Boecker S., Zahoor A., Schramm T., Link H., Klamt S. (2019). Broadening the scope of enforced ATP wasting as a tool for metabolic engineering in *Escherichia coli*. *Biotechnol J*, **14**, 9, 1800438.
[104] Zahoor A., Messerschmidt K., Boecker S., Klamt S. (2020). ATPase-based implementation of enforced ATP wasting in *Saccharomyces cerevisiae* for improved ethanol production. *Biotechnol Biofuels*, **13**, 1, 185.
[105] Liu J., Kandasamy V., Würtz A., Jensen P. R., Solem C. (2016). Stimulation of acetoin production in metabolically engineered *Lactococcus lactis* by increasing ATP demand. *Appl Microbiol Biotechnol*, **100**, 22, 9509–9517.
[106] Semkiv M. V., Dmytruk K. V., Abbas C. A., Sibirny A. A. (2016). Activation of futile cycles as an approach to increase ethanol yield during glucose fermentation in *Saccharomyces cerevisiae*. *Bioengineered*, **7**, 2, 106–111.
[107] Wichmann J., Behrendt G., Boecker S., Klamt S. (2023). Characterizing and utilizing oxygen-dependent promoters for efficient dynamic metabolic engineering. *Metab Eng*, **77**, 199–207.
[108] Boecker S. et al. (2021). Deciphering the physiological response of *Escherichia coli* under high ATP demand. *Mol Syst Biol*, **17**, 12, e10504.
[109] Kim J., Kim Y.-M., Lebaka V. R., Wee Y.-J. (2022). Lactic acid for green chemical industry: recent advances in and future prospects for production technology, recovery, and applications. *Fermentation*, **8**, 11, 609.
[110] Reimers A.-M., Understanding metabolic regulation and cellular resource allocation through optimization, Ph.D. dissertation, Freie Universität Berlin, Berlin, Germany, (2017).

[111] Lindhorst H., Reimers A.-M., Waldherr S. (2018). Dynamic modeling of enzyme controlled metabolic networks using a receding time horizon. *IFAC-PapersOnLine*, **51**, 18, 203–208.

[112] Lindhorst H., Lucia S., Findeisen R., Waldherr S. (2019). Modeling enzyme controlled metabolic networks in rapidly changing environments by robust optimization. *IEEE Control Syst Lett*, **3**, 2, 248–253.

[113] Koebmann B. J., Westerhoff H. V., Snoep J. L., Nilsson D., Jensen P. R. (2002). The glycolytic flux in *Escherichia coli* is controlled by the demand for ATP. *J Bacteriol*, **184**, 14, 3909–3916.

[114] Camacho E. F., Alba C. B. (2007). Model predictive control, ISBN: 978-1-85233-694-3.

[115] Rawlings J., Mayne D., Diehl M. (2020). Model predictive control: theory, computation and design, ISBN: 978-0-9759377-5-4.

[116] Reimers A.-M., Lindhorst H., Waldherr S. (2017). A protocol for generating and exchanging (genome-scale) metabolic resource allocation models. *Metabolites*, **7**, 3, 47.

[117] Erdrich P., Steuer R., Klamt S. (2015). An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC Syst Biol*, **9**, 1, 48.

[118] Hädicke O., Klamt S. (2017). EColiCore2: a reference network model of the central metabolism of *Escherichia coli* and relationships to its genome-scale parent model. *Sci Rep*, **7**, 1, 39647.

[119] Jeske L., Placzek S., Schomburg I., Chang A., Schomburg D. (2019). BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res*, **47**, D1, D542–D549.

[120] Wittig U. et al. (2012). SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Res*, **40**, D1, D790–D796.

[121] Milo R., Jorgensen P., Moran U., Weber G., Springer M. (2010). BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res*, **38**, suppl_1, D750–D753.

[122] The UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*, **47**, D1, D506–D515.

[123] Goelzer A., Fromion V., Scorletti G. (2011). Cell design in bacteria as a convex optimization problem. *Automatica*, **47**, 6, 1210–1218.

[124] Dempe S., Franke S. (2019). Solution of bilevel optimization problems using the KKT approach. *Optimization*, **68**, 8, 1471–1489.

[125] Dempe S. (2020). Bilevel optimization: theory, algorithms, applications and a bibliography, *Bilevel Optimization*, **161**, 581–672, ISBN: 978-3-030-52118-9 978-3-030-52119-6.

[126] Carøe H. B., Bilevel optimization with application in energy, Ph.D. dissertation, University of Copenhagen, Copenhagen, Denmark, (2018).

[127] Andersson J. A. E., Gillis J., Horn G., Rawlings J. B., Diehl M. (2019). CasADi: a software framework for nonlinear optimization and optimal control. *Math Program Comput*, **11**, 1, 1–36.

[128] Wächter A., Biegler L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program*, **106**, 1, 25–57.

[129] Poznyak A. S. (2008). Variational calculus and optimal control, *Advanced mathematical tools for automatic control engineers: deterministic techniques*, 647–711, ISBN: 978-0-08-044674-5.

[130] Olson E. J., Hartsough L. A., Landry B. P., Shroff R., Tabor J. J. (2014). Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat Methods*, **11**, 4, 449–455.

[131] Senoo S., Tandar S. T., Kitamura S., Toya Y., Shimizu H. (2019). Light-inducible flux control of triosephosphate isomerase on glycolysis in *Escherichia coli*. *Biotechnol Bioeng*, **116**, 12, 3292–3300.

[132] Benito B., Moreno E., Lagunas R. (1991). Half-life of the plasma membrane ATPase and its activating system in resting yeast cells. *Biochim Biophys Acta - Biomembr*, **1063**, 2, 265–268.

[133] Macauley-Patrick S., Finn B. (2008). Modes of fermenter operation, *Practical Fermentation Technology*, 69–95, ISBN: 978-0-470-72530-6 978-0-470-01434-9.

[134] Santos-Navarro F. N., Vignoni A., Boada Y., Picó J. (2021). RBS and promoter strengths determine the cell-growth-dependent protein mass fractions and their optimal synthesis rates. *ACS Synth Biol*, **10**, 12, 3290–3303.

[135] Doran P. M. (2013). Bioprocess engineering principles, ISBN: 978-0-12-220851-5.

[136] Liu S. (2020). Bioprocess engineering: kinetics, sustainability, and reactor design, ISBN: 978-0-12-821012-3.

[137] Liu Z., Zhang J., Jin J., Geng Z., Qi Q., Liang Q. (2018). Programming bacteria with light—sensors and applications in synthetic biology. *Front Microbiol*, **9**, 2692.

[138] Lindner F., Diepold A. (2021). Optogenetics in bacteria – applications and opportunities. *FEMS Microbiol Rev*, fuab055.

[139] Yang S. et al. (2019). Transcription and translation contribute to gene locus relocation to the nucleoid periphery in *E. coli*. *Nat Commun*, **10**, 1, 5131.

[140] Scull C. E., Dandpat S. S., Romero R. A., Walter N. G. (2021). Transcriptional riboswitches integrate timescales for bacterial gene expression control. *Front Mol Biosci*, **7**, 607158.

[141] Espinel-Ríos S., Morabito B., Bettenbrock K., Klamt S., Findeisen R. (2022). Soft sensor for monitoring dynamic changes in cell composition. *IFAC-PapersOnLine*, **55**, 23, 98–103.

[142] Elsheikh M., Hille R., Tatulea-Codrean A., Krämer S. (2021). A comparative review of multi-rate moving horizon estimation schemes for bioprocess applications. *Comput Chem Eng*, **146**, 107219.

[143] Chanquia S. N., Vernet G., Kara S. (2021). Photobioreactors for cultivation and synthesis: specifications, challenges, and perspectives. *Eng Life Sci*, elsc.202100070.

[144] Hofmann A., Simon A., Grkovic T., Jones M. (2014). Spectroscopic methods, *Methods of Molecular Analysis in the Life Sciences*, 10–94.

[145] Pfaffinger C. E., Schöne D., Trunz S., Löwe H., Weuster-Botz D. (2016). Model-based optimization of microalgae areal productivity in flat-plate gas-lift photobioreactors. *Algal Res*, **20**, 153–163.

[146] Koller A. P., Wolf L., Brück T., Weuster-Botz D. (2018). Studies on the scale-up of biomass production with *Scenedesmus* spp. in flat-plate gas-lift photobioreactors. *Bioprocess Biosyst Eng*, **41**, 2, 213–220.

[147] Blanken W., Postma P. R., Winter L. de, Wijffels R. H., Janssen M. (2016). Predicting microalgae growth. *Algal Res*, **14**, 28–38.

[148] Kim H., Ju J., Lee H. N., Chun H., Seong J. (2021). Genetically encoded biosensors based on fluorescent proteins. *Sensors*, **21**, 3, 795.

[149] Rasmussen C. E., Williams C. K. I. (2006). Gaussian processes for machine learning, ISBN: 978-0-262-18253-9.

[150] Schmidl S. R., Sheth R. U., Wu A., Tabor J. J. (2014). Refactoring and optimization of light-switchable *Escherichia coli* two-component systems. *ACS Synth Biol*, **3**, 11, 820–831.

[151] Tabor J. J., Levskaya A., Voigt C. A. (2011). Multichromatic control of gene expression in *Escherichia coli*. *J Mol Biol*, **405**, 2, 315–324.
[152] Pohlodek J., Morabito B., Schlauch C., Zometa P., Findeisen R. (2022). Flexible development and evaluation of machine-learning-supported optimal control and estimation methods via HILO-MPC, *arXiv:2203.13671*.
[153] Himmel A., Matschek J., Kok R., Morabito B., Nguyen H. H., Findeisen R. (2023). Machine learning for process control of (bio)chemical processes, *arXiv:2301.06073*.
[154] Heijnen J. J., Romein B. (1995). Derivation of kinetic equations for growth on single substrates based on general properties of a simple metabolic network. *Biotechnol Prog*, **11**, 6, 712–716.
[155] Pirt S. J. (1965). The maintenance energy of bacteria in growing cultures. *Proc Royal Soc B*, **163**, 991, 224–231.
[156] Luedeking R., Piret E. L. (1959). A kinetic study of the lactic acid fermentation. Batch process at controlled pH. *J Biochem Microbiol Technol Eng*, **1**, 4, 393–412.
[157] Hill A. V. (1910). The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *J Physiol*, **40**, iv–vii.
[158] Hoops S. et al. (2006). COPASI—a COmplex PAthway SImulator. *Bioinformatics*, **22**, 24, 3067–3074.
[159] Blazejewski T., Ho H.-I., Wang H. H. (2019). Synthetic sequence entanglement augments stability and containment of genetic information in cells. *Science*, **365**, 6453, 595–598.
[160] Ingram D., Stan G.-B. (2023). Modelling genetic stability in engineered cell populations. *Nat Commun*, **14**, 1, 3471.
[161] Espinel-Ríos S., Slaviero G., Bettenbrock K., Klamt S., Findeisen R. (2023). Monitoring intracellular metabolite concentrations by moving horizon estimation based on kinetic modeling. *IFAC-PapersOnLine*, **56**, 2, 4608–4613.
[162] Sgobba E., Wendisch V. F. (2020). Synthetic microbial consortia for small molecule production. *Curr Opin Biotechnol*, **62**, 72–79.
[163] Rafieenia R., Atkinson E., Ledesma-Amaro R. (2022). Division of labor for substrate utilization in natural and synthetic microbial communities. *Curr Opin Biotechnol*, **75**, 102706.
[164] Espinel-Ríos S., Bettenbrock K., Klamt S., Avalos J. L., Findeisen R. (2023). Machine learning-supported cybergenetic modeling, optimization and control for synthetic microbial communities, *Comput Aided Chem Eng*, **52**, 2601–2606, ISBN: 978-0-443-15274-0.
[165] Espinel-Ríos S., Kok R., Klamt S., Avalos J. L., Findeisen R., "Batch-to-batch optimization with model adaptation leveraging Gaussian processes: the case of optogenetically assisted microbial consortia", in *2023 23rd International Conference on Control, Automation and Systems (ICCAS)*, Yeosu, Korea, Republic of: IEEE, (2023), pp. 1292–1297.
[166] Tinafar A., Jaenes K., Pardee K. (2019). Synthetic biology goes cell-free. *BMC Biol*, **17**, 1, 64.
[167] Lim H. J., Kim D.-M. (2022). Cell-free synthesis of industrial chemicals and biofuels from carbon feedstocks. *Curr Opin Biotechnol*, **73**, 158–163.
[168] Espinel-Ríos S. et al. (2022). Cell-free biosynthesis meets dynamic optimization and control: a fed-batch framework. *IFAC-PapersOnLine*, **55**, 23, 92–97.
[169] Kai L., Schwille P. (2019). Cell-free protein synthesis and its perspectives for assembling cells from the bottom-up. *Adv Biosyst*, **3**, 6, 1800322.

[170] Khambhati K., Bhattacharjee G., Gohil N., Braddick D., Kulkarni V., Singh V. (2019). Exploring the potential of cell-free protein synthesis for extending the abilities of biological systems. *Front Bioeng Biotechnol*, **7**, 248.

[171] Choi K. H. et al. (2008). Genetic tools for select-agent-compliant manipulation of *Burkholderia pseudomallei*. *Appl Environ Microbiol*, **74**, 4, 1064–1075.

[172] Choi K.-H. et al. (2005). A Tn7-based broad-range bacterial cloning and expression system. *Nat Methods*, **2**, 6, 443–448.

[173] Cherepanov P. P., Wackernagel W. (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*, **158**, 1, 9–14.

[174] Tanaka S., Lerner S. A., Lin E. C. (1967). Replacement of a phosphoenolpyruvate dependent phosphotransferase by a nicotinamide adenine dinucleotide-linked dehydrogenase for the utilization of mannitol. *J Bacteriol*, **93**, 2, 642–648.