



RESEARCH ARTICLE

Process Systems Engineering

Simultaneous design of fermentation and microbe

Anita L. Ziegler¹ | Ashutosh Manchanda¹ | Marc-Daniel Stumm¹  |
Lars M. Blank² | Alexander Mitsos^{1,3,4} 

¹Process Systems Engineering (AVT.SVT),
RWTH Aachen University, Aachen, Germany

²Institute of Applied Microbiology (iAMB),
Aachen Biology and Biotechnology (ABBt),
RWTH Aachen University, Aachen, Germany

³JARA-ENERGY, Aachen, Germany

⁴Institute of Energy and Climate Research,
Energy Systems Engineering (IEK-10),
Forschungszentrum Jülich GmbH, Jülich,
Germany

Correspondence

Alexander Mitsos, Process Systems
Engineering (AVT.SVT), RWTH Aachen
University, 52074 Aachen, Germany.
Email: amitsos@alum.mit.edu

Funding information

RWTH Aachen University, Grant/Award
Number: thes1376; Deutsche
Forschungsgemeinschaft, Grant/Award
Number: 390919832

Abstract

Constraint-based optimization of microbial strains and model-based bioprocess design have been used extensively to enhance yields in biotechnological processes. However, strain and process optimization are usually carried out in sequential steps, causing underperformance of the biotechnological process when scaling up to industrial fermentation conditions. Herein, we propose the optimization formulation *SimulKnock* that combines the optimization of a fermentation process with metabolic network design in a bilevel optimization program. The upper level maximizes space-time yield and includes mass balances of a continuous fermentation, while the lower level is based on flux balance analysis. *SimulKnock* predicts optimal gene deletions and finds the optimal trade-off between growth rate and product yield. Results of a case study with a genome-scale metabolic model of *Escherichia coli* indicate higher space-time yields than a sequential approach using *OptKnock* for almost all target products considered. By leveraging *SimulKnock*, we reduce the gap between strain and process optimization.

KEYWORDS

computational strain design, constraint-based metabolic modeling, metabolic engineering, process optimization

1 | INTRODUCTION

Industrial microbiology promises product synthesis from renewable feedstocks representing a sustainable alternative to petro-chemical synthesis.¹ Often, these biotechnological products are produced by high-performing microbial strains that have been designed using metabolic engineering. Strain design is supported by computational methods to reduce the experimental effort.² In the first place, constraint-based metabolic modeling formulations predict the influence of a given genetic modification on the microorganism.³ The best-known formulation is flux balance analysis (FBA).^{4–6} Using linear programming, FBA predicts the internal fluxes of a microorganism based on its genome-scale metabolic model (GEM). Variations of FBA

refine the cellular objective of the organism⁷ or better suit a genetically modified microorganism.^{8–10} Constraint-based strain optimization formulations go one step further and predict targets for genetic modification.^{3,11} The first formulation, *OptKnock*, was presented by Burgard et al.¹² *OptKnock* proposes optimal gene deletions by solving a bilevel optimization problem. The upper level represents the bioengineering perspective to maximize the product yield on the substrate. The lower level, based on FBA, represents the microorganism with the cellular objective of maximizing biomass production. *OptKnock* has been extended, for example, to account for worst-case predictions,¹³ insertion of genes,¹⁴ and up and down regulation of genes.^{15,16} It was also modified to suit a genetically modified organism better.¹⁷ For strain design, the predicted

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *AIChE Journal* published by Wiley Periodicals LLC on behalf of American Institute of Chemical Engineers.

modifications are experimentally tested and evaluated at a laboratory scale, and the best-performing strain is chosen.

For industrial production, industrial-scale process conditions come into play. Process design involves the analysis of mass and energy balances and production costs, including up- and downstream processing.¹⁸ In bioprocess design, a suitable kinetic represents the microorganism, for example, its growth.^{19,20} Computational process optimization supports process design by adjusting the process parameters to achieve maximal space-time yield or minimal production cost.²⁰⁻²²

Typically, strain and process design are separate steps, performed in sequence. Process conditions, however, influence the behavior and performance of the microorganism. Hence, a high-performing strain in the laboratory may underperform in an industrial process. For example, a strain is first designed under small-scale batch fermentation conditions in the laboratory. Next, the industrial process is designed for the designated strain, whereby industrial processes are envisaged to run on a large scale in continuous mode, and example processes already exist in the pharmaceutical and food sectors. This switch in conditions entails a significant adjustment for the organism. Reasons for the under-performance can be steep gradients in the spatial distribution of substrates (e.g., oxygen) in the fermenter, high downstream processing costs, or lacking genetic and phenotypic stability by the microbe.^{23,24} Strain and process design coupling is needed to overcome these scale-up difficulties.^{24,25} There exist different approaches to do so. In their attempt to capture spatial changes in a stirred bioreactor, Lapin et al.²⁶ coupled computational fluid dynamics and metabolic modeling. To model diauxic growth, Mahadevan et al.²⁷ connected the mass balance equations of batch fermentation with metabolic modeling in their dynamic flux balance analysis (dFBA). Structurally, dFBA are differential-algebraic equations with embedded optimization criteria. For fed-batch fermentation, Oliveira et al.²⁸ replaced the metabolic network in dFBA with a surrogate model. In an iterative process, Zhuang et al.²⁹ connected dFBA with computational strain optimization techniques to suit the organism to the future process conditions of batch fermentation. Ploch et al.³⁰ developed a differential-algebraic equation system with embedded dFBA optimization to model a biorefinery process under changing conditions. Jabariyelisdeh et al.³¹ introduced a bilevel dynamic optimization framework, where a previously found genetic modification can be switched on during batch fermentation. In a step-wise approach, Tafur Rangel et al.³² selected a microbial strain regarding the required downstream process units. Dimitriou et al.³³ considered downstream process optimization simultaneous to strain optimization, namely in a superstructure optimization program. However, the implications of the switch from laboratory (batch or fed-batch) fermentation conditions to industrial (continuous) fermentation conditions have yet to be considered.

We propose a new formulation combining strain optimization with process optimization. Our simultaneous approach *SimulKnock*, suggests optimal gene deletions that suit a strain to its future industrial-scale conditions, and at the same time, the process conditions are customized to the strain. *SimulKnock* is a bilevel optimization program where the upper level maximizes the space-time yield subject to the mass balances of fermentation, and the lower level maximizes biomass production based on the FBA formulation. In this article we

consider continuous fermentation, an envisaged fermentation mode in industry. *SimulKnock*, however, can readily be extended to account for other fermentation modes, too. Certainly, extensions may result in different optimization classes, for example, simultaneous optimization of batch operation and strain design would be a dynamic bilevel optimization problem, which is challenging.^{30,34} Michaelis-Menten or Monod kinetics are employed to connect the two levels. Compared to existing strain optimization formulations, *SimulKnock* can be interpreted as an extension of *OptKnock* by process equations or as a variation of dFBA with a switch to continuous fermentation and implementation of a process objective and knockout predictions. Similarly, *SimulKnock* can be seen as an extension of process optimization by the strain optimization level (lower level). For our case studies, we use strong dualization to reformulate the bilevel optimization program into a single-level program and solve it globally. We apply our formulation to a GEM of *Escherichia coli* for the production of ethanol, succinate, acetate, formate, fumarate, and lactate and compare the results with sequential optimization results and experimental results for succinate production from the literature. The proposed formulation *SimulKnock* will help close the gap in profitability between petrochemical and biochemical processes.

2 | METHOD: SIMULKNOCK

2.1 | Mass balances for a continuous process optimization

For optimal process design, the fermentation process was modeled with mathematical equations. These equations may include mass and energy balances of different reactor types and process units. In *SimulKnock*, we chose the mass balance equations of a continuous stirred-tank reactor. In principle, however, *SimulKnock* can easily be extended to account for other reactor types or even full processes. We assumed that the reactor is ideally mixed and contains only one homogeneous phase, that only one substrate limits growth and production, and that only one population of organisms is in the tank. Furthermore, we assumed that the volumetric inflow equals the volumetric outflow. The biomass, the substrate, and the product were chosen as the representative compounds of the fermentation. Similar to standard literature, mass conservation leads to

$$\begin{aligned} \frac{dc_{\text{bio}}}{dt} &= (\mu - D) \cdot c_{\text{bio}} \\ \frac{dc_S}{dt} &= -\frac{1}{Y_{\text{bio}/S}} \cdot \mu \cdot c_{\text{bio}} - \frac{1}{Y_{P/S}} \cdot q_P \cdot c_{\text{bio}} - m_S \cdot c_{\text{bio}} + D \cdot (c_{S,\text{Feed}} - c_S) \quad (1) \\ \frac{dc_P}{dt} &= q_P \cdot c_{\text{bio}} - c_P \cdot D, \end{aligned}$$

where D is the dilution rate in h^{-1} ; c_{bio} is the biomass concentration in gram Cell Dry Weight (CDW) per liter ($\text{g}_{\text{CDW}} \text{L}^{-1}$); $c_{S,\text{Feed}}$, c_S and c_P are the substrate feed concentration, substrate concentration, and product concentration, respectively (all in g L^{-1}). $Y_{\text{bio}/S}$ and $Y_{P/S}$, both with the unit g g^{-1} , are the yields of biomass and product per substrate, respectively. The maintenance factor is denoted as m_S in $\text{g g}^{-1} \text{h}^{-1}$

and describes the required substrate uptake rate for cellular maintenance. Two kinetics were used in the equations: the product kinetics q_p in $g\ g^{-1}\ h^{-1}$, describing the rate of formation of the product, and the growth rate kinetics μ in h^{-1} , describing the rate of formation of the biomass. The dilution rate D is the quotient of the volumetric inflow/outflow and the culture volume within the bioreactor. The dilution rate is also the inverse of the residence time. As is typical, we assumed that a flow equilibrium is achieved after a sufficiently long period of operation at a constant dilution rate. It follows that the continuous fermentation is at a steady state, such that $\frac{dc_{bio}}{dt} = 0$, $\frac{dc_s}{dt} = 0$, and $\frac{dc_p}{dt} = 0$ apply. Thereby, it directly follows that $D = \mu$, which means that biomass is neither washed out nor accumulated in the process.

2.2 | Flux balance analysis

FBA analyzes the internal fluxes within a cell using linear programming.⁴⁻⁶ The analysis is based on the metabolic network of an organism that includes the metabolites and the stoichiometry of the internal reactions. In FBA, based on the steady state assumption, the metabolites' net accumulation rates are constrained to zero, and the reaction fluxes are constrained to upper and lower bounds. One cellular objective is defined, for example, the maximization of the biomass flux. The mathematical formulation of FBA reads

$$\begin{aligned} \max_{\mathbf{v} \in \mathbb{R}^n} \quad & v_{bio} \\ \text{s.t.} \quad & \mathbf{S}\mathbf{v} = \mathbf{0} \\ & \mathbf{v}^{lower} \leq \mathbf{v} \leq \mathbf{v}^{upper}, \end{aligned}$$

where \mathbf{v} denotes the vector of reaction fluxes in $mmol\ g_{CDW}^{-1}\ h^{-1}$ and where n is the number of reactions in a metabolic network, where all reversible reactions were split into one forward and one backward reaction, further called irreversible network. The biomass flux is denoted by v_{bio} in h^{-1} and is an element of \mathbf{v} ; \mathbf{S} is the stoichiometric matrix of the irreversible network, and lower and upper are lower and upper bounds of the flux values, respectively. These bounds may include thresholds on the biomass flux or the ATP maintenance reaction.

2.3 | Proposed combined optimization: SimulKnock

We designed SimulKnock by embedding the FBA into the optimization of a continuous fermentation accounting for cellular degrees of freedom, resulting in a bilevel optimization formulation (Figure 1), following the paradigm of OptKnock.¹² Using the steady-state mass balances of metabolites stemming from FBA, we assume that the cellular metabolism adapts infinitesimally quickly to shifting environmental conditions.³⁵

SimulKnock maximizes the space-time yield in the upper-level program like other bioprocess optimization examples.^{36,37} The connection between upper- and lower-level program is achieved via the expression of process parameters with metabolic variables, based on Ploch et al.³⁰ and Mahadevan et al.²⁷ The growth rate directly transforms to $\mu = v_{bio}$, and due to steady-state continuous fermentation conditions, it holds $D = v_{bio}$. Moreover, instead of considering that everything taken up goes into product formation, biomass formation, or maintenance, the substrate uptake is now directly expressed by $-\frac{1}{Y_{X/S}} \cdot \mu - \frac{1}{Y_{P/S}} \cdot q_p - m_s = -v_s \cdot M_s$, where v_s is the substrate uptake flux in $mmol\ g_{CDW}^{-1}\ h^{-1}$ and M_s denotes the molar mass of the substrate in $g\ mmol^{-1}$. Thus, the maintenance is only directly considered at the lower level in SimulKnock by setting a threshold on the ATP maintenance reaction; there is no doubling at the upper level. The rate of product formation can also be directly considered now by setting $q_p = v_p \cdot M_p$, where v_p is the product flux in $mmol\ g_{CDW}^{-1}\ h^{-1}$ and is an element of the flux vector \mathbf{v} and M_p is the molar mass of the product in $g\ mmol^{-1}$. Hence, (1) transforms to

$$\begin{aligned} D &= v_{bio} \\ 0 &= -v_s \cdot M_s \cdot c_{bio} + D \cdot (c_{s,Feed} - c_s) \\ 0 &= v_p \cdot M_p \cdot c_{bio} - c_p \cdot D. \end{aligned}$$

Note that instead of fixing the dilution rate in the upper level and, thereby, fixing the growth rate, the dilution rate D will be set according to the optimal value of v_{bio} after the lower-level optimization. Hence, in the following, D will be replaced with v_{bio} .

Up to this point, we replaced process parameters with metabolic variables, using the mechanistic, FBA-based model in the cellular level. However, the process variable c_s does not yet have a connection to

Maximize Space-time yield
through gene knockouts, feed substrate concentration

Subject to

- Number of knockouts \leq limit
- Mass balance constraints of continuous fermentation including reformulated Monod kinetics for substrate concentration

Maximize Cellular objective
through fluxes

Subject to

- Mass balance constraints of metabolites
- Low threshold on biomass
- Michaelis-Menten kinetics for substrate uptake
- Knockout constraints

FIGURE 1 The bilevel optimization formulation of SimulKnock. Note that either Monod or Michaelis-Menten kinetics are applied at the upper level or the lower level.

the cellular level. Without this connection, c_S will always be set to zero during optimization (for analysis and exemplary studies, see the Appendix S1). Whether this setting is a valid assumption, however, can not be known in advance, or requires expert knowledge of the specific strain in application. In order to overcome this issue, we embedded an empirical model of kinetics. The combination of mechanistic and empirical models in our optimization formulation is necessary, because the metabolic model can only replace process variables originating from the cellular level, as we have executed above. We considered two alternative kinetics to connect the upper and lower levels: Monod and Michaelis-Menten. In the Appendix S1, we discuss why including both kinetics simultaneously is not advisable.

Monod is an empirical, widely known, and easy-to-measure model of microbial growth. It links the substrate concentration with the growth rate and reads

$$v_{\text{bio}} = v_{\text{bio}}^{\max} \frac{c_S}{c_S + K_S} \Rightarrow c_S = K_S \frac{v_{\text{bio}}}{v_{\text{bio}}^{\max} - v_{\text{bio}}}, \quad (\text{Monod}) \quad (2)$$

where v_{bio}^{\max} denotes the maximum growth rate in h^{-1} , and K_S denotes the Monod affinity constant in g L^{-1} for the specified substrate. We assumed that gene deletions would not affect the kinetic parameters, i.e., we assumed that the kinetic parameters were constants for a given organism and substrate. Thus, both v_{bio}^{\max} and K_S are parameters taken from literature or, in the case of v_{bio}^{\max} , can be approximated with FBA. The growth rate results from the lower-level program, and thus, the Monod kinetics need to be imposed on the upper-level program (reactor level). As depicted on the right-hand side in (2), we, therefore, reformulated the kinetics such that the substrate concentration is a function of the growth rate: $c_S = f(v_{\text{bio}})$. Then, we implemented the kinetics in the upper level by replacing c_S with the kinetic term. If we would keep c_S , the growth rate would be fixed by the kinetic term (equality constraint) instead of being optimized at the lower level. SimulKnock with included Monod kinetics reads

$$\begin{aligned} & \max_{\mathbf{y}, c_{S,\text{Feed}}, c_P, c_{\text{bio}}, v_S, v'_{\text{bio}}, v'_P} && c_P \cdot v'_{\text{bio}} \\ & \text{s.t.} && \sum_{i=1}^r (1 - y_i) \leq K \\ & && 0 = -v'_S \cdot M_S \cdot c_{\text{bio}} \\ & && \quad + v'_{\text{bio}} \cdot \left(c_{S,\text{Feed}} - K_S \frac{v'_{\text{bio}}}{v_{\text{bio}}^{\max} - v'_{\text{bio}}} \right) \\ & && 0 = v'_P \cdot M_P \cdot c_{\text{bio}} - c_P \cdot v'_{\text{bio}} \\ & && v' \in \arg \max_{\mathbf{v} \in \mathbb{R}^n} v_{\text{bio}} \\ & && \text{s.t. } \mathbf{Sv} = \mathbf{0} \\ & && v_{\text{bio}} \geq f \cdot v_{\text{bio,WT}} \\ & && \mathbf{v} \geq \mathbf{v}^{\text{lower}} \circ (\mathbf{By}) \\ & && \mathbf{v} \leq \mathbf{v}^{\text{upper}} \circ (\mathbf{By}), \end{aligned} \quad (3)$$

where \circ denotes the element-wise product, $\mathbf{y} \in \{0,1\}^r$ denotes the binary knockout vector, and r denotes the number of reactions in a metabolic network, which contains irreversible and reversible

reactions, further referred to as a reversible network. The parameter K is the number of maximal allowed knockouts, and the parameter f denotes a value between zero and one. The matrix \mathbf{B} maps the reactions in the irreversible metabolic network with those in the reversible network. The apostrophe ' denotes the cellular variables in the upper level. Note that, in the algorithm, they are fixed by the lower-level program. As comes clear from (3), we define “knockout” as “reaction elimination” in this manuscript. The corresponding genes to the suggested reactions can be identified via the gene-protein-reaction rule. Note that several genes can correspond to one reaction.

As described above, we chose to eliminate D and c_S from the formulation. In theory, the process variables c_{bio} and c_P could also be eliminated, as the equality constraints at the upper level are fixing them. Nevertheless, we decided on this formulation because preliminary tests and our intuition suggest that it results in simpler convergence. Note that our recent research indicates that an intermediate between full-space (keeping all variables) and reduced-space formulation (eliminating all variables except the degrees of freedom) is most beneficial in some cases.^{38,39} With c_{bio} and c_P being fixed by equality constraints, the substrate concentration in the feed $c_{S,\text{Feed}}$ is the only degree of freedom in the process conditions. At the cellular level, v_S will reach its upper bound based on an experimental value, as it is standard in other strain optimization examples (e.g., Reference 12). Even if the Monod kinetics constrain v_{bio} and $c_{S,\text{Feed}}$ is fixed during optimization; spare carbon atoms will go into product formation.

In this article, the Monod constant was set to 0.044 g L^{-1} , and v_{bio}^{\max} was set to 0.73 h^{-1} , according to Wick et al.⁴⁰ The wild-type biomass flux $v_{\text{bio,WT}}$ was determined by a previously performed FBA, and the parameter f was set to 0.1. Furthermore, the ATP maintenance reaction flux threshold was set to $6.86 \text{ mmol g}_{\text{CDW}}^{-1} \text{ h}^{-1}$, according to the default value in iML1515.⁴¹ The upper bound of the glucose exchange reaction was set to $10 \text{ mmol g}_{\text{CDW}}^{-1} \text{ h}^{-1}$.

Alternatively, we used Michaelis-Menten, the simplest form of enzyme kinetics. Enzymes catalyze reactions within a cell, with metabolites as reactants. An enzyme-catalyzed reaction can be described using Michaelis-Menten kinetics. In alignment with Mahadevan et al.²⁷ and Ploch et al.,³⁰ we chose to apply Michaelis-Menten kinetics to the substrate uptake reaction. Thus, the kinetics link the substrate concentration in the bioreactor with the substrate uptake rate. More precisely, the substrate uptake rate is a function of the substrate concentration: $v_S = g(c_S)$. The kinetics read

$$v_S = v_S^{\max} \frac{c_S}{c_S + K_{S,\text{MM}}}. \quad (\text{Michaelis-Menten})$$

The maximal substrate uptake is denoted with v_S^{\max} in $\text{mmol g}_{\text{CDW}}^{-1} \text{ h}^{-1}$, and the Michaelis constant is $K_{S,\text{MM}}$ in g L^{-1} . Again, we assumed that the kinetic parameters were constants and, thus, can be retrieved from the literature. As v_S is a cellular variable, the kinetics were implemented in the lower level. SimulKnock with embedded Michaelis-Menten kinetics reads

$$\begin{aligned}
& \max_{Y, c_S, \text{Feed}, c_S, c_P, c_{\text{bio}}, v_S', v_{\text{bio}}, v_P'} c_P \cdot v_{\text{bio}} \\
& \text{s.t.} \quad \sum_{i=1}^F (1 - y_i) \leq K \\
& \quad 0 = -v_S' \cdot M_S \cdot c_{\text{bio}} + v_{\text{bio}}' \cdot (c_S, \text{Feed} - c_S) \\
& \quad 0 = v_P' \cdot M_P \cdot c_{\text{bio}} - c_P \cdot v_{\text{bio}} \\
& \quad v' \in \arg \max_{v \in \mathbb{R}^n} v_{\text{bio}} \\
& \quad \text{s.t.} \quad Sv = 0 \\
& \quad v_{\text{bio}} \geq f \cdot v_{\text{bio}, \text{WT}} \\
& \quad v_S = v_S^{\max} \frac{c_S}{c_S + K_{S, \text{MM}}} \\
& \quad v \geq v^{\text{lower}} \circ (\mathbf{By}) \\
& \quad v \leq v^{\text{upper}} \circ (\mathbf{By}).
\end{aligned} \tag{4}$$

Again, c_{bio} and c_P are being fixed by the equality constraints in the upper level. However, unlike the formulation including Monod, c_S is a degree of freedom at the process level now. It is included in the upper-level equality constraint in the summand $v_{\text{bio}} \cdot (c_S, \text{Feed} - c_S)$. With v_{bio} being part of the upper-level objective, this term also has an optimum, and c_S is determined during the optimization. Another difference is that v_S is now constrained by the kinetics instead of reaching the upper bound set by the user.

In our article, the Michaelis constant was calculated with 0.53 mmol L^{-1} multiplied by the molar weight of glucose, and v_S^{\max} was set to $10 \text{ mmol g}_{\text{CDW}}^{-1} \text{ h}^{-1}$, both values according to Meadows et al.⁴²

To simplify the comparison of Monod and Michaelis-Menten, Table 1 summarizes the different aspects of both kinetics applied in SimulKnock.

Apart from the aspects already discussed above, we found that, in general, the parameter availability for Monod is higher than for Michaelis-Menten. Furthermore, the inclusion of inhibitions is supposedly a straightforward extension with Monod, whereas including

TABLE 1 Comparison of Monod and Michaelis-Menten in SimulKnock.

	Monod	Michaelis-Menten
Functionality	$c_S = f(v_{\text{bio}})$	$v_S = g(c_S)$
Level	Whole-cell kinetics → process level	Enzyme kinetics → cellular level
v_S	Unrestricted, reaches its upper bound	Restricted by kinetics and c_S
c_S	Calculated by kinetics	Determined by equality constraint and objective in process level
Assumption	Gene deletion does not change growth behavior	Gene deletion does not change substrate uptake behavior
Parameter availability	Higher	Lower
Possible extensions	Inclusion of inhibitions	Consideration of multiple substrates

Abbreviations: c_S : substrate concentration; f, g : notation for functions; v_{bio} : growth rate; v_S : substrate uptake rate.

multiple substrates would potentially be easier with Michaelis-Menten (cf. Reference 30).

2.4 | Reformulation of the bilevel program to a single-level quadratically constrained quadratic program

We used the software packages libALE⁴³ and libDIPS⁴⁴ to implement SimulKnock in both versions—Monod (3) and Michaelis-Menten (4)—and successfully solved the bilevel programs for the *E. coli* core network⁴⁵ using the algorithm⁴⁶ without KKT-based tightening and with Gurobi v10.0⁴⁷ as the subsolver. However, this algorithm was too expensive for the genome-scale metabolic model iML1515.⁴¹ Consequently, we reformulated the bilevel programs to single-level mixed-integer quadratically constrained quadratic programs (MIQCQP). We conducted two reformulation steps for numerical reasons: first, we reformulated SimulKnock as a single-level program. Second, we eradicated nonlinear terms to achieve an MIQCQP formulation of SimulKnock.

To the first reformulation: In both versions of SimulKnock (Monod or Michaelis-Menten embedded), the cellular level is linear (in the lower-level variables, as the process level fixes the process variables). Thus, we can apply strong duality and reformulate to a single-level program. The reformulations for SimulKnock with Monod kinetics embedded are identical to those presented in OptKnock,¹² as the Monod kinetics are in the upper-level program. In the following, we will show the reformulation of SimulKnock with Michaelis-Menten kinetics embedded. For ease of notation, we introduced

$$\tilde{\mathbf{C}} = \begin{bmatrix} \mathbf{I}^{|\mathbf{n}|} \\ \mathbf{I}^{|\mathbf{n}|} \\ \mathbf{c}_{\text{kin}}^T \end{bmatrix}, \tilde{\mathbf{b}} = \begin{bmatrix} v^{\text{lower}} \circ (\mathbf{By}) \\ -v^{\text{upper}} \circ (\mathbf{By}) \\ d \end{bmatrix}, c_i = \begin{cases} 0, & \text{if } i \neq \text{bio} \\ 1, & \text{if } i = \text{bio}, \end{cases} \quad \forall i \in 1, \dots, n,$$

where c describes the position of the biomass reaction and \mathbf{I} is the unity matrix. The vector $\mathbf{c}_{\text{kin}} \in \{0, 1\}^{|\mathbf{n}|}$ has exactly one nonzero entry at the index of the glucose uptake reaction. The scalar d describes the right-hand side of the kinetics, that is, $d = v_S^{\max} \frac{c_S}{c_S + K_{S, \text{MM}}}$. By applying the definitions above, the lower-level program of (4) can be rewritten as

$$\begin{aligned}
& \max_v \quad c^T v \\
& \text{s.t.} \quad S \cdot v = 0 \\
& \quad \tilde{\mathbf{C}} \cdot v \leq \tilde{\mathbf{b}} \\
& \quad v \geq 0.
\end{aligned} \tag{5}$$

Converting program (5) to its dual and applying the strong duality theorem yields the system of equations

$$\begin{aligned}
& c^T v = \tilde{\mathbf{b}}^T \cdot \tilde{\mu} \\
& S \cdot v = 0 \\
& \tilde{\mathbf{C}} \cdot v \leq \tilde{\mathbf{b}} \\
& c = S^T \cdot \lambda + \tilde{\mathbf{C}}^T \cdot \tilde{\mu} \\
& \tilde{\mu} \geq 0,
\end{aligned} \tag{6}$$

with $\lambda \in \mathbb{R}^{|\mathcal{M}|}$ being the dual variables corresponding to the mass balances for the metabolites and $\bar{\mu} \in \mathbb{R}_+^{2n+1}$ being the dual variables corresponding to the bounds set on the reaction fluxes through the inequality constraints and the kinetics. The set of Equations (6) can readily replace the cellular level in (4). The resulting program is a single-level mixed-integer nonlinear program. The nonlinearity is introduced by the fact that process and cellular variables are at the same level now. The nonlinear term is the kinetic term in d . These considerations about nonlinearity are also applicable to the reformulated SimulKnock with Monod embedded.

In the second reformulation step, we reformulated the nonlinear kinetic term. We introduced an additional variable for the fraction on the right-hand side of the kinetics, thereby introducing an additional equality constraint in the problem. For example, in the Michaelis-Menten kinetics, the fraction is replaced with the optimization variable σ_{MM} , and the constraints

$$\begin{aligned} v_S &= v_S^{\max} \sigma_{MM} \\ \sigma_{MM}(c_S + K_{S,MM}) &= c_S \end{aligned} \quad (7)$$

are introduced, which results in the single-level optimization problem becoming an MIQCQP. This reformulation is beneficial because MIQCQPs are solved directly by commercial solvers, such as Gurobi.⁴⁷

Replacing the lower level in (4) with the set of constraints of (6) and applying the reformulation of the kinetic term suggested in (7) yields the single-level mixed-integer quadratic form of SimulKnock with Michaelis-Menten embedded:

$$\begin{aligned} \max_{y, c_S, c_S, \text{Feed}, \sigma} \quad & c_P \cdot v_{\text{bio}} \\ \text{s.t.} \quad & \sum (1 - y_i) \leq 1 \\ & 0 = -v_S \cdot M_S \cdot c_{\text{bio}} + v_{\text{bio}} \cdot (c_{S, \text{Feed}} - c_S) \\ & 0 = v_P \cdot M_P \cdot c_{\text{bio}} - c_P \cdot v_{\text{bio}} \\ & v_S = v_S^{\max} \sigma_{MM} \\ & c_S = \sigma_{MM}(c_S + K_{S,MM}) \\ & v_{\text{bio}} = [-v^{\text{lower}} \circ (\mathbf{B}y) \quad v^{\text{upper}} \circ (\mathbf{B}y) \quad v_S^{\max} \sigma] \cdot \boldsymbol{\mu} \\ & \mathbf{S} \cdot \mathbf{v} = \mathbf{0} \\ & v \geq v^{\text{lower}} \circ (\mathbf{B}y) \\ & v \leq v^{\text{upper}} \circ (\mathbf{B}y) \\ & \mathbf{S}^T \cdot \boldsymbol{\lambda} + [\mathbf{I}^{|\mathcal{N}|} \quad \mathbf{I}^{|\mathcal{N}|} \quad c_S] \cdot \boldsymbol{\mu} = \mathbf{c} \\ & v_{\text{bio}} \geq f \cdot v_{\text{bio}}^{\max} \\ & \mathbf{v}, \boldsymbol{\mu} \geq \mathbf{0}. \end{aligned} \quad (8)$$

Similar reformulations also reduce SimulKnock with Monod kinetics to an MIQCQP.

The reformulated programs were implemented using Pyomo.^{48,49} In a preprocessing step, the search space of reaction eliminations in the metabolic network was reduced. All exchange reactions, diffusive transport reactions and reactions without gene-protein-reaction rule were excluded from potential elimination. Note that users of our open-source code may also decide to exclude transporters and the corresponding reactions from elimination, which represents a small extension of our preprocessing. The software package COBRAPy ver.0.26.0⁵⁰ was used to also exclude blocked reactions and lethal reactions from the search space. Similar to Larhlmi et al.⁵¹ and based

on flux coupling analysis,⁵² lastly, the metabolic network was divided into sections. This last step of network reduction was applied in our case studies with GEMs but is optional for users of the open-source code. All programs were solved with Gurobi v10.0.⁴⁷

3 | CASE STUDIES

We performed three case studies with SimulKnock. The first case study illustrates SimulKnock's mode of action compared to OptKnock and a sequential optimization approach. The second case study employs the genome-scale metabolic model iML1515⁴¹ of *E. coli* for one to three knockout predictions and elaborates on the differences between Monod and Michaelis-Menten kinetics. Note again that the term knockout refers to reaction elimination in this article and that one knockout can result in multiple gene deletions, if several genes correspond to the targeted reaction. The third case study compares the results of SimulKnock with the published results of an experimental study on *E. coli* continuous fermentation.⁵³ The first case study was performed on up to 8 Intel Xeon E5-2640 CPU threads and was solved within less than a second. Case studies two and three were solved within 5 h on the RWTH high-performance computing cluster using 48 threads and up to 4 GB of memory per thread.

3.1 | SimulKnock predicts different knockouts than OptKnock with embedded illustrative network

To show SimulKnock's mode of action compared to sequential optimization using OptKnock, we constructed an illustrative network (Figure 2).

We performed four optimizations with the illustrative network embedded: (i) an FBA with maximization of biomass as the objective (referred to as wild-type), (ii) an OptKnock prediction, (iii) a sequential approach using OptKnock (see below), and (iv) SimulKnock. Table 2 shows the results of the case study.

The sequential approach denotes a two-step procedure. First, OptKnock is performed. Second, the predicted knockouts are applied to the network, and a continuous process optimization is performed. The continuous process optimization looks like the formulation of SimulKnock but with fixed knockouts, which were determined from OptKnock. Thus, OptKnock is included in the sequential approach. Still, we added the results of OptKnock for completeness to allow for a direct comparison of functionalities with SimulKnock. Michaelis-Menten kinetics were applied in the sequential approach and SimulKnock. The FBA and OptKnock predictions do not imply process variables, for example, the substrate or product concentration, which is also why the kinetics were not applied in OptKnock or FBA.

The results indicate that SimulKnock and OptKnock predict different optimal knockouts for one maximum allowable knockout. The reason for this difference is that OptKnock optimizes for the target chemical production, whereas SimulKnock optimizes for space-time yield. Indeed, the product flux v_P is higher for OptKnock than for SimulKnock. The applied kinetics in SimulKnock and the sequential approach force the substrate uptake lower than the maximum allowed value. Namely, in the sequential approach, the substrate uptake is set

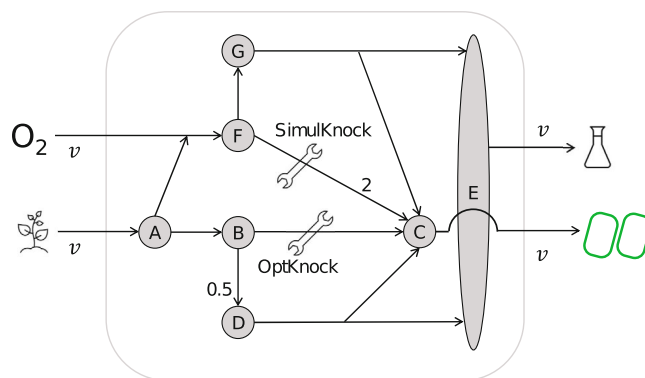


FIGURE 2 Illustrative network. The grey circles A to G denote metabolites, v_O and v_S denote the flux of oxygen and the substrate uptake reaction, respectively; v_P and v_{bio} are the target chemical and biomass flux, respectively. All stoichiometric coefficients are one, except where indicated with numbers. For one maximum allowable knockout, the wrenches indicate the knockout prediction of SimulKnock and OptKnock.

TABLE 2 Results of the case study with the illustrative network (Figure 2) embedded.

	Knockout (-)	v_S ($\text{mmol g}_{CDW}^{-1} \text{h}^{-1}$)	v_O ($\text{mmol g}_{CDW}^{-1} \text{h}^{-1}$)	v_{bio}	v_P	c_{bio} (g L^{-1})	c_P	STY ($\text{g L}^{-1} \text{h}^{-1}$)
Wild-type	—	10	3	13	0	—	—	—
OptKnock	B-C	10	3	9.5	3.5	—	—	—
Sequential	B-C	7.8	3	8.4	2.4	8.7	2.5	21.0
SimulKnock	F-C	3	3	3	3	9.8	9.8	29.3

Notes: The substrate is glucose, and the maximum substrate flux allowed is set to $v_S^{upper} = 10 \text{ mmol g}_{CDW}^{-1} \text{h}^{-1}$; the upper bounds on A-B and A-F are set to 7 and 3 $\text{mmol g}_{CDW}^{-1} \text{h}^{-1}$, respectively; the maximum oxygen flux allowed is $v_O^{upper} = 3 \text{ mmol g}_{CDW}^{-1} \text{h}^{-1}$. Michaelis-Menten kinetics were applied in SimulKnock and the sequential approach, with the Michaelis constant $K_{S,MM} = 0.53 \text{ mmol L}^{-1}$.⁴² In the sequential approach, 1. OptKnock, and 2. a process optimization is performed with embedded knockouts from OptKnock. Abbreviations: c_{bio} , biomass concentration; c_P , product concentration; STY, space-time yield; v_S , substrate uptake flux; v_O , oxygen flux; v_{bio} , growth rate.

to $8 \text{ mmol g}_{CDW}^{-1} \text{h}^{-1}$ by the process optimization, whereas OptKnock predicted it to be $10 \text{ mmol g}_{CDW}^{-1} \text{h}^{-1}$. In SimulKnock, only the reactions in the upper part of the network, that is, the reactions going via the metabolites A, F, G, E, and C, are active and used for biomass and target chemical production. In contrast, in OptKnock and the sequential approach, the reactions going via B, D, and E are used for target chemical production, while the route via F and C produces additional biomass. The space-time yield, which is the product of the growth rate v_{bio} and the product concentration c_P , is higher for SimulKnock than for the sequential approach. Note that both formulations predict the same two knockouts, B-C and F-C, and the same space-time yield for two maximum allowable knockouts. However, this does not generally indicate that with more allowed knockouts, the results of OptKnock and SimulKnock would be similar. Instead, the illustrative network is constructed so small that no other options exist.

3.2 | SimulKnock can achieve higher space-time yields than sequential optimization with embedded iML1515

To further investigate the computational tractability of the SimulKnock approach, we chose a genome-scale metabolic model, iML1515,⁴¹ which describes the metabolism of *E. coli*. The metabolic

network includes 1516 genes, resulting in 1877 metabolites and 2712 metabolic reactions. The objective was set to maximize the space-time yield of six target chemicals for a glucose-limited medium.

Furthermore, to highlight the superiority of the simultaneous strain and process optimization, SimulKnock was compared against the sequential approach on the space-time yield of the target chemical.

3.2.1 | One knockout

For one knockout prediction, SimulKnock and the sequential approach furnished identical results. The knockout predictions and space-time yields are in the Appendix S1 (cf. Figure S1 and Table S2).

3.2.2 | Two knockouts

In this case study, we allowed for two gene knockouts. The comparison results of Michaelis-Menten and Monod kinetics are plotted in Figure 3.

Note that for *E. coli*, the growth rate is higher with oxygen than without. The lower-level optimization problem always activates the oxygen uptake to maximize growth. Hence, in SimulKnock, the oxygen

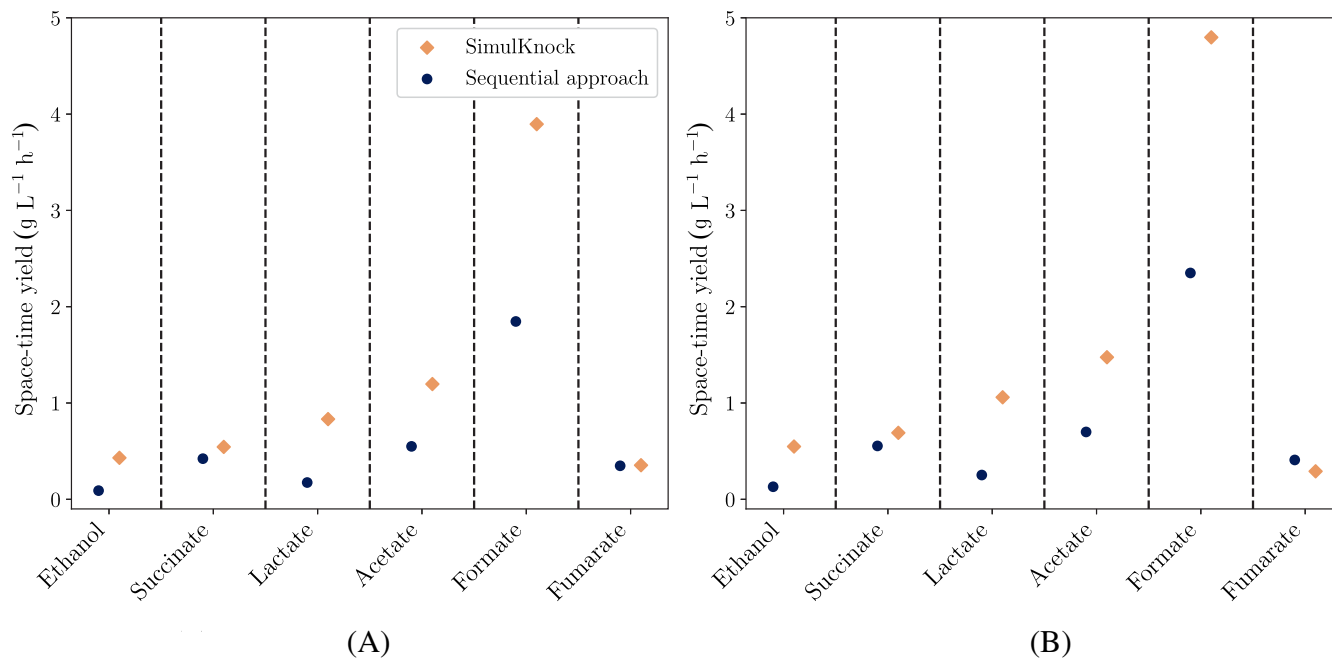


FIGURE 3 Comparison of space-time yield using Michaelis-Menten and Monod kinetics for two reaction eliminations on the genome-scale metabolic model iML1515.⁴¹ Ethanol, succinate, and lactate were produced via anaerobic pathways; acetate, formate, and fumarate were produced via aerobic pathways. (A) Michaelis-Menten, (B) Monod.

supply must be set by the modeler. We performed runs with and without an oxygen supply. Figure 3 shows the results where SimulKnock achieved the higher absolute value.

SimulKnock predicts significantly higher space-time yields for four of the six considered target chemicals with embedded Michaelis-Menten kinetics. The difference in the space-time yield from the two approaches is due to different reaction knockouts predicted (see Table S3 for reference). For example, SimulKnock targets the 6-phosphogluconolactonase and the acetate reversible transport for higher ethanol space-time yield instead of the ATP synthase and the triose-phosphate isomerase. The most significant difference becomes visible for formate. Instead of the phosphoglycerate mutase, SimulKnock predicts the phosphofructokinase will be knocked out. Even if SimulKnock and the sequential approach do not predict the same knockouts for fumarate, they reach the same result with Michaelis-Menten. Interestingly, the predicted reaction eliminations for ethanol and lactate production are equal. The reason for this result is the optimistic formulation of SimulKnock: pathways which are competing are not recognized.

The kinetics also influence the knockout predictions. SimulKnock with embedded Michaelis-Menten kinetics suggests blocking acetaldehyde dehydrogenase and D-alanine-D-alanine dipeptidase for succinate production. With Monod embedded, dytosine deaminase is knocked out instead of the dipeptidase. Interestingly, SimulKnock achieves a lower space-time yield for fumarate than the sequential approach with embedded Monod kinetics.

The sequential formulation maximizes the target chemical concentration at the process level, with the biomass growth rate being used only in the metabolism level problem. Therefore, the targeted reaction knockouts decrease the growth rate to achieve the maximum target chemical concentration. In contrast, the SimulKnock formulation

recognizes the trade-off between a higher growth rate and a higher target chemical flux. It is, therefore, not surprising that the sequential approach predicts a higher target chemical concentration and lower biomass concentrations than the SimulKnock approach.

The different modes of action become even more visible when comparing the molar product yield on the substrate. We computed the yield by dividing the target chemical flux v_P by the substrate flux v_S (cf. Table S3 for the data). In all cases, the molar product yield was lower with SimulKnock than with the sequential approach. At maximum, in the case of acetate production, the yield was 45% lower, whereas the space-time yield was 2.6 times higher. With the same underlying uptake rate, the production envelopes of succinate and acetate (Figure 4A,B, respectively) give a detailed picture on product yield.

In both cases of succinate and acetate, the envelope of SimulKnock lies right-hand of the envelope of OptKnock, which was to be expected. This established form of production envelope, however, is not able to show the mode of action and strength of SimulKnock. Therefore, we created a new form of production envelope, where space-time yield is plotted over biomass growth (Figure 4C,D). We compare SimulKnock with the sequential approach. It now comes visible that the space-time yield which is reachable with the mutant created with SimulKnock is larger than the one with the sequential approach. Also the optimistic nature of both formulations comes visible, with both formulations exhibiting a vertical descent in their optimal point.

3.2.3 | Three knockouts

Upon increasing the possible reaction eliminations to three, the SimulKnock formulation with Michaelis-Menten kinetics could not be

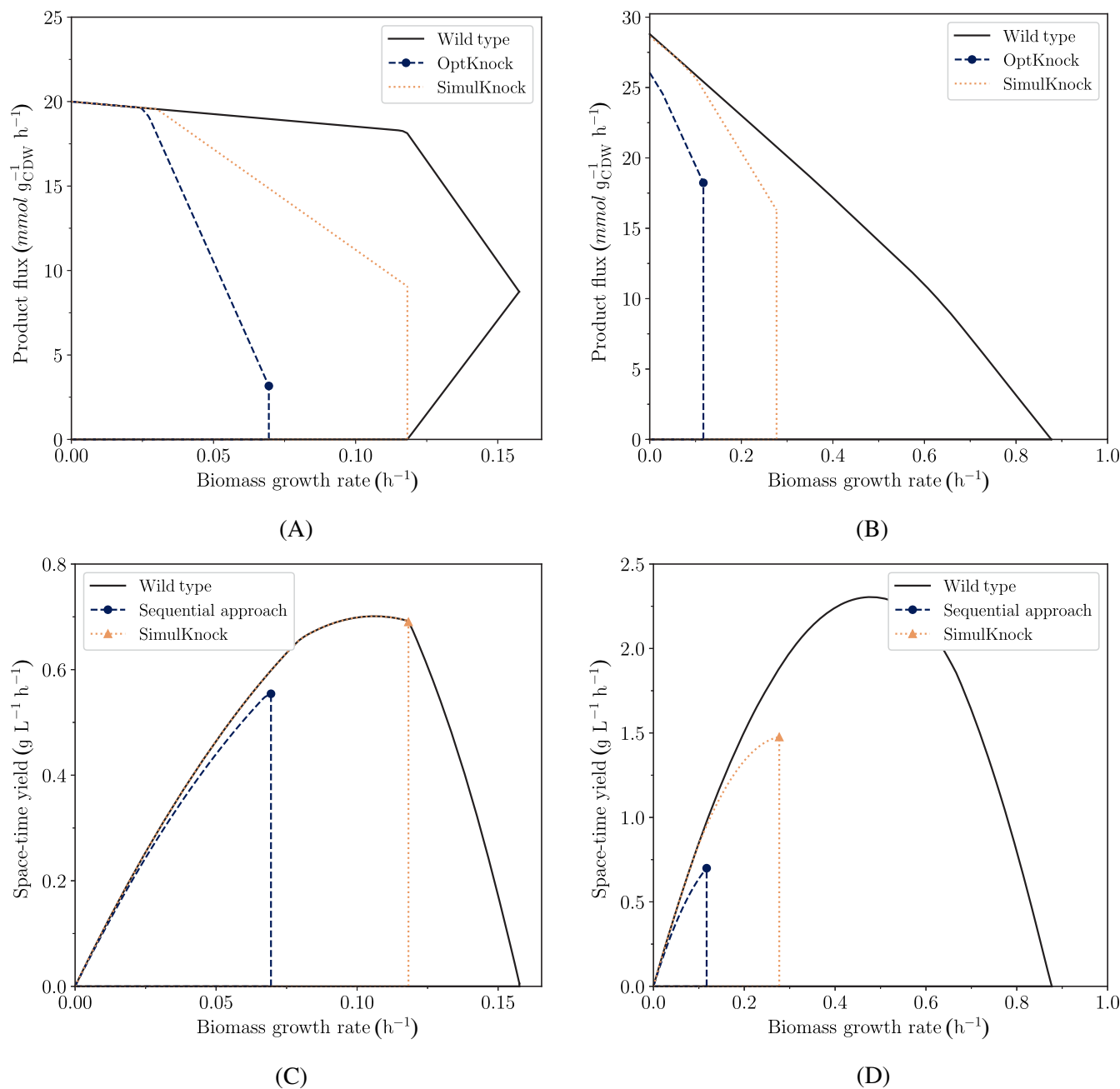


FIGURE 4 Production envelopes for succinate and acetate production with Monod kinetics for the wild type and the mutants (two knockouts) predicted by SimulKnock and OptKnock/Sequential approach. Succinate (A,C) is produced anaerobically; acetate (B,D) is produced aerobically. (C,D) represent a new form of production envelope, where space-time yield is plotted against biomass growth rate. The plots are generated by repeatedly maximizing and minimizing product flux or space-time yield, with a growth rate fixed to a fraction of its maximum. The points and triangles denote the optimal value of the respective formulation. Note that no optimal value can be given for SimulKnock in (A) and (B), due to SimulKnock not optimizing for product flux. (A) Succinate, (B) Acetate, (C) Succinate, (D) Acetate.

solved within a feasible time limit. Thus, only the Monod kinetics results are presented in Figure 5.

Note that, again, we performed runs with and without an oxygen supply. Figure 5 shows the results where SimulKnock achieved the higher absolute value, all achieved using aerobic pathways.

Acetate and formate space-time yield is increased by about 100% when using SimulKnock, whereas ethanol, succinate, and fumarate increase marginally, and lactate remains the same. Once again, the

reason for this difference is the different reaction knockouts predicted by the two approaches (cf. Table S4 for data). The knockouts for lactate and fumarate are similar, often targeting neighboring reactions, thereby predicting similar yields. When ethanol is considered, two of the three reaction eliminations target the same reactions with SimulKnock when compared to the sequential approach. When the target chemical is acetate, all three knockouts are different; with formate, two of the three knockouts are different. This observation hints

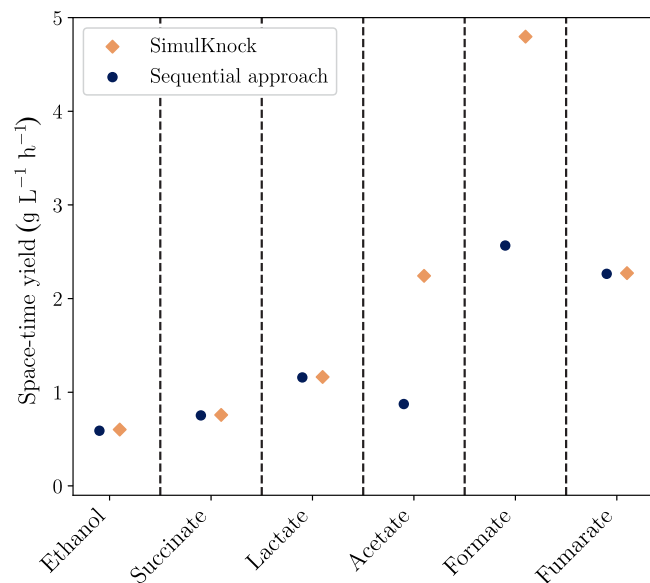


FIGURE 5 Space-time yield using Monod kinetics for three reaction eliminations on the genome-scale metabolic model iML1515.⁴¹ All six target chemicals were produced via aerobic pathways.

toward the possibility of missing possible knockout strategies with the sequential approach.

It is worth noting that the assumption made in this article about the biomass reaction being the rate-limiting step has yet to be experimentally validated. Therefore, reaction eliminations could, theoretically, target the rate-limiting step on which the Monod kinetic parameters were fitted. In that case, a new parameter fitting using the modified organism would be required. In contrast, the Michaelis-Menten kinetics do not require parameter refitting due to reaction eliminations, as the substrate uptake reaction is always active.

When we line up space-time yield and growth rate with an increasing number of knockouts, we observe different behaviors: For ethanol, succinate, lactate, and formate, space-time yield and growth rate with SimulKnock stay more or less constant over one to three knockouts. Exemplary, Figure 6A shows the course for succinate.

In the sequential approach, space-time yield decreases with an increasing number of knockouts, as depicted at two gene knockouts in Figure 6A. This decrease in space-time yield comes due to the mathematical structure of OptKnock. In OptKnock, high product yields are achieved at the expense of reduced growth (cf. Table S4 for data). In turn, space-time yield (equals product concentration times growth rate) is linearly dependent on the growth rate. The figure depicts the aerobic pathway for three gene knockouts, following our decision to show the results for the highest SimulKnock space-time yield. Hence, the fermentation conditions change compared to the anaerobic conditions that apply for one and two gene knockouts. While the space-time yield of the sequential approach would decline for anaerobic conditions (not shown in the figure), it reaches similar results as SimulKnock for aerobic conditions.

Acetate and fumarate exhibit a different behavior, as depicted in an example in Figure 6B. With SimulKnock, the space-time yield rises

significantly at three knockouts, in alignment with the increase in growth. With the sequential approach, the space-time yield is lowest for two knockouts and then increases again, decreasing the growth rate. This interesting behavior stems from the two-step procedure. Only in the second step are the process conditions optimized, but they can only influence one of the two factors, namely the product concentration c_p . Thus, there is no monotonous behavior with the sequential approach. This finding displays the problems often occurring during scale-up when sequentially optimizing the microbial strain and the process conditions. On the other hand, the acetate case underpins the potential of SimulKnock.

3.3 | SimulKnock can furnish meaningful knockout predictions but also exhibits model-experiment mismatch

In the last case study, we took the experimental study of van Heerden et al.⁵³ as a test case to elaborate on how far SimulKnock reproduces their results regarding space-time yield, dilution rate, and knockouts. We aimed to see whether SimulKnock furnishes experimentally meaningful results. In their experimental study, van Heerden et al.⁵³ produced succinic acid from *E. coli* KJ134 in a continuous fermentation with two different glucose feed concentrations. We applied SimulKnock with embedded Monod kinetics to the *E. coli* GEMs iML1515⁴¹ as well as iEC1349_Crooks⁵⁴ with three maximum allowable knockouts. While van Heerden et al.⁵³ tested different dilution rates and measured the space-time yield for each, SimulKnock predicted the optimal space-time yield and the corresponding dilution rate. Figure 7 shows the case study results, depicted as the space-time yield over the dilution rate. Note that, in continuous fermentation, the dilution rate equals the growth rate μ and the biomass flux v_{bio} .

In the experimental study, 13 genetic modifications were applied to the organism to reach a maximum space-time yield of around $1 \text{ g L}^{-1} \text{ h}^{-1}$ on 20 g L^{-1} glucose feed. SimulKnock predicted a maximum space-time yield of about 4 and $6 \text{ g L}^{-1} \text{ h}^{-1}$ with iML1515 and iEC1349_Crooks, respectively, on 50 g L^{-1} . The space-time yields were reached with three knockouts. The reactions to be eliminated, predicted with iML1515 embedded, were acetate kinase, ATP synthase, and fumarase. With iEC1349_Crooks embedded, the predicted reaction eliminations were phosphotransacetylase, ATP synthase, and succinyl-CoA synthetase. Among them, acetate kinase and phosphotransacetylase were also eliminated in the experimental strain. Hence, SimulKnock proves to predict experimentally meaningful results. However, the figure reveals a substantial model-experiment mismatch. Especially the results of iEC1349_Crooks are higher than any experimental value. In iEC1349_Crooks, the high growth rate might result from the biomass function difference. For illustration, for an FBA with a glucose limitation of $10 \text{ mmol g}_{\text{CDW}}^{-1} \text{ h}^{-1}$, iML1515 has a growth rate of 0.87 h^{-1} , whereas iEC1349_Crooks has a growth rate of 1.02 h^{-1} . The mismatch in space-time yield that comes visible with both metabolic networks might result from SimulKnock not considering inhibition and repression effects and being

FIGURE 6 Change of space-time yield and growth rate over the number of knockouts for iML1515⁴¹ with embedded Monod kinetics. Succinate (A) was produced anaerobically for one and two knockouts and aerobically for three knockouts. Acetate (B) was produced via an aerobic pathway. STY: space-time yield, Seq. Appr.: Sequential approach. (A) Succinate; (B) Acetate.

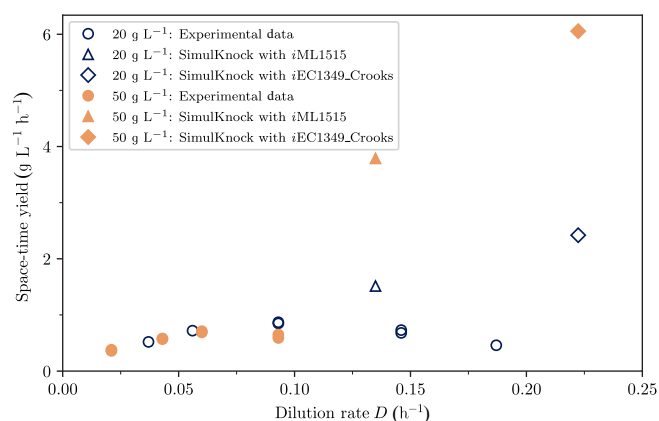
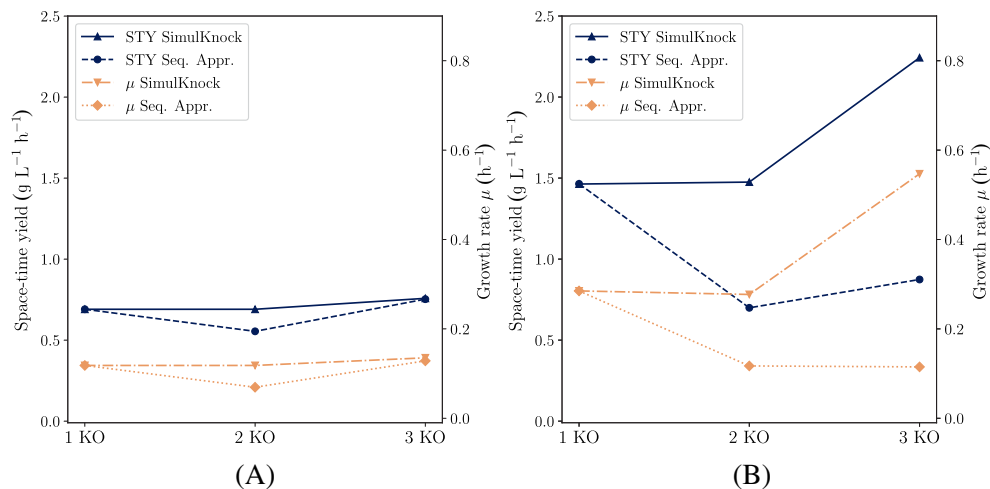


FIGURE 7 Comparison of experimental laboratory data for succinic acid production using *E. coli* KJ134⁵³ with results from SimulKnock, based on the metabolic networks iML1515⁴¹ and iEC1349_Crooks.⁵⁴ Two glucose feed concentrations were studied: 20 g L^{-1} and 50 g L^{-1} .

overly optimistic by not considering byproduct formation unless the byproduct formation hinders biomass formation. The mismatch could be reduced by including experimental flux data of the organism in question, for example, from Fischer et al.⁵⁵ for *E. coli*, or by using additional information from a protein allocation model.⁵⁶ In both cases, the bounds of specific fluxes would be fixed in a preprocessing step, which can readily be done in the SimulKnock code. These refinements of the metabolic network and adaptations at the cellular level are interesting in case of application but were not in the scope of this article.

4 | CONCLUSION

We presented SimulKnock, an optimization formulation combining continuous fermentation optimization with strain optimization via gene knockouts. Such an optimization formulation allows us to consider industrial fermentation conditions already in the strain design,

thereby overcoming current scale-up difficulties. SimulKnock is a bilevel optimization formulation, which we transformed into a single-level mixed-integer nonlinear optimization program and solved globally for an illustrative example network as well as for the *E. coli* GEM iML1515.⁴¹ We applied Monod or Michaelis-Menten kinetics to connect the process level to the cellular level in SimulKnock. SimulKnock with applied Monod kinetics showed similar results as Michaelis-Menten kinetics, with Monod being less computationally intensive. SimulKnock predicted different knockouts than OptKnock.¹² Also, the space-time yield was significantly higher with SimulKnock compared to a sequential approach of OptKnock plus process optimization for aerobic and anaerobic conditions. Compared to experimental data,⁵³ SimulKnock indicated that higher space-time yields could be achieved with fewer knockouts. SimulKnock is readily applicable to different strains by exchanging the metabolic network and to different target products measurable in a fermenter, represented as exchange reactions in the respective metabolic network.

The computations for SimulKnock with embedded GEM were performed on a high-performance computing cluster. A decrease in the computation demand and reduction of computation time could be achieved by applying further network reduction methods⁵⁷ or by considering essential genes.⁵⁸ Adaptation of the OptForce¹⁶ framework for SimulKnock could also lead to speed-up and a higher number of possible knockouts, as Chowdhury et al.⁵⁹ have already shown significantly lower computation time of OptForce compared to OptKnock with increasing number of interventions in the metabolic network. Future work should consider that kinetics would have to be adjusted when knockouts are applied to the organism because Michaelis-Menten, describing the substrate uptake, and Monod, describing the growth behavior, are affected by genetic modifications. Especially for altered growth behavior, the update of the kinetics should be considered.

Our results of the case study with experimental data indicated that SimulKnock furnishes overly optimistic results. Furthermore, SimulKnock does not account for competing pathways, as came clear when SimulKnock predicted the same reaction eliminations for ethanol and lactate production. The conversion to a robust optimization

formulation, as suggested in RobustKnock¹³ and OptForce,¹⁶ could tackle these two observations. Moreover, experimental flux data (for *E. coli*, see Pharkya et al.¹⁵ and Fischer et al.⁵⁵) could reduce the model-experiment mismatch by fixing the bounds of specific fluxes in a preprocessing step. With available experimental data, SimulKnock could also be integrated in the Design section of a Design-Build-Test-Learn pipeline, as it has already been suggested for OptKnock.^{60,61} An interesting future study could also include determining the complete knockout set of SimulKnock by iteratively solving SimulKnock and excluding the previously found solution from the search space of the next iteration. In this scope, a comparison with OptKnock appears of interest, for example, to identify the number of iterations needed, when such procedure is applied to OptKnock, to gain the same solution as suggested by SimulKnock in the first place.

To make a more accurate prediction, SimulKnock could be extended both at the process level and at the cellular level, for example, by more elaborate microbial formulations,^{2,8,9} as depicted by OptKnock¹² and its extensions.^{62,63} These formulations include a reference flux distribution, which could be furnished using a protein allocation model⁵⁶ or experimental *in vivo* flux data.⁶⁴ More elaborate microbial formulations or an adapted OptForce framework¹⁶ could also tackle the current limitation of SimulKnock to growth-coupled production, which is due to the maximization of growth in the cellular level. The extension could also comprise other genetic modifications, that is, gene insertion¹⁴ and regulation,^{15,16} other compounds, for example, oxygen, as well as other process operation modes, for example, continuous mode with cell retention, batch, and fed-batch. The three key metrics for an efficient, economic fermentation are yield, titer, and productivity, that is, space-time yield. Optimizing these three factors results in optimal substrate consumption, low downstream separation effort, and optimal fermentation reactor size, respectively. Yield is maximized by OptKnock,¹² and SimulKnock maximizes space-time yield. With yield and space-time yield being conflicting objectives, our results depicted that OptKnock outperforms SimulKnock with respect to yield, and SimulKnock outperforms OptKnock with respect to space-time yield. An idea to gain the full picture would be to extend SimulKnock to a multi-objective optimization, targeting all three key metrics at the same time.

AUTHOR CONTRIBUTIONS

ALZ designed the SimulKnock formulation, the reformulation, the implementation, and the case studies. MS performed the formulation setup, the reformulation, the implementation, and the case studies under the supervision of ALZ and AsMa. ALZ and MS visualized the data. ALZ, MS, and AsMa analyzed and discussed the data. ALZ and AsMa wrote the manuscript draft. AIMi and LMB had the initial idea of SimulKnock, discussed the data, and reviewed the draft. AIMi secured funding. All authors read and approved the final manuscript.

Anita L. Ziegler: conceptualization; methodology; software; validation; investigation; writing – original draft; writing – review and editing; visualization; project administration. **Ashutosh Manchanda:** methodology; software; investigation; writing – review and editing. **Marc-Daniel Stumm:** writing – review and editing; methodology; software; validation; investigation; visualization. **Lars M. Blank:**

writing – review and editing; conceptualization. **Alexander Mitsos:** conceptualization; funding acquisition; supervision; methodology, writing – review and editing.

ACKNOWLEDGMENTS

This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – Cluster of Excellence 2186 “The Fuel Science Center” – ID: 390919832. This project benefited from the work of Clemens Kortmann, who improved the modularity and the user-friendliness of our git repository. Computations were performed with computing resources granted by RWTH Aachen University under project thes1376. Open Access funding enabled and organized by Projekt DEAL.

CONFLICTS OF INTEREST STATEMENT

No conflicts to declare.

DATA AVAILABILITY STATEMENT

The implementation of the SimulKnock approach, the data preprocessing, and the interface with the solver Gurobi are openly available in our GitLab repository “SimulKnock: Simultaneous design of fermentation and microbe” at <https://git.rwth-aachen.de/avt-svt/public/simulknock>. Data from the article's figures are tabulated in the Appendix S2.

ORCID

Marc-Daniel Stumm  <https://orcid.org/0009-0000-7464-4122>

Alexander Mitsos  <https://orcid.org/0000-0003-0335-6566>

REFERENCES

1. Clomburg JM, Crumbley AM, Gonzalez R. Industrial biomanufacturing: the future of chemical production. *Science*. 2017;355:aag0804.
2. Landon S, Rees-Garbutt J, Marucci L, Grierson C. Genome-driven cell engineering review: in vivo and in silico metabolic and genome engineering. *Essays Biochem*. 2019;63:267-284.
3. Maia P, Rocha M, Rocha I. In silico constraint-based strain optimization methods: the quest for optimal cell factories. *Microbiol Mol Biol Rev*. 2016;80:45-67.
4. Savinell JM, Palsson BØ. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *J Theor Biol*. 1992;154:421-454.
5. Watson MR. A discrete model of bacterial metabolism. *Comput Appl Biosci*. 1986;2:23-27.
6. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*. 2010;28:245-248.
7. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U. Multidimensional optimality of microbial metabolism. *Science*. 2012;336:601-604.
8. Segrè D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A*. 2002;99:15112-15117.
9. Brochado AR, Andrejev S, Maranas CD, Patil KR. Impact of stoichiometry representation on simulation of genotype-phenotype relationships in metabolic networks. *PLoS Comput Biol*. 2012;8:e1002758.
10. Shlomi T, Berkman O, Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A*. 2005;102:7695-7700.
11. Valderrama-Gomez MA, Kreitmayer D, Wolf S, Marin-Sanguino A, Kremling A. Application of theoretical methods to increase succinate

- production in engineered strains. *Bioprocess Biosyst Eng.* 2017;40:479-497.
12. Burgard AP, Pharkya P, Maranas CD. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng.* 2003;84:647-657.
 13. Tepper N, Shlomi T. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics.* 2010;26:536-543.
 14. Pharkya P, Burgard AP, Maranas CD. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.* 2004;14:2367-2376.
 15. Pharkya P, Maranas CD. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab Eng.* 2006;8:1-13.
 16. Ranganathan S, Suthers PF, Maranas CD. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol.* 2010;6:e1000744.
 17. Kim J, Reed JL, Maravelias CT. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *PLoS One.* 2011;6:e24162.
 18. Biegler LT. *Systematic Methods of Chemical Process Design.* Prentice Hall International Series in the Physical and Chemical Engineering sciences. Prentice Hall PTR; 1997.
 19. Chmiel H, Takors R, Weuster-Botz D, eds. *Bioprozesstechnik.* Springer Spektrum4. auflage ed; 2018.
 20. Villadsen J, Nielsen J, Lidén G. *Bioreaction Engineering Principles.* 3rd ed. Springer Science+Business Media; 2011.
 21. Edgar TF, Himmelblau DM, Lasdon LS. *Optimization of Chemical Processes.* McGraw-Hill Chemical Engineering Series. 2nd ed. McGraw-Hill; 2001.
 22. Gordeeva YL, Gordeev LS. Optimization of continuous microbiological synthesis processes with nonlinear microbial growth kinetics. *Theor Found Chem Eng.* 2015;49:829-835.
 23. Wehrs M, Tanjore D, Eng T, Lievens J, Pray TR, Mukhopadhyay A. Engineering robust production microbes for large-scale cultivation. *Trends Microbiol.* 2019;27:524-537.
 24. Olsson L, Rugbjerg P, Torello Pianale L, Trivellini C. Robustness: linking strain design to viable bioprocesses. *Trends Biotechnol.* 2022;40:918-931.
 25. Richelle A, David B, Demaegd D, et al. Towards a widespread adoption of metabolic modeling tools in biopharmaceutical industry: a process systems biology engineering perspective. *NPJ Syst Biol Appl.* 2020;6:6.
 26. Lapin A, Müller D, Reuss M. Dynamic behavior of microbial populations in stirred bioreactors simulated with Euler–Lagrange methods: traveling along the lifelines of single cells. *Ind Eng Chem Res.* 2004;43:4647-4656.
 27. Mahadevan R, Edwards JS, Doyle FJ. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J.* 2002;83:1331-1340.
 28. Oliveira RD, Guedes MN, Matias J, Le Roux GAC. Nonlinear predictive control of a bioreactor by surrogate model approximation of flux balance analysis. *Ind Eng Chem Res.* 2021;60:14464-14475.
 29. Zhuang K, Yang L, Cluett WR, Mahadevan R. Dynamic strain scanning optimization: an efficient strain design strategy for balanced yield, titer, and productivity. DySSCo strategy for strain design. *BMC Biotechnol.* 2013;13:8.
 30. Ploch T, Zhao X, Hüser J, et al. Multiscale dynamic modeling and simulation of a biorefinery. *Biotechnol Bioeng.* 2019;116:2561-2574.
 31. Jabarivelisdeh B, Waldherr S. Optimization of bioprocess productivity based on metabolic-genetic network models with bilevel dynamic programming. *Biotechnol Bioeng.* 2018;115:1829-1841.
 32. Tafur Rangel AE, Oviedo AG, Mojica FC, Gómez JM, González Barrios AF. Development of an integrating systems metabolic engineering and bioprocess modeling approach for rational strain improvement. *Biochem Eng J.* 2022;178:108268.
 33. Konstantinos D, Antonis K. Simultaneous synthesis of metabolic and process engineering for the production of Muconic acid. *Comput Aid Chem Eng.* 2022;51:889-894.
 34. Mitsos A, Chachuat B, Barton PI. Towards global bilevel dynamic optimization. *J Global Optim.* 2009;45:63-93.
 35. Stephanopoulos G, Aristidou AA, Nielsen J. *Metabolic Engineering: Principles and Methodologies.* Academic Press[nachdr.] ed; 1998.
 36. Gordeeva EL, Ravichev LV, Borodkin AG, Gordeeva YL. Mathematical Modeling of a biotechnological continuous fermentation process for lactic acid production: a review. *Theor Found Chem Eng.* 2021;55:1192-1203.
 37. Sinner P, Kager J, Daume S, Herwig C. Model-based analysis and optimisation of a continuous *Corynebacterium glutamicum* bioprocess utilizing lignocellulosic waste. *IFAC-PapersOnLine.* 2019;52:181-186.
 38. Bongartz D, Mitsos A. Deterministic global flowsheet optimization: between equation-oriented and sequential-modular methods. *AIChE J.* 2019;65:1022-1034.
 39. Najman J, Bongartz D, Mitsos A. Linearization of McCormick relaxations and hybridization with the auxiliary variable method. *J Global Optim.* 2021;80:731-756.
 40. Wick LM, Quadroni M, Egli T. Short- and long-term changes in proteome composition and kinetic properties in a culture of *Escherichia coli* during transition from glucose-excess to glucose-limited growth conditions in continuous culture and vice versa. *Environ Microbiol.* 2001;3:588-599.
 41. Monk JM, Lloyd CJ, Brunk E, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol.* 2017;35:904-908.
 42. Meadows AL, Karnik R, Lam H, Forestell S, Snedecor B. Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. *Metab Eng.* 2010;12:150-160.
 43. Djelassi H, Mitsos A. libALE – a library for algebraic-logical expression trees. 2019.
 44. Jungen D, Zingler A, Djelassi H, Mitsos A. libDIPS - a library for discretization-based semi-infinite programming solvers. 2023.
 45. Orth JD, Fleming RMT, Palsson BØ. Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. *EcoSal Plus.* 2010;4:1-48.
 46. Mitsos A, Lemonidis P, Barton PI. Global solution of bilevel programs with a nonconvex inner program. *J Global Optim.* 2008;42:475-513.
 47. Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual.* Gurobi Optimization, LLC; 2023.
 48. Bynum ML, Hackebeil GA, Hart WE, et al. *Pyomo–Optimization Modeling in Python.* Vol 67. 3rd ed. Springer Science & Business; 2021.
 49. Hart WE, Watson J-P, Woodruff DL. Pyomo: modeling and solving mathematical programs in python. *Math Program Comput.* 2011;3:219-260.
 50. Ebrahim A, Lerman JA, Palsson BØ, Hyduke DR. COBRApy: COConstraints-based reconstruction and analysis for python. *BMC Syst Biol.* 2013;7:74.
 51. Larhlmi A, David L, Selbig J, Bockmayr A. F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. *BMC Bioinformatics.* 2012;13:57.
 52. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* 2004;14:301-312.
 53. van Heerden CD, Nicol W. Continuous and batch cultures of *Escherichia coli* KJ134 for succinic acid fermentation: metabolic flux distributions and production characteristics. *Microb Cell Fact.* 2013;12:80.
 54. Monk JM, Koza A, Campodonico MA, et al. Multi-omics quantification of species variation of *Escherichia coli* links molecular features with strain phenotypes. *Cell Syst.* 2016;3:238-251.e12.
 55. Fischer E, Zamboni N, Sauer U. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. *Anal Biochem.* 2004;325:308-316.

56. Alter TB, Blank LM, Ebert BE. Proteome regulation patterns determine *Escherichia coli* wild-type and mutant phenotypes. *mSystems*. 2021;6:e00625-20.
57. Erdrich P, Steuer R, Klamt S. An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC Syst Biol*. 2015;9:48.
58. Goodall ECA, Robinson A, Johnston IG, et al. The essential genome of *Escherichia coli* K-12. *MBio*. 2018;9:e02096-17.
59. Chowdhury A, Zomorodi AR, Maranas CD. Bilevel optimization techniques in computational strain design. *Comput Chem Eng*. 2015;72:363-372.
60. Liu R, Bassalo MC, Zeitoun RI, Gill RT. Genome scale engineering techniques for metabolic engineering. *Metab Eng*. 2015;32:143-154.
61. Carbonell P, Currin A, Jervis AJ, et al. Bioinformatics for the synthetic biology of natural products: integrating across the design-build-test cycle. *Nat Prod Rep*. 2016;33:925-932.
62. Apaydin M, Xu L, Zeng B, Qian X. Robust mutant strain design by pessimistic optimization. *BMC Genomics*. 2017;18:677.
63. Kim J, Reed JL. OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst Biol*. 2010;4:53.
64. Kuepfer L, Sauer U, Blank LM. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res*. 2005;15:1421-1430.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ziegler AL, Manchanda A, Stumm M-D, Blank LM, Mitsos A. Simultaneous design of fermentation and microbe. *AICHE J*. 2024;70(9):e18501. doi:[10.1002/aic.18501](https://doi.org/10.1002/aic.18501)