

Toward a modeling, optimization, and predictive control framework for fed-batch metabolic cybergenetics

Sebastián Espinel-Ríos¹  | Bruno Morabito² | Johannes Pohlodek³ |
Katja Bettenbrock¹ | Steffen Klamt¹  | Rolf Findeisen³ 

¹Analysis and Redesign of Biological Networks, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

²Yokogawa Insilico Biotechnology GmbH, Stuttgart, Germany

³Control and Cyber-Physical Systems Laboratory, Technical University of Darmstadt, Darmstadt, Germany

Correspondence

Rolf Findeisen, Control and Cyber-Physical Systems Laboratory, Technical University of Darmstadt, Darmstadt, Germany.

Email: rolf.findeisen@iat.tu-darmstadt.de

Funding information

European Regional Development Fund; International Max Planck Research Graduate School Proeng; International Max Planck Research School for Advanced Methods in Process and Systems Engineering

Abstract

Biotechnology offers many opportunities for the sustainable manufacturing of valuable products. The toolbox to optimize bioprocesses includes *extracellular* process elements such as the bioreactor design and mode of operation, medium formulation, culture conditions, feeding rates, and so on. However, these elements are frequently insufficient for achieving optimal process performance or precise product composition. One can use metabolic and genetic engineering methods for optimization at the intracellular level. Nevertheless, those are often of *static* nature, failing when applied to *dynamic* processes or if disturbances occur. Furthermore, many bioprocesses are optimized empirically and implemented with little-to-no feedback control to counteract disturbances. The concept of cybergenetics has opened new possibilities to optimize bioprocesses by enabling online modulation of the gene expression of metabolism-relevant proteins via external inputs (e.g., light intensity in optogenetics). Here, we fuse cybergenetics with model-based optimization and predictive control for optimizing dynamic bioprocesses. To do so, we propose to use dynamic constraint-based models that integrate the dynamics of metabolic reactions, resource allocation, and inducible gene expression. We formulate a model-based optimal control problem to find the optimal process inputs. Furthermore, we propose using model predictive control to address uncertainties via online feedback. We focus on fed-batch processes, where the substrate feeding rate is an additional optimization variable. As a simulation example, we show the optogenetic control of the ATPase enzyme complex for dynamic modulation of enforced ATP wasting to adjust product yield and productivity.

KEYWORDS

constraint-based modeling, dynamic metabolic control, metabolic cybergenetics, model predictive control, optogenetics, state estimation

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Biotechnology and Bioengineering* published by Wiley Periodicals LLC.

1 | INTRODUCTION

The demand for sustainable biotechnological products has grown significantly in recent years (Hughes & Jones, 2020; Wohlgemuth et al., 2021). Although several bioprocesses are commercially successful (Julleson et al., 2015; Sanford et al., 2016), many are still discarded at early stages because they are not as competitive as traditional technologies. A natural question that arises is how bioprocesses' efficiency can be optimized.

The toolbox of bioprocess optimization at the *extracellular* or macro-level includes selection of the bioreactor mode of operation, bioreactor design, optimization of cultivation conditions (pH, temperature, etc.), formulation of culture media, determination of optimal feeding profiles and initial concentrations, among others (cf. e.g., Azimi et al., 2019; Behera et al., 2019; Vandermies & Fickers, 2019). These optimization strategies can influence the overall cell metabolism. Still they *alone* tend to fail at targeting specific metabolic elements, such as key metabolic fluxes, without affecting other cell functionalities.

Dynamic model-based optimization and predictive control strategies can be used to exploit the dynamic potential of bioprocesses. Dynamic optimization allows finding the optimal dynamic operation conditions, for example, in del Rio-Chanona et al. (2019), Jabarivelisdeh et al. (2018), Jabarivelisdeh and Waldherr (2016), Nimmegeers et al. (2018), Ryu et al. (2019). Feedback control schemes, especially predictive control approaches as in Jabarivelisdeh et al. (2020), Jabarivelisdeh and Waldherr (2018), Morabito et al. (2019, 2021, 2022), allow one to counteract unknown disturbances such as changes in feed conditions or nonmodeled dynamics, while maximizing the production efficiency and rendering a consistent process performance.

At the *intracellular* or micro-level, the bioprocess optimization toolbox includes metabolic and genetic engineering methods for rewiring metabolic pathways. Classical *static* metabolic engineering aims at increasing the cell's product yield, often at the expense of lower biomass yield as the substrate flux diverges from biomass-producing reactions to the product-of-interest pathway. This inevitably decreases the volumetric productivity rates in batch-type bioreactors (Lalwani et al., 2018; Venayak et al., 2015). Furthermore, designing dynamic processes based on static metabolic control principles, usually derived under steady-state assumptions, can lead to metabolic imbalances (Cui et al., 2021).

Inducible expression of metabolism-relevant proteins via external inputs has emerged as a promising *dynamic* degree of freedom for bioprocess optimization at the micro-level (Hartline et al., 2021; Lalwani et al., 2018; Shen et al., 2019). Of increasing popularity is the application of optogenetics, the use of light to modulate gene expression (Hoffman et al., 2022). With optogenetics, one can switch on/off fluxes along metabolic pathways via modulation of enzyme expression (Lalwani, Ip, et al., 2021; Tandar et al., 2019; Zhao et al., 2021). One can also directly influence cell growth via modulation of the expression of (anti)toxin proteins (Lalwani, Kawabe, et al., 2021). In the latter optogenetic examples, the *optimal*

light input values were determined using factorial experiments and similar heuristic approaches, resulting in, for example, two-stage or three-stage fermentations. Furthermore, the inputs were applied in an open-loop fashion, that is, without online feedback or corrective actions. Considering the often present stochasticity of gene expression (De Vrieze et al., 2020) and the possible presence of process disturbances and batch-to-batch variability, bioprocesses operated in this manner may portray poor reproducibility, moderate-to-poor product quality and a higher risk of failure.

Motivated by these challenges, some authors have proposed *cybergenetic* schemes whereby computer-aided feedback control is used to compensate for uncertainties. In such cases, the corrective actions are calculated outside the cell, for example, by a computer-aided controller (Hsiao et al., 2018; Khammash, 2022). To the best of our knowledge, the *biotechnological* applications of cybergenetics have been so far limited to, for instance, controlling the expression of fluorescence proteins and growth-regulatory proteins (e.g., enzymes involved in essential amino acid synthesis or antibiotic-resistance conferring proteins) via optogenetics (GutiérrezMena et al., 2022; Miliás-Argeitis et al., 2016).

It is believed that the next step in this direction, bearing considerable potential, is to implement *metabolic* cybergenetic systems, that is, emphasizing dynamic metabolic engineering applications (Carrasco-López et al., 2020). Therefore, we seek to extend the scope of cybergenetics to scenarios where metabolic fluxes are to be dynamically manipulated (e.g., toward maximizing the volumetric productivity, achieving a target product yield, rendering a given ratio of products, etc.), while being able to compensate for disturbances and process changes. We aim to use model-based optimization and predictive control methods to exploit the full potential of metabolic cybergenetic systems, considering both cybergenetic inputs and traditional process inputs such as feeding rates simultaneously.

A reasonably good model capable of relating inducible gene expression to changes in the metabolic flux distribution and potential resource burden thus becomes fundamental for advancing in our quest. Unfortunately, so far only very simple models have been used in the context of cybergenetics, often based on phenomenological relations (cf. e.g., Gutiérrez Mena et al., 2022; Lovelett et al., 2021; Miliás-Argeitis et al., 2016). In our opinion, these models do not allow capturing all the important phenomena required for model-based control of *metabolic* cybergenetic systems.

Thus, as the core contribution of this work, we propose a modeling framework for metabolic cybergenetics, which is combined with model-based optimization and predictive control to dynamically modulate intracellular metabolic fluxes for bioprocess optimization. Without loss of generality, we focus on fed-batch processes due to their advantages compared to pure batch setups. Fed-batch processes include a concentrated feed that supplies fresh medium to the bioreactor, thus extending the production phase. This provides additional dynamic inputs (feed rates) to the system, and allows for higher productivity and more concentrated product streams. It furthermore provides an efficient way to handle processes with substrate inhibition (Doran, 2013; Liu, 2020). The proposed fed-batch

metabolic cybergenetic platform comprises four major components (see Figure 1): (1) a cybergenetic input capable of inducing gene expression dynamically, (2) a manipulatable substrate feeding stream, (3) online (bio)sensors and state estimators that monitor and estimate the state of the process, (4) and model-based optimization that operates in a closed-loop and fully automated fashion.

The remainder of the paper is structured as follows. In Section 2 we outline a dynamic constraint-based cybergenetic modeling approach that integrates metabolism, resource allocation, and inducible gene expression. The derived model is used to support model-based optimization, feedback control, and state estimation of metabolic cybergenetics (Sections 3 and 4). In Section 5, we evaluate our framework considering the optogenetic modulation of the ATPase F₁-subunit¹ in the anaerobic lactate fermentation by *Escherichia coli* for improved yield and productivity. We consider a *fed-batch* regime with *nonhomogeneous* light penetration, applying both open-loop and feedback control. Note that in a previous study, we covered only open-loop optimization in classic batch processes with homogeneous light penetration (Espinel-Ríos, Morabito, Pohlodek, et al., 2022).

2 | MODELING FOR DYNAMIC OPTIMIZATION AND CONTROL OF FED-BATCH METABOLIC CYBERGENETICS

We use an extended constraint-based modeling approach for capturing the combined dynamics of metabolism, resource allocation, and inducible gene expression. Constraint-based models are usually underdetermined (Gottstein et al., 2016; Klamt & von Kamp, 2022). Therefore, they are often formulated as optimization problems with biologically sound objective functions and subject to constraints.

Constraint-based models generally require a smaller number of parameters (e.g., flux bounds) and can be structurally simpler compared to kinetic models (Saa & Nielsen, 2017; Yasemi & Jolicoeur, 2021).

Without loss of generality, we consider that cells are composed of metabolic enzymes, ribosomes, and quota elements.² These biomass components are contained in the molar vector $\mathbf{p} \in \mathbb{R}^{n_p}$ such that the amount of biomass $B \in \mathbb{R}$ can be computed as

$$B(t) = v_L(t)(\mathbf{b}^T \mathbf{p}(t)), \quad (1)$$

where t is the time, $v_L \in \mathbb{R}$ is the bioreactor volume and $\mathbf{b} \in \mathbb{R}^{n_p}$ is a vector of the corresponding molecular weights of \mathbf{p} . Therefore, $\mathbf{b}^T \mathbf{p}(t)$ is a scalar quantity corresponding to the biomass concentration in g/L. Remark that we use bold fonts for vectors and matrices, and nonbold fonts for scalar variables and parameters.

We collect in \mathbf{p} both the concentrations of regulated proteins $\mathbf{p}_{\text{reg}} \in \mathbb{R}^{n_{\text{reg}}}$ and the concentrations of the remaining unregulated biomass components $\mathbf{p}_{\text{unr}} \in \mathbb{R}^{n_{\text{unr}}}$, such that $\mathbf{p}(t) = [\mathbf{p}_{\text{reg}}(t)^T, \mathbf{p}_{\text{unr}}(t)^T]^T$. In this text, the term “regulated” refers to the fact that the protein expression is under cybergenetic control, externally modulated via a suitable genetic system such as a light-inducible gene expression system (Lindner & Diepold, 2021; Liu et al., 2018). Note that the regulated proteins can comprise enzymes directly involved in metabolic pathways, which is the main focus of this paper, but can in principle also include (anti)toxin proteins or antibiotic-resistance conferring proteins for growth modulation. The “unregulated” biomass components typically contain, for example, the metabolic enzymes that are *not* under cybergenetic control, ribosomes, and quota elements. Hereafter, we will omit writing the dependency of the variables with respect to time when clear from the context.

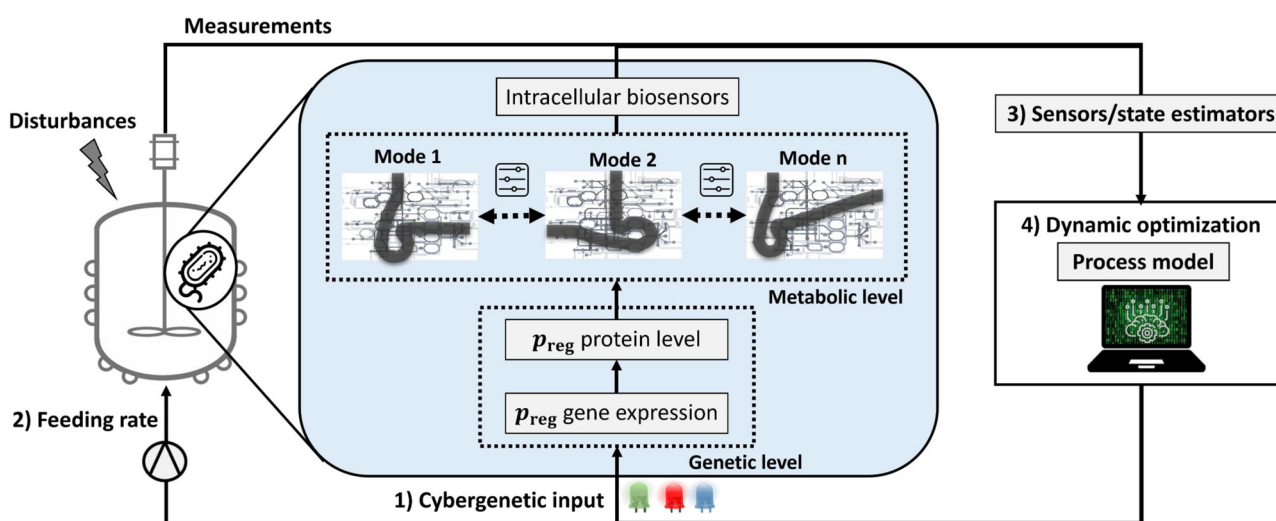


FIGURE 1 Overview of metabolic cybergenetics in fed-batch regime. Key metabolism-relevant proteins such as enzymes $\mathbf{p}_{\text{reg}} \in \mathbb{R}^{n_{\text{reg}}}$ are under the regulation of inducible gene expression systems to enable different metabolic modes over time using external inputs (e.g., light intensity). Model-based optimization finds the optimal inputs to the plant. The process outcome can be monitored via (bio)sensors and state estimators. Repeated solution of the optimization leads to feedback control.

With regard to the tunable gene expression systems, we differentiate between the inputs manipulated by the controller $\mathbf{u}_s \in \mathbb{R}^{n_u}$ (e.g., light intensity) and the values perceived by the cells inside the bioreactor $\mathbf{u}_c \in \mathbb{R}^{n_u}$. The distinction is necessary, as depending on the characteristics of the input and bioreactor, $\mathbf{u}_s = \mathbf{u}_c$ might not hold. This is especially relevant in large-scale setups where conditions tend to be less homogeneous, or where the input values received by the cells might depend on the cell density, for example, due to turbidity.

In Section 5, we will derive \mathbf{u}_c to account for light penetration in the context of optogenetics. For the sake of generality, we consider that the input perceived by the cells is given by a function $f_u: \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_{\theta_u}} \rightarrow \mathbb{R}^{n_u}$ which maps the input at the source \mathbf{u}_s to an average input $\bar{\mathbf{u}}_c$. Hence,

$$\bar{\mathbf{u}}_c = f_u(\mathbf{u}_s, \mathbf{x}, \boldsymbol{\theta}_u), \quad (2)$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ can include in principle all the model states, $\boldsymbol{\theta}_u \in \mathbb{R}^{n_{\theta_u}}$ comprises possible parameters of $f_u(\cdot)$ and $\bar{\mathbf{u}}_c$ is the average value of \mathbf{u}_c in the bioreactor following well-mixed conditions. Introducing $\bar{\mathbf{u}}_c$ simplifies the model as it is limited to changes in time and not in space, for example, while still accounting for average input gradients.

We describe the resulting change in the amount of regulated protein as,

$$\frac{d(v_L p_{\text{reg}})}{dt} = F_{\text{reg}}(\mathbf{B}, \bar{\mathbf{u}}_c) - D_{\text{reg}}(p_{\text{reg}}), \quad (3)$$

where $F_{\text{reg}}: \mathbb{R} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_{p_{\text{reg}}}}$ and $D_{\text{reg}}: \mathbb{R}^{n_{p_{\text{reg}}}} \rightarrow \mathbb{R}^{n_{p_{\text{reg}}}}$ are functions representing protein production and degradation, respectively, in units of mole per time.

Cells possess transcription factors that can switch between active and inactive states at a rate dictated by a specific signal. When active transcription factors bind the promoter region of a regulated gene, they can activate or repress the transcription process. Ribosomes catalyze the translation of the messenger ribonucleic acid, resulting from the transcription process, into proteins. In bacteria such as *E. coli*, transcription and translation are highly coupled, meaning that translation occurs at the same time as active transcription (Scully et al., 2021; Yang et al., 2019). Therefore, we propose to combine these two processes into lumped dose-response functions $\boldsymbol{\eta}: \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_{p_{\text{reg}}}}$, hence,

$$F_{\text{reg}}(\mathbf{B}, \bar{\mathbf{u}}_c) = \mathbf{B}\boldsymbol{\eta}(\bar{\mathbf{u}}_c). \quad (4)$$

For the degradation of regulated proteins, we consider both the effect of cell dilution due to growth and intrinsic protein turnover. The latter is captured by D_{reg} . On the other hand, the dilution of the regulated proteins due to cell growth is implicitly considered in our modeling framework. That is, as cellular components are modeled separately, the relative concentration of regulated proteins per biomass dry weight is reduced by the production of the remaining biomass components (such as unregulated enzymes, ribosomes, and quota elements).

We connect the dynamics of the regulated proteins to the overall metabolism and cell resource allocation via dynamic enzyme-cost flux balance analysis (deFBA), a constraint-based metabolic framework that considers resource allocation constraints (Jabarivelisdeh et al., 2020; Jabarivelisdeh & Waldherr, 2018; Waldherr et al., 2015).

The amount of extracellular metabolites is modeled as,

$$\frac{d(v_L \mathbf{z})}{dt} = F_{\text{in}} \mathbf{z}_{\text{in}} + v_L (\mathbf{S}_z \mathbf{V}) - \mathbf{D}_z(\mathbf{z}), \quad (5)$$

where $\mathbf{z} \in \mathbb{R}^{n_z}$ is the molar vector of extracellular metabolites, $\mathbf{S}_z \in \mathbb{R}^{n_z \times n_V}$ is the stoichiometric matrix of \mathbf{z} and $\mathbf{V} \in \mathbb{R}^{n_V}$ includes the fluxes for transport, metabolic and biomass-producing reactions in molar amount per time. $\mathbf{D}_z: \mathbb{R}^{n_z} \rightarrow \mathbb{R}^{n_z}$ captures the degradation of \mathbf{z} , $F_{\text{in}} \in \mathbb{R}$ is the feeding rate and $\mathbf{z}_{\text{in}} \in \mathbb{R}^{n_z}$ comprises the feed concentrations of \mathbf{z} . In the model, we assume that the feed contains only substrates.

The cell needs to invest resources to manufacture its components. Thus, changing the expression of a regulated protein is expected to influence the production rate of other biomass components and the resulting metabolic flux distribution since resources are limited and shared within the cell. Note that including biomass-producing reactions in the network is a way to capture the resource cost because we explicitly consider the required stoichiometric precursor and energy equivalents for the synthesis of all biomass components. With this in mind, the amount of unregulated biomass components follows

$$\frac{d(v_L \mathbf{p}_{\text{unr}})}{dt} = v_L (\mathbf{S}_{\text{p}_{\text{unr}}} \mathbf{V}) - \mathbf{D}_{\text{unr}}(\mathbf{p}_{\text{unr}}), \quad (6)$$

where $\mathbf{S}_{\text{p}_{\text{unr}}} \in \mathbb{R}^{n_{\text{p}_{\text{unr}}} \times n_V}$ is the stoichiometric matrix of \mathbf{p}_{unr} and $\mathbf{D}_{\text{unr}}: \mathbb{R}^{n_{\text{p}_{\text{unr}}}} \rightarrow \mathbb{R}^{n_{\text{p}_{\text{unr}}}}$ describes the degradation of \mathbf{p}_{unr} . For the regulated proteins, we add the following constraint

$$\mathbf{V}_{\text{p}_{\text{reg}}} - \frac{d\mathbf{p}_{\text{reg}}}{dt} = \mathbf{0}, \quad (7)$$

where $\mathbf{V}_{\text{p}_{\text{reg}}} \in \mathbb{R}^{n_{p_{\text{reg}}}}$ contains the corresponding regulated protein-producing reaction fluxes.

We consider quasi-steady-state dynamics for the amount of intracellular metabolites

$$\frac{d(v_L \mathbf{m})}{dt} = v_L (\mathbf{S}_m \mathbf{V}) - \mathbf{D}_m(\mathbf{m}) = \mathbf{0}, \quad (8)$$

where $\mathbf{m} \in \mathbb{R}^{n_m}$ is the molar vector of intracellular metabolites, $\mathbf{S}_m \in \mathbb{R}^{n_m \times n_V}$ is the stoichiometric matrix of \mathbf{m} , and $\mathbf{D}_m: \mathbb{R}^{n_m} \rightarrow \mathbb{R}^{n_m}$ describes the degradation of \mathbf{m} .

The metabolic fluxes of reactions catalyzed by enzymes in \mathbf{p}_{unr} are constrained by the corresponding catalytic enzyme concentration and catalytic constant ($\mathbf{k}_{\text{cat}} \in \mathbb{R}^{n_{\text{cat}}}$)

$$\sum_{j \in \text{cat}_{\text{unr}}} \left| \frac{V_j}{k_{\text{unr},j}} \right| \leq p_{\text{unr},j}, \quad \forall i \in [1, n_{\text{p}_{\text{unr}}}], \quad (9)$$

where $\text{cat}_{\text{unr}}(i)$ are all the reactions catalyzed by an enzyme p_{unr_i} and $|\cdot|$ refers to the absolute value operator.

The metabolic fluxes of reactions catalyzed by enzymes in p_{reg} are constrained by

$$\sum_{j \in \text{cat}_{\text{reg}}(i)} \left| \frac{V_j}{k_{\text{reg},j}} \right| = p_{\text{reg}_i}, \forall i \in [1, n_{p_{\text{reg}}}], \quad (10)$$

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

We consider enzyme saturation conditions. Thus, Equation (9) takes the product of the enzyme concentrations and the catalytic constants as upper bounds for the metabolic fluxes. While Equation (9) is an inequality constraint, we use an equality constraint in Equation (10) under the assumption that we have control over the fluxes catalyzed by p_{reg_i} . In other words, we *shift* this degree of freedom from the cell to an external controller.

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

$$\varphi_Q (\mathbf{b}^T \mathbf{p}) \leq p_Q, p_Q \in p_{\text{unr}}. \quad (11)$$

The bioreactor liquid volume changes over time due to the substrate feeding rate F_{in}

$$\frac{dV_L}{dt} = F_{\text{in}}. \quad (12)$$

Metabolic fluxes are constrained by biologically feasible lower and upper bounds

$$V_{\text{min}} \leq V \leq V_{\text{max}}. \quad (13)$$

Similarly, we consider feasible bounds for the dynamic states

$$p_{\text{min}} \leq \mathbf{p} \leq p_{\text{max}}, \mathbf{z}_{\text{min}} \leq \mathbf{z} \leq \mathbf{z}_{\text{max}}, v_{L_{\text{min}}} \leq v_L \leq v_{L_{\text{max}}}. \quad (14)$$

The conditions of the system at the initial process time t_0 are

$$\mathbf{p}(t_0) = \mathbf{p}_0, \mathbf{z}(t_0) = \mathbf{z}_0, v_L(t_0) = v_{L0}. \quad (15)$$

Summarizing, we express the resulting dynamic constraint-based model for fed-batch metabolic cybergenetic systems in terms of the following general dynamic optimization problem³

$$\max_{V(\cdot)} \int_{t_0}^{t_0 + \Delta t_{\text{bio}}} F_V(\cdot) dt, \quad (16a)$$

$$\text{s.t. Equations (1)–(15),} \quad (16b)$$

where $F_V(\cdot)$ is the objective function that the cell *optimizes*, usually one assumes it is the maximization of cell growth, and $V(\cdot)$ is a function of the resulting metabolic flux distribution. Solving this dynamic optimization problem allows one to simulate and predict the cell's behavior, as demonstrated in Section 5. We will use this

dynamic constraint-based model as a basis for optimizing and controlling the process.

3 | OPTIMAL CONTROL FOR METABOLIC CYBERGENETICS

Based on the derived dynamic constraint-based model, we aim to find the optimal input trajectories to drive the cell metabolism toward maximizing a desired performance criterion described by a cost function $J(\cdot)$. Let us collect all the process inputs (manipulated variables) in \mathbf{u}_p , that is, $\mathbf{u}_p := [\mathbf{u}_s(\cdot)^T, F_{\text{in}}(\cdot)]^T$, and all the model parameters in the vector θ . Recall that \mathbf{x} contains all the dynamic states, that is, $\mathbf{x} := [\mathbf{p}_{\text{reg}}^T, \mathbf{z}^T, \mathbf{p}_{\text{unr}}^T, v_L]^T$. To find the optimal inputs to the plant, we formulate an optimal control problem

$$\max_{\mathbf{u}_p} J(\cdot), \quad (17a)$$

$$\text{s.t. (16),} \quad (17b)$$

$$\mathbf{0} \leq \mathbf{g}(\mathbf{x}, \mathbf{u}_p, \theta). \quad (17c)$$

Solving (17) is a bilevel optimal control problem as the dynamic constraint-based model in (16) involves an optimization on its own. Equation (17c) captures additional state and input constraints.

$J(\cdot)$ can be defined in several ways based on specific goals. One may want to maximize production, maintain a desired set-point, and follow a reference trajectory, among other possibilities. Equation (17c) can include, for example, physical-, safety-, or economic-related process constraints. If the process is run in batch mode, then $\mathbf{u}_s(\cdot)$ can be set as the only optimization degree of freedom. The optimal control problem in (17) is an open-loop optimization, as only the initial conditions of the process states are used to compute an optimal input trajectory which is then applied to the plant without feedback. Doing so, would not allow reacting to unknown disturbances or model-plant mismatch.

4 | MODEL PREDICTIVE CONTROL FOR METABOLIC CYBERGENETICS

As we consider fed-batch processes, we use shrinking horizon model predictive control (MPC) to mitigate the effects of process uncertainty such as model-plant mismatch and disturbances (Findeisen & Allgöwer, 2002; Rawlings et al., 2020), that is, to mitigate the challenges of open-loop control.

4.1 | Shrinking horizon model predictive control

In MPC, the optimal control problem is evaluated repetitively at given sampling times. At these sampling instances, the states of the system

are measured or estimated with an observer. This introduces feedback since the information on the current system states is passed to the controller and corrective control actions can be taken. Let t_k be the sampling times at which measurements are taken. Without loss of generality, we assume that state measurements are available at equidistant sampling times, that is, $t_k := kh_s$ where $k \in \mathbb{N}_0$ and h_s is a fixed sampling interval. Furthermore, we assume that the controller predicts up to the final time $t_f := Nh_s$, where $N \in \mathbb{N}$ is the number of steps in the horizon. Therefore, the prediction horizon *shrinks* at every sampling time. The shrinking horizon MPC at time t_k reads

$$\max_{u_p} J(\cdot), \quad (18a)$$

$$\text{s.t. } \max_{V(\cdot)} \int_{t_k}^{t_k + \Delta t_{\text{bio}}} F_V(\cdot) dt, \quad (18b)$$

$$\text{s.t. } \text{Eqs. (1)–(14)}, \quad (18c)$$

$$\mathbf{x}(t_k) = \bar{\mathbf{x}}_k, \quad (18d)$$

$$\text{Equation (17c)}, \quad (18e)$$

where $t \in [t_k, t_f]$ and $\bar{\mathbf{x}}_k$ indicates the measured value of \mathbf{x} .

It is well known that introducing feedback increases the robustness of the controlled system even though the uncertainties are not explicitly taken into account in the controller (Findeisen & Allgöwer, 2005; Yu et al., 2014). More advanced robust MPC approaches (Mayne, 2014), such as stochastic MPC (Heirung et al., 2018; Mesbah et al., 2014), can take explicitly the system uncertainty into account. For simplicity, in this paper, we do not elaborate further on these approaches.

We assume that the culture volume can be monitored straightforwardly based on the applied feeding rate, and there is a range of online sensors available for the extracellular metabolite concentrations (Fung Shek & Betenbaugh, 2021; Reardon, 2021; Reyes et al., 2022). Therefore, monitoring v_L and \mathbf{z} is technically possible with the present technologies. However, typically there are no commercial sensors for the complete intracellular biomass composition. To circumvent this challenge, some state estimators have been proposed for *reconstructing* the biomass composition (Espinell-Ríos, Morabito, Bettenbrock, et al., 2022; Jabarivelisdeh et al., 2020). In the next section, we briefly describe the use of a full information estimator—an optimization-based estimator that considers the process dynamics and process constraints, as well as past and current measurements. For more details, we refer the reader to Espinell-Ríos, Morabito, Bettenbrock, et al. (2022).

4.2 | Reconstructing unmeasured cell components

Let $(\cdot)_i$ be a general optimization variable calculated at time t_i . We collect the dynamic Equations (3), (5), (6), and (12) in the vector

function $\mathbf{f}(\mathbf{x}, \mathbf{u}_p, \boldsymbol{\theta})$. As with the MPC, we assume for simplicity of presentation equidistant sampling times for the estimator, although nonequidistant sampling times are also possible. At time t_k , a full information estimator can be formulated by solving the following optimization problem

$$\min_{\mathbf{x}_0, \boldsymbol{\theta}, \boldsymbol{\omega}} \left\| \begin{bmatrix} \mathbf{x}_0 \\ \boldsymbol{\theta} \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{x}}_0 \\ \hat{\boldsymbol{\theta}} \end{bmatrix} \right\|_{\mathbf{P}}^2 + \sum_{i=0}^k \|\mathbf{y}(t_i) - \hat{\mathbf{y}}_i\|_{\mathbf{R}}^2 + \|\mathbf{w}_i\|_{\mathbf{Q}}^2, \quad (19a)$$

$$\text{s.t. } \max_{V(\cdot)} \int_{t_i}^{t_i + \Delta t_{\text{bio}}} F_V(\cdot) dt, \quad (19b)$$

$$\text{s.t. } \mathbf{x}(t_i + h_s) = \mathbf{x}(t_i) + \int_{t_i}^{t_i + h_s} \mathbf{f}(\mathbf{x}, \mathbf{u}_p, \boldsymbol{\theta}) dt + \mathbf{w}_i, \quad (19c)$$

$$\mathbf{0} \leq \mathbf{c}(\mathbf{x}(t_i), \mathbf{u}_p(t_i), \boldsymbol{\theta}), \quad (19d)$$

$$\mathbf{y}(t_i) = \mathbf{h}(\mathbf{x}(t_i), \mathbf{u}_p(t_i), \boldsymbol{\theta}), \quad (19e)$$

$$\mathbf{0} \leq \mathbf{g}(\mathbf{x}(t_i), \mathbf{u}_p(t_i), \boldsymbol{\theta}), \quad (19f)$$

where $\|\mathbf{a}\|_{\mathbf{A}}^2 := \mathbf{a}^T \mathbf{A} \mathbf{a}$ and $i \in [0, \dots, k]$. \mathbf{P} , \mathbf{R} , and \mathbf{Q} are weighting matrices of appropriate dimensions, $\mathbf{c} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_{u+1}} \times \mathbb{R}^{n_{\theta}} \rightarrow \mathbb{R}^{n_c}$ are the model constraints and $\mathbf{h} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_{u+1}} \times \mathbb{R}^{n_{\theta}} \rightarrow \mathbb{R}^{n_y}$ are the measurement equations. The optimization variables are the initial condition \mathbf{x}_0 , the parameter $\boldsymbol{\theta}$ and the state noise $\mathbf{w} := [\mathbf{w}_0^T, \dots, \mathbf{w}_k^T]^T$. We indicate with $(\cdot)^*$ the solution of the full information estimation problem and with $(\cdot)^\circ$ the *prior information* of a variable. With \mathbf{x}_0^* , \mathbf{w}^* , and $\boldsymbol{\theta}^*$ we reconstruct the states at t_k which can be used in the MPC. The full information estimator considers *all* the measurements; instead, if only the measurements in a given time window are used, one refers to a moving horizon estimator (Elsheikh et al., 2021; Rawlings et al., 2020).

It is worth noting that other state estimation methods, such as Kalman filters, have been proposed in the literature for inferring unmeasured states (Elsheikh et al., 2021; Haseltine & Rawlings, 2005; Tuveri et al., 2021, 2022). Kalman filters, however, are most effective with unconstrained systems. Consequently, they might not be well-suited for dynamic *constraint*-based models like the one outlined in this work. In contrast, our optimization-based soft sensor naturally accommodates constraints on states and inputs. Furthermore, Kalman filters have a memory of just one-time step, while our soft sensor considers a *trajectory* of states and inputs, which theoretically offers improved state estimation performance and robustness.

5 | EXAMPLE: OPTOGENETIC CONTROL OF ATPase IN ANAEROBIC LACTATE FERMENTATION BY *E. coli*

We consider the anaerobic lactate fermentation by *E. coli* using glucose as substrate, with optogenetic control of the ATPase enzyme complex, c.f. Figure 2. We only have one regulated protein, hence,

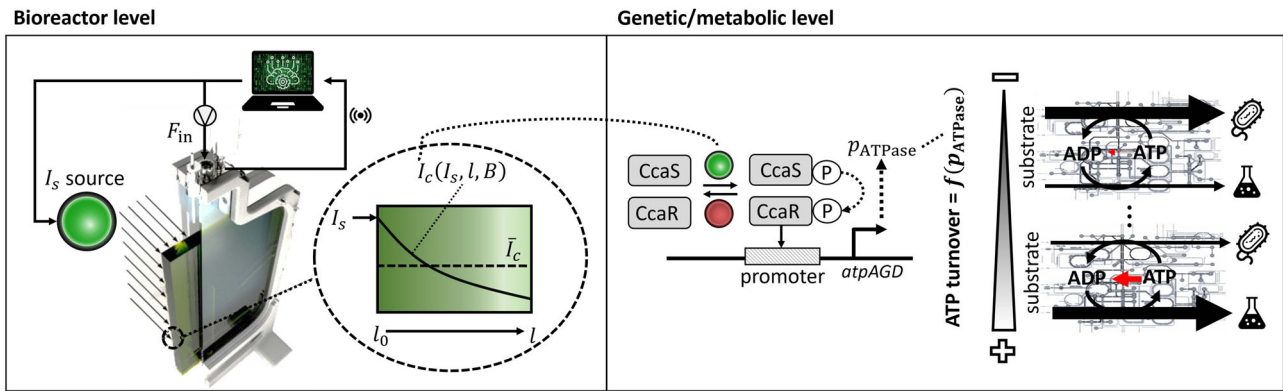


FIGURE 2 Overview of the considered example using the Ccas/CcaR optogenetic system for modulating the expression of the ATPase. The left side shows the considered flat-panel (photo)bioreactor and the metabolic cybergenetic control scheme, including the average light intensity inside the culture. On the right side, we show the effect of different expression levels of ATPase on the cell's metabolism. Picture of the bioreactor adapted from Pfaffinger et al. (2016).

$p_{reg} := p_{ATPase}$. The latter enzyme is responsible for catalyzing the hydrolysis reaction of ATP into ADP. We focus on the *E. coli* KBM10111 strain, engineered with gene deletions of *adhE* (aldehyde-alcohol dehydrogenase), *ackA* (acetate kinase), and *pta* (phosphate acetyltransferase) (Hädicke et al., 2015). Under these conditions, lactate synthesis from pyruvate is required to balance the redox cofactors generated during glycolysis. Since glycolysis renders net ATP gain, lactate production is linked to ATP synthesis (see Figure 3). In such cases, where the product pathway is linked to net ATP formation, it has been shown that an enforced ATP turnover or wasting can lead to an increase in the substrate uptake and the metabolic flux through the ATP-producing pathway as a way to counterbalance the ATP loss (cf. e.g., Boecker et al., 2021; Hädicke et al., 2015; Zahoor et al., 2020). Dynamic manipulation of the ATPase expression, and thereby the ATPase flux, can thus be exploited to modulate the product yield and volumetric productivity in bioprocesses (Espinel-Ríos, Bettenbrock, et al., 2022; Espinel-Ríos, Morabito, Bettenbrock, et al. 2022).

We consider the Ccas/CcaR optogenetic system (Olson et al., 2014) for modulating the ATPase expression. CcaS is a sensor histidine kinase that is activated with green light ($\lambda_{535\text{ nm}}$). Active CcaS phosphorylates the cognate response regulator CcaR. Phosphorylated (active) CcaR enables the transcription of the target genes. In contrast, CcaS is inactivated with red light ($\lambda_{650\text{ nm}}$), thereby blocking transcription. From now on, let $u_s := I_s$ and $\bar{u}_c := \bar{I}_c$, where I_s is the green light intensity manipulated by the controller and \bar{I}_c is the average of I_c , that is, of the green light intensity perceived by the cells inside the bioreactor. Therefore, the process inputs comprise one cybergenetic input plus the substrate feeding rate. Furthermore, we consider a flat-panel photobioreactor, consisting of two flat surfaces joint by a thin gap, thereby creating a rectangular channel (Chanquia et al., 2021). The bioreactor is illuminated from one side by a green light source. This geometry is known to maximize the illumination area per culture volume, hence it is appealing for optogenetics.

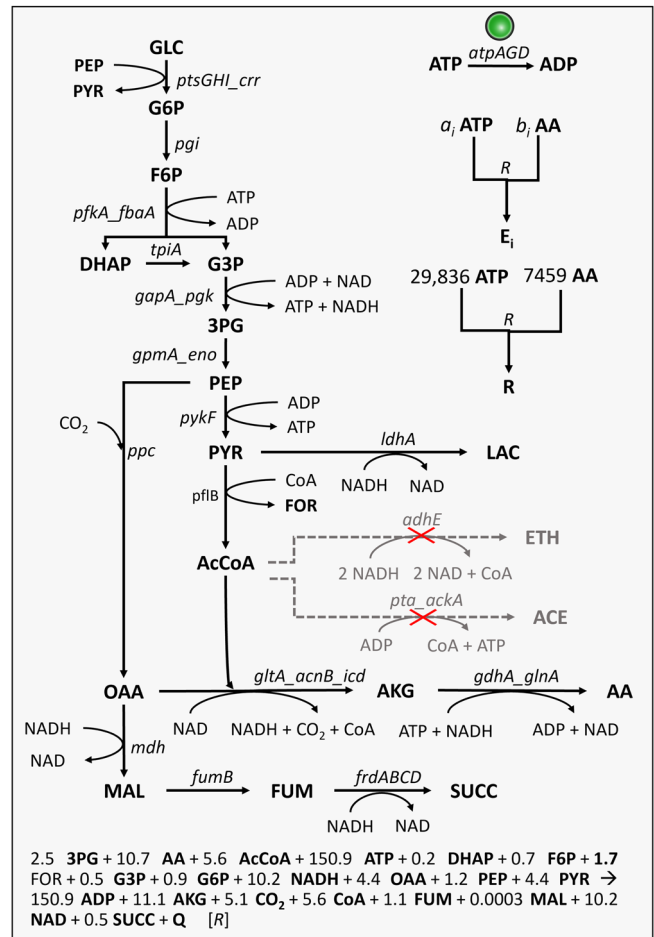


FIGURE 3 Scheme of the resource allocation model for the anaerobic lactate fermentation by *Escherichia coli* KBM10111. Catalytic species are shown in italics (e.g., *atpAGD* refers to the genes of the ATPase enzyme). Some enzymes are lumped via underscore symbols. In gray we depict the blocked pathways. Adapted from Espinel-Ríos, Bettenbrock, et al. (2022).

5.1 | Model and process considerations

We assume that the dose–response function for ATPase expression follows a Hill function (Olson et al., 2014)

$$\eta_{\text{ATPase}}(\bar{l}_c) = \alpha + \beta \frac{\bar{l}_c^\delta}{K^\delta + \bar{l}_c^\delta}, \quad (20)$$

where α is an input-independent basal rate of production (e.g., due to promoter *leakage* or constitutive expression), β is an input-dependent maximum rate of production, K is a saturation constant and δ is the Hill coefficient.

We assume that light penetration inside the bioreactor is not homogeneous as the cells interfere with the light beam. Let l be the length between the two plates of the bioreactor. We set up a balance over an infinitesimally small distance dl , assuming that the light hits perpendicularly with respect to the illuminated flat surface and that the culture is well-mixed. After integrating from l_0 to l we obtain

$$l_c(l, B) = l_c(l_0, B) e^{-a_\lambda B(l-l_0)}, \quad (21)$$

where $l_c(l_0, B) = l_s$, and a_λ is a lumped biomass-specific constant that accounts for light scattering and absorption effects. Note that the latter equation follows a similar derivation as the Lambert-Beer law (Hofmann et al., 2014). We obtain \bar{l}_c from the mean integral of $l_c(l, B)$ from 0 to l

$$\bar{l}_c(B) = \frac{l_s}{a_\lambda B l} (1 - e^{-a_\lambda B l}). \quad (22)$$

The dynamics of the lactate fermentation are based on an existing deFBA model (Espinel-Ríos, Bettenbrock, et al., 2022). See Figure 3 for a summary of the resource allocation model. In general, the model contains 34 fluxes: 16 metabolic reactions and 18 biomass-producing reactions. From the latter, 16 reactions are for the synthesis of catalytic enzymes, one for ribosomes and another one for a lumped quota compound. It considers five species in \mathbf{z} (glucose, lactate, formate, succinate, and carbon dioxide), 18 species in \mathbf{m} and 18 species in \mathbf{p} . The cell composition (g/g biomass) is 0.06 catalytic enzymes, 0.38 noncatalytic enzymes, 0.27 ribosomes, and 0.29 other components (DNA, lipids, carbohydrates, etc.), hence $\varphi_Q = 0.38 + 0.29$. The model already includes an ATPase-producing reaction. The cost of the biomass-producing reactions are expressed in terms of amino acids and ATP equivalents. The model parameters include 34 catalytic constants and 18 molecular weights. Furthermore, the model takes $F_V = B(t)$ in (16).

Additionally, we consider negligible degradation in Equations (5),(6) and (8), while for the ATPase enzyme

$$D_{\text{ATPase}} = d_{\text{ATPase}} p_{\text{ATPase}} V_L, \quad (23)$$

where d_{ATPase} is a constant ATPase degradation rate.

In Table 1, we summarize the model parameters for the CcaS/CcaR module and flat-panel bioreactor, as well as the process initial conditions. The cost function is chosen to maximize the lactate

TABLE 1 Relevant parameters and initial conditions of the nominal model.

Item	Value	Unit	Ref./Note
δ	2.490	1	Olson et al. (2014)
K	0.138	W/m ²	Olson et al. (2014)
α	2×10^{-6}	mmol/g/h	Note 1
β , S1	1×10^{-4}	mmol/g/h	Note 1
β , S2	2.5×10^{-5}	mmol/g/h	Note 1
β , S3	1×10^{-5}	mmol/g/h	Note 1
d_{ATPase}	6.3×10^{-2}	1/h	Note 2
l	0.022	m	Note 3
$v_{L\text{max}}$	45	L	Note 3
a_λ	1×10^{-2}	m ² /g	Note 4
$z_{\text{GLC},\text{in}}$	2220	mM	–
$x_{\text{GLC}}(0)$	139	mM	–
$x_{\text{LAC}}(0)$	0	mM	–
$x_{\text{CO}_2}(0)$	0	mM	–
$x_{\text{FOR}}(0)$	0	mM	–
$x_{\text{SUCC}}(0)$	0	mM	–
$B(0)$	0.59	g/L	–
$\mathbf{p}(0)$	Note 5	mM	–
$v_L(0)$	30	L	–

Note: (1) Assumed biologically sound values inferred from feasible deFBA simulations (Espinel-Ríos, Bettenbrock, et al., 2022) for different induction strength scenarios (S_i). (2) Estimated as $d_{\text{ATPase}} = \frac{\ln(2)}{t_{0.5}}$, where $t_{0.5}$ is the ATPase protein half-life time (Benito et al., 1991). (3) Based on a pilot-scale flat-panel photobioreactor design (Koller et al., 2018). (4) Assumed biologically sound order of magnitude. Estimated as ca. 1/30 of typical parameter values for microalgae (Blanken et al., 2016). (5) Estimated from $B(0)$ using resource balance analysis (Jabarivelisdeh et al., 2020).

concentration at the end time of the process, that is, $J = z_{\text{LAC}}(t_f)$, with $t_f = 30$ h and 12 control actions ($N = 12$). We furthermore consider box constraints for the inputs, namely $l_s = [0, 1]$ W/m² and $F_{\text{in}} = [0, 1]$ L/h. We add a further constraint to the optimization, $z_{\text{GLC}}(t_f) = 0$, to ensure that all glucose, the feeding substrate, is fully consumed. Finally, the bioreactor volume should not surpass the maximum working volume capacity $v_{L\text{max}}$, thus we add the constraint $v_L \leq v_{L\text{max}}$.

Remark on the numerical solution of the optimization problem: the solution to problems (16)–(18) are optimal input functions. This renders these problems infinite-dimensional, hence generally impractical to solve. One way to obtain a solution is via a finite-dimensional approach (Findeisen & Allgöwer, 2002; Rawlings et al., 2020). In our case, this is achieved by assuming piece-wise constant inputs and by discretizing the ordinary differential equations using orthogonal collocation based on Lagrange interpolation polynomials as motivated by Waldherr et al. (2015). The bilevel

optimizations in (17)–(19) were transformed into mathematical programs with complementarity constraints (single-level optimizations) by applying the Karush-Kuhn-Tucker conditions to the inner optimization problems, following an optimistic approach (Dempe & Franke, 2019). The resulting optimizations were solved in Python using CasADi (Andersson et al., 2019) and IPOPT (Wächter & Biegler, 2006).

5.2 | Open-loop optimal optogenetic control

Figure 4 shows the open-loop optimization results for the fed-batch fermentation considering *no model-plant mismatch*. We depict four scenarios:

1. S1: high-strength inducible CcaS/CcaR system—high β value.
2. S2: medium-strength inducible CcaS/CcaR system—medium β value.
3. S3: low-strength inducible CcaS/CcaR system—low β value.
4. NI: no inducible ATPase enzyme—the CcaS/CcaR system is absent.

The NI case rendered a final lactate concentration of 1434.3 mM, whereas S1 achieved 1572.3 mM ($\uparrow 10\%$), S2 1538.5 mM ($\uparrow 7\%$), and S3 1498.4 mM ($\uparrow 4\%$). Note that, by the end of all fermentations, the maximum allowed bioreactor volume was reached and all the glucose

was fully depleted. This implies that overall the same net amount of glucose was fed and consumed. Consequently, in the previous scenarios, the relative gains in product titer also correspond, proportionally, to increments in product yield and volumetric productivity.

Furthermore, in S1 the maximum ATPase enzyme concentration was 12.1% of the cell dry weight, in contrast to 8.5% and 3.8% in S2 and S3, respectively. As foreseeable, the higher the induction capacity strength of the CcaS/CcaR system, the higher the net ATPase enzyme expression, and therefore the higher the net increase in product yield. In previous works dealing with dynamic ATP turnover in one-stage batch fermentations, the increase in product yield was correlated with a loss in volumetric productivity (Espinel-Ríos, Bettenbrock, et al., 2022; Espinel-Ríos, Morabito, Bettenbrock, et al., 2022b). Here, we show that with a fed-batch system, it is possible to increase both the product yield via the ATP turnover mechanism *and* the volumetric productivity via the introduction of a feeding rate.

Compared to NI, scenarios S1–S3 resulted in 63%, 46%, and 28% lower final biomass concentrations. This can be explained by the combined effect of the lower biomass yields due to the ATP turnover and the potential resource burden related to the cost of producing the ATPase enzyme. Note that there is also a dilution effect from the feeding of the substrate. Overall, the increased ATP turnover rates managed to enhance the final lactate titer despite the lower biomass growth rates.

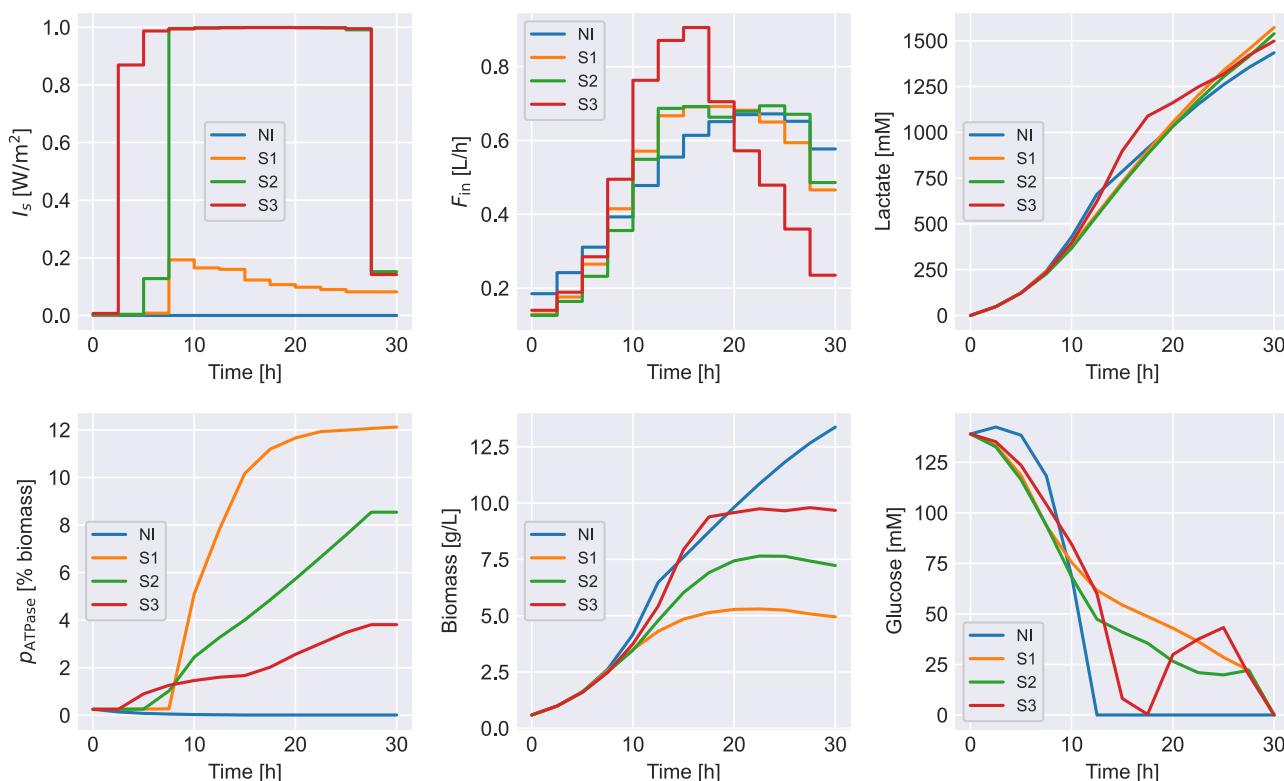


FIGURE 4 Open-loop fed-batch simulations *without* model-plant mismatch. Scenarios: NI—no induction, S1—high-strength, S2—medium-strength, and S3—low-strength inducible CcaS/CcaR systems.

In all induction scenarios, there was at first a gradual increase in the feeding rate, followed by a continuous decrease after around the midterm of the fermentation. This allowed for making up sufficient biomass while keeping low induction levels of the ATPase enzyme. Then, to avoid excessive dilution of the biomass, the feeding rate decreased at increasing ATPase expression levels. A benefit of our model-based optimization is that it takes into account resource allocation constraints. The resource allocation phenomena associated with the expression of the ATPase enzyme is presented in Figure 5, where we show the dynamic enzyme composition profiles throughout the open-loop fermentations. Note that the induction of the ATPase enzyme led to a re-accommodation of the unregulated enzymes. For instance, let us compare the profiles of enzymes *frdABCD*, *fumB*, *mdh*, *ppc*, *ldhA*, *tpiA*, and *pgi*. While they seem to slightly accumulate in the NI fermentation, they are kept at lower concentrations in the ATPase induction cases. The effect is clearer in scenario S1 because there the ATPase was expressed at higher levels.

5.3 | Counteracting model uncertainties and disturbances—optogenetic closed-loop control

Nominal open-loop control does not account for model uncertainties and unforeseen disturbances and process changes. Thus, we now evaluate the performance of shrinking horizon MPC for addressing system uncertainty. We limit our analysis to the high-strength inducible CcaS/CcaR system which provided the best results. We introduced model-plant mismatch by scaling the catalytic constants of the enzymes *pfkA_fbaA*, *gpmA_eno*, *gapA_pgk*, *gltA_acnB_icd*, and *gdhA_glnA* by a factor of 0.98, which slightly decreases the fermentation rates. We also scaled down δ and d_{ATPase} by 0.97 and 0.98, respectively; the latter decreases the steepness of the Hill function and the former makes the ATPase enzyme slower-degrading. The *modified* model was used for the plant simulations while the nominal model was given to the controller. Two MPC cases are considered:

1. MPC 1: all the states can be measured online without measurement noise.
2. MPC 2: the concentrations of the ATPase enzyme,⁴ biomass dry weight and extracellular metabolites can be measured online. Gaussian white noise (1% standard deviation) is added to the measurements. The cell composition is estimated via full information estimation.⁵ The *reconstructed* cell composition, along with the online measurements, are passed to the MPC.

The MPC simulations are shown in Figure 6. We also plot the open-loop scenario (*with model-plant mismatch*) as a reference case. The open-loop controller resulted in a final lactate concentration of 1449.5 and 64.3 mM net unconsumed glucose. The applied light intensity brought the ATPase enzyme concentration up to 12.6% of the cell dry weight, but then it decreased slightly to 11.9%.

The applied light intensity in MPC 1 allowed the ATPase enzyme fraction in the cell to eventually surpass the value achieved in the open-loop fermentation. The combined effect of the corrected light intensities and feeding rates rendered a final lactate titer of 1567.8 mM. The latter represents an 8% improvement with respect to the open-loop optimization. Also, with MPC 1 there was no unconsumed glucose by the end of the process. Note that MPC 1 scenario is a very optimistic result as the full state measurement is assumed and there is no measurement noise present.

MPC 2 results are more realistic regarding practical implementation, that is, with state estimation and measurement noise. The estimation of the cell composition for selected species at the different sampling times is presented in Figure 7. We also calculated the standard error (SE_FIE) of the estimates.⁶ Overall, the full information estimator tracked well the concentration trends of the biomass components. However, it should be noted that, in general, the estimation improved as the process proceeded. That is, the estimations were less accurate during the first one-third of the process (cf. e.g., the estimation profiles of enzymes *pfkA_fbaA*, *gltA_acnB_icd*, and *gdhA_glnA*). The progressive improvement of the estimation is explained by the

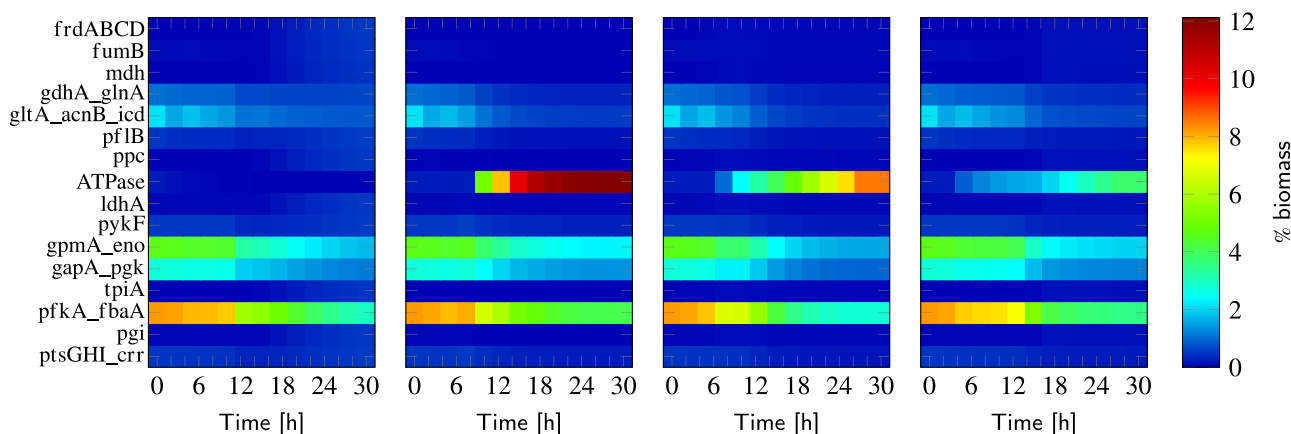


FIGURE 5 Enzyme expression heat map relative to the biomass dry weight for the fed-batch fermentations in Figure 4. From left to right, NI—no induction, S1—high-strength, S2—medium-strength, and S3—low-strength inducible CcaS/CcaR systems.

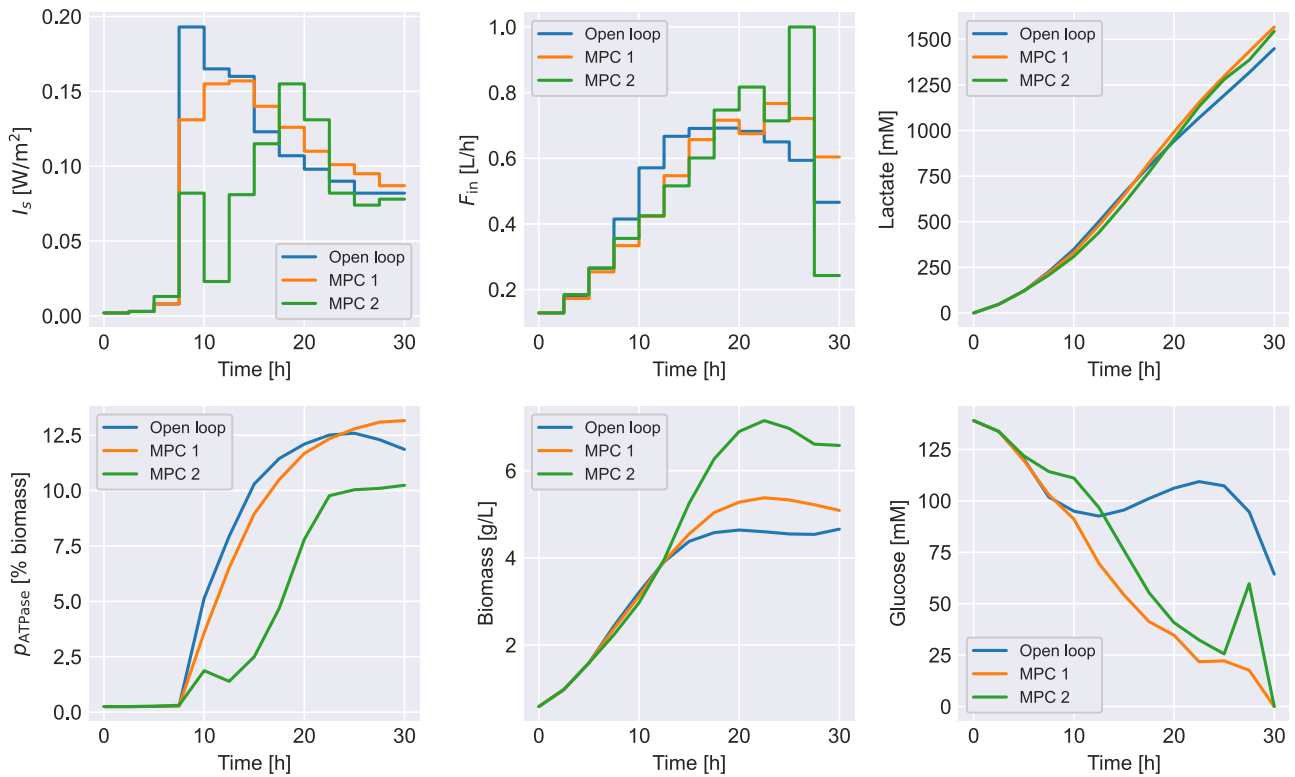


FIGURE 6 Closed-loop fed-batch simulations with model uncertainty for the high-strength inducible CcaS/CcaR system. MPC 1: with full state measurement and no measurement noise. MPC 2: with measurements of p_{ATPase} , B , and z in the presence of measurement noise; full information estimator used for estimating p . The open-loop case (without online corrective actions) is also shown.

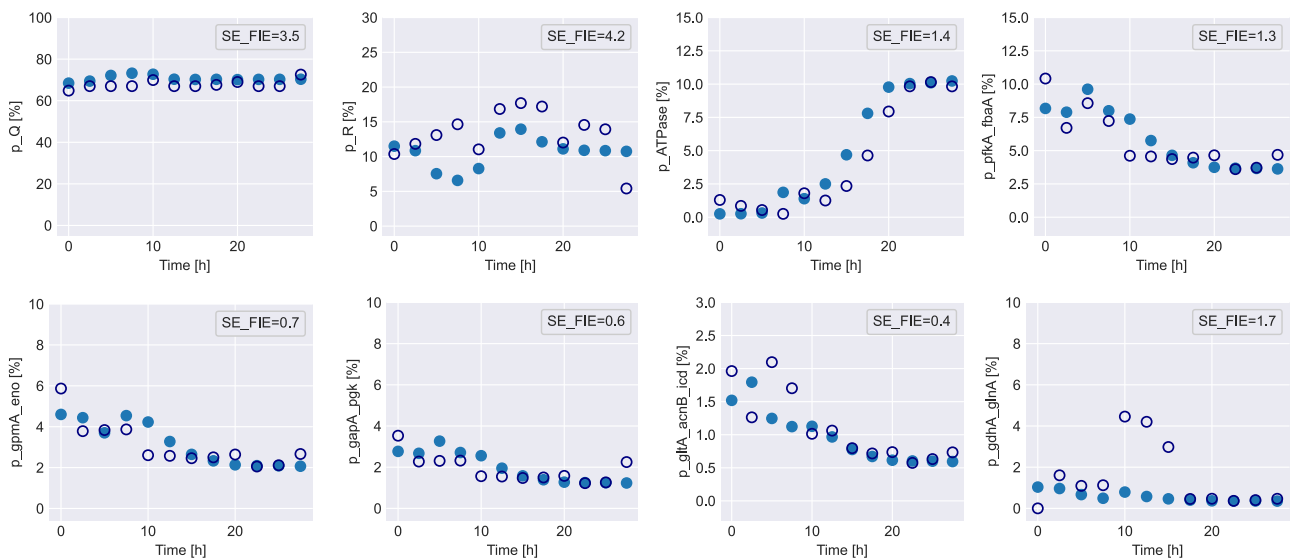


FIGURE 7 Online estimation of the cell components in percentage of cell dry weight. Only the eight most abundant species are shown. Filled circle: exact value. Empty circle: estimated state. The standard error of the estimate (SE_FIE) is presented.

growing estimation horizon and thus the increasing number of available measurements. This furthermore explains why at the beginning of the process the controller's predictions were comparatively *off* with respect to MPC 1 and the open-loop optimization. MPC 2 reached an intracellular ATPase enzyme

concentration of about 10.2%, leading to more biomass accumulation. The controller adjusted the feeding rates to avoid having unconsumed glucose by the end of the process. MPC 2 rendered a final lactate concentration of 1544.4 mM, very close to the value achieved in the MPC 1 scenario.

The MPC simulations demonstrate that using model-based feedback control, optionally coupled to state estimation methods, can improve the process performance of metabolic cybergenetic systems in the presence of system uncertainty.

6 | CONCLUSIONS AND OUTLOOK

We propose to fuse cybergenetics with model-based optimization and predictive control for dynamic metabolic engineering applications. The proposed metabolic cybergenetic framework exploits the concept of online metabolic regulation by dynamically modulating the gene expression of metabolism-relevant intracellular proteins. To do so, we developed a dynamic constraint-based modeling framework that integrates the dynamics of metabolic reactions, resource allocation, and external gene expression regulation. The model is combined with model-based optimization, predictive control, and estimation methods to facilitate the implementation of metabolic cybergenetic systems.

The potential of this technology is highlighted considering the dynamic control of the cellular ATP turnover via optogenetic regulation of the ATPase gene expression. We show that optimal control of the light intensity and the substrate feeding rate can enhance the process performance in terms of product titer and volumetric productivity. Furthermore, we demonstrated that introducing feedback via model predictive control can help to counteract system uncertainty.

We believe that the outlined metabolic cybergenetic framework opens the door to new and more advanced biotechnological applications where manipulating metabolic fluxes throughout the process is required. This is actually in line with dynamic metabolic engineering approaches and goes beyond traditional cybergenetic schemes where the regulated proteins (e.g., fluorescence reporters) are not directly involved in metabolic pathways. Moreover, the model-based feature of the presented framework can contribute to shortening and reducing the cost of process development, and obtaining a more robust, consistent, and flexible operation.

Note that the constraint-based dynamic model outlined in this work constitutes an optimization problem on its own. Consequently, model-based optimization using such models is of a bilevel nature. Here, we applied, for simplicity of presentation, the Karush-Kuhn-Tucker conditions to the inner optimization problem to obtain a single-level optimization that can be solved with conventional nonlinear solvers. However, this transformation renders nonconvex mathematical programs with complementarity constraints, which can be difficult to solve due to the nonlinearity of these constraints. Furthermore, the inclusion of Lagrange multipliers or dual variables, increases the size of the optimization problem. One might employ more tailored optimization approaches to overcome this challenge. We are currently working on simpler mathematical modeling approaches for metabolic cybergenetic systems, augmented with machine learning, to reduce the complexity associated with bilevel optimizations and facilitate practical implementations. Ultimately, our

goal is to experimentally validate the proposed framework considering the presented case study alongside other relevant bioprocesses. We are also working on extending the scope of the presented metabolic cybergenetic framework to include synthetic microbial communities (cf. e.g., Espinel-Ríos et al., 2023, n.d).

AUTHOR CONTRIBUTIONS

Sebastián Espinel-Ríos: Conceptualization; methodology; software; formal analysis; writing—original draft; writing—review and editing; visualization. **Bruno Morabito:** Methodology; writing—original draft; writing—review and editing. **Johannes Pohlodek:** Methodology; writing—original draft; writing—review and editing; visualization. **Katja Bettenbrock:** Conceptualization; writing—original draft; writing—review and editing; supervision; funding acquisition. **Steffen Klamt:** Conceptualization; writing—original draft; writing—review and editing; supervision; funding acquisition. **Rolf Findeisen:** Conceptualization; methodology; writing—original draft; writing—review and editing; supervision; funding acquisition.

ACKNOWLEDGMENTS

This work was supported by the International Max Planck Research School for Advanced Methods in Process and Systems Engineering (IMPRS ProEng) and by the EU-EFRE funded project DIGIPOL. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Sebastián Espinel-Ríos  <https://orcid.org/0000-0002-1582-6169>

Steffen Klamt  <http://orcid.org/0000-0003-2563-7561>

Rolf Findeisen  <http://orcid.org/0000-0002-9112-5946>

ENDNOTES

- ¹ From now on, we will refer to the F_1 -subunit of the ATPase enzyme/gene as the “ATPase enzyme/gene.”
- ² Quota elements include, for example, DNA, lipids, carbohydrates, noncatalytic proteins plus other small molecules.
- ³ The model allows describing batch systems by setting $\dot{v}_l = 0$. For continuous processes, one can include an additional flow rate leaving the bioreactor.
- ⁴ The ATPase concentration could be measured, for example, using a fluorescence-based biosensor (Kim et al., 2021).
- ⁵ For simplicity, in the full information estimation, we only consider the second term of the objective function in (19a). We neglect the state noise and assume constant model parameters, hence they are not estimated. The matrix \mathbf{R} is chosen as the identity matrix.

$$^6 SE_FIE = \sqrt{\frac{\sum_{j=0}^{N-1} (p_{i,j} - \hat{p}_{i,j})^2}{E_i}}, \forall i \in [1, n_p], E_i: \text{total number of } p_i \text{ estimates.}$$

REFERENCES

- Andersson, J. A. E., Gillis, J., Horn, G., Rawlings, J. B., & Diehl, M. (2019). CasADI: A software framework for nonlinear optimization and

- optimal control. *Mathematical Programming Computation*, 11(1), 1–36.
- Azimi, H., Tezel, H., & Thibault, J. (2019). Optimization of the in situ recovery of butanol from ABE fermentation broth via membrane pervaporation. *Chemical Engineering Research and Design*, 150, 49–64.
- Behera, B., Acharya, A., Gargey, I. A., Aly, N., & Balasubramanian, P. (2019). Bioprocess engineering principles of microalgal cultivation for sustainable biofuel production. *Bioresource Technology Report*, 5, 297–316.
- Benito, B., Moreno, E., & Lagunas, R. (1991). Half-life of the plasma membrane ATPase and its activating system in resting yeast cells. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1063(2), 265–268.
- Blanken, W., Postma, P. R., de Winter, L., Wijffels, R. H., & Janssen, M. (2016). Predicting microalgae growth. *Algal Research*, 14, 28–38.
- 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*. *Microbial Cell Factories*, 20(1), 63.
- Carrasco-López, C., García-Echauri, S. A., Kichuk, T., & Avalos, J. L. (2020). Optogenetics and biosensors set the stage for metabolic cybergenetics. *Current Opinion Biotechnology*, 65, 296–309.
- Chanquia, S. N., Vernet, G., & Kara, S. (2021). Photobioreactors for cultivation and synthesis: Specifications, challenges, and perspectives. *Engineering in Life Science*, 22(12), 202100070.
- Cui, S., Lv, X., Xu, X., Chen, T., Zhang, H., Liu, Y., Li, J., Du, G., Ledesma-Amaro, R., & Liu, L. (2021). Multilayer genetic circuits for dynamic regulation of metabolic pathways. *ACS Synthetic Biology*, 10(7), 1587–1597.
- del Rio-Chanona, E. A., 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. *Biotechnology Bioengineering*, 116(11), 2971–2982.
- Dempe, S., & Franke, S. (2019). Solution of bilevel optimization problems using the KKT approach. *Optimization*, 68(8), 1471–1489.
- Doran, P. M. (2013). *Bioprocess engineering principles*. (2nd ed.). Elsevier/Academic Press.
- Elsheikh, M., Hille, R., Tatulea-Codrean, A., & Krämer, S. (2021). A comparative review of multi-rate moving horizon estimation schemes for bioprocess applications. *Computers & Chemical Engineering*, 146, 107219.
- 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. *Computer Aided Chemical Engineering*, 52, 2601–2606.
- 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 Journal*, 68(4), e17555.
- EspinelRíos, S., Kok, R., Klamt, S., Avalos, J. L., & Findeisen, R. (n.d). Batch-to-batch optimization with model adaptation leveraging gaussian processes: the case of optogenetically assisted microbial consortia. IEEE Xplore.
- 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., 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.
- Findeisen, R., & Allgöwer, F. (2002). An introduction to nonlinear model predictive control. *21st Benelux Meeting on Systems and Control*, 11, 119–141.
- Findeisen, R., & Allgöwer, F. (2005). Robustness properties and output feedback of optimization based sampled-data open-loop feedback. In Proceedings of the 44th IEEE conference on decision and control (CDC) (pp. 54–59). IEEE.
- FungShek, C., & Betenbaugh, M. (2021). Taking the pulse of bioprocesses: At-line and in-line monitoring of mammalian cell cultures. *Current Opinion Biotechnology*, 71, 191–197.
- Gottstein, W., Olivier, B. G., Bruggeman, F. J., & Teusink, B. (2016). Constraint-based stoichiometric modelling from single organisms to microbial communities. *The Journal of the Royal Society Interface*, 13, 124, 20160627.
- GutiérrezMena, J., Kumar, S., & Khammash, M. (2022). Dynamic cybergenetic control of bacterial co-culture composition via optogenetic feedback. *Nature Communication*, 13(1), 4808.
- Hartline, C. J., Schmitz, A. C., Han, Y., & Zhang, F. (2021). Dynamic control in metabolic engineering: Theories, tools, and applications. *Metabolic Engineering*, 63, 126–140.
- Haseltine, E. L., & Rawlings, J. B. (2005). Critical evaluation of extended Kalman filtering and moving-horizon estimation. *Industrial & Engineering Chemistry Research*, 44(8), 2451–2460.
- 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. *Biotechnology and Bioengineering*, 112(10), 2195–2199.
- Heirung, T. A. N., Paulson, J. A., O'Leary, J., & Mesbah, A. (2018). Stochastic model predictive control—How does it work? *Computers & Chemical Engineering*, 114, 158–170.
- Hofmann, A., Simon, A., Grkovic, T., & Jones, M. (2014). *Spectroscopic methods*, (pp. 10–94). Cambridge University Press.
- Hoffman, S. M., Tang, A. Y., & Avalos, J. L. (2022). Optogenetics illuminates applications in microbial engineering. *Annual Review of Chemical and Biomolecular Engineering*, 13(1), 373–403.
- 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 System*, 38(3), 32–62.
- Hughes, S. R., & Jones, M. A. (2020). The global demand for biofuels and biotechnology-derived commodity chemicals: Technologies, markets, and challenges. In A. A. Vertè, N. Qureshi, H. P. Blaschek, & H. Yukawa (Eds.) *Green Energy to Sustainability*. (1st ed., pp. 137–156). Wiley.
- Jabarivelisdeh, B., Carius, L., Findeisen, R., & Waldherr, S. (2020). Adaptive predictive control of bioprocesses with constraint-based modeling and estimation. *Computers & Chemical Engineering*, 135, 106744.
- 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.
- Jabarivelisdeh, B., & Waldherr, S. (2016). Improving bioprocess productivity using constraint-based models in a dynamic optimization scheme. *IFAC-PapersOnLine*, 49(26), 245–251.
- Jabarivelisdeh, B., & Waldherr, S. (2018). Optimization of bioprocess productivity based on metabolic-genetic network models with bilevel dynamic programming. *Biotechnology and Bioengineering*, 115(7), 1829–1841.
- Julleson, D., David, F., Pflieger, B., & Nielsen, J. (2015). Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnology Advance*, 33(7), 1395–1402.
- Khammash, M. H. (2022). Cybergenetics: Theory and applications of genetic control systems. *Proceeding of IEEE*, 110(5), 631–658.
- Kim, H., Ju, J., Lee, H. N., Chun, H., & Seong, J. (2021). Genetically encoded biosensors based on fluorescent proteins. *Sensors*, 21(3), 795.
- Klamt, S., & von Kamp, A. (2022). Analyzing and resolving infeasibility in flux balance analysis of metabolic networks. *Metabolites*, 12(7), 585.
- 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 Biosystem Engineering*, 41(2), 213–220.

- Lalwani, M. A., Ip, S. S., Carrasco-López, C., Day, C., Zhao, E. M., Kawabe, H., & Avalos, J. L. (2021). Optogenetic control of the lac operon for bacterial chemical and protein production. *Nature Chemical Biology*, 17(1), 71–79.
- 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 Synthetic Biology*, 10(8), 2015–2029.
- Lalwani, M. A., Zhao, E. M., & Avalos, J. L. (2018). Current and future modalities of dynamic control in metabolic engineering. *Current Opinion Biotechnology*, 52, 56–65.
- Lindner, F., & Diepold, A. (2021). Optogenetics in bacteria—Applications and opportunities. *FEMS Microbiology Reviews*, 46(2), fuab055.
- Liu, S. (2020). *Bioprocess engineering: Kinetics, sustainability, and reactor design*. (3rd ed.) Elsevier.
- Liu, Z., Zhang, J., Jin, J., Geng, Z., Qi, Q., & Liang, Q. (2018). Programming bacteria with light—sensors and applications in synthetic biology. *Frontiers in Microbiology*, 9, 2692.
- Lovelett, R. J., Zhao, E. M., Lalwani, M. A., Toettcher, J. E., Kevrekidis, I. G., & Avalos, J. L. (2021). Dynamical modeling of optogenetic circuits in yeast for metabolic engineering applications. *ACS Synthetic Biology*, 10(2), 219–227.
- Mayne, D. Q. (2014). Model predictive control: Recent developments and future promise. *Automatica*, 50(12), 2967–2986.
- Mesbah, A., Streif, S., Findeisen, R., & Braatz, R. D. (2014). Stochastic nonlinear model predictive control with probabilistic constraints. In *2014 American control conference*, (pp. 2413–2419). IEEE.
- Miliás-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. *Nature Communication*, 7(1), 12546.
- 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.
- 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.
- 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.
- Nimmegeers, P., Vallerio, M., Telen, D., Impe, J., & Logist, F. (2018). Interactive multi-objective dynamic optimization of bioreactors under parametric uncertainty. *Chemie Ingenieur Technik*, 91(3), 201800082.
- 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. *Nature Methods*, 11(4), 449–455.
- 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 Research*, 20, 153–163.
- Rawlings, J., Mayne, D., & Diehl, M. (2020). *Model predictive control: Theory, computation and design*. (2nd ed.) Nob Hill Publishing, LLC.
- Reardon, K. F. (2021). Practical monitoring technologies for cells and substrates in biomanufacturing. *Current Opinion Biotechnology*, 71, 225–230.
- 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.
- Ryu, K. H., Kim, B., & Lee, J. H. (2019). A model-based optimization of microalgal cultivation strategies for lipid production under photo-autotrophic condition. *Computers & Chemical Engineering*, 121, 57–66.
- Saa, P. A., & Nielsen, L. K. (2017). Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnology in Advances*, 35(8), 981–1003.
- Sanford, K., Chotani, G., Danielson, N., & Zahn, J. A. (2016). Scaling up of renewable chemicals. *Current Opinion Biotechnology*, 38, 112–122.
- Scull, C. E., Dandpat, S. S., Romero, R. A., & Walter, N. G. (2021). Transcriptional riboswitches integrate timescales for bacterial gene expression control. *Frontiers in Molecular Biosciences*, 7, 607158.
- Shen, X., Wang, J., Li, C., Yuan, Q., & Yan, Y. (2019). Dynamic gene expression engineering as a tool in pathway engineering. *Current Opinion Biotechnology*, 59, 122–129.
- Tandar, S. T., Senoo, S., Toya, Y., & Shimizu, H. (2019). Optogenetic switch for controlling the central metabolic flux of *Escherichia coli*. *Metabolic Engineering*, 55, 68–75.
- Tuveri, A., Holck, H. E., Nakama, C. S., Matias, J., Jäschke, J., Imsland, L., & Bar, N. (2022). Bioprocess monitoring: A moving horizon estimation experimental application. *IFAC-PapersOnLine*, 55(7), 222–227.
- Tuveri, A., Pérez-García, F., Lira-Parada, P. A., Imsland, L., & Bar, N. (2021). Sensor fusion based on extended and unscented Kalman filter for bioprocess monitoring. *Journal of Process Control*, 106, 195–207.
- Vandermies, M., & Fickers, P. (2019). Bioreactor-scale strategies for the production of recombinant protein in the yeast *Yarrowia lipolytica*. *Microorganisms*, 7(2), 40.
- Venayak, N., Anesiadis, N., Cluett, W. R., & Mahadevan, R. (2015). Engineering metabolism through dynamic control. *Current Opinion Biotechnology*, 34, 142–152.
- De Vrieze, J., De Mulder, T., Matassa, S., Zhou, J., Angenent, L. T., Boon, N., & Verstraete, W. (2020). Stochasticity in microbiology: Managing unpredictability to reach the Sustainable development goals. *Microbial Biotechnology*, 13(4), 829–843.
- Waldherr, S., Oyarzún, D. A., & Bockmayr, A. (2015). Dynamic optimization of metabolic networks coupled with gene expression. *Journal of Theoretical Biology*, 365, 469–485.
- 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.
- Wohlgemuth, R., Twardowski, T., & Aguilar, A. (2021). Bioeconomy moving forward step by step—A global journey. *New Biotechnology*, 61, 22–28.
- Yang, S., Kim, S., Kim, D.-K., Jeon An, H., Bae Son, J., HedénGynnå, A., & Ki Lee, N. (2019). Transcription and translation contribute to gene locus relocation to the nucleoid periphery in *E. coli*. *Nature Communication*, 10(1), 5131.
- Yasemi, M., & Jolicoeur, M. (2021). Modelling cell metabolism: A review on constraint-based steady-state and kinetic approaches. *Processes*, 9(2), 322.
- Yu, S., Reble, M., Chen, H., & Allgöwer, F. (2014). Inherent robustness properties of quasi-infinite horizon nonlinear model predictive control. *Automatica*, 50(9), 2269–2280.
- Zahoor, A., Messerschmidt, K., Boecker, S., & Klamt, S. (2020). ATPase-based implementation of enforced ATP wasting in *Saccharomyces cerevisiae* for improved ethanol production. *Biotechnology Biofuels*, 13(1), 185.
- 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 Synthetic Biology*, 10(5), 1143–1154.

How to cite this article: Espinel-Ríos, S., Morabito, B., Pohlodek, J., Bettenbrock, K., Klamt, S., & Findeisen, R. (2024). Toward a modeling, optimization, and predictive control framework for fed-batch metabolic cybergenetics. *Biotechnology and Bioengineering*, 121, 366–379. <https://doi.org/10.1002/bit.28575>