



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Multiobjective optimization of gene circuits for metabolic engineering

Citation for published version:

Otero-Muras, I, Mannan, AA, Banga, JR & Oyarzún, DA 2019, 'Multiobjective optimization of gene circuits for metabolic engineering', *IFAC-PapersOnLine*, vol. 52, no. 26, pp. 13-16.
<https://doi.org/10.1016/j.ifacol.2019.12.229>

Digital Object Identifier (DOI):

[10.1016/j.ifacol.2019.12.229](https://doi.org/10.1016/j.ifacol.2019.12.229)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

IFAC-PapersOnLine

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Multiobjective optimization of gene circuits for metabolic engineering^{*}

Irene Otero-Muras^{*} Ahmad A. Mannan^{**} Julio R. Banga^{*}
Diego A. Oyarzún^{***,****,†}

^{*} *BioProcess Engineering Group. Spanish National Research Council (CSIC). Vigo, Spain (ireneotero@iim.csic.es)*

^{**} *Warwick Integrative Synthetic Biology Centre & School of Engineering, University of Warwick, Coventry, UK*

^{***} *School of Informatics, University of Edinburgh, UK*

^{****} *School of Biological Sciences, University of Edinburgh, UK*

[†] *SynthSys – Centre for Synthetic and Systems Biology, University of Edinburgh, UK (d.oyarzun@ed.ac.uk)*

Abstract: Metabolic engineering has enabled the production of a wealth of chemicals with microorganisms. Classic strategies for pathway engineering rely on the expression of heterologous enzymes in a host that convert native intermediates into target products. Although traditional implementations are based on open-loop control, recent advances in gene circuit engineering offer opportunities for building feedback systems that dynamically control pathway activity. Here we present a framework for the design of metabolic control circuits based on multiobjective optimization. We show that positive and negative feedback loops produce a range of optimal dynamics along a Pareto front. Such regulatory loops define connectivities between pathway intermediates and enzymatic genes that trade-off metabolic production against the burden to the host. Our results lay the groundwork for the automated design of gene circuitry in applications at the interface of synthetic biology and metabolic engineering.

Keywords: Metabolic engineering; Synthetic Biology; multiobjective optimization

1. INTRODUCTION

Metabolic engineering has been extremely successful in the production of valuable chemicals with genetically modified microorganisms (Lee et al., 2019). Microbial strains can be equipped with enzymatic genes that catalyze native metabolites into target chemical products. A common strategy in metabolic engineering is the “push-pull-block” approach (Liu et al., 2018). This involves up-regulation of native enzymes to push flux from precursors, expression of heterologous enzymes to pull flux towards the desired product, and knockdown of native pathways to block the diversion of flux away from production.

In traditional pathway engineering, foreign enzymes are expressed at constant levels from inducible or constitutive promoters. This corresponds to an open-loop strategy where pathway expression cannot adapt to perturbations or changes in bioreactor conditions (Venayak et al., 2015). Moreover, the diversion of metabolic flux from native processes towards chemical production alters the energetic budget of the host. This causes metabolic imbalances that often impair growth and limit production. Such imbalances arise from e.g. the accumulation of toxic intermediates or the depletion of key metabolites for survival. Pathway expression also draws resources from

the genetic machinery of the host, thus affecting the biosynthesis of native proteins needed for growth. As a result, a central challenge in metabolic engineering is to determine enzyme expression levels that maximise production with a reduced footprint on the host.

Last decade has witnessed the birth of *dynamic metabolic engineering*, a new technology that aims to use gene regulatory circuits in conjunction with engineered pathways (Zhang et al., 2012; Xu et al., 2014; Doong et al., 2018). The core principle is to design feedback control circuits that adapt pathway expression in response to the metabolic state of the host (Mannan et al., 2017; Stevens and Carothers, 2015). These control circuits can potentially resolve many of the challenges typically encountered in pathway engineering (Oyarzún and Stan, 2013). Yet to date there are no systematic strategies to navigate the design space and single out candidate circuits that can achieve given performance specifications.

Here we present a circuit design strategy based on multi-objective optimization. Our approach automatically finds circuit architectures and parameters that optimally trade-off the production flux against the metabolic burden on the host. This is a first step towards automated design of control circuits for engineered pathways and can speed up the iterations between circuit design, implementation and testing. As a proof-of-principle, we illustrate our approach with a toy pathway that contains some of the key elements encountered in real applications. We show

^{*} IOM and JRB acknowledge funding from the Spanish Ministry of Science, Innovation and Universities and the European Union FEDER under grant SYNBIOCNTROL (DPI2017-82896-C2-2-R).

that optimal solutions produce Pareto-optimal fronts with different feedback loops between pathway intermediates and metabolic genes.

2. DYNAMIC MODEL OF AN ENGINEERED PATHWAY

We consider a heterologous pathway that branches from a native pathway that is essential for growth (see Fig. 1). We denote by x_0 and x_1 the concentration of the metabolites of the host and synthetic pathway, respectively. The pool of native metabolite (x_0) is synthesized at a rate V_{in} and consumed by a native enzyme e_0 . The synthetic pathway contains two enzymes, denoted by e_1 and e_2 .

We model pathway kinetics with mass balance equations:

$$\begin{aligned}\dot{x}_0 &= V_{\text{in}} - V_0 - V_1 - \lambda x_0, \\ \dot{x}_1 &= V_1 - V_2 - \lambda x_1,\end{aligned}\quad (1)$$

where V_i are the fluxes of each reaction in Fig. 1B and λ is the growth rate of the culture. In particular, the reaction flux V_2 represents the flux through the engineered pathway. We model the expression of the pathway enzymes by:

$$\begin{aligned}\dot{e}_1 &= u_1 - \lambda e_1, \\ \dot{e}_2 &= u_2 - \lambda e_2,\end{aligned}\quad (2)$$

with u_i being the expression rate of the pathway genes. The reaction rates are assumed to follow irreversible Michaelis-Menten kinetics of the form:

$$V_0 = e_0 \cdot \frac{k_{\text{cat}0} \cdot x_0}{K_{M0} + x_0}, \quad (3)$$

$$V_1 = e_1 \cdot \frac{k_{\text{cat}1} \cdot x_0}{K_{M1} + x_0}, \quad (4)$$

$$V_2 = e_2 \cdot \frac{k_{\text{cat}2} \cdot x_1}{K_{M2} + x_1}, \quad (5)$$

with kinetic parameters ($k_{\text{cat}i}, K_{Mi}$) that are specific to each enzyme.

In dynamic pathway engineering, gene regulatory circuits are employed to adapt enzyme expression levels in response to a metabolic signal in the host. A common strategy is to use pathway intermediates to modulate the expression of enzymatic genes through metabolite-responsive transcription factors. The dose-response curve of such transcription factors can be modelled by a sigmoidal dose-response curve (Mannan et al., 2017) of the form

$$\begin{aligned}\sigma_i^+(x) &= a_i \frac{x^n}{\theta_i^n + x^n} \quad (\text{gene activation}), \\ \sigma_i^-(x) &= a_i \frac{\theta_i^n}{\theta_i^n + x^n} \quad (\text{gene repression}),\end{aligned}$$

where the parameter a_i represents the maximal expression, θ_i is a regulatory threshold, and n is an effective Hill coefficient.

For the system in Fig. 1B, we seek to find gene control circuits that maximize the production flux V_2 with a low footprint on enzyme expression. To formulate these specifications as a multiobjective optimization problem, we parameterize the enzyme expression rates u_i by the general function:

$$u_i(x_1) = b_i + |\eta_i| \cdot \left(\frac{1 + \eta_i}{2} \cdot \sigma_i^+(x_1) + \frac{1 - \eta_i}{2} \cdot \sigma_i^-(x_1) \right), \quad (6)$$

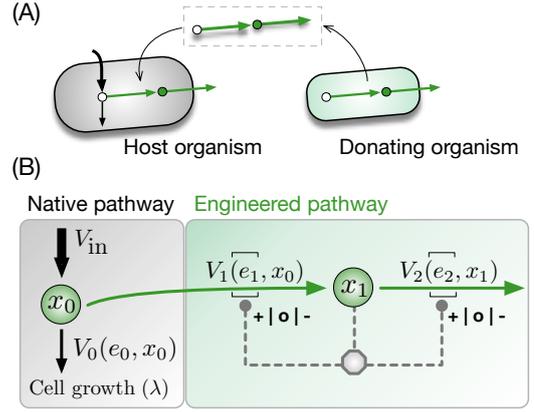


Fig. 1. Design of gene circuits for metabolic engineering. **(A)** Engineered pathways are built by expressing foreign enzymes into a target host. **(B)** Exemplar pathway described in Eqs. (1)–(2). We seek gene circuits that optimally trade-off the production flux V_2 against the burden caused by expression of enzymes e_1 and e_2 .

where $\eta_i \in \{-1, 0, 1\}$ is an integer variable and b_i describes the basal expression rate for each enzyme. The parameterization in (6) encodes all the regulatory circuits shown in Fig. 1B into a single function that is amenable for simultaneous optimization of regulatory parameters, comprised in the Hill functions σ_i^+ and σ_i^- , and regulatory architectures, described by the integers in η_i . This formulation accounts for two regulated genes and three modes of regulation (activation, repression, and no regulation) exerted by the metabolite x_1 . Each regulatory circuit can thus be described by a two dimensional integer vector $\eta = [\eta_1, \eta_2]$ and a six dimensional, real positive vector of regulatory parameters $w = [a_1, a_2, \theta_1, \theta_2, b_1, b_2]$.

3. CIRCUIT DESIGN AS AN OPTIMIZATION PROBLEM

Our goal is to design a gene circuit that maximizes the production of the metabolite of interest while keeping the burden of enzyme production at a minimum. To this aim we define two integral objective functions for our optimization problem. The first objective accounts for the flux through the production pathway:

$$J_1 = \int_0^T |V_2(t) - V_{\text{in}}| dt. \quad (7)$$

The second objective describes the genetic burden caused by pathway expression:

$$J_2 = \sum_i \int_0^T u_i dt. \quad (8)$$

In both objectives, T is a pre-specified final time. The objective function J_1 decreases when the production flux V_2 is high and close to the feed flux V_{in} . But since this requires high enzyme expression, a low J_1 can be only achieved at the expense of a large value for objective J_2 .

To find the optimal circuit architectures that trade-off production against enzyme expression, we formulate the following optimization problem:

$$\min_{\eta, w} [J_1(\eta, w), J_2(\eta, w)]$$

subject to:

- i) The dynamics of the pathway in eq. (1)–(2) and a given initial condition.
- ii) Upper and lower bounds for the decision variables:

$$w^L \leq w \leq w^U$$

$$-1 \leq \eta \leq 1$$

- iii) Inequality constraints to avoid the potential toxic effects caused by high concentration of product x_1 :

$$x_1(t) \leq x_{1\max}, \text{ for all } t. \quad (9)$$

To solve the resulting mixed-integer multiobjective optimization problem, we employ an ε -constraint strategy combined with mixed-integer hybrid global optimization algorithms. We employ the global heuristic eSS proposed in Egea et al. (2010) combined with MISQP by Exler and Schittkowski (2007) as a local solver. This optimization strategy has been shown to perform well for mixed-integer problems involving gene circuit design and allow for an efficient search over the space of circuit architectures and parameters (Otero-Muras and Banga, 2017).

4. EXAMPLE APPLICATION

As a proof of concept we solved the multiobjective problem for a representative set of kinetic parameters $k_{\text{cat } i} = 12\text{s}^{-1}$ and $K_{M i} = 10\mu\text{M}$ for all enzymes ($i = 0, 1, 2$), with a growth rate $\lambda = 3 \cdot 10^{-4}\text{s}^{-1}$, influx $V_{\text{in}} = 1\mu\text{Ms}^{-1}$, and concentration of the native enzyme $e_0 = 0.05\mu\text{M}$. We considered the parameter bounds given in Table 1 and included an additional upper bound for the product concentration $x_{1\max} = 180\mu\text{M}$.

Table 1. Parameter bounds for the optimization problem.

	a_1	a_2	θ_1	θ_2	b_1	b_2
w^L	10^{-7}	10^{-7}	10^{-2}	10^{-2}	10^{-5}	10^{-5}
w^U	10^{-4}	$9 \cdot 10^{-4}$	10	10	10^{-4}	10^{-4}
units	μMs^{-1}	μMs^{-1}	μM	μM	μMs^{-1}	μMs^{-1}

We implemented the ε -constraint strategy by setting objective J_2 as a constraint in a single objective problem to minimize J_1 . To this end we consider seven intervals corresponding to $J_2 \leq 1$, $1 < J_2 \leq 2$, $2 < J_2 \leq 3$, $3 < J_2 \leq 4$, $5 < J_2 \leq 6$ and $J_2 > 6$. The resulting Pareto front of solutions is shown in Fig. 2, together with the titer of product x_1 for each optimal circuit, defined as:

$$\text{titer} = \int_0^T V_2(t) dt. \quad (10)$$

The extremes points of the Pareto front, labelled as P1 and P7 in Fig. 2, correspond to solutions that account for the single objectives J_2 and J_1 , respectively. The average computational time needed by the solver for each constrained single objective problem is $\sim 650\text{s}$ and no solution exceeded 1500s using Matlab 2017b on a PC with Intel 2.8 GHz Xeon processor.

The resulting Pareto front contains three different circuit architectures (Fig. 2): solution P1 corresponds to the open loop case ($\eta_1 = \eta_2 = 0$). Solutions P2 to P5 require

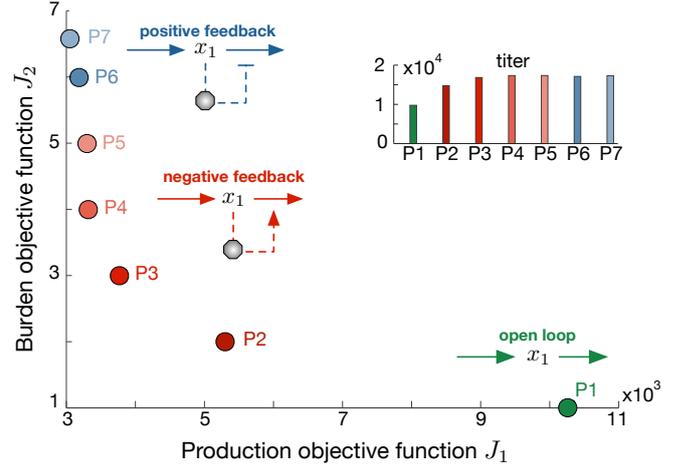


Fig. 2. Pareto front and optimal regulatory architectures. The inset shows the titer for each solution in μM , computed according to Eq. (10). Low values of J_1 correspond to better production performance. Optimal parameters are shown in Table 2.

activation of the consumption enzyme e_2 ($\eta_1 = 0$ and $\eta_2 = 1$), which corresponds to a negative feedback loop of x_1 onto itself. Solutions P6 and P7 require repression of the consumption enzyme e_2 ($\eta_1 = 0$ and $\eta_2 = -1$), and thus correspond to a positive feedback loop of x_1 onto itself. In all cases, optimal solutions do not need control on the expression of the first pathway enzyme (e_1). In Figs. 3-4 we show the optimal trajectories of metabolic product (x_1), pathway enzymes (e_1 and e_2) and metabolic fluxes for all optimal solutions.

Table 2. Optimal parameter values for each solution along the Pareto front in Fig. 2. The Hill coefficient was fixed to $n = 2$. Units are given in Table 1.

	a_1	a_2	θ_1	θ_2	b_1	b_2
P1	$9.5 \cdot 10^{-5}$	$3.2 \cdot 10^{-4}$	$6.3 \cdot 10^{-1}$	1.9	$1.9 \cdot 10^{-5}$	$1.7 \cdot 10^{-5}$
P2	$6.9 \cdot 10^{-7}$	$5.2 \cdot 10^{-5}$	8.4	9.9	$5 \cdot 10^{-5}$	$1 \cdot 10^{-5}$
P3	$1.9 \cdot 10^{-5}$	$9.2 \cdot 10^{-5}$	9.9	9.9	$8.8 \cdot 10^{-5}$	$1.9 \cdot 10^{-5}$
P4	$9.8 \cdot 10^{-5}$	$5.3 \cdot 10^{-4}$	3.0	9.9	10^{-4}	$6.6 \cdot 10^{-5}$
P5	$2.8 \cdot 10^{-5}$	$8.9 \cdot 10^{-4}$	6.9	5.7	$9.9 \cdot 10^{-5}$	$9.1 \cdot 10^{-5}$
P6	$6.1 \cdot 10^{-5}$	$2.6 \cdot 10^{-4}$	7.3	9.9	$9.4 \cdot 10^{-5}$	$1 \cdot 10^{-5}$
P7	$8.3 \cdot 10^{-6}$	$2.7 \cdot 10^{-4}$	5.1	10	$9.9 \cdot 10^{-5}$	$1 \cdot 10^{-4}$

Our results suggest that both positive and negative feedback can lead to an optimal trade-off between production and burden to the host. Solutions from P3 to P7 perform significantly better in terms of titer than solutions P1 and P2. Note that solution P1 accounts for one objective only, resulting in a slow, 1st-order like response. In contrast, solutions that account for the second objective (from P2 to P5) display increasingly faster and nonlinear responses as we move along the Pareto front.

5. DISCUSSION

Here we have addressed the regulation of heterologous pathways using a multiobjective dynamic optimization approach. We show that encoding the design problem in a mixed-integer framework enables an efficient search through the topology and parameter spaces simultaneously. Our multiobjective perspective allows us to find

REFERENCES

- Doong, S.J., Gupta, A., and Prather, K.L.J. (2018). Layered dynamic regulation for improving metabolic pathway productivity in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 115(12), 2964–2969.
- Egea, J.A., Marti, R., and Banga, J.R. (2010). An evolutionary method for complex-process optimization. *Comput Oper Res*, 37, 315–324.
- Exler, O. and Schittkowski, K. (2007). A trust region sqp algorithm for mixed-integer nonlinear programming. *Optim Lett*, 1(3), 269–280.
- Lee, S.Y., Kim, H.U., Chae, T.U., Cho, J.S., Kim, J.W., Shin, J.H., Kim, D.I., Ko, Y.S., Jang, W.D., and Jang, Y.S. (2019). A comprehensive metabolic map for production of bio-based chemicals. *Nature Catalysis*, 2(1), 18–33.
- Liu, D., Mannan, A., Han, Y., Oyarzún, D., and Zhang, F. (2018). Dynamic metabolic control: towards precision engineering of metabolism. *Journal of Industrial Microbiology and Biotechnology*.
- Mannan, A.A., Liu, D., Zhang, F., and Oyarzún, D.A. (2017). Fundamental design principles for transcription-factor-based metabolite biosensors. *ACS synthetic biology*, 6(10), 1851–1859.
- Otero-Muras, I. and Banga, J. (2017). Automated design framework for synthetic biology exploiting pareto optimality. *ACS Synthetic Biology*, 6(7), 1180–1193.
- Oyarzún, D.A. and Stan, G.B. (2013). Synthetic gene circuits for metabolic control: design trade-offs and constraints. *Journal of the Royal Society Interface*, 10(78), 20120671.
- Stevens, J.T. and Carothers, J.M. (2015). Designing RNA-based genetic control systems for efficient production from engineered metabolic pathways. *ACS Synthetic Biology*, 4(2), 107–115.
- Venayak, N., Anesiadis, N., Cluett, W.R., and Mahadevan, R. (2015). Engineering metabolism through dynamic control. *Curr. Opin. Biotechnol.*, 34, 142–152.
- Xu, P., Li, L., Zhang, F., Stephanopoulos, G., and Kofas, M. (2014). Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proceedings of the National Academy of Sciences*, 111(31), 11299–11304.
- Zhang, F., Carothers, J.M., and Keasling, J.D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology*, 30(4), 354–359.

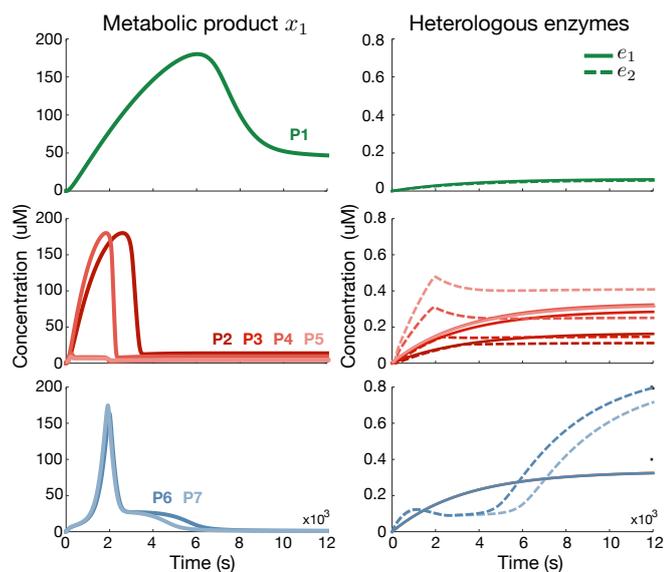


Fig. 3. Optimal time courses of the pathway product and heterologous enzymes along the Pareto front.

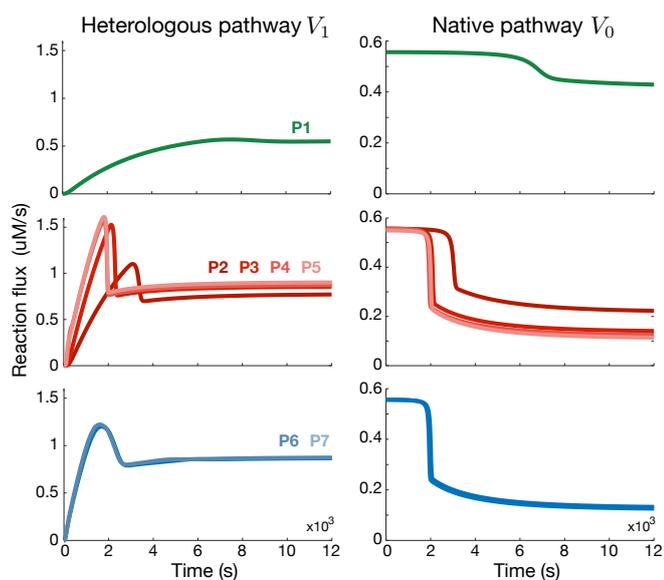


Fig. 4. Optimal metabolic fluxes in the heterologous and native pathway for each solution along the Pareto front.

solutions that optimally trade-off performance in terms of production and burden to the host, as quantified by the cost of protein production. The resulting Pareto front suggests that open loop control imposes the lowest burden at the expense of poor production. A positive feedback architecture lies at the other extreme of the Pareto front, showing high production but at the cost of high protein production. A negative feedback architecture lies in between these two extremes, with a good performance and intermediate values of burden. Our results show promise for future applications of multiobjective optimization in the design of complex circuitry that combine gene expression with metabolic activity.