Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-23 (2022) 92–97

Cell-free biosynthesis meets dynamic optimization and control: a fed-batch framework

Sebastián Espinel-Ríos [*,**] **Nicolas Huber** [**] **Edgar Alberto Alcalá-Orozco** [***] **Bruno Morabito** [*] **Thomas F.T. Rexer** [***] **Udo Reichl** [***] **Steffen Klamt** [**] **Rolf Findeisen** [****]

[*] *Laboratory for System Theory and Automatic Control, Otto von Guericke University Magdeburg, Germany*

[**] *Analysis and Redesign of Biological Networks, Max Planck Institute for Dynamics of Complex Technical Systems, Germany*

[***] *Bioprocess Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Germany*

[****] *Control and Cyber-Physical Systems Laboratory, TU Darmstadt, Germany (e-mail: rolf.findeisen@iat.tu-darmstadt.de)*

Abstract: Cell-free biosynthesis uses the machinery of cells, such as the metabolic reactions, to carry out conversion processes *in vitro*. This can be more beneficial than *in vivo* approaches like fermentations. Some advantages of these synthetic biology processes include higher product yields, rates and titers, and more flexibility in pathway design. Cell-free biosynthesis is still in early stages and, unlike *in vivo* production, there are very few examples of model-based optimization. Moreover, we encounter static optimizations in most cases, neglecting the dynamic nature of the processes. We present an optimal control framework to maximize the efficiency of cell-free biosynthesis. We focus on fed-batch setups as they allow enhancing the reaction rates via the feeding, extending production processes for longer times, and minimizing the potential negative effects of enzyme kinetics with substrate inhibition. Our framework can in principle handle several cost functions and exploit both static and dynamic degrees of freedom. An aspect that can hinder model-based optimization is model uncertainty, which can arise due to uncertain parameters, oversimplified model assumptions or unknown reaction mechanisms. To counteract this, we propose the use of model predictive control during the process operation. In addition, we outline the use of moving horizon estimation as an observer in the case of unmeasured states. We consider the *de novo* cell-free synthesis of uridine diphosphate-N-acetylglucosamine as a biomedical relevant case study, where we were able to maximize the volumetric productivity in simulations, and indirectly also the titer and enzyme efficiency use.

Keywords: Cell-free, fed-batch, optimization, control, state estimation, UDP-GlcNAc.

1. INTRODUCTION

Biotechnological manufacturing is mainly dominated by cell-based approaches, meaning that whole cells or microorganisms catalyze the conversion reactions. However, employing cells as biocatalysts carries intrinsic disadvantages. For example, certain metabolites or products can be toxic to the cell, thus limiting the achievable product titers. Furthermore, there is a well-known trade-off between biomass and product yield, i.e., increasing the product yield often comes at the expense of decreasing the biomass yield, and with that, the volumetric productivity. Also, the downstream processing can be very complex because the product of interest tends to be highly diluted, and one also needs to deal with biomass waste. In addition, the design-build-test-learn cycles are frequently slow since pathway engineering usually involves genetic modifications of the cells (Tinafar et al., 2019; Lim and Kim, 2022).

The cell-based paradigm has been increasingly challenged by cell-free biosynthesis. The latter consists in taking the metabolic reactions, and potentially other cellular machinery, out of the cells to perform metabolic conversions *in vitro*. A great advantage is that there are no competing pathways, thus higher yields are possible. Higher fluxes and volumetric productivities can be also reached because the processes can operate at more favourable conditions. Cell toxicity issues can be neglected, hence higher product concentrations can be obtained. More concentrated product streams and no biomass waste lead to a simpler downstream processing. Pathway engineering becomes easier since the designed pathways do not need to sustain life, and they can be tested *in vitro*. Consequently, cell-free systems can bring tremendous opportunities in the fields of synthetic biology and biotechnology. They can be applied in the manufacturing of therapeutics, antigens, virus-like particles, antimicrobials, platform chemicals, biofuels, etc. (Lim and Kim, 2022; Tinafar et al., 2019).

We identify three main challenges in the area of cell-free biosynthesis. First, catalytic enzymes must be produced *a*

⋆ This work was supported by the International Max Planck Research School for Advanced Methods in Process and Systems Engineering (IMPRS ProEng).

10.1016/j.ifacol.2023.01.021

priori, which can be expensive. Secondly, enzyme cofactors often need to be added continuously to keep the pathways active, also increasing the costs. However, cofactor regeneration schemes (Mordhorst and Andexer, 2020) have the potential to greatly solve this issue. Finally, optimization strategies are not always performed in a structured and efficient way. That is, many processes are designed based on trial and error and recipes, requiring a lot of experimental effort, time and resources. In that sense, most of the work has focused on improving the enzymatic cascade, enzyme optimization, best reaction conditions, and process design (Siedentop et al., 2021).

Despite the potential benefits of model-based optimization, such as less time, costs and resources needed, surprisingly very little has been done in the field of cell-free biosynthesis. Furthermore, static optimizations seem to be the dominant approach (Siedentop et al., 2021), overlooking the fact that cell-free reactions are dynamic processes, operating in a dynamic environment. An aspect that may limit model-based optimization is model uncertainty. It is common when modeling cell-free systems that some parameters are not observable, i.e., they cannot be uniquely inferred from the available measurements. This problem can also arise due to structurally non-identifiable parameters (Bellman and Åström, 1970). Another reason for model uncertainty lies in wrong model assumptions, meaning that the exact reaction mechanisms are not always known. Commonly used enzyme kinetics are often based on steady state and reactant stationary assumptions underlying the Michaelis-Menten equation (Schnell, 2014), or on mass action kinetics where enzyme-specific parameters are ignored (Du et al., 2016). Overall, this may lead to a mismatch between the actual plant behaviour and the model approximation.

In this work, we present a systems approach to optimize cell-free biosynthesis. To do so, we propose the implementation of a fed-batch setup where materials such as substrates and enzymes can be in principle continuously added to the system to dynamically influence the reaction rates (Section 2). Fed-batch processes could also help to extend production for longer periods, thus bringing a reduction in downtimes (non-production time between batches). It can also be very useful in case of enzyme kinetics with substrate inhibition. In such cases, continuous substrate feeding might be more beneficial than adding all the substrate at once from the beginning. To maximize the production efficiency, we formulate a flexible model-based optimization problem (Section 3) that can handle several cost functions and both static and dynamic degrees of freedom. We also outline the use of shrinking horizon model predictive control (MPC), an advanced feedback control scheme, to tackle possible model uncertainty (Section 4). Furthermore, we show how moving horizon estimation (MHE) can be used to infer unavailable state measurements during the MPC implementation (Section 5). As a case study, we consider the uridine diphosphate (UDP)-N-acetylglucosamine (GlcNAc) cell-free biosynthesis. This sugar nucleotide can act as an activated sugar donor in glycoengineering with promising applications for the biomanufacturing of therapeutic proteins, enzymes, glycan-based delivery systems, and drug molecules (Ma et al., 2020; Gadekar et al., 2020).

2. FED-BATCH MODEL FOR CELL-FREE SYSTEMS

We express the dynamics of cell-free biosynthesis in fed-batch as follows

$$\frac{\mathrm{d}m(t)}{\mathrm{d}t} = SV(E,m,p) + \left(\frac{1}{v_\mathrm{L}}\right)(F_\mathrm{m} \odot m_\mathrm{in} - F_\mathrm{T} m)\,, \quad (1a)$$

$$\frac{\mathrm{d}E(t)}{\mathrm{d}t} = \left(\frac{1}{v_\mathrm{L}}\right)(F_\mathrm{E} \odot E_\mathrm{in} - F_\mathrm{T} E) - d_\mathrm{E} \odot E, \quad (1b)$$

$$\frac{\mathrm{d}v_\mathrm{L}(t)}{\mathrm{d}t} = F_\mathrm{T}, \quad (1c)$$

$$F_\mathrm{T} = \sum_{j=1}^{n} F_{\mathrm{m,j}} + \sum_{j=1}^{n} F_{\mathrm{E,j}}, \quad (1d)$$

with initial conditions

$$m(t_0) = m_0, \quad (2)$$

$$E(t_0) = E_0, \quad (3)$$

$$v_\mathrm{L}(t_0) = v_{L0}, \quad (4)$$

where m is the molar vector of metabolite species in the pathway, S is the stoichiometric matrix of m, V is a vector function describing the reaction fluxes, E is the molar vector of enzyme species, p is the vector of flux-related parameters, v_L is the liquid volume of the reactor, F_m is the vector of metabolite feeding rates, $\odot$ is the Hadamard product (i.e. element-wise product), m_in is the vector of molar concentrations in the feed, F_T is the net feed flow rate into the reactor, F_E is the vector of enzyme feeding rates, E_in is the vector of enzyme concentrations in the feed, and d_E is a vector of enzyme degradation rates. Furthermore, m_0, E_0 and v_{L0} are the concentrations of m, E and v_L, respectively, at the initial time t_0. Note that when F_m and F_E are zero-vectors, the system behaves as a batch process. It is also worth noting that V can be typically modeled using kinetic relations which are functions of E and m; therefore, both the enzyme and metabolite concentrations can be exploited to influence the reaction rates.

3. OPTIMAL CONTROL

We formulate an optimal control problem that maximizes a given cost functional that captures the efficiency of the process

$$\min_{u(\cdot),\, x_0,\, t_\mathrm{f}} \quad \int_{t_0}^{t_f} l(\tau, x(\tau), u(\tau))\mathrm{d}\tau + e(t_f, x(t_f)), \quad (5a)$$

$$\text{s.t.} \quad \text{Eqs.(1a)} - \text{(1d)}, \quad (5b)$$

$$0 \geq g(t, x(t), u(t)), \quad (5c)$$

$$u = [F_\mathrm{m}, F_\mathrm{E}]^\mathsf{T}, \quad (5d)$$

$$x = [m, E, v_\mathrm{L}]^\mathsf{T}, \quad (5e)$$

$$\tau \in [t_0, t_f] \subset \mathbb{R}, \quad (5f)$$

$$x(t_0) = x_0, \quad (5g)$$

where $u(\cdot)$ is the input function, $x \in \mathbb{R}^{n_x}$ is the state, $l : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}$ is the stage cost, $e : \mathbb{R} \times \mathbb{R}^{n_x} \to \mathbb{R}$ is the terminal cost, $g : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_g}$ are additional state and input constraints, and τ is the time that spans from t_0 to the final time t_f. Note that t_f can also be optimized. Depending on the specific process and context, one can maximize, e.g., the profit of the plant (i.e., revenues minus operational costs), or certain economy-related metrics such as the final product titer or

volumetric productivity. In that regard, many variables can affect the efficiency and economics of fed-batch cell-free systems, which can thus be optimized. Therefore, for the sake of generality, we consider in the above formulation the feeding rates, the initial state concentrations, and the final time of the process as potential degrees of freedom for the optimization.

4. MODEL PREDICTIVE CONTROL

To counteract for possible model uncertainty such as model-plant mismatch, we propose the use of shrinking horizon MPC during the process operation. MPC consists in solving the optimal control problem described in (5) at every sampling time t_k (Rawlings et al., 2020). The optimization problem reads

$$\begin{aligned}
\min_{u(\cdot)} \quad & \int_{t_k}^{t_f} l(\tau, x(\tau), u(\tau))\mathrm{d}\tau + e(t_f, x(t_f)), && \text{(6a)}\\
\text{s.t.} \quad & \text{Eqs.(1a)} - \text{(1d)}, && \text{(6b)}\\
& 0 \geq g(t, x(t), u(t)), && \text{(6c)}\\
& u = [F_\mathrm{m}, F_\mathrm{E}]^\mathsf{T}, && \text{(6d)}\\
& x = [m, E, v_\mathrm{L}]^\mathsf{T}, && \text{(6e)}\\
& \tau \in [t_0, t_f] \subset \mathbb{R}, && \text{(6f)}\\
& x(t_k) = \tilde{x}_k. && \text{(6g)}
\end{aligned}$$

Here, we assumed equidistant sampling times with interval Δt, i.e., $t_{k+1} = t_k + \Delta t$, $t_0 = 0$. Note that, also non-equidistant sampling time can be used. The integral of the stage cost spans from t_k to t_f, $\tau \in [t_k, t_\mathrm{f}]$, and the prediction horizon shrinks in time. Note that non-equidistant sampling time can be also used. In this case, we need the estimated or measured state at every sampling time, $x(t_k) = \tilde{x}_k$. As we will see in the next section, the moving horizon estimator can be used for this purpose.

5. MOVING HORIZON ESTIMATION

Solving the MPC requires knowing the state $\tilde{x}_k$ at each sampling time to re-solve the optimization problem. Depending on the size and complexity of the considered metabolic pathway, the number of states to be measured can be considerably high. Therefore, we propose to use MHE as an observer for inferring unmeasured states based on available measurements (Rawlings, 2013). Similarly to optimal control and MPC, the MHE solves an optimal control problem along a horizon but projected into the *past*. To consider state and measurement uncertainty, the state noise w and measurement noise v are added to the problem. The measurements are taken at discrete time intervals. Here, without loss of generality, we assume they are taken at constant time intervals Δt. For compactness, let us indicate with $(\cdot)_i$ and $(\hat{\cdot})$ the optimization and measured variables at time t_i, respectively. Since the measurements are taken at discrete sampling times, it is more natural to write the MHE formulation using the dynamic model in discrete form. At time t_k MHE reads

$$\begin{aligned}
\min_{x_{k-\bar{N}}, p_k, \mathbf{w}_k} \quad & J(x_{k-\bar{N}}, p_k, \mathbf{w}_k), && \text{(7a)}\\
\text{s.t.} \quad & x_{i+1} = F(x_i, \hat{u}_i, p_k) + w_i, && \text{(7b)}\\
& y_i = h(t_i, x_i, \hat{u}_i, p_k) + v_i, && \text{(7c)}\\
& 0 \geq c(t_i, x_i, w_i, y_i), && \text{(7d)}\\
& \text{for } i \in [k - \bar{N}, k], \quad k, \bar{N} \in \mathbb{N}, && \text{(7e)}
\end{aligned}$$

where $\bar{N}$ is the MHE horizon length, $x_{k-\bar{N}}$ is the state at the beginning of the MHE horizon, p_k is the vector of the system parameters to be estimated, $c : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_c}$ are additional constraints, $h : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_h}$ are the measurements equations, and $\mathbf{w}_k = [w_{k-\bar{N}}^\mathsf{T}, w_{k-\bar{N}+1}^\mathsf{T}, ..., w_k^\mathsf{T}]^\mathsf{T}$ is the sequence of state noise. Note that the parameters are considered constant in the horizon. The objective function is defined as

$$\begin{aligned}
J(x_{k-\bar{N}}, \mathbf{w}_k) = & \left\| \begin{bmatrix} x_{k-\bar{N}} - \tilde{x}_{k-N} \\ p_k - \tilde{p} \end{bmatrix} \right\|_{P_k}^2 + \\
& \sum_{i=k-\bar{N}}^{k} \|\hat{y}_i - y_i\|_R^2 + \|w_i\|_W^2,
\end{aligned} \tag{8}$$

where $\|a\|_A^2 \triangleq a^\mathsf{T} A a$, P_k is the *arrival cost* weighting matrix, R the measurements error matrix, and W the state error matrix. The matrix of the arrival cost can be updated at every time step, to better approximate the information that is "outside" the horizon (Rawlings et al., 2020). $\tilde{p}$ and $\tilde{x}_{k-N}$ are the best guesses available of the parameters and states at the beginning of the horizon, respectively. Usually, they are initialized by the user and then can be substituted with the estimated values after the first estimation. Furthermore

$$F(x_i, \hat{u}_i, p) = x(t_i) + \int_{t_i}^{t_i+\Delta t} f(x(t), \hat{u}(t), p)\mathrm{d}t, \tag{9}$$

where $f(x(t), \hat{u}(t), p)$ is a shorthand notation for Eqs.(1a)-(1d). The results of (7) are an optimal initial state $x_{k-\bar{N}}^*$, an optimal parameter p_k^* and an optimal state noise sequence $\mathbf{w}_k^*$. With these optimal values we can obtain an optimal state estimate $\tilde{x}_k$ that is needed by the MPC.

6. APPLICATION STUDY

The considered cell-free UDP-GlcNAc pathway, based on the process and mathematical model proposed by Mahour et al. (2018), is depicted in Fig. 1. UDP-GlcNAc is produced from uridine monophosphate (UMP), GlcNAc and catalytic amounts of adenosine triphosphate (ATP). Note that the pathway has an *in situ* ATP regeneration cycle based on polyphosphate, avoiding the continuous addition of this expensive cofactor.

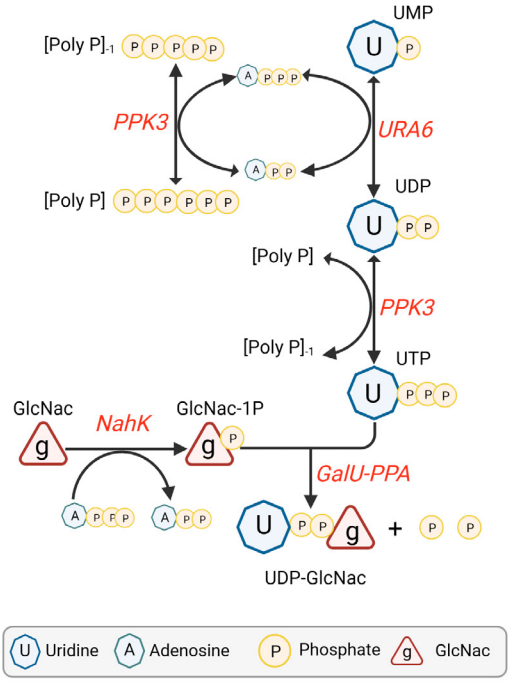
The rate laws $V(E, m, p)$ of the model are defined as (Mahour et al., 2018)

$$\begin{aligned}
V_\mathrm{NahK} &= k_1 E_\mathrm{NahK} m_\mathrm{GlcNAc} m_\mathrm{ATP}, && \text{(10a)}\\
V_\mathrm{GalU\text{-}PPA} &= k_2 E_\mathrm{GalU} m_\mathrm{GlcNAc\text{-}1P} m_\mathrm{UTP}, && \text{(10b)}\\
V_{\mathrm{PPK3}_{\mathrm{ADP/ATP}}} &= k_3 E_\mathrm{PPK3} \left(m_\mathrm{ADP} - \frac{m_\mathrm{ATP}}{k_\mathrm{eq1}} \right), && \text{(10c)}\\
V_{\mathrm{PPK3}_{\mathrm{UDP/UTP}}} &= k_4 E_\mathrm{PPK3} \left(m_\mathrm{UDP} - \frac{m_\mathrm{UTP}}{k_\mathrm{eq2}} \right), && \text{(10d)}\\
V_\mathrm{URA6} &= k_5 E_\mathrm{URA6} \left(m_\mathrm{UMP} m_\mathrm{ATP} - \frac{m_\mathrm{UDP} m_\mathrm{ADP}}{k_\mathrm{eq3}} \right), && \text{(10e)}
\end{aligned}$$

where k_i and $k_{\mathrm{eq}i}$ are kinetic parameters. The parameter values used in this study were obtained from Mahour et al. (2018). Furthermore, we assumed negligible enzyme degradation. Hereafter we will refer to this model as the *nominal model*.

6.1 Open-loop optimization

As mentioned before, several alternatives for the selection of the cost function and decision variables of the optimizations are in principle possible. For example, a high



NahK: N–acetylhexosamine kinase, URA6: uridine monophosphate kinase, PPK3: polyphosphate kinase, GalU-PPA: lumped Glc–1P uridyltransferase and inorganic diphosphatase.

Fig. 1. Cell-free UDP-GlcNAc pathway. Figure made in BioRender.com.

volumetric productivity usually correlates with shorter production times and smaller facilities needed to achieve a given production target. Therefore, it is very often an economic parameter to be optimized towards maximizing the plant profitability (Woodley, 2020). As a proof of concept, we consider in our case study the maximization of the volumetric productivity in a fed-batch configuration, hence $l = -m_{\text{UDP-GlcNAc}}$ in the optimal control problem in (5). The chosen decision variables are the UMP and GlcNAc feeding rates in a 10 h-process. By doing so, we indirectly also maximize the product titer in the given process time frame, and the catalytic potential of enzymes per process run, defined as the amount of UDP-GlcNAc produced per total initial enzyme amount. Note that a step size $\Delta t = 1/60$ h is used and that t_f is fixed to 10 h. All optimization problems were solved using HILO-MPC (Pohlodek et al., 2022), an optimization and control toolbox developed in our group.

First, we solved the optimal control problem using the nominal model for the fed-batch system, and then applied the inputs to the plant in an open-loop fashion (Fig. 2-A), assuming no model-plant mismatch. The results are shown in Fig. 3, where we compare the batch against the fed-batch case for the same initial conditions and time frame. The optimized fed-batch rendered a final UDP-GlcNAc concentration of 0.85 mM compared to 0.37 mM of the batch process, which represents around 2.3-fold improvement in titer and thus volumetric productivity. Given that both processes started with the same initial enzyme concentrations, one could also state that the fed-batch had 2.3-fold higher enzyme efficiency use per process run.

Regarding the dynamics of the process, the fed-batch system had significantly higher production rates compared to the batch system. Moreover, while the rates in the batch

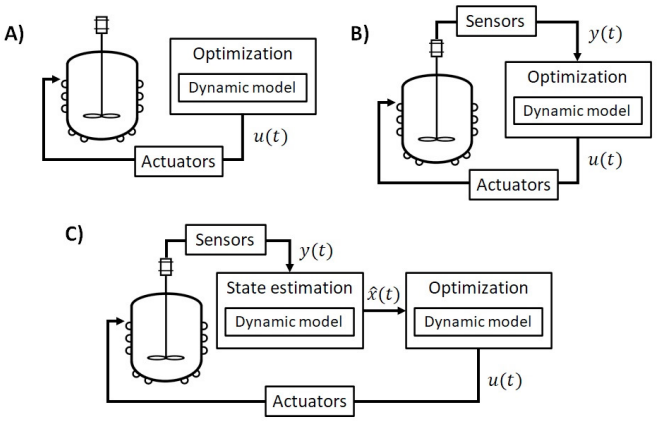


Fig. 2. Control configurations considered in this study. A) Open-loop control. B) Model predictive control. C) Model predictive control with state estimation.

setup gradually decreased over time, the rates in the fed-batch remained comparatively unaffected, only starting to slowly decrease by the end of the process. To explain this difference, let us analyze the rate laws in Eqs. (10a)-(10e). As can be seen, the reaction rate V_{E_i} can be enhanced by either increasing the substrate or enzyme concentrations. In our case study, the optimization predicted a GlcNAc feeding rate profile such that the GlcNAc concentration in the fed-batch reactor was kept always above the batch one; therefore, this translated into higher reaction rates through the cascade.

A similar statement can be made about the predicted UMP feeding rates. However, the main difference is that UMP does not accumulate in the reactor, in fact, it is rapidly consumed and then reaches a sort of quasi-steady state. This can be explained by the fast kinetics of the first reaction in the UMP to UTP pathway branch. In contrast, the metabolite that does accumulate is UTP. One can see that the controller managed to keep the UTP concentration in the fed-batch overall higher than in the batch process, thus also boosting the reaction rates downstream in the cascade.

Moreover, the optimizer managed to find a trade-off between adding more substrates to the system and excessive enzyme dilution. This is relevant because too much substrate feeding could bring the enzyme concentrations down to levels that would negatively impact the conversion rates. We should insist that here we only show the results for maximizing the volumetric productivity, hence the input profiles look relatively straightforward (on-off). Nevertheless, the results would probably differ if we considered more complex cost functions (e.g., taking into account enzyme, substrate and product costs, upstream and downstream processing expenditures, etc.). Also, the simplified UDP-GlcNAc model does not consider substrate inhibition or substrate-limiting kinetics such as the Michaelis-Menten equation, which would otherwise further constraint the optimal input space, thus making the use of model-based optimization even more pertinent towards enhancing the process performance.

6.2 Closed-loop control in case of model uncertainty

As a second step, to show the potential of feedback control to address model uncertainty, we assume that the model used by the controller has a mismatch with respect to the

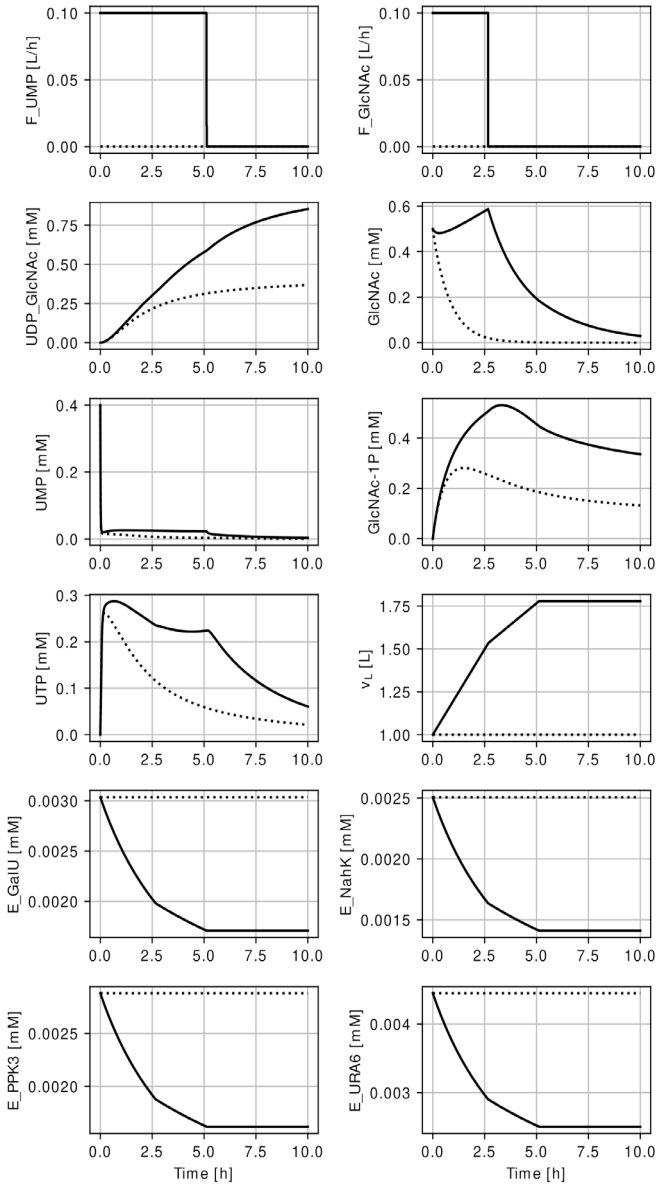


Fig. 3. Open-loop optimization results for the fed-batch UDP-GlcNAc biosynthesis (——). The batch process (····) is presented for comparison. Only selected intermediate metabolites are shown for simplicity of presentation. No model uncertainty was considered.

actual plant. To do so, we disturb some parameters of the nominal model and employ the modified model for the controller and observer predictions. In contrast, the plant simulations are based on the (unmodified) nominal model. In Fig. 4 we present the results of two shrinking horizon MPC configurations, one with full-state measurement and no measurement noise (Fig. 2-B), and another one with measurement noise and using MHE as an observer to infer unmeasured states (Fig. 2-C). For the MHE/MPC case, we consider that only measurements of GlcNAc, UMP, UDP-GlcNAc, GlcNAc-1P and UTP are available. These measurements are affected by Gaussian noise with a relative standard deviation of 3 %. We use a growing horizon length up the end of the fedbatch, i.e., to 10 h [1]. Also, remark that the MHE was limited to estimate unmeasured states and did not include model parameters as optimization variables. As a reference, we also plot in Fig. 4 the batch process and fed-batch with open-loop control for the same initial conditions and time frame. A step size $\Delta t = 0.5$ h is used in these control studies to consider manageable sampling times.

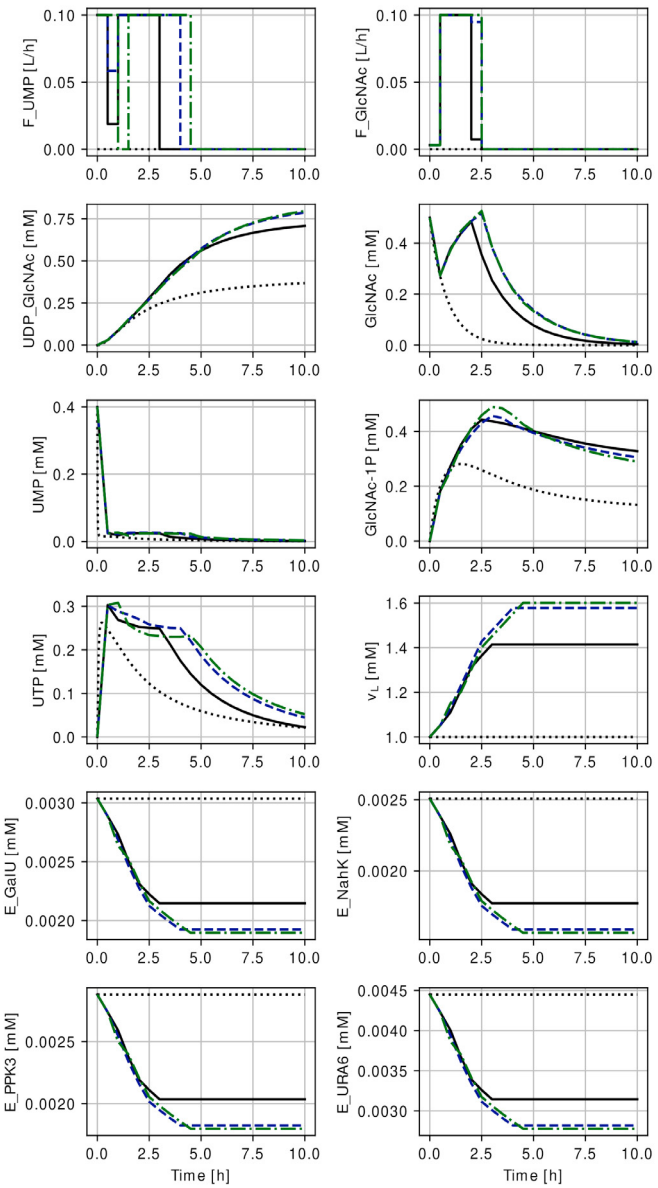


Fig. 4. Control results for the fed-batch UDP-GlcNAc biosynthesis. Case 1: open-loop control (——), case 2: model predictive control with full-state measurement (– – –), case 3: model predictive control with moving horizon estimation (–·–). The batch process (····) is presented for comparison. Only selected intermediate metabolites are shown for simplicity of presentation. Model uncertainty was considered.

The batch process reached a maximum UDP-GlcNAc titer of 0.37 mM, while the open-loop fed-batch system achieved

[1] Since we use all the available measurements from the beginning of the process, the estimation window actually grows but does not *move*. This special type of configuration can be referred to as a full information estimation (Rawlings et al., 2020).

0.71 mM, about 1.9-fold enhancement in volumetric productivity and enzyme efficiency use. In line with the discussion from the previous section, the GlcNAc and UMP feeding rates kept the production pathway working at higher conversion rates compared to the batch. As expected, the improvement with this open-loop optimization was lower than in the scenario presented in Fig. 3 since there we assumed no model-plant mismatch, while now the controller uses a model that does not exactly match the plant behaviour.

Using MPC with full state measurement increased the UDP-GlcNAc concentration up to 0.79 mM, by adapting the feeding rates online to compensate for the model-plant mismatch. Compared to the open-loop profile, the feeding rate of UMP was extended longer, thereby increasing the turnover of the UMP to UTP subpathway, and in turn the overall production rates. Similarly, the MPC delivered a slightly more prolonged GlcNAc feeding. This led to an improvement close to 11.3 % with respect to the open-loop fed-batch. However, one should be aware that this MPC scenario is naturally very optimistic because we assume that all state measurements are available and that there is no measurement noise.

The MPC coupled to MHE, in principle a more realistic scenario with limited state measurements and measurement noise, reached a product titer of 0.80 mM. That is, the MHE/MPC configuration was able to match –and slightly surpass– the performance of the MPC with full state information and no measurement noise, with the added benefit of reducing the number of model states needed to be measured online, from 14 to only five. It is also worth noting that, in particular at the beginning of the process, the UMP feeding rate predictions with MHE/MPC were different compared to those of the MPC with full state information. As the estimation horizon grew, the MHE/MPC predictions started to get better and the controller managed to compensate for the previous suboptimal control actions. The GlcNAc feeding profile was, in contrast, practically the same for both scenarios.

Although the feedback control cases did improve the cell-free bioprocess compared to the system without corrective actions, remark that they still could not fully tackle all model uncertainty. In other words, they did not manage to reach the 0.85 mM UDP-GlcNAc achieved in the most idealistic case with no model-plant mismatch, as presented in Section 6.1. To further improve the performance of the MPC, one could consider an adaptive MPC scheme where the model parameters –along with the unmeasured states– are re-estimated online using the MHE algorithm.

7. CONCLUSION AND OUTLOOK

In this work, we showed a general model-based optimization framework to maximize the production efficiency of cell-free processes, with emphasis on fed-batch regime. The framework can in principle consider several economic objectives and optimization variables, as well as input and state constraints if necessary. Furthermore, we outlined the use of MPC to counteract model uncertainty. We also discussed how an observer such as MHE can be integrated with MPC to infer unmeasured states based on available process information.

Using the *de novo* cell-free UDP-GlcNAc biosynthesis as a case study, we demonstrated the benefits of our optimization strategy for improving the volumetric productivity, product titer, and efficiency of enzyme use per process run compared to conventional batch schemes. This was achieved by optimizing the substrate feeding rates of UMP and GlcNAc. Moreover, we have shown that MPC alone, but optionally also coupled to MHE, can be in general very beneficial to tackle model uncertainty online.

Based on these results, we would like to encourage the synthetic biology and biotechnology community to consider the benefits of control theory and process engineering applied to cell-free biosynthesis. We are convinced that such a systems engineering perspective could make these emerging types of synthetic biology processes more economically efficient, competitive and robust, thus facilitating their commercial implementation. Future work involves a detailed (machine learning-supported) kinetic modeling strategy towards model-based optimization and control of cell-free synthesis, and the application of our framework to different case studies.

REFERENCES

Bellman, R. and Åström, K. (1970). On structural identifiability. *Math. Biosci.*, 7(3), 329–339.

Du, B., Zielinski, D.C., Kavvas, E.S., Dräger, A., Tan, J., Zhang, Z., Ruggiero, K.E., Arzumanyan, G.A., and Palsson, B.O. (2016). Evaluation of rate law approximations in bottom-up kinetic models of metabolism. *BMC Evol. Biol.*, 10(1), 40.

Gadekar, A., Bhowmick, S., and Pandit, A. (2020). A glycotherapeutic approach to functionalize biomaterials-based systems. *Adv. Funct. Mater.*, 30(44), 1910031.

Lim, H.J. and Kim, D.M. (2022). Cell-free synthesis of industrial chemicals and biofuels from carbon feedstocks. *Curr. Opin. Biotechnol.*, 73, 158–163.

Ma, B., Guan, X., Li, Y., Shang, S., Li, J., and Tan, Z. (2020). Protein glycoengineering: an approach for improving protein properties. *Front. Chem.*, 8, 622.

Mahour, R., Klapproth, J., Rexer, T.F., Schildbach, A., Klamt, S., Pietzsch, M., Rapp, E., and Reichl, U. (2018). Establishment of a five-enzyme cell-free cascade for the synthesis of uridine diphosphate N-acetylglucosamine. *J. Biotechnol.*, 283, 120–129.

Mordhorst, S. and Andexer, J.N. (2020). Round, round we go – strategies for enzymatic cofactor regeneration. *Nat. Prod. Rep.*, 37(10), 1316–1333.

Pohlodek, J., Schlauch, B.M.C., Zometa, P., and Findeisen, R. (2022). Flexible development and evaluation of machine-learning-supported optimal control and estimation methods via HILO-MPC. Publisher: arXiv Version Number: 1.

Rawlings, J., Mayne, D., and Diehl, M. (2020). *Model predictive control: theory, computation and design*. Nob Hill Publishing, LLC, Santa Barbara, 2nd edition edition.

Rawlings, J.B. (2013). Moving horizon estimation. In J. Baillieul and T. Samad (eds.), *Encyclopedia of Systems and Control*, 1–7. Springer London, London.

Schnell, S. (2014). Validity of the michaelis–menten equation – steady-state or reactant stationary assumption: that is the question. *FEBS J.*, 281(2), 464–472.

Siedentop, R., Claaßen, C., Rother, D., Lütz, S., and Rosenthal, K. (2021). Getting the most out ofenzyme cascades: strategies to optimize in vitro multi-enzymatic reactions. *Catalysts*, 11(10), 1183.

Tinafar, A., Jaenes, K., and Pardee, K. (2019). Synthetic biology goes cell-free. *BMC Biol.*, 17(1), 64.

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