

Model-based optimization of cell-free enzyme cascades exemplified for the production of GDP-fucose

Nicolas Huber^{a,1}, Edgar Alberto Alcalá-Orozco^{a,1}, Thomas Rexer^{a,b}, Udo Reichl^a, Steffen Klamt^{a,*}

^a Max Planck Institute for Dynamics of Complex Technical Systems, 39106, Magdeburg, Germany

^b eversyn, 39106, Magdeburg, Germany

ARTICLE INFO

Keywords:

Cell-free production
Enzyme cascade
Nucleotide sugar
GDP-Fucose
Kinetic modeling
Model-based optimization
Parameter uncertainty

ABSTRACT

Cell-free production systems are increasingly used for the synthesis of industrially relevant chemicals and biopharmaceuticals. Cell-free systems often utilize cell lysates, but biocatalytic cascades based on recombinant enzymes have emerged as a promising alternative strategy. However, implementing efficient enzyme cascades is a non-trivial task and mathematical modeling and optimization has become a key tool to improve their performance. In this work, we introduce a generic framework for the model-based optimization of cell-free enzyme cascades based on a given kinetic model of the system. We first formulate and systematize seven optimization problems relevant in the context of cell-free production processes including, for example, the maximization of productivity or product yield and the minimization of overall costs. We then present an approach that accounts for parameter uncertainties, not only during model calibration and model analysis but also when performing the actual optimization. After constructing a kinetic model of the enzyme cascade, experimental data are used to generate an ensemble of kinetic parameter sets reflecting their variabilities. For every parameter set, systems optimization is then performed and the resulting solution subsequently cross-validated for all other parameterizations to identify the solution with the highest overall performance under parameter uncertainty. We exemplify our approach for the cell-free synthesis of GDP-fucose, an important sugar nucleotide with various applications. We selected and solved three optimization problems based on a constructed dynamic model and validated two of them experimentally leading to significant improvements of the process (e.g., 50% increase of titer under identical total enzyme load).

Overall, our results demonstrate the potential of model-driven optimization for the rational design and improvement of cell-free production systems. The developed approach for systems optimization under parameter uncertainty could also be relevant for the metabolic design of cell factories.

1. Introduction

Using genetically modified cells as ‘cell factories’ for the production of value-added chemicals is a traditional method in the field of biotechnology. However, cell-based methods are often limited by several factors, including (1) a narrow range of process conditions that ensures the viability of the cells, (2) potential accumulation of cytotoxic metabolites, and (3) reduced product yields due to substrate demands for cell growth and maintenance. Additionally, cellular metabolism is a complex system featuring redundancies and various interconnected regulatory mechanisms making it difficult to identify and implement

suitable intervention strategies to maximize the flux to the desired product. Therefore, cell-free methods are increasingly used as an alternative based on isolated cellular parts to create *in vitro* synthetic reaction networks. In particular, cell-free systems enable a modular and open design and are easier to modify than entire cells. On the other hand, implementation of cell-free production systems, especially for industrial processes, has been hampered primarily by the comparatively high cost of preparative steps such as the production, isolation and purification of pathway enzymes (for a detailed comparison of cell-based and cell-free methods see Claassens et al., 2019). Accordingly, there is an increasing interest in experimental and mathematical optimization methods to

* Corresponding author.

E-mail address: klamt@mpi-magdeburg.mpg.de (S. Klamt).

¹ Equal contribution.

<https://doi.org/10.1016/j.ymben.2023.10.007>

Received 15 June 2023; Received in revised form 18 October 2023; Accepted 22 October 2023

Available online 30 October 2023

1096-7176/© 2023 The Authors. Published by Elsevier Inc. on behalf of International Metabolic Engineering Society. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

enhance the efficiency of cell-free production systems (a comprehensive review can be found in (Siedentop et al., 2021)). Experimentally, cell-free systems can be optimized by a multitude of approaches including process and cascade design strategies, improved enzyme selection and identification of optimal reaction parameters such as substrate and cofactor concentrations, reaction temperature and pH value, and the selection of appropriate solvents. However, experimental optimization can be time-consuming as well as labor- and cost-intensive. Therefore, model-based mathematical optimization can be used as a systemic approach to generate qualitative and quantitative predictions for an improved operation of the cell-free system.

In this work, we consider a cell-free enzyme cascade for the synthesis of the nucleotide sugar GDP-fucose, the combination of guanosine diphosphate (GDP) and the hexose sugar fucose. Nucleotide sugars are the main substrates for Leloir glycosyltransferases that act as energetically activated building blocks for the synthesis of glycans. The latter are oligosaccharides that are often conjugated to other biomolecules such as proteins and lipids. This process is called glycosylation and it is a crucial post-translational modification for the bioactivity of some enzymes. A nucleotide sugar consists of a nucleotide such as cytidine or guanosine connected through one or two phosphate groups to a monosaccharide, the latter being the group that is transferred by the glycosyltransferases. GDP-fucose is a nucleotide sugar acting as donor for fucosylation reactions. Its relevance concerns, for example, the immune system, fertilization and development of mammals (Schneider et al., 2017). In bacteria, fucosylated carbohydrates are part of the synthesis of lipopolysaccharides of cell walls (Ma et al., 2006). Fucose is transferred by fucosyltransferases from its donor GDP-fucose to acceptor biomolecules such as proteins, lipids, or oligosaccharides (Rexer et al., 2021). Such fucosylated biomolecules have many industrial applications, which require a large-scale and cost-efficient synthesis of GDP-fucose. For example, human milk oligosaccharides (HMOs) can be added to infant formula, where they are highly sought after because it has been shown that they play an important role in the healthy development of infants that cannot be breastfed (Bode, 2012).

There have been significant research activities on the synthesis of GDP-fucose including cell-based approaches (Koizumi et al., 2000; Mattila et al., 2000; Ruffing and Chen, 2006; Byun et al., 2007; Lee et al., 2011; Chin et al., 2013; Zhai et al., 2015) and cell-free systems, either with cell lysate (Prohaska and Schenkel-Brunner, 1975; Yamamoto et al., 1984; Stiller and Thiem, 1992) or with purified enzymes (Wang et al., 2009; Zhao et al., 2010; Mahour et al., 2021; Frohnmeyer et al., 2022). In particular, Mahour et al. established a cell-free multi-enzyme cascade for the production of GDP-fucose from the inexpensive substrates fucose, guanosine (or guanosine monophosphate (GMP); see below) and polyphosphate. Polyphosphate, together with a polyphosphate kinase, enables the regeneration of ATP (necessary for the phosphorylation of fucose and guanosine or GMP) avoiding the costly addition of large amounts of ATP (Mahour et al., 2021). In this work, we aimed to use mathematical optimization techniques to improve the efficiency of this cascade with respect to different objectives. As starting point, we considered a slightly modified version of the cascade of Mahour et al. (Fig. 1) where GMP instead of guanosine was chosen as substrate to avoid experimental difficulties related to the solubility of guanosine.

Apart from empirical and statistical modeling and optimization approaches (Chen et al., 2013; Onyeogaziri and Papanephytous, 2019), mechanistic (kinetic) models have been frequently utilized for the model-based optimization of cell-free production systems and will also be used herein. Kinetic models are based on first principles and usually comprise a set of ordinary differential equations (ODEs) describing the dynamics of metabolite concentrations by rate laws being functions of reaction stoichiometries and reaction kinetics. Generally, model-based optimization requires four steps: 1) model construction, 2) model parameterization, 3) optimization, and 4) validation. One common challenge is the selection of proper parameters. Some parameters can be

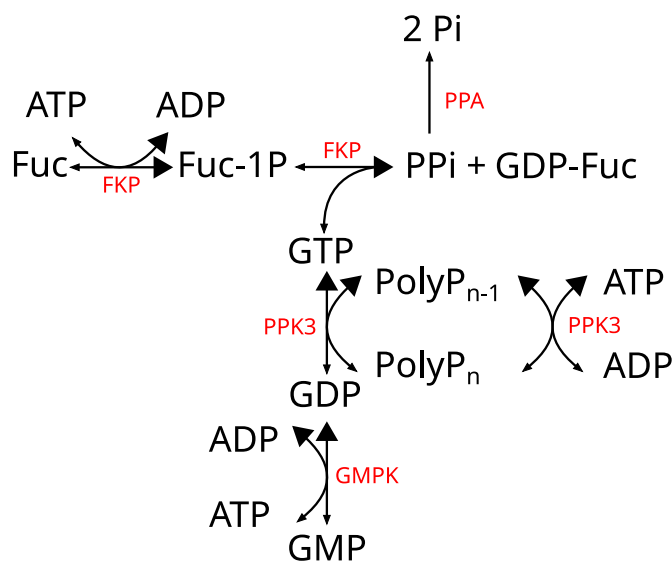


Fig. 1. Enzyme cascade for the production of GDP-fucose (adapted from Mahour et al., 2021). Enzymes are shown in red, metabolites in black. Abbreviations: Fuc: fucose; Fuc-1P: fucose-1-phosphate; ATP: adenosine triphosphate; ADP: adenosine diphosphate; PPi: inorganic pyrophosphate; GMP: guanosine monophosphate; GDP: guanosine diphosphate; GTP: guanosine triphosphate; PolyPn: polyphosphate of chain length n; Pi: inorganic phosphate; FKP: bifunctional fucokinase/L-fucose-1-P-guanylyltransferase; PPA: inorganic diphosphatase; GMPK: guanylate kinase; PPK3: polyphosphate kinase.

derived from literature and databases or must be estimated from experimental data. However, parameters are often neither available from knowledge bases nor precisely identifiable from available experimental data and tackling the associated uncertainties should therefore be a crucial part of model parameterization. Suitable approaches to deal with those uncertainties are, for example, Bayesian modeling (Wilkinson, 2007; Linden et al., 2022), ensemble modeling (Tran et al., 2008; Jia et al., 2012; Tan and Liao, 2012; Villaverde et al., 2022), robust optimization (Bertsimas et al., 2011; Gabrel et al., 2014; García and Peña, 2018; Puschke et al., 2018) or combinations thereof (Liu and Gunawan, 2017). However, to the best of our knowledge, those methods have not been used in the context of cell-free enzyme cascades so far. Once an appropriate kinetic model has been constructed and parameterized, the model is utilized to identify possible interventions to optimize the process. In the majority of previous works, this was based, for example, on *in silico* experiments (e.g., simulation of selected changes of enzyme and substrate concentrations (Korman et al., 2017; Česnik et al., 2020)) or on an analysis of parameter sensitivities (e.g., by means of metabolic control analysis (Shen et al., 2020)) to identify promising optimization targets. An alternative to those manual optimization approaches is a rigorous systematic mathematical optimization featuring well-defined objective functions (e.g., maximization of titer or of productivity or minimization of costs) subject to certain various constraints in the design variables (e.g., bounded concentrations of substrates or enzymes). With some exceptions (e.g., Ardao and Zeng, 2013; Rollin et al., 2015; Hold et al., 2016; Finnigan et al., 2019; Paschalidis et al., 2022), those rigorous directed optimization approaches have been rarely applied in the context of cell-free processes and we could not find any study in this field that took parameter uncertainties into account when optimizing the system.

Due to these reasons, in this work we present a general workflow for (a) setting up relevant optimization problems in the context of cell-free enzyme cascades that (b) enables the consideration of parameter uncertainties when performing the actual optimization. As a showcase we employed our approach for optimizing the enzyme cascade of GDP-fucose synthesis: we constructed a kinetic model of this cascade and parameterized it repeatedly with experimental time course data

resulting in an ensemble of models each associated with one estimated parameter set. This model ensemble then served as basis for the optimization of the enzyme cascade, again accounting for uncertainty arising when optimizing the ensemble of models. We selected three of seven proposed objective functions to optimize efficiency and/or costs of the system under parameter uncertainty and validated two of the three predictions experimentally leading to significant improvements in the performance of the cascade.

2. Methods

2.1. Recombinant protein expression

All enzymes were produced as described previously in Mahour et al. (2021). Briefly, each enzyme's DNA sequence was cloned into a pET28a (+) vector (pET100/D-TOPO for FKP), modified with a 6x His-tag at the N-terminus for ease of purification and codon optimized for expression in *E. coli* by BioCat GmbH. The host chosen for recombinant expression was an *E. coli* BL21 (DE3) strain from New England Biolabs GmbH and their recommended high efficiency transformation protocol was used.

Protein expression was carried out in 1 L shake flasks containing 200 mL media supplemented with antibiotics. When an OD₆₀₀ of 0.8–1 was observed, protein expression was induced by adding isopropyl β -D-thiogalactopyranoside (IPTG) to the shaker. The culture was harvested after 8–12 h. The biomass was separated from the media by centrifugation at 6000 \times g for 10 min.

2.2. Enzyme purification

Cells were collected and lysed by a high-pressure homogenizer. Cell debris was removed by centrifugation and the supernatant was subjected to affinity chromatography for target protein purification. To remove imidazole, standard buffer exchanges were carried out using centrifugal filters. All enzyme stocks were stored at $-20\text{ }^{\circ}\text{C}$ with 50% glycerol.

SDS-PAGE (BIORAD Laboratories Inc.) was carried out to identify presence and purity of enzymes. The Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific) was used to estimate protein concentration.

2.3. Enzyme cascade reactions

All enzyme cascade reactions were set up by mixing MgCl₂, GMP, fucose, ATP and polyphosphate with the corresponding amount of enzymes (GMPK, PPK3, FKP and PPA). The reactions were performed in biological triplicates at 37 °C Thermoblock (Eppendorf AG). The reaction volume was 200 μ L in 1.5 mL Eppendorf micro tubes. Sampling was done by recovering 6 μ L from the reaction and diluting it with 1500 μ L Ultra-pure Milli-Q water, the samples were analyzed by HPAEC-UV shortly after and stored at $-20\text{ }^{\circ}\text{C}$.

2.4. HPAEC-UV

High Performance Anion-Exchange Chromatography with UV detection (HPAEC-UV) was used for detection and quantification of the UV-active components of the reaction. A Dionex IC5000 chromatography system (ThermoFisher Scientific Inc.) was used with a CarboPac PA200 analytical (3 mm \times 250 mm) and guard column (3 mm \times 50 mm). An elution gradient reaching 1 M NaOAc and 1 mM NaOH over 18 min was used with an injection volume of 25 μ L. Analytical standards were prepared for identification and quantification of GMP, GDP, GTP, AMP, ADP, ATP and GDP-fucose in the range of 2–100 μ M.

2.5. Model and simulation

The kinetic model and its implementation are described in the

Appendix. Model building, simulation, parameter estimation, and optimization have been performed using the COPASI software (version 4.35, build 258) (Hoops et al., 2006) available at www.copasi.org. For model simulation we used COPASI's deterministic LSODA method (Hindmarsh, 1983), a single-shooting approach that selects between Adams and BDF methods depending on the detected stiffness of the system. For parameter estimation and for solving the formulated optimization problems, we used COPASI's implementations of Evolutionary Strategy (SRES) (Runarsson and Yao, 2000) and Genetic Algorithm (Bäck and Schwefel, 1993; Michalewicz, 1994; Mitchell, 1998), respectively. Both numerical methods attempt to find the global minimum of a given objective function and their calculation schemes are inspired by the eponymous biological phenomena and feature stochastic elements that facilitate the escape of local minima during the search. Due to the stochastic nature of the parameter estimation algorithm different optimal parameter sets are generated in different runs (even when starting with the same seed).

All model files and scripts relevant to this work and enabling reproduction of the calculated results can be found in the following repository: https://github.com/klamt-lab/GDP-Fucose_OptMet2023_Paper.

3. Results

In the following, we present an integrated framework for the model-based optimization of cell-free systems under parameter uncertainty and exemplify each step with the enzyme cascade for GDP-fucose synthesis as an example.

3.1. Model construction and first round of parameter estimation

We implemented the cell-free reaction network for synthesis of GDP-fucose (Fig. 1) as a kinetic (ODE) model $\dot{x}(t) = f(x(t), t)$, which is described in detail in the Appendix. The model consists of 14 state variables ($x(t) \in \mathbb{R}^{14}$) and 29 parameters. Several parameter values were obtained from databases and literature, but only three equilibrium constants were considered to be certain and fixed, while the other 26 parameter values were estimated from a parameter fitting procedure where parameter values found in literature or databases served as start values for the estimation. For the latter we used data of time course measurements from multiple one-pot assays with varying initial concentrations of substrates and enzymes (for available experimental data see Supplementary Table 1). Notably, we did not perform exhaustive experimental characterizations of the kinetics of individual enzymes due to time and cost limitations and the fact that the kinetics of a single enzyme often change when operating in an enzyme cascade (e.g., due to substrate channeling, molecular crowding, protein/protein interactions, allosteric modifications of enzymes by pathway intermediates etc.). Parameter estimation and all model simulations were performed with the COPASI software (Hoops et al., 2006); also see section 2.5.

3.2. Exploration of model variability due to parameter uncertainty

Kinetic models with a large set of (mainly unknown) kinetic parameters fitted against experimental data often hold uncertainties due to the non-identifiability of some parameters. To deal with this problem, we pursued an ensemble modeling approach. We derived an ensemble of kinetic model parameter sets from 100 repeated parameter estimations from the available time course data. Due to the stochastic nature of the used global solution algorithm (see section 2.5) this led to 100 different optimal parameter sets (Fig. 2). The sum of squared residues (SSR), which is the objective to be minimized during the estimation (see Appendix), is shown in the last histogram in Fig. 2 indicating that all identified parameter sets are relatively close to the optimal SSR value of 90.2; we therefore considered all parameter sets as equally relevant.

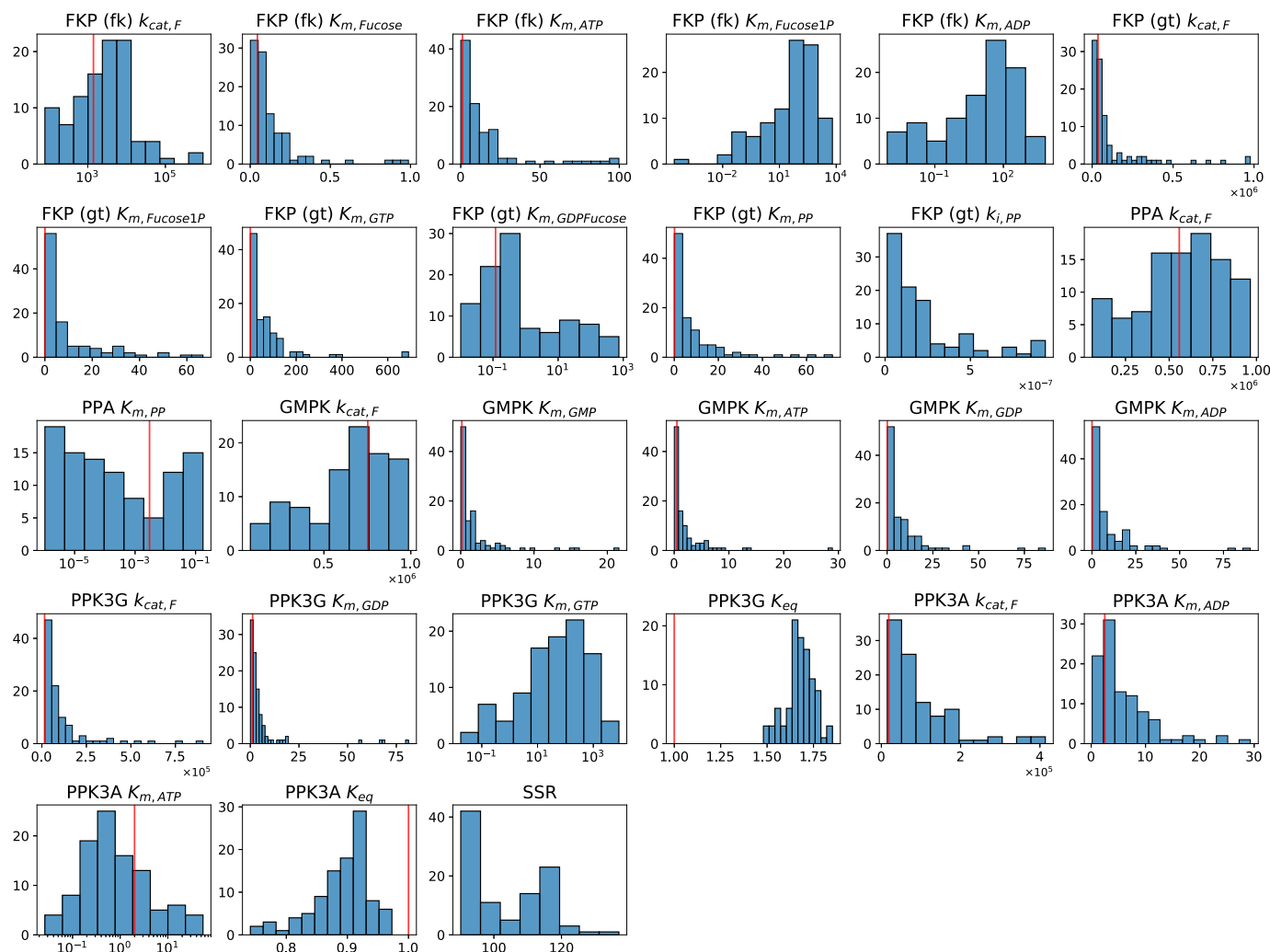


Fig. 2. Distributions of the 26 unknown parameter values of the kinetic model derived from repeated global parameter estimations ($n = 100$). Parameter estimations were carried out using the Evolutionary Strategy (SRES) global solution algorithm as implemented in COPASI with default settings (see section 2.5). Red lines show literature values, which, if available, were used as start values for the estimations. Histogram x-axes of parameters where estimated values cover multiple orders of magnitude were logarithmized to improve legibility. The sum of squared residues (SSR) is the objective value of the least-squares parameter estimation to be minimized (see Appendix). Y-axes show absolute frequency, x-axes show parameter values with different units according to parameter type (K_m, k_i : [mM], k_{cat} : [1/h], K_{eq} : dimensionless quantity).

Overall, most distributions of the estimated kinetic parameters cluster around well-defined values, which are in many (but not all) cases close to the respective literature values (if the latter were available). After the 100 rounds of parameter estimations, the model was simulated with each of the 100 parametrizations. For these simulations, the initial substrate and enzyme concentrations were set according to experiment 5, one of the five experiments included in the parameter estimation procedure (see Supplementary Table 1 and Supplementary Fig. 5) and which was selected as representing the (non-optimized) baseline performance (Fig. 3). The qualitative trends of the simulated time courses correspond well to the experimental data for all measured species, especially for the initial substrate GMP and the final product GDP-fucose. A larger deviation can be observed for ADP and ATP. This can be attributed to the fact that the sum of ADP and ATP is fixed in the model, while we found that some AMP occurs as an unintended by-product of the cascade affecting the adenosine balance. The spread across all 100 simulated trajectories per species is greatest for fucose and fucose-1-P, which was to be expected since they have not been measured causing non-identifiability of their trajectories. Despite these uncertainties, the model provides a good description of the underlying

process (in particular for the relevant readout of product GDP-fucose) and as such it was later utilized for model-based optimization.

3.3. Model-based optimization of cell-free systems: a zoo of relevant optimization problems

Similar to traditional chemical processes, cell-free synthesis can be optimized with respect to certain objectives such as the maximization of production rate (productivity; in [mmol/L/h] or [g/L/h]), product titer (e.g. [mM]) or [g/L]) and product yield (e.g., mmol product per mmol substrate), or the minimization of certain costs. There are often various constraining factors, which limit the space of available solutions. For example, the productivity can usually be enhanced by increasing the enzyme concentrations but this will in turn increase the costs for which an upper bound might be given. The formulation of mathematical optimization problems can help clarify the exact objectives and constraints and these problems can be solved numerically to identify the changes necessary to get the optimal objective value. Due to the many possible combinations of objectives, optimization variables, and constraints, a plethora of relevant optimization problems can theoretically

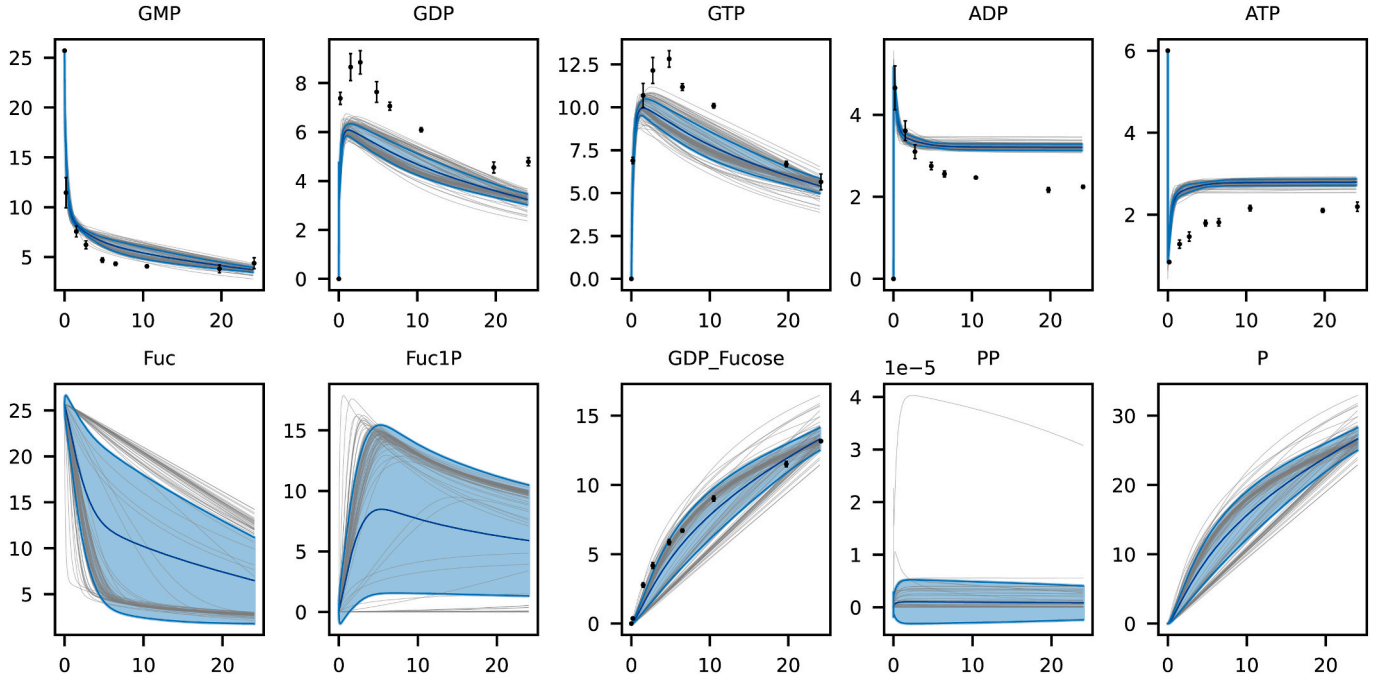


Fig. 3. Model simulations based on the ensemble of 100 estimated sets of kinetic parameters. Black dots represent the measured concentrations of the baseline experiment 5 (see [Supplementary Table 1](#)) as averages of triplicates with error bars showing the associated standard deviation. Grey lines: individual simulation trajectories, central blue line: average across all simulation trajectories, outer blue lines: average trajectory plus and minus associated standard deviation respectively. X-axes show time in hours, y-axes show concentration in mM.

be constructed. In the following section, we formulate and systematize optimization problems (operating on a given kinetic model of the reaction network), which we consider most relevant for improving the performance of cell-free systems operating in a batch process. In a later section, we then compute solutions of selected problems within the specific context of the GDP-fucose model and test the obtained predictions experimentally.

A parameterized kinetic model of a cell-free enzyme cascade can be used to solve dynamic optimization problems with the following general formulation:

$$\max_{\mathbf{x}^0} \text{ or } \min_{\mathbf{x}^0} F(\mathbf{x}^0, \mathbf{x}(t_{end})) \quad (1a)$$

s.t.

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), t), \forall t \in [t_0, t_{end}], \mathbf{x}(t_0) = \mathbf{x}^0 \quad (1b)$$

$$lb_i \leq x_i^0 \leq ub_i, \forall i = 1, \dots, n_x \quad (1c)$$

$$g_j(x_i^0, \mathbf{x}(t_{end})) \leq \alpha_j, \forall j = 1, \dots, n_g \quad (1d)$$

F is the actual objective function that is either maximized or minimized (e.g., maximization of product titer at the end of the process). The vector $\mathbf{x}(t)$ represents the state variables of the system at time point t determined by the kinetic model (differential equations) together with the initial state \mathbf{x}^0 at the predefined starting time t_0 . In our context, $\mathbf{x}(t)$ typically comprises the enzymes $\mathbf{x}_E(t)$ and substrates $\mathbf{x}_S(t)$. Importantly, the initial values \mathbf{x}^0 are the free (design) variables of the entire optimization problem, which can be adjusted to optimal values minimizing/maximizing the objective function F . The latter is a function of the initial concentrations \mathbf{x}^0 (as cost values) and the final concentrations $\mathbf{x}(t_{end})$, which themselves depend on the start values \mathbf{x}^0 . The initial values are usually constrained by lower (lb_i) and upper (ub_i) bounds for each of the n_x entries in \mathbf{x}^0 as formulated by the inequalities in (1c). In the following, we split the boundary constraints for enzymes and substrate concentrations:

$$lb_{E_i} \leq x_{E_i}^0 \leq ub_{E_i}, \forall i = 1, \dots, n_E \quad (2)$$

$$lb_{S_j} \leq x_{S_j}^0 \leq ub_{S_j}, \forall j = 1, \dots, n_S \quad (3)$$

Without specifying upper bounds, the optimization solver will, in many cases, choose infinite substrate and enzyme concentrations when maximizing, for example, product titer or yield. Furthermore, there are often natural or experimental limitations for substrate and enzyme concentrations that can be used to specify the boundary values.

Apart from the starting time t_0 , the second relevant time point is t_{end} , the predefined endpoint of the process. Accordingly, $\mathbf{x}(t_{end})$ contains the values of \mathbf{x} at the end of the process which are obtained by simulating the ODE system with a given \mathbf{x}^0 for the time interval $[t_0, t_{end}]$. During the optimization, the solver will use this information to adjust the free optimization variable \mathbf{x}^0 eventually optimizing the system. Although the discrete set of optimization variables contained in \mathbf{x}^0 are time-invariant, it should be noted that the optimization problem itself is usually classified as a dynamic optimization problem since it includes the dynamic (ODE) model as dynamic constraint (Biegler, 2010).

Importantly, apart from allowed ranges for the initial concentration values specified by (2) and (3), additional constraints on these design variables as well as on the states at the end of the process ($\mathbf{x}(t_{end})$) can be included. In eq. (1d), this is indicated by n_g many functions g_j of \mathbf{x}^0 and $\mathbf{x}(t_{end})$, whose values must lie below specified bounds α_j . Constraints that are often additionally (or alternatively) used to the boundary constraints (2) and (3) are the maximal overall enzyme (or biocatalyst) load E_{tot} .

$$E_{tot} = \sum_{i=1}^{n_E} k_i x_{E_i}^0 \leq E_{tot,max} \quad (4)$$

as well as the maximal substrate load S_{tot} .

$$S_{tot} = \sum_{j=1}^{n_S} k_j x_{S_j}^0 \leq S_{tot,max} \quad (5)$$

Clearly, the specific upper bounds in (2) and (3) may already define

an implicit upper bound for the overall substrate and enzyme loads, but often more constrained overall bounds are used in (4) and (5), e.g., to reflect available substrate and enzyme budgets. The weights k_i and k_j can either all be set to 1 so that the sum describes the total amount of enzyme or substrate, respectively, or they may represent known costs. We denote constraints (2)–(5) as *basic capacity constraints* as they are relevant for all optimization problems (though possibly only a (non-empty) subset of them may be active).

Other often relevant constraints are a demanded minimum final concentration (titer) of product P

$$x_P(t_{end}) \geq x_{P,min} \quad (6)$$

and/or a minimum demanded product yield

$$Y_{P/S_Y} \geq Y_{P/S_Y,min}, \quad (7)$$

where the product yield Y_{P/S_Y} is defined as the ratio of the product concentration at the given endpoint of the process and the initial concentration of the relevant (reference) substrate x_{S_Y} :

$$Y_{P/S_Y} = \frac{x_P(t_{end})}{x_{S_Y}^0}. \quad (8)$$

Instead of a single reference substrate, a (weighted) sum of substrates could be used in the denominator. Note that a multiplication of inequalities (6) and (7) with -1 transforms these constraints to the form $g_j(x_i^0, x(t_{end})) \leq \alpha_j$ used in (1d).

In the following, we will now formulate seven relevant optimization problems for cell-free enzyme cascades, which are all characterized by their specific objective functions and constraints.

3.3.1. Objective O1: maximization of product titer (and productivity)

A common objective of any (bio)chemical production system is to maximize the final amount (usually concentration) of the target compound, i.e., the final titer of the product ($x_P(t_{end})$). In the context of a cell-free enzyme cascade, the possible titer primarily depends on the initial concentrations of substrates and enzymes and on the chosen time point t_{end} where the cascade is stopped and the product harvested. Depending on the kinetic and regulatory mechanisms of the system, an increase of the initial concentration of a substrate could increase the titer linearly but also lead to a reduction of the titer, e.g., due to substrate inhibition. Using the base constraints as only relevant side conditions, the optimization problem for maximizing the product titer thus reads:

$$\max_{x_E^0, x_S^0} x_P(t_{end}) \quad (9)$$

s.t. ODE model (eq. (1b)) and basic capacity constraints (eqs. (2)–(5)).

It should be noted that, due to the fixed endpoint of the process (t_{end}), the solution found for the maximization of the titer $x_P(t_{end})$ always corresponds to the solution for maximization of the productivity (volumetric rate or space-time yield) given by $x_P(t_{end})/t_{end}$.

3.3.2. Objective O2: maximization of product yield

Another relevant performance measure of cell-free systems is the product yield Y_{P/S_Y} (eq. (8)) quantifying the conversion efficiency. As for objective O2, the base constraints of eqs. (2)–(5) are relevant here as well, but in addition we demand a minimum product titer (eq. (6)). As the yield is a quantity relative to the initial substrate concentration, high yields do not automatically guarantee high titers. Without this constraint, it may happen that a high product yield is reached but with a very low titer because only little substrate has been converted. The optimization problem for maximizing the product yield thus reads:

$$\max_{x_E^0, x_S^0} Y_{P/S_Y} \quad (10)$$

s.t. ODE model (eq. (1b)), basic capacity constraints (eqs. (2)–(5)), and

minimum product titer (eq. (6)).

3.3.3. Objective O3: minimization of enzyme load

A major cost factor of cell-free production processes relates to the preparative procedures necessary to supply cell lysate or purified enzymes. Hence, there is an interest to minimize the total enzyme amount of the process, i.e., the enzyme load, for a minimum required product titer $x_{P,min}$ and/or yield $Y_{P/S_Y,min}$:

$$\min_{x_E^0, x_S^0} E_{tot} = \sum_{i=1}^{n_E} k_i x_{E_i}^0 \quad (11)$$

s.t. ODE model (eq. (1b)), basic capacity constraints (eq. (2), (3), (5)), minimum product titer (eq. (6)) and/or yield (eq. (7)).

The objective function represents the enzyme load when all k_i are set to 1, but alternative weights (e.g., reflecting enzyme costs) can be used. The basic capacity constraint (4) is not usually needed here since the total enzyme load is minimized by the objective function.

3.3.4. Objective O4: combined maximization of product yield and titer

Optimization of product yield and titer (or productivity) may lead to different solutions due to some inherent trade-offs. Therefore, it might be of interest to optimize a combination of both in a multi-objective optimization problem. A suitable objective function could be a weighted sum of both individual targets and the weights w_1 and w_2 can be chosen according to which goal is deemed to be more important. In order for the weights to be effective, the value ranges of the objectives need to be accounted for.

$$\max_{x_E^0, x_S^0} w_1 \cdot Y_{P/S_Y} + w_2 \cdot x_P(t_{end}) \quad (12)$$

s.t. ODE model (eq. (1b)) and basic capacity constraints (eqs. (2)–(5)).

In principle, a weighted sum in the objective function can be applied to any combination of optimization targets. Another approach to multi-objective optimization is based on Pareto optimization, where a set of best compromises between different objectives is identified (Paschalidis et al., 2022).

3.3.5. Objective O5: minimization of overall costs

The high prices of many components necessary for cell-free production makes minimizing the process costs a crucial target in industrial applications. A suitable cost term could be defined as follows

$$C = \left(\sum_{i=1}^{n_E} p_{E_i} \cdot x_{E_i}^0 \right) + \left(\sum_{j=1}^{n_S} p_{S_j} \cdot x_{S_j}^0 \right) + \epsilon \quad (13)$$

where p_E and p_S are the prices of enzymes and substrates, respectively, and ϵ represents additional fixed process costs. A minimum titer $x_{P,min}$ needs to be defined to exclude the trivial solution of an inactive process. Accordingly, the costs then refer to this amount of product synthesized. The overall optimization problem reads:

$$\min_{x_E^0, x_S^0} C \quad (14)$$

s.t. ODE model (eq. (1b)), basic capacity constraints (eqs. (2)–(5)), and minimum product titer (eq. (6)).

3.3.6. Objective O6: minimization of overall costs normalized by product titer

An extension of the cost optimization O5 is to normalize the costs by the obtained product titer in the objective function:

$$\min_{x_E^0, x_S^0} \frac{C}{x_P(t_{end})} \quad (15)$$

s.t. ODE model (eq. (1b)), basic capacity constraints (eq. (2), (3), (5)),

minimum product titer (eq. (6)) and yield (eq. (7)).

In addition to the constraints used for objective O5 (eq. (14)), a minimal product yield should be demanded. Without the latter, a solution may be found where the minimum demanded titer is reached by adding a large amount of an inexpensive substrate leading to an undesirably low conversion rate.

3.3.7. Objective O7: maximization of profit

The goal of this optimization problem is to maximize the overall profit of the process Pf , which can be defined as

$$Pf = (I - C) \cdot V \quad (16)$$

with income $I = p_P \cdot x_P(t_{end})$, costs C (according to eq. (13)) and the volume of the reaction vessel V . The complete optimization problem reads:

$$\max_{x_E^0, x_S^0} Pf \quad (17)$$

s.t. ODE model (eq. (1b)), basic capacity constraints (eq. (2), (3), (5)), minimum product titer (eq. (6)) and yield (eq. (7)).

3.4. Solving optimization problems under parameter uncertainty and experimental validation

We selected three of the optimization problems defined above and solved them for our enzyme cascade for the synthesis of GDP-fucose. As a conceptual contribution, we propose a method how to deal with parameter uncertainties when determining the optimal solution. For two of the three optimizations we compare the predictions with data from validation experiments.

3.4.1. Maximization of product titer (objective O1)

In the first optimization we used objective O1 (eq. (9)) to maximize the product titer in our cascade for a given maximal enzyme load. As mentioned earlier, with a fixed end time t_{end} this is equivalent to maximizing the productivity. We used the following specifications: the end time t_{end} was set to 24 h and all initial substrate concentrations were considered fixed according to the initial conditions of the baseline experiment (25.7 mM fucose, 25.7 mM GMP, 6 mM ATP). Finally, the enzyme concentrations were used as free optimization variables (Table 1). Note that, in our setup, the enzyme concentrations remain constant over the course of the process, hence, the initial enzyme concentrations correspond to the enzyme concentrations throughout the process. In the following, we will thus only refer to “enzyme concentrations”. The maximal total enzyme concentration $E_{tot,max}$ was fixed to 0.0469 mM, which corresponds to the total enzyme load of the baseline experiment. All weights in the total enzyme load constraint (4) were set to 1. With this setting, what is actually optimized is the *enzyme ratio* to obtain the maximal product titer.

To account for the parameter uncertainty, we solved the optimization problem for each of the 100 parameter sets determined in section 3.2 resulting in 100 predictions representing optimal sets of enzyme concentrations (optimization was performed using the genetic algorithm as implemented in COPASI; see section 2.5). The distributions of the identified optimal enzyme concentrations are shown in Fig. 4.

All predicted solutions feature a strong increase of the FKP

Table 1

Constraints for the enzyme concentrations used for optimization of product titer.

	Lower boundary [mM]	Start value [mM]	Upper boundary [mM]
FKP	0.000210	0.00663	0.0149
GMPK	0.00241	0.00636	0.507
PPA	0.00319	0.0311	0.224
PPK3	0.000180	0.00288	0.0129
E_{tot} [mM]	–	0.0469	0.0469

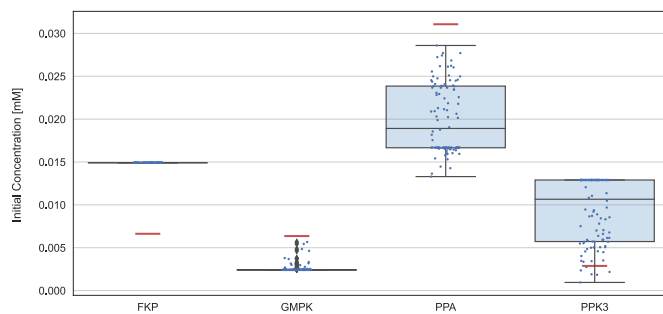


Fig. 4. Distribution of predicted optimal enzyme concentrations according to objective O1 for each of the 100 parameter sets contained in the ensemble computed in section 3.2. Box plots show 25%, 50%, and 75% quartiles, scatter plots show the complete set. Red lines mark the starting value for each enzyme (as used in the baseline experiment).

concentration to its upper boundary. To allow for this substantial increase in FKP concentration while satisfying the maximum enzyme load constraint, the optimization algorithm chose to reduce the concentrations of GMPK (all solutions at or close to the lower boundary of GMPK) and PPA (with varying concentrations within the boundaries). Predicted optimization results of PPK3 are more ambiguous yet a trend is visible with most solutions recommending an increase and only few recommending a slight reduction. These suggested optimizations of the enzyme concentrations are in line with the importance of the enzymes for the performance of the cascade. In particular, the bifunctional enzyme FKP is involved in both the phosphorylation of fucose and the formation of the final product GDP-fucose. Therefore, it has a strong influence on the flux of the entire sugar branch of the cascade and as such directly on the product titer. The enzyme PPK3 is responsible for maintaining a sufficiently high level of ATP by regenerating it from ADP and polyphosphate while also being involved in the phosphorylation of GDP thereby supplying GTP, a direct substrate of the product-forming reaction.

While some trends can be observed from the results shown in Fig. 4, it still remains unclear which of the 100 optimal solutions (one for each of the 100 parameter sets) should be selected since it is not known which parameter set extrapolates the real systems dynamics best. In order to identify the most suitable solution for experimental testing, a cross validation was carried out: the performance of each optimal solution was tested across all 100 model instances associated with the ensemble of 100 estimated parameter sets. Thus, each optimal set of enzyme concentrations $x_E^{0,h}$ (obtained from titer optimization under parameter set with index h) was tested for all 100 parameterizations - including the parameter set h itself, for which the optimal enzyme concentration set $x_E^{0,h}$ was originally determined - and the resulting titer of GDP-fucose was determined by simulating the respective system. Accordingly, 100 by 100 simulations were run and the obtained simulated GDP-fucose titers at 24 h are displayed in Fig. 5.

Fig. 5 shows that the performance of the 100 parameter sets (rows) with respect to the optimal enzyme concentration sets (columns) is subject to a substantial amount of variance underlining the need to analyze the performance of optimization results across different parameter sets. The simulated titers also vary depending on the selected optimization result (column), albeit to a much lesser degree. We used the following scoring scheme to select the best result across all columns i :

$$Score(i) = \frac{M_i}{M_{max}} + \frac{T_i}{T_{max}} + \frac{S_i}{S_{max}}, \quad (18)$$

where M_i is the median titer of column i , which is normalized to the overall highest median titer M_{max} , T_i is the minimum titer of column i , which is normalized by the overall highest minimum titer T_{max} , and S_i is

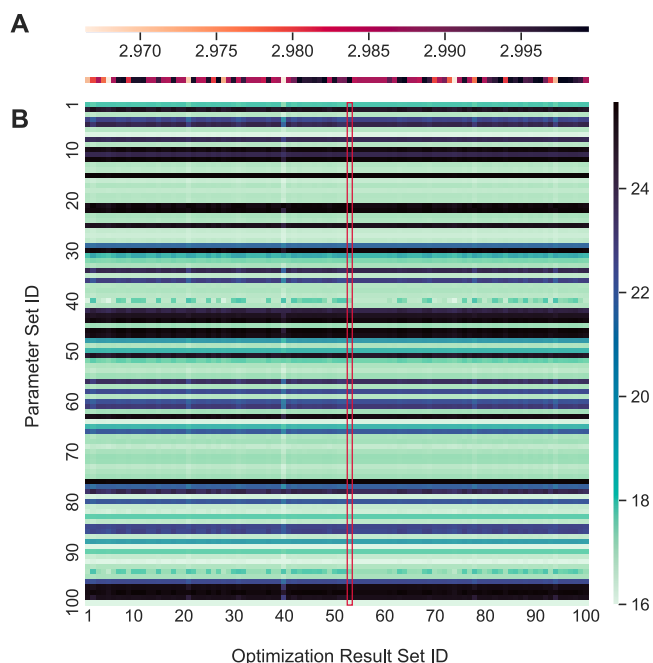


Fig. 5. (A) Scores of the columns of the titer heatmap shown in (B) according to equation (18). The best (highest) scoring column (optimization result ID 53) is highlighted in (B). (B) Heatmap of simulated GDP-fucose titers resulting from subsequent simulations of the model with (a) enzyme concentrations set to the optimized enzyme concentrations determined for each of the 100 parameter sets (x-axis) and (b) the kinetic parameters set to the 100 parameter sets of the estimated ensemble (y-axis). As an example, the color of the matrix element in row 32 and column 49 indicates the obtained titer when using the optimization result (optimal enzyme concentrations) calculated with the model set to parameter set 49 in a model parameterized with parameter set 32. The coloring is based on the simulated GDP-fucose titer at 24 h in mM.

the sum over all titers of column i , which is normalized by the overall highest titer sum S_{max} . This score was calculated for each column of the heat map (Figure 5A) identifying the column with ID 53 as the optimization result performing best across all parameter sets. It uses the following enzyme concentrations: FKP 0.014910 mM, GMPK 0.002413 mM, PPA 0.020148 mM, and PPK3 0.009456 mM.

We used this optimal enzyme concentration set and simulated the time courses for all 100 estimated parameter sets (Fig. 6). We then performed an experiment where the enzyme concentrations were set to the identified optimal values. The corresponding measurements of this experiment are also shown in Fig. 6. As initial substrate concentrations in the simulations, we used the actual (measured) start concentrations of these compounds (fucose 22 mM, GMP 22 mM, ATP 5 mM).

Note that the initial concentrations of the primary substrates fucose and GMP were not exactly the same for the unoptimized baseline experiment (e.g., 25.7 mM GMP) and the validation experiment (22 mM GMP) due to variations inherent to the experimental set-up. However, even with the reduced amount of the substrate GMP (−14.4 %), the measured titer could be improved from 13.16 mM to 19.86 mM (+50.1 %). Additionally, the yield with respect to the initial GMP concentration could be increased from 51.2 % to 90.3 %.

3.4.2. Minimization of enzyme load (objective O3)

Next, we selected optimization problem O3 to minimize the enzyme load according to eq. (11) using the following specifics. The factors k_j were set to the molecular weights of each enzyme to convert the enzyme load from mM to grams per liter. A minimum titer of 15 mM was demanded, which is higher than the titer obtained in the baseline experiment (13.16 mM) and corresponds (with the initial fucose

concentration of 25.7 mM in the baseline experiment) to a minimum yield of 58 %. The start value as well as the lower and upper bounds of the enzyme concentrations were the same as in Table 1 (in the changed units, the start value of the total enzyme load corresponds to 1.55 g/l and no bounds were given for the enzyme load for this optimization). As in the previous scenario, the initial substrate concentrations were fixed (25.7 mM fucose, 25.7 mM GMP, and 6 mM ATP) and the end time of the process was again set to 24 h.

As before, we solved the optimization problem for each of the 100 estimated parameter sets of the ensemble leading again to 100 optimal predictions visualized in Fig. 7 (distribution of enzymes in the optimal solutions) and Fig. 8C (resulting optimal total enzyme load for each of the 100 parametrizations).

Despite the important role of the enzyme in removing pyrophosphate (the inhibitor of the guanylyltransferase activity of FKP), all predicted optimal PPA concentrations are reduced, in many cases significantly (Fig. 7). Even more apparent, GMPK is reduced in all cases to its lower boundary. The predicted optimal enzyme concentration of FKP is increased in the vast majority of cases, though to a lower extent compared to the objective of titer optimization (cf. Fig. 4), and the concentration of PPK3 is mostly close to the start value with a clear trend towards a slight reduction. Overall, these results underline the importance of FKP for the production of GDP-fucose since it is the only enzyme, which is not decreased when attempting to minimize the overall enzyme load despite its significantly higher molecular weight compared to the other enzymes (molecular weights: FKP 105,660 g/mol, GMPK 23,592 g/mol, PPA 19,313 g/mol, PPK3 34,740 g/mol).

We simulated the model with all combinations of (100) optimized enzyme concentrations and (100) kinetic parameter sets which resulted in the heat map in Fig. 8B showing the obtained product titers at 24 h. The simulated titers vary with the selected parameter set and with the selected optimization result. In order to identify the best prediction, which should take into account the enzyme load as well as the achievable titer, the following score was calculated for each column i

$$Score(i) = 10 \cdot \frac{E_{min}}{E_i} + 1 \cdot (T_i - 15), \quad (19)$$

where E_{min} is the lowest of all optimized enzyme loads, E_i is the enzyme load of column i of the enzyme load heatmap, and T_i is the minimum titer of column i of the titer heat map compared to the titer constraint of the optimization setup (15 mM). This comparison is reasonable because titers of the heat map are not guaranteed to fulfill the minimum titer constraint of the preceding optimization since the model is now simulated with combinations of enzyme concentrations and kinetic parameters for which no optimization was performed. Since identifying a low enzyme load was the main objective of this optimization and because a minimum titer was already demanded in the optimization problem, the first term of the scoring was weighted ten times higher than the second. Computation of the score for each column revealed the optimization result with ID 88 to be the best choice. Indeed, it combines a low enzyme load of 0.98 g/l with product titers at 24 h ranging between 11.3 and 15 mM, depending on the chosen parameter set (see Fig. 9). This highest scoring optimization result was then tested experimentally. Fig. 9 displays the simulations and the measurements of the validation experiment. Both validation experiment and simulations used the following initial substrate concentrations: 22 mM fucose, 22 mM GMP, 5 mM ATP.

The validation experiment reached a GDP-fucose titer of 11.26 mM which is lower than the titer of the baseline experiment (13.16 mM). However, again, the initial concentrations of the most cost-relevant substrate GMP was different between the unoptimized baseline experiment (25.7 mM GMP) and the validation experiment (22 mM; the same stock solution as for the validation experiment for O1 was used) complicating a comparison of the two processes. Despite its lower value, the obtained titer of the validation experiment corresponds to a yield that is almost identical to the one obtained in the baseline experiment

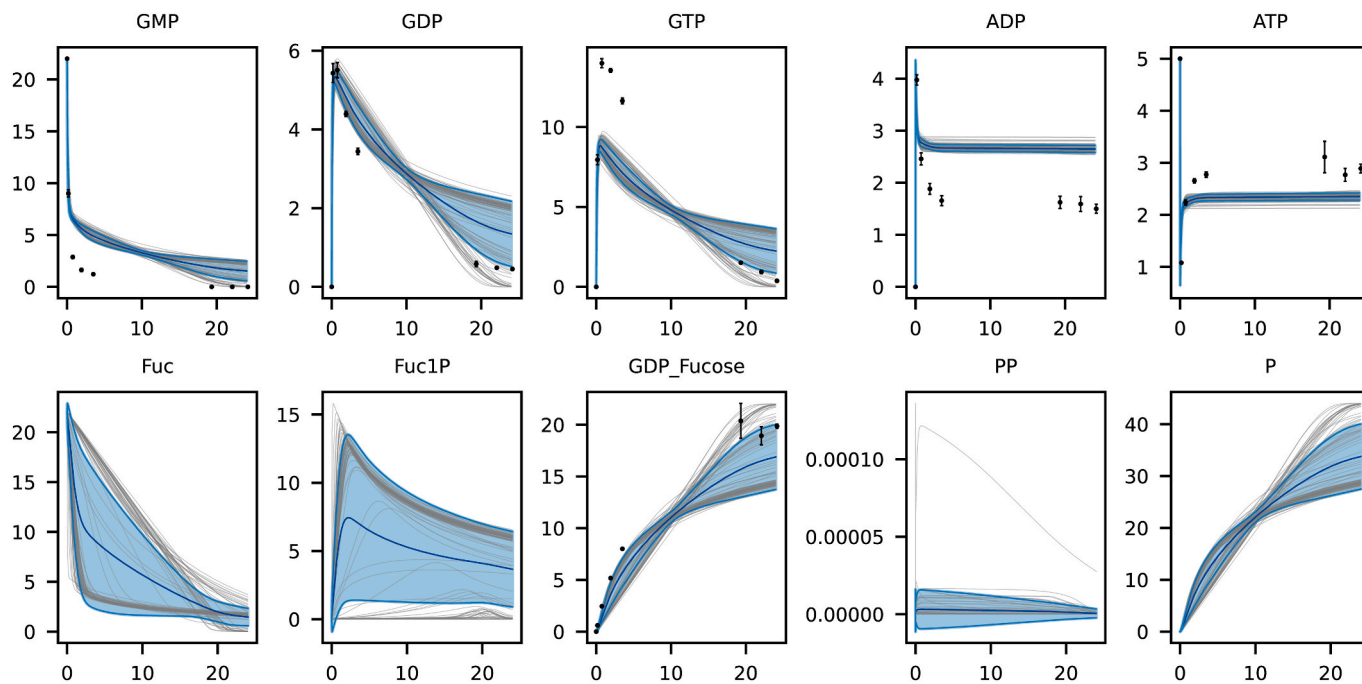


Fig. 6. Time courses of model simulations for the ensemble of 100 sets of estimated kinetic parameters using the identified set (ID 53) of optimal enzyme concentrations (maximizing the titer of GDP-fucose). For comparison, measurements from validation experiments are shown: black dots represent the measured concentrations of the validation experiment (Exp. 6; see [Supplementary Table 1](#) and [Supplementary Fig. 6](#)) as averages of triplicates with error bars showing the associated standard deviation. Grey lines: individual simulation trajectories, central blue line: average across all simulation trajectories, outer blue lines: average trajectory plus and minus associated standard deviation respectively. X-axes show time in hours, y-axes show concentration in mM.

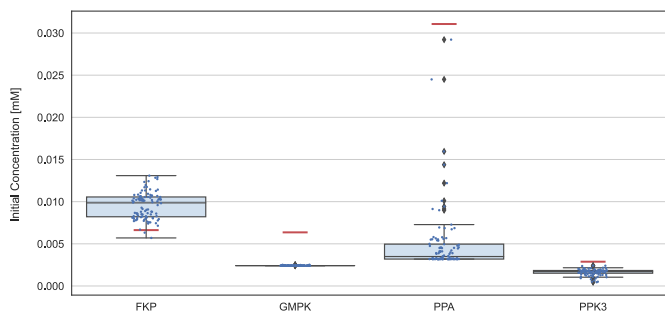


Fig. 7. Distribution of predicted optimal enzyme concentrations for objective O3 for each of the 100 parameter sets of the ensemble. Box plots show 25%, 50%, and 75% quartiles, scatter plots show the complete set. Red lines mark the starting values for each enzyme (as used in the baseline experiment).

(51.2 %) while using an enzyme load that is reduced by 36.2 % (from 1.55 g/l to 0.98 g/l). Model simulations across all estimated parameter sets using the optimized enzyme concentrations together with the (higher) baseline initial substrate concentrations (25.7 mM GMP) revealed an average GDP-fucose titer of 13.16 mM, which exactly equals the titer of the baseline experiment despite its significantly higher enzyme load.

3.4.3. Minimization of overall costs normalized by product titer (objective O6)

In a third application example we aimed to minimize the (relative) overall costs of the process normalized by the product titer, i.e., we optimized the cascade based on objective O6 defined in eq. (15). In contrast to the previous examples, both the enzyme and substrate concentrations were included as optimization variables and can thus be adjusted to minimize the relative costs. Start and boundary values of the enzyme concentrations were the same as listed in [Table 1](#) with the only difference that the upper boundary of FKP was increased from 0.0149 to

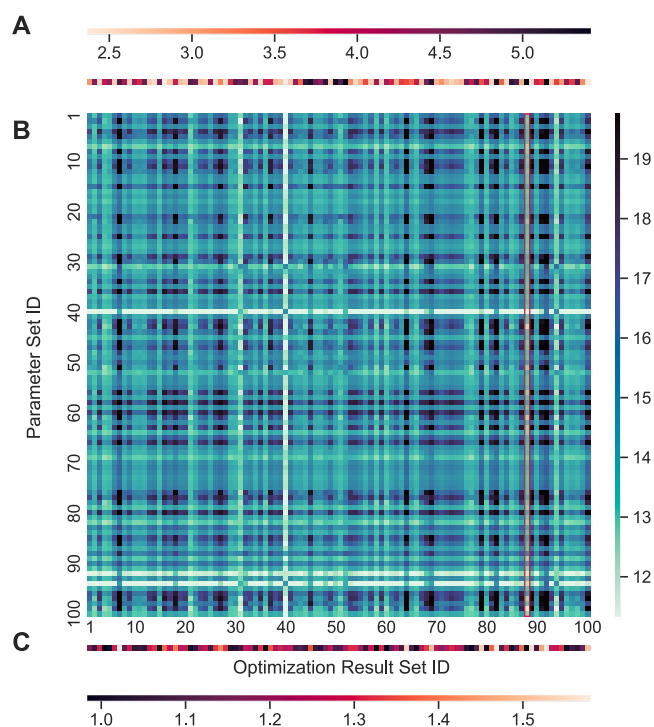


Fig. 8. (A) Scores of the columns of the heatmap shown in (B) according to equation 19. (B) Simulated GDP-fucose titers in mM at 24 h. (C) Optimized enzyme loads in g/l. Titters result from model simulations that combine optimal enzyme loads determined for each of the 100 parameter sets (optimization result sets) with the 100 parameter sets of the estimated ensemble (see also [Fig. 5](#) for further explanations). The best (highest) scoring column of heatmap B (optimization result ID 88) is highlighted.

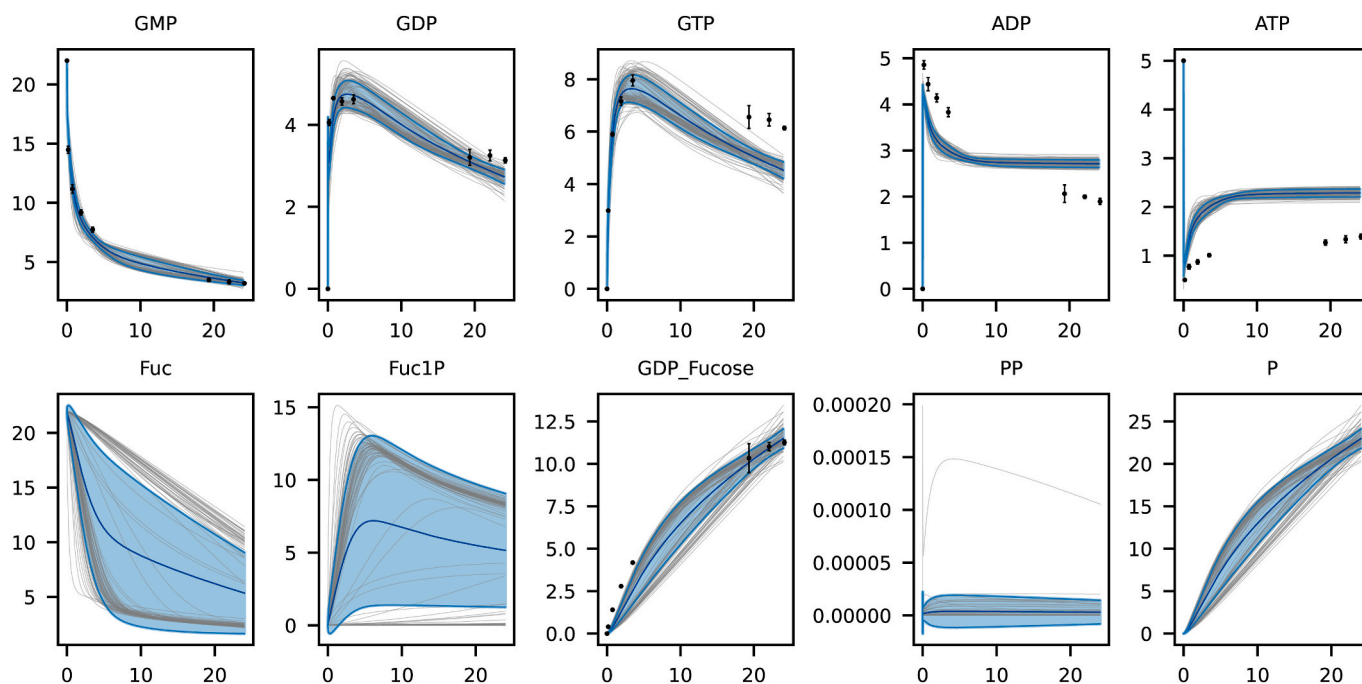


Fig. 9. Time courses of model simulations for the ensemble of 100 kinetic parameter sets when using the identified set (ID 88) of optimal enzyme concentrations (minimizing the enzyme load) and comparison with measurements from validation experiments. Black dots represent the measured concentrations of the validation experiment (Exp. 7, see [Supplementary Table 1](#) and [Supplementary Fig. 7](#)) as averages of triplicates with error bars showing the associated standard deviation. Grey lines: individual simulation trajectories, central blue line: average across all simulation trajectories, outer blue lines: average trajectory plus and minus associated standard deviation respectively. X-axes show time in hours, y-axes show concentration in mM.

0.05 mM to give the optimization more flexibility. Substrate start values were set to the initial concentrations of the baseline experiment (experiment 5) with all lower boundaries set to 0 mM and upper boundaries as follows: ATP 20 mM, fucose 40 mM, GMP 40 mM. The process end time was again set to 24 h. The enzyme and substrate prices of the cost term were based on cost estimates of our own enzyme production process and market prices (taken from Biosynth, formerly Carbosynth) respectively. The cost estimates for the production of each enzyme were 200.45 €/g (FKP), 58.65 €/g (GMPK), 162.44 €/g (PPA), and 541.75 €/g (PPK3) resulting in an overall enzyme cost of 300.75 €/l given the start values used for the optimization. The substrate market prices were 0.25 €/g (GMP), 5.66 €/g (fucose), and 0.51 €/g (ATP) which led to substrate costs of 27.76 €/l for the chosen start values. The fixed cost term was set to an estimate of 500 €/l. Overall, the normalized

process costs at the start of the optimization were 62.96 €/mmol for the baseline product titer of 13.16 mM at 24 h. The optimization results for the model ensemble with the 100 parameterizations are shown in [Fig. 10](#).

All optimization results predict an increase of the FKP concentration as well as reductions of the PPK3 and GMPK concentrations with the latter hitting its lower boundary in the vast majority of cases. PPA results are more ambiguous: most of them suggest reductions with various magnitudes while two results recommend a significant increase compared to the start value. Substrate predictions are in most cases hitting the upper boundaries with only some results staying below them. This was to be expected since substrate costs are relatively low compared to the enzymes and higher initial substrate concentrations enable higher product titers at 24 h and thus lower normalized process

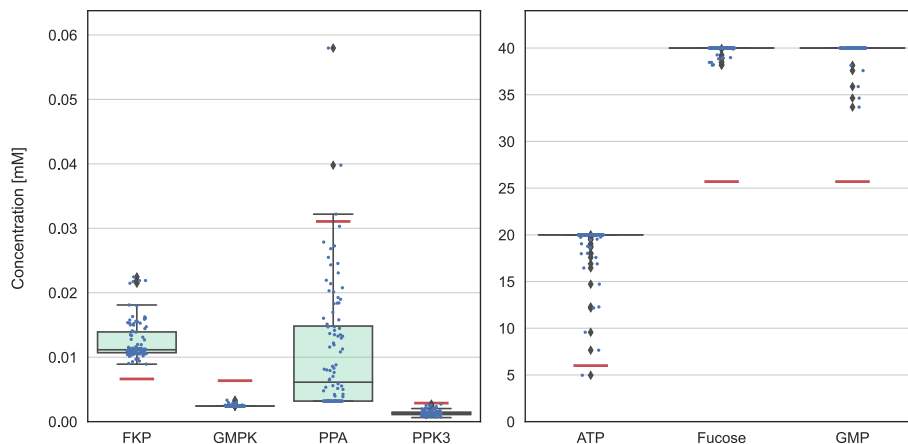


Fig. 10. Distribution of predicted optimal enzyme and substrate concentrations when minimizing the relative costs of GDP-fucose production (objective O6) for each of the 100 parameter sets of the ensemble. Box pots show 25%, 50%, and 75% quartiles, scatter plots show the complete set. Red lines mark the starting values for each enzyme and substrate (as used in the baseline experiment).

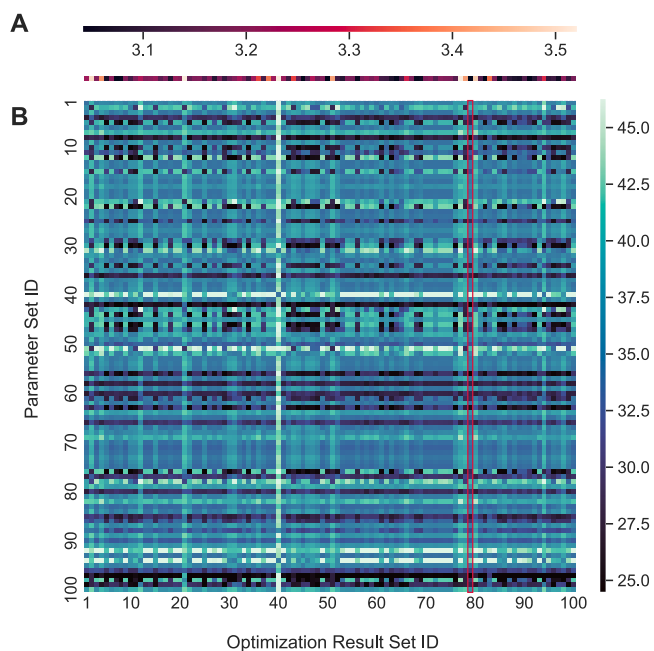


Fig. 11. (A) Column scores of the normalized process cost shown in the heatmap in (B) according to equation (20). The best (here lowest) scoring column (optimization result ID 79) is highlighted in (B). (B) Heat map of optimized normalized process costs in €/mmol. Normalized costs are calculated using titers from model simulations that combine optimal enzyme and substrate concentrations determined for each of the 100 parameter sets (optimization result sets) with the 100 parameter sets of the estimated ensemble (see also Fig. 5 for further explanations).

costs. We used the following scoring scheme to select the best result across all columns i :

$$Score(i) = \frac{M_i}{M_{min}} + \frac{C_i}{C_{min}} + \frac{S_i}{S_{min}}, \quad (20)$$

where M_i is the median cost of column i , which is normalized to the overall lowest median cost M_{min} , C_i is the minimum cost of column i , which is normalized by the overall lowest minimum cost C_{min} , and S_i is the sum over all costs of column i , which is normalized by the overall lowest costs sum S_{min} . Cross validation (Fig. 11) revealed the optimization result with the ID 79 as best performing across the entire ensemble (FKP 0.011 mM, GMPK 0.0024 mM, PPA 0.0076, PPK3 0.0013 mM, ATP 20 mM, fucose 40 mM, GMP 40 mM). Its average normalized process cost of 33.71 ± 4.71 €/mmol (min: 22.53 €/mmol, max: 46.57 €/mmol) is far below the costs of the baseline experiment (62.96 €/mmol).

4. Discussion

In the present work, we formulated a theoretical framework for constrained optimization of cell-free production systems operating on a kinetic model with parametric uncertainty (related to non-identifiabilities arising from parameter estimation). We demonstrated how this model-based optimization approach led to significant improvements in the efficiency of an enzyme cascade for the production of GDP-fucose. Several previously published works leveraged model-based optimization to improve the performance of cell-free enzyme cascades. In many of those studies, the constructed kinetic models were thoroughly analyzed, e.g., by means of sensitivity analysis (including metabolic control analysis; MCA) or by testing the effect of certain perturbations (e.g., altered initial substrate and enzyme concentrations) to eventually identify promising optimization targets. Some recent examples of such approaches include works of Korman et al. (2017),

Česnik et al. (2020), Shen et al. (2020), and Martin et al. (2023), in which suitable enzyme levels were determined leading to higher yields and product titers. For example, Martin et al. (2023) utilized MCA results including local elasticities and global control coefficients to derive predictions about how to optimize the production of butanol in a cell-free version of the corresponding pathway from *Escherichia coli*. However, even if the model would represent the reality perfectly, such local analyses do not necessarily lead to the globally optimal solution. An alternative could be an exhaustive approach as followed by Dvorak et al. (2014), where the solution space is discretized and then evaluated at all possible value combinations of the optimization variables. While this approach will certainly find the optimal solution (or one that is close to it), such a brute-force strategy will only be possible in very small systems with few optimization variables (e.g., three enzymes in Dvorak et al., 2014).

In contrast, herein we used rigorously defined mathematical optimization problems, which are based on well-defined objective functions and constraints (the latter including the dynamic model). Available solvers can tackle these problems by repeatedly simulating the kinetic model to determine the output performance of the system for particular instances of the free variables to eventually identify the optimal solution. After providing a general form for such optimization problems for cell-free systems we formulated seven specific optimization problems which we deem most relevant in this context (but further optimization problems could easily be formulated in this way). We emphasize that there are few other works that indeed used a rigorous optimization approach (e.g., Ardao and Zeng, 2013; Hold et al., 2016; Finnigan et al., 2019; Paschalidis et al., 2022), in most cases to optimize the enzyme ratio for maximizing the product titer, corresponding to our optimization problem O1. However, we believe that other optimization problems as stated (and partially also applied) herein could be of high relevance as well, depending on the application.

Furthermore, as a methodological development, we presented an approach to tackle parameter uncertainty (arising from non-identifiabilities during parameter estimation) not only during model construction and analysis but also in the subsequent model-based optimization of the cell-free system. Since most of our model parameters are estimated from experimental data, issues of parameter uncertainty and parameter non-identifiability arise due to interdependencies between the parameters. Therefore, alternative methods of optimization under parameter uncertainty such as robust optimization (Klamroth et al., 2017; Puschke et al., 2018) or stochastic programming (Birge and Louveaux, 2011) are not directly applicable here because those methods usually consider, separately for each parameter, either a known probability distribution or simply a bounded parametric uncertainty (with known bounds) (Puschke et al., 2018) and thus implicitly assume parameters to be independent from each other. To capture the uncertainty of dependent parameters of a kinetic model we followed an ensemble modeling approach, which has been used in different variants for modeling other systems, including signaling and metabolic networks in cells (Tran et al., 2008; Jia et al., 2012; Tan and Liao, 2012; Villaverde et al., 2022). By repeating the global parameter estimation many times, different parameter sets are found leading to an approximation of the underlying distributions of each parameter. We note, however, that our ensemble approach is different to the one of Tran et al. (2008) and Tan and Liao (2012) since the latter are based on reference steady-state metabolic flux distributions in cells, while in batch operation of cell-free systems we usually have to deal with transient systems behavior (and thus changing fluxes and metabolite concentrations). Moreover and most importantly, here we go one step further and not only integrate but also account for this uncertainty when optimizing the system. Our approach is based on cross validation, i.e., we test the performance of the optimal initial enzyme or substrate concentrations identified for one parameter set also for the resulting performance under all other parameterizations and then select the optimal solution that performs best on average in terms of a suitable metric. Alternative metrics, such as the

maximization of certain performance measures under worst-case conditions (Misener et al., 2018) (related to min-max optimizations often employed in robust optimization; Puschke et al., 2018) could be used as well. Taken together, it should be noted that our optimization framework involves three separated optimization steps: (1) repeated parameter estimation based on experimental data resulting in a parameter ensemble (section 3.2); (2) optimization of the kinetic model (here with the initial enzyme/substrate concentrations as free variables) for a given objective function for every parameter set (section 3.3); and (3) using cross-validation to identify the best optimization result from (2) that maximizes a given metric (section 3.4). It should be noted that the best optimization result identified in this manner has indeed a superior metrics over all other optimization results (that were obtained when optimizing the model with each parametrization of the ensemble), but it is not necessarily globally optimal. For example, in our application where we maximized the product titer for a given maximal enzyme load (objective O1), there might exist a (valid) enzyme ratio that leads to a higher score in eq. (18). However, direct optimization determining the globally optimal solution maximizing this score would require significantly higher computational effort than our approach and we do not expect larger differences in the results if the ensemble is large enough.

We consider our developed framework to be generic and therefore applicable to a wide variety of cell-free batch processes and even to other systems including the optimization of metabolic pathways or networks in (whole-)cell factories. Clearly, certain modules, e.g., the generation of an ensemble of parameter sets (models), could be replaced with others. For example, an alternative approach to estimating model parameters is Bayesian inference where initial distributions of parameter values based on prior knowledge are updated by incorporating information from available data, leading to improved (posterior) parameter distributions (Wilkinson, 2007; Linden et al., 2022; Villaverde et al., 2022). Some of those methods have been used in the context of optimization of bioprocesses (but not cell-free enzyme cascades). There might be cases where such a Bayesian approach to determine parameter sets is more appropriate, especially if there are well-defined priors of the unknown parameters (i.e., more precise previous knowledge than just a uniform distribution over an interval of possible parameter values as also used herein). For a review and comparison of different methods accounting for (parametric) uncertainty, we refer the reader to Villaverde et al. (2022). Generally, the probability to find a (close to) optimal parameter set decreases with increasing system dimension. Since cell-free systems usually contain a relatively low number of state variables (often 3–10, sometimes up to 30), there is a reasonable chance to obtain a good ensemble of meaningful parameter sets. In our application, we used an ensemble of one hundred estimated parameter sets, each exploring the parameter space by going through at least 25,000 iterations of testing different parameter sets while searching for the global minimum. Hence, the total number of parameter sets evaluated is at least 2.5 million, letting us believe that the aggregate of all 100 global estimation runs is representative for the underlying unknown parameter space. Clearly, the number of estimated parameter sets included in the ensemble can be increased, e.g., in larger systems, which, however, will also increase the computational demand. For our application example, starting with the kinetic model and parameter estimation, it takes ca. 9 h on a standard desktop PC to run all steps of our workflow (parameter estimation, optimization for each parameter set, cross validation) to eventually obtain the final optimal solution for objective O1. This indicates that there is room to deal with significantly larger systems or networks when using, for example, compute clusters.

Application of our optimization framework to improve the cell-free system for the production of GDP-fucose was very successful. The calculated optimal enzyme ratio maximizing the product titer (objective O1) suggests an increase of the initial FKP concentration, which indeed led to a significant increase of the titer in validation experiments. This finding is also in line with insights by Frohnmeyer et al. (2022) who identified the pyrophosphorylase activity of FKP as most likely to limit

the process of GDP-fucose synthesis. Additionally, the enzyme exerts a disproportional amount of control on the cascade due to its bifunctional activity, making it an important target when optimizing the performance of the cascade. Notably, FKP is much heavier than all other enzymes. Hence, increasing its molar amounts [mM] may more profoundly affect the gram enzyme load [g/l]. In the minimization of the enzyme load (objective O3), we therefore defined the enzyme load with respect to total mass [gram], i.e., it corresponds to a weighted sum of initial enzyme concentrations with conversion factors based on the molecular weight of the respective enzymes. Other studies usually focus on molar enzyme load, but we consider the mass-based enzyme load to be more relevant since it is the enzyme weight per volume that matters when setting up the process and assessing its costs.

A limitation of a kinetic model as used herein for the optimization is that important process parameters such as temperature, pH value, and concentrations of cofactors are not incorporated because of their often poorly defined relationship with the kinetic parameters of the model. Here, statistical methods such as design of experiment (DoE) approaches are often used to identify optimal values for these parameters (Mandeni and Brundin, 2008; Onyeogaziri and Papanephytous, 2019). These methods, however, are limited in taking prior quantitative knowledge of the reaction mechanisms into account. We therefore consider a combination of our optimization approach based on kinetic models with DoE approaches as an interesting topic for future work. Furthermore, the experimental setup of the underlying biochemical process modeled and optimized in this work was a batch process where all substrates and enzymes are provided at the beginning. In order to potentially reach even higher titers, the process could be operated in fed-batch mode, e.g., using a certain substrate feeding rate. This could also increase the overall cost efficiency if the substrate feeding enables a longer process time as the expensive purified enzymes would then be used for additional substrate conversions. Optimizing such a fed-batch process would involve more advanced dynamic optimization approaches (known as optimal control, Kapadi and Gudi, 2004), which attempt to find an optimal trajectory of the substrate (or enzyme) feeding rate instead of a set of optimal initial enzyme and substrate concentrations (one recent study in this direction was presented by Paschalidis et al., 2022). The extension of our framework to this kind of dynamic optimization, including answering the question of when choosing a fed-batch approach over a simple batch process is advisable in the context of cell-free production systems, remains as a topic for future work.

Author statement

Nicolas Huber: Methodology, Investigation, Data Curation, Validation, Writing – Original Draft, Writing - Review & Editing, Visualization. Edgar Alberto Alcalá-Orozco: Investigation, Data Curation, Validation, Writing - Review & Editing. Thomas Rexer: Validation, Writing - Review & Editing, Supervision, Project Administration, Funding Acquisition. Udo Reichl: Validation, Writing - Review & Editing, Supervision, Project Administration, Funding Acquisition. Steffen Klamt: Conceptualization, Methodology, Validation, Writing – Original Draft, Writing - Review & Editing, Supervision, Project Administration, Funding Acquisition.

Declaration of competing interest

T.R. and E.A. are co-founders of eversyn - a company focused on the production and commercial sales of nucleotide sugars including GDP-fucose.

Data availability

Experimental data relevant to this study are part of the manuscript and code/models are available at the following repository: https://github.com/klamt-lab/GDP-Fucose_OptMet2023_Paper.

Acknowledgements

N.H. and E. A. received financial support by the International Max

Planck Research School for Advanced Methods in Process and Systems Engineering (IMPRS ProEng).

Appendix

Construction of a kinetic model of the enzyme cascade for GDP-fucose synthesis

We implemented the cell-free reaction network for synthesis of GDP-fucose (Fig. 1) as a kinetic model. The general form of kinetic models based on a set of ordinary differential equations (ODEs) is given by the following formula

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), t) = \mathbf{N} \cdot \mathbf{r}(\mathbf{x}(t), \mathbf{P}, t),$$

where $\mathbf{x}(t)$ is the vector of species concentrations, \mathbf{N} the stoichiometric matrix of the network, and \mathbf{r} the vector of reaction rates which are defined by rate laws as functions of the species concentrations $\mathbf{x}(t)$, reaction parameters \mathbf{P} , and time t . In this work, the rate law equations are based on the convenience kinetics formulation by Liebermeister and Klipp (2006), a generalized version of Michaelis-Menten kinetics, which we extended by the Haldane relationship to include terms incorporating the equilibrium constants for each reaction. This allows a simple yet realistic description of the dynamics of all enzymatically catalyzed reactions including the affinities of participating compounds (K_m), the turnover rates (k_{cat}), information about equilibria (K_{eq}), and the actions of potential regulators. The kinetic rate laws and ODEs of the model are shown below. It should be noted that polyphosphate (PP) was considered to be at saturation levels during the whole process and thus was not added as a dynamical species to the model.

Kinetic rate laws

Rate Law for the FKP (fucokinase) reaction FKP_{gt}	$E_{FKP} \frac{k_{cat,F} \frac{Fuc}{K_m Fuc} \frac{ATP}{K_m ATP} - k_{cat,F} \frac{1}{K_{eq_FKP_fk}} \frac{Fuc1P \cdot ADP}{K_m Fuc \cdot K_m ATP}}{\left(1 + \frac{Fuc}{K_m Fuc}\right) \cdot \left(1 + \frac{ATP}{K_m ATP}\right) + \left(1 + \frac{Fuc1P}{K_m Fuc1P}\right) \cdot \left(1 + \frac{ADP}{K_m ADP}\right) - 1}$
Rate Law for the FKP (guanylyl-transferase) reaction FKP_{fk}	$\frac{k_i_{PP}}{k_i_{PP} + PP} E_{FKP} \frac{k_{cat,F} \frac{Fuc1P}{K_m Fuc1P} \frac{GTP}{K_m GTP} - k_{cat,F} \frac{1}{K_{eq_FKP_gt}} \frac{GDP \cdot Fuc \cdot PP}{K_m Fuc1P \cdot K_m GTP}}{\left(1 + \frac{Fuc1P}{K_m Fuc1P}\right) \cdot \left(1 + \frac{GTP}{K_m GTP}\right) + \left(1 + \frac{GDP \cdot Fuc}{K_m GDP \cdot Fuc}\right) \cdot \left(1 + \frac{PP}{K_m PP}\right) - 1}$
Rate Law for GMPK reaction $GMPK$	$E_{GMPK} \frac{k_{cat,F} \frac{GMP}{K_m GMP} \frac{ATP}{K_m ATP} - k_{cat,F} \frac{1}{K_{eq_GMPK}} \frac{GDP \cdot ADP}{K_m GMP \cdot K_m ATP}}{\left(1 + \frac{GMP}{K_m GMP}\right) \cdot \left(1 + \frac{ATP}{K_m ATP}\right) + \left(1 + \frac{GDP}{K_m GDP}\right) \cdot \left(1 + \frac{ADP}{K_m ADP}\right) - 1}$
Rate Law for PPA reaction PPA	$E_{PPA} \cdot k_{cat,F} \frac{1}{1 + \frac{K_m PP}{PP}}$
Rate Law for the PPK3 (ADP-dependent) reaction $PPK3_A$	$E_{PPK3} \frac{k_{cat,F} \frac{ADP}{K_m ADP} - k_{cat,F} \frac{1}{K_{eq_PPK3_A}} \frac{ATP}{K_m ADP}}{\left(1 + \frac{ADP}{K_m ADP}\right) + \left(1 + \frac{ATP}{K_m ATP}\right) - 1}$
Rate Law for the PPK3 (GDP-dependent) reaction $PPK3_G$	$E_{PPK3} \frac{k_{cat,F} \frac{GDP}{K_m GDP} - k_{cat,F} \frac{1}{K_{eq_PPK3_G}} \frac{GTP}{K_m GDP}}{\left(1 + \frac{GDP}{K_m GDP}\right) + \left(1 + \frac{GTP}{K_m GTP}\right) - 1}$

ODEs

$$d[PP]/dt = +FKP_{gt} - PPA$$

$$d[P]/dt = +2 \cdot (PPA)$$

$$d[ATP]/dt = -FKP_{fk} - GMPK + PPK3_A$$

$$d[ADP]/dt = +FKP_{fk} + GMPK - PPK3_A$$

$$d[E_{PPA}]/dt = 0$$

$$d[E_{PPK3}]/dt = 0$$

$$d[E_{FKP}]/dt = 0$$

$$d[Fuc]/dt = -FKP_{fk}$$

$$d[Fuc1P]/dt = +FKP_{fk} - FKP_{gt}$$

$$d[\text{GDP_Fucose}]/dt = +\text{FKP_gt}$$

$$d[\text{GTP}]/dt = -\text{FKP_gt} + \text{PPK3_G}$$

$$d[\text{GDP}]/dt = +\text{GMPK} - \text{PPK3_G}$$

$$d[\text{GMP}]/dt = -\text{GMPK}$$

$$d[\text{E_GMPK}]/dt = 0$$

Reaction parameters

We searched for relevant reaction parameters k_{cat} (turnover rate), K_m (Michaelis-Menten constant), k_i (inhibitory constant), and K_{eq} (equilibrium constant) in the literature and databases. All parameters that could not be found or were considered to be uncertain were estimated from a parameter fitting procedure based on a weighted least-squares estimation with available time course data \mathbf{y} from multiple experiments (see experiments 1–5 in Supplementary Table 1). For the latter, the following objective function was used

$$SSR = \min \sum_k \sum_{ij} \omega_{j,k} \cdot (y_{i,j,k} - x_{i,j,k}(\mathbf{P}))^2,$$

where SSR is the sum of squared residues, $y_{i,j,k}$ is the concentration of species j measured at time point i in experiment k , and $x_{i,j,k}(\mathbf{P})$ is the concentration of the corresponding model species at the same time point derived from a simulation of the model (based on the parameter set \mathbf{P} and with initial concentrations used in the experiment k). $\omega_{j,k}$ is the normalized weight of species j for experiment k and is calculated with the following equation (as implemented in COPASI):

$$\omega_{j,k} = \frac{1 / \langle y_{j,k}^2 \rangle}{\max_{j \in \{1 \dots n_j\}} 1 / \langle y_{j,k}^2 \rangle}$$

Here, $\langle y_{j,k}^2 \rangle$ expresses the mean square of all measurements y of species j across all time points of experiment k and n_j is the number of species measured in experiment k .

An overview of the kinetic parameters is provided in the table below. Apart from three K_{eq} values, most parameters have been included in the estimation procedure with literature values serving as start values if available.

Parameter name	Literature value	Reference	Included in parameter estimation	Estimation Boundaries
(FKP (fucokinase)).kcat_F	1404 [1/h]	Liu et al. (2019)	Yes	[1e0, 1e6]
(FKP (fucokinase)).Km_Fuc	0.045 [mM]	Wang et al. (2009)	Yes	[1e-5, 1e0]
(FKP (fucokinase)).Km_ATP	1.08 [mM]	Wang et al. (2009)	Yes	[1e-2, 1e2]
(FKP (fucokinase)).Km_Fuc1P	–	–	Yes	[1e-7, 1e4]
(FKP (fucokinase)).Km_ADP	–	–	Yes	[1e-6, 1e4]
(FKP (fucokinase)).K_eq_FKP_fk	4 [-]	Equilibrator (Beber et al., 2022)	No	–
(FKP (guanylyltransferase)).kcat_F	36,252 [1/h]	Liu et al. (2019)	Yes	[1e0, 1e6]
(FKP (guanylyltransferase)).Km_Fuc1P	0.03295 [mM]	Liu et al. (2019)	Yes	[1e-4, 1e2]
(FKP (guanylyltransferase)).Km_GTP	0.012 [mM]	Wang et al. (2009)	Yes	[1e-4, 1e3]
(FKP (guanylyltransferase)).Km_GDP_Fuc	0.12 [mM]	Pastuszak et al. (1998)	Yes	[1e-2, 1e3]
(FKP (guanylyltransferase)).Km_PP	0.135 [mM]	Pastuszak et al. (1998)	Yes	[1e-2, 1e2]
(FKP (guanylyltransferase)).ki_PP	–	–	Yes	[1e-10, 1e-6]
(FKP (guanylyltransferase)).K_eq_FKP_gt	20	Equilibrator (Beber et al., 2022)	No	–
(PPA).kcat_F	558,000 [1/h]	Kapyla et al. (1995)	Yes	[1e0, 1e6]
(PPA).Km_PP	0.003 [mM]	Rodina et al. (2007)	Yes	[1e-6, 1e1]
(GMPK).kcat_F	752,400 [1/h]	Nomura et al. (2014)	Yes	[1e0, 1e6]
(GMPK).Km_GMP	0.156 [mM]	Nomura et al. (2014)	Yes	[1e-4, 1e2]
(GMPK).Km_ATP	0.5 [mM]	Moriguchi et al. (1981)	Yes	[1e-3, 1e2]
(GMPK).Km_GDP	0.097 [mM]	Li et al. (1996)	Yes	[1e-4, 1e2]
(GMPK).Km_ADP	0.017 [mM]	Li et al. (1996)	Yes	[1e-4, 1e2]
(GMPK).K_eq_GMPK	1	Equilibrator (Beber et al., 2022)	No	–
(PPK3_G).kcat_F	15,048 [1/h]	Achbergerová and Nahalka (2014)	Yes	[1e0, 1e6]
(PPK3_G).Km_GDP	1.2 [mM]	Achbergerová and Nahalka (2014)	Yes	[1e-4, 1e2]
(PPK3_G).Km_GTP	–	–	Yes	[1e-6, 1e4]
(PPK3_G).K_eq_PPK3_G	1 [-]	Equilibrator (Beber et al., 2022)	Yes	[0.5, 2]
(PPK3_A).kcat_F	18,780 [1/h]	Achbergerová and Nahalka (2014)	Yes	[1e0, 1e6]
(PPK3_A).Km_ADP	2.5 [mM]	Achbergerová and Nahalka (2014)	Yes	[1e-4, 1e2]
(PPK3_A).Km_ATP	2 [mM]	Ahn and Kornberg (1990)	Yes	[1e-4, 1e2]
(PPK3_A).K_eq_PPK3_A	1 [-]	Equilibrator (Beber et al., 2022)	Yes	[0.5, 2]

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymben.2023.10.007>.

References

- Achbergerová, L., Nahalka, J., 2014. Degradation of polyphosphates by polyphosphate kinases from *Ruegeria pomeroyi*. *Biotechnol. Lett.* 36, 2029–2035. <https://doi.org/10.1007/s10529-014-1566-6>.
- Ahn, K., Kornberg, A., 1990. Polyphosphate kinase from *Escherichia coli*. Purification demonstration of a phosphoenzyme intermediate. *J. Biol. Chem.* 265 (20), 11734–11739. [https://doi.org/10.1016/S0021-9258\(19\)38459-5](https://doi.org/10.1016/S0021-9258(19)38459-5).
- Ardao, I., Zeng, A.-P., 2013. In silico evaluation of a complex multi-enzymatic system using one-pot modular approaches: application to the high-yield production of hydrogen from a synthetic metabolic pathway. *Chem. Eng. Sci.* 87, 183–193. <https://doi.org/10.1016/j.ces.2012.10.005>.
- Bäck, T., Schwefel, H.-P., 1993. An overview of evolutionary algorithms for parameter optimization. *Evol. Comput.* 1 (1), 1–23. <https://doi.org/10.1162/evco.1993.1.1.1>.
- Beber, M.E., Gollub, M.G., Mozaffari, D., Shebek, K.M., Flamholz, A.I., Milo, R., Noor, E., 2022. eQuilibrator 3.0: a database solution for thermodynamic constant estimation. *Nucleic Acids Res.* 50 (D1), D603–D609. <https://doi.org/10.1093/nar/gkab1106>.
- Bertsimas, D., Brown, D.B., Caramanis, C., 2011. Theory and applications of robust optimization. *SIAM Rev.* 53 (3), 464–501. <https://doi.org/10.1137/080734510>.
- Biegler, L.T., 2010. Introduction to dynamic process optimization. In: *Nonlinear Programming. Society for Industrial and Applied Mathematics*, pp. 213–249. <https://doi.org/10.1137/1.9780898719383.ch8>.
- Birge, J.R., Louveaux, F., 2011. Introduction to stochastic programming. In: *Springer Series in Operations Research and Financial Engineering*. Springer, New York. <https://doi.org/10.1007/978-1-4614-0237-4>.
- Bode, L., 2012. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* 22 (9), 1147–1162. <https://doi.org/10.1093/glycob/cws074>.
- Byun, S.G., Kim, M.D., Lee, W.H., Lee, K.J., Han, N.S., Seo, J.H., 2007. Production of GDP-L-fucose, L-fucose donor for fucosyloligosaccharide synthesis, in recombinant *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 74, 768–775. <https://doi.org/10.1007/s00253-006-0730-x>.
- Česník, M., Sudar, M., Hernández, K., Charnock, S., Vasić-Rački, D., Clapés, P., Findrik Blažević, Z., 2020. Cascade enzymatic synthesis of l-homoserine - mathematical modelling as a tool for process optimisation design. *React. Chem. Eng.* 5 (4), 747–759. <https://doi.org/10.1039/C9RE00453J>.
- Chen, X., Zhang, C., Zou, R., Zhou, K., Stephanopoulos, G., Too, H.P., 2013. Statistical experimental design guided optimization of a one-pot biphasic multienzyme total synthesis of amorpha-4,11-diene. *PLoS One* 8 (11), e79650. <https://doi.org/10.1371/journal.pone.0079650>.
- Chin, Y.-W., Park, J.-B., Park, Y.-C., Kim, K.H., Seo, J.-H., 2013. Metabolic engineering of *Corynebacterium glutamicum* to produce GDP-L-fucose from glucose mannose. *Bioproc. Biosyst. Eng.* 36 (6), 749–756. <https://doi.org/10.1007/s00449-013-0900-z>.
- Claassens, N.J., Burgener, S., Vögeli, B., Erb, T.J., Bar-Even, A., 2019. A critical comparison of cellular cell-free bioproduction systems. *Curr. Opin. Biotechnol.* 60, 221–229. <https://doi.org/10.1016/j.copbio.2019.05.003>, 31207555.
- Dvorak, P., Kurumbang, N.P., Bendl, J., Brezovsky, J., Prokop, Z., Damborsky, J., 2014. Maximizing the efficiency of multienzyme process by stoichiometry optimization. *Chembiochem: A European Journal of Chemical Biology* 15 (13), 1891–1895. <https://doi.org/10.1002/cbic.201402265>.
- Finnigan, W., Cutlan, R., Snajdrova, R., Adams, J.P., Littlechild, J.A., Harmer, N.J., 2019. Engineering a seven enzyme biotransformation using mathematical modelling characterized enzyme parts. *ChemCatChem* 11 (15), 3474–3489. <https://doi.org/10.1002/cctc.201900646>.
- Frohnmeyer, H., Rueben, S., Elling, L., 2022. Gram-scale production of GDP-β-L-fucose with multi-enzyme cascades in a repetitive-batch mode. *ChemCatChem*, e202200443. <https://doi.org/10.1002/cctc.202200443> n/a(n/a).
- Gabrel, V., Murat, C., Thiele, A., 2014. Recent advances in robust optimization: an overview. *Eur. J. Oper. Res.* 235 (3), 471–483. <https://doi.org/10.1016/j.ejor.2013.09.036>.
- García, J., Peña, A., 2018. Robust optimization: concepts and applications. In: Ser, J.D., Osaba, E. (Eds.), *Nature-inspired Methods for Stochastic, Robust and Dynamic Optimization*. <https://doi.org/10.5772/intechopen.75381>. Ch. 2). IntechOpen.
- Hindmarsh, A.C., 1983. ODEPACK, A systematized collection of ODE solvers. *Sci. Comput.* 55–64. https://computing.llnl.gov/sites/default/files/ODEPACK_pub1_u88007.pdf.
- Hold, C., Billerbeck, S., Panke, S., 2016. Forward design of a complex enzyme cascade reaction. *Nat. Commun.* 7 (1), 12971. <https://doi.org/10.1038/ncomms12971>.
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U., 2006. COPASI—a Complex Pathway Simulator. *Bioinformatics* 22 (24), 3067–3074. <https://doi.org/10.1093/bioinformatics/btl485>.
- Jia, G., Stephanopoulos, G., Gunawan, R., 2012. Ensemble kinetic modeling of metabolic networks from dynamic metabolic profiles. *Metabolites* 2 (4), 891–912. <https://doi.org/10.3390/metabo2040891>.
- Kapadi, M.D., Gudi, R.D., 2004. Optimal control of fed-batch fermentation involving multiple feeds using Differential Evolution. *Process Biochem.* 39 (11), 1709–1721. <https://doi.org/10.1016/j.procbio.2003.07.006>.
- Kapyla, J., Hyytiä, T., Lahti, R., Goldman, A., Baykov, A.A., Cooperman, B.S., 1995. Effect of D97E substitution on the kinetic thermodynamic properties of *Escherichia coli* inorganic pyrophosphatase. *Biochemistry* 34 (3), 792–800. <https://doi.org/10.1021/bi00003a012>.
- Klamroth, K., Köbis, E., Schöbel, A., Tammer, C., 2017. A unified approach to uncertain optimization. *Eur. J. Oper. Res.* 260 (2), 403–420. <https://doi.org/10.1016/j.ejor.2016.12.045>.
- Koizumi, S., Endo, T., Tabata, K., Nagano, H., Ohnishi, J., Ozaki, A., 2000. Large-scale production of GDP-fucose Lewis X by bacterial coupling. *J. Ind. Microbiol. Biotechnol.* 25 (4), 213–217. <https://doi.org/10.1038/sj.jim.7000055>.
- Korman, T.P., Opgenorth, P.H., Bowie, J.U., 2017. A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nat. Commun.* 8 (1), 15526. <https://doi.org/10.1038/ncomms15526>.
- Lee, W.-H., Chin, Y.-W., Han, N.S., Kim, M.-D., Seo, J.-H., 2011. Enhanced production of GDP-L-fucose by overexpression of NADPH regenerator in recombinant *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 91 (4), 967–976. <https://doi.org/10.1007/s00253-011-3271-x>.
- Li, Y., Zhang, Y., Yan, H., 1996. Kinetic thermodynamic characterizations of yeast guanylate kinase. *J. Biol. Chem.* 271 (45), 28038–28044. <https://doi.org/10.1074/jbc.271.45.28038>.
- Liebermeister, W., Klipp, E., 2006. Bringing metabolic networks to life: convenience rate law thermodynamic constraints. *Theor. Biol. Med. Model.* 3, 41. <https://doi.org/10.1186/1742-4682-3-41>.
- Linden, N.J., Kramer, B., Rangamani, P., 2022. Bayesian parameter estimation for dynamical models in systems biology. *PLoS Comput. Biol.* 18 (10), e1010651. <https://doi.org/10.1371/journal.pcbi.1010651>.
- Liu, Y., Gunawan, R., 2017. Bioprocess optimization under uncertainty using ensemble modeling. *J. Biotechnol.* 244, 34–44. <https://doi.org/10.1016/j.jbiotec.2017.01.013>.
- Liu, Y., Hu, H., Wang, J., Zhou, Q., Wu, P., Yan, N., Wang, H.-W., Wu, J.-W., Sun, L., 2019. Cryo-EM structure of L-fucokinase/GDP-fucose pyrophosphorylase (FKP) in *Bacteroides fragilis*. *Protein & Cell* 10 (5), 365–369. <https://doi.org/10.1007/s13238-018-0576-x>.
- Ma, B., Simala-Grant, J.L., Taylor, D.E., 2006. Fucosylation in prokaryotes eukaryotes. *Glycobiology* 16 (12), 158R–184R. <https://doi.org/10.1093/glycob/cw040>.
- Mahour, R., Marichal-Gallardo, P.A., Rexer, T.F.T., Reichl, U., 2021. Multi-enzyme cascades for the in vitro synthesis of guanosine diphosphate L-fucose. *ChemCatChem* 13 (8), 1981–1989. <https://doi.org/10.1002/cctc.202001854>.
- Mandenius, C.-F., Brundin, A., 2008. Bioprocess optimization using design-of-experiments methodology. *Biotechnol. Prog.* 24 (6), 1191–1203. <https://doi.org/10.1002/btpr.67>.
- Martin, J.P., Rasor, B.J., DeBonis, J., Karim, A.S., Jewett, M.C., Tyo, K.E.J., Broadbelt, L. J., 2023. A dynamic kinetic model captures cell-free metabolism for improved butanol production. *Metab. Eng.* 76, 133–145. <https://doi.org/10.1016/j.ymben.2023.01.009>.
- Mattila, P., Rabinä, J., Hortling, S., Helin, J., Renkonen, R., 2000. Functional expression of *Escherichia coli* enzymes synthesizing GDP-L-fucose from inherent GDP-D-mannose in *Saccharomyces cerevisiae*. *Glycobiology* 10 (10), 1041–1047. <https://doi.org/10.1093/glycob/10.10.1041>.
- Michalewicz, Z., 1994. Genetic Algorithms + Data Structures = Evolution Programs, 2nd, Extended Ed. Springer-Verlag <https://dl.acm.org/doi/book/10.5555/184675>.
- Misener, R., Allenby, M.C., Fuentes-Garí, M., Gupta, K., Wiggins, T., Panoskaltis, N., Pistikopoulos, E.N., Mantalaris, A., 2018. Stem cell biomanufacturing under uncertainty: a case study in optimizing red blood cell production. *AIChE J.* 64 (8), 3011–3022. <https://doi.org/10.1002/aic.16042>.
- Mitchell, M., 1998. An Introduction to Genetic Algorithms. The MIT Press. <https://doi.org/10.7551/mitpress/3927.001.0001>.
- Moriguchi, M., Kohno, H., Kamei, M., Tochikura, T., 1981. Purification properties of guanylate kinase from baker's yeast. *Biochim. Biophys. Acta* 662 (1), 165–167. [https://doi.org/10.1016/0005-2744\(81\)90239-4](https://doi.org/10.1016/0005-2744(81)90239-4).
- Nomura, Y., Izumi, A., Fukunaga, Y., Kusumi, K., Iba, K., Watanabe, S., Nakahira, Y., Weber, A.P.M., Nozawa, A., Tozawa, Y., 2014. Diversity in guanosine 3',5'-bisphosphate (ppGpp) sensitivity among guanylate kinases of bacteria plants. *J. Biol. Chem.* 289 (22), 15631–15641. <https://doi.org/10.1074/jbc.M113.534768>.
- Onyeogaziri, F.C., Papanephytous, C., 2019. A general guide for the optimization of enzyme assay conditions using the design of experiments approach. *SLAS Discovery* 24 (5), 587–596. <https://doi.org/10.1177/2472555219830084>.
- Paschalidis, L., Beer, B., Sutiono, S., Sieber, V., Burger, J., 2022. Design of enzymatic cascade reactors through multi-objective dynamic optimization. *Biochem. Eng. J.* 181, 108384. <https://doi.org/10.1016/j.bej.2022.108384>.
- Pastuszak, I., Ketchum, C., Hermanson, G., Sjöberg, E.J., Drake, R., Elbein, A.D., 1998. GDP-fucose Pyrophosphorylase: PURIFICATION, cDNA CLONING, PROPERTIES OF THE ENZYME. *J. Biol. Chem.* 273 (46), 30165–30174. <https://doi.org/10.1074/jbc.273.46.30165>.
- Prohaska, R., Schenkel-Brunner, H., 1975. A simple efficient method for the preparation of GDP-fucose. *Anal. Biochem.* 69 (2), 536–544. [https://doi.org/10.1016/0003-2697\(75\)90158-X](https://doi.org/10.1016/0003-2697(75)90158-X).
- Puschke, J., Djelassi, H., Kleinekorte, J., Hannemann-Tamás, R., Mitsos, A., 2018. Robust dynamic optimization of batch processes under parametric uncertainty: utilizing

- approaches from semi-infinite programs. *Comput. Chem. Eng.* 116, 253–267. <https://doi.org/10.1016/j.compchemeng.2018.05.025>.
- Rexer, T., Laaf, D., Gottschalk, J., Frohnmeyer, H., Rapp, E., Elling, L., 2021. Enzymatic synthesis of glycans glycoconjugates. *Advances in Biochemical Engineering/biotechnology* 175, 231–280. https://doi.org/10.1007/10_2020_148.
- Rodina, E.V., Vorobyeva, N.N., Kurilova, S.A., Sitnik, T.S., Nazarova, T.I., 2007. ATP as effector of inorganic pyrophosphatase of *Escherichia coli*. The role of residue Lys112 in binding effectors. *Biochemistry. Biokhimiia* 72 (1), 100–108. <https://doi.org/10.1134/s0006297907010129>.
- Rollin, J.A., Martin del Campo, J., Myung, S., Sun, F., You, C., Bakovic, A., Castro, R., Chandrayan, S.K., Wu, C.-H., Adams, M.W.W., Senger, R.S., Zhang, Y.-H.P., 2015. High-yield hydrogen production from biomass by in vitro metabolic engineering: mixed sugars coutilization kinetic modeling. *Proc. Natl. Acad. Sci. USA* 112 (16), 4964–4969. <https://doi.org/10.1073/pnas.1417719112>.
- Ruffing, A., Chen, R.R., 2006. Metabolic engineering of microbes for oligosaccharide polysaccharide synthesis. *Microb. Cell Factories* 5 (1), 25. <https://doi.org/10.1186/1475-2859-5-25>.
- Runarsson, T.P., Yao, X., 2000. Stochastic ranking for constrained evolutionary optimization. *IEEE Trans. Evol. Comput.* 4 (3), 284–294. <https://doi.org/10.1109/4235.873238>.
- Schneider, M., Al-Shareffi, E., Haltiwanger, R.S., 2017. Biological functions of fucose in mammals. *Glycobiology* 27 (7), 601–618. <https://doi.org/10.1093/glycob/cwx034>.
- Shen, L., Kohlhaas, M., Enoki, J., Meier, R., Schönenberger, B., Wohlgemuth, R., Kourist, R., Niemeyer, F., van Niekerk, D., Bräsen, C., Niemeyer, J., Snoep, J., Siebers, B., 2020. A combined experimental modelling approach for the Weimberg pathway optimisation. *Nat. Commun.* 11 (1), 1098. <https://doi.org/10.1038/s41467-020-14830-y>.
- Siedentop, R., Claaßen, C., Rother, D., Lütz, S., Rosenthal, K., 2021. Getting the most out of enzyme cascades: strategies to optimize in vitro multi-enzymatic reactions. *Catalysts* 11 (10). <https://doi.org/10.3390/catal11101183>.
- Stillier, R., Thiem, J., 1992. Enzymatic synthesis of β -L-Fucose-1-phosphate GDP-fucose. *Liebigs Ann. Chem.* 1992 (5), 467–471. <https://doi.org/10.1002/jlac.199219920183>.
- Tan, Y., Liao, J.C., 2012. Metabolic ensemble modeling for strain engineers. *Biotechnol. J.* 7 (3), 343–353. <https://doi.org/10.1002/biot.201100186>.
- Tran, L.M., Rizk, M.L., Liao, J.C., 2008. Ensemble modeling of metabolic networks. *Biophys. J.* 95 (12), 5606–5617. <https://doi.org/10.1529/biophysj.108.135442>.
- Villaverde, A.F., Raimúndez, E., Hasenauer, J., Banga, J.R., 2022. Assessment of prediction uncertainty quantification methods in systems biology. *IEEE ACM Trans. Comput. Biol. Bioinf.* 1. <https://doi.org/10.1109/TCBB.2022.3213914>. –12.
- Wang, W., Hu, T., Frantom, P.A., Zheng, T., Gerwe, B., del Amo, D.S., Garret, S., Seidel, R.D., Wu, P., 2009. Chemoenzymatic synthesis of GDP-L-fucose the Lewis X glycan derivatives. *Proc. Natl. Acad. Sci. USA* 106 (38), 16096–16101. <https://doi.org/10.1073/pnas.0908248106>.
- Wilkinson, D.J., 2007. Bayesian methods in bioinformatics computational systems biology. *Briefings Bioinf.* 8 (2), 109–116. <https://doi.org/10.1093/bib/bbm007>.
- Yamamoto, K., Maruyama, T., Kumagai, H., Tochikura, T., Seno, T., Yamaguchi, H., 1984. Preparation of GDP-L-fucose by using microbial enzymes. *Agric. Biol. Chem.* 48 (3), 823–824. <https://doi.org/10.1271/bbb1961.48.823>.
- Zhai, Y., Han, D., Pan, Y., Wang, S., Fang, J., Wang, P., Liu, X.-w., 2015. Enhancing GDP-fucose production in recombinant *Escherichia coli* by metabolic pathway engineering. *Enzym. Microb. Technol.* 69, 38–45. <https://doi.org/10.1016/j.enzmictec.2014.12.001>.
- Zhao, G., Guan, W., Cai, L., Wang, P.G., 2010. Enzymatic route to preparative-scale synthesis of UDP-GlcNAc/GalNAc, their analogues GDP-fucose. *Nat. Protoc.* 5 (4), 636–646. <https://doi.org/10.1038/nprot.2010.3>.