Design of modular cells by goal attainment optimization

Sergio Garcia and Cong T. Trinh

Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, TN, United States and Center for Bioenergy Innovation, Oak Ridge National Laboratory Oak Ridge, TN, United States

ctrinh@utk.edu | https://github.com/trinhlab

Abstract

Motivation. Microbial biocatalysis can in principle be engineered to obtain all types of chemicals, including material precursors, fuels, and drugs, from renewable feedstocks, e.g., plant biomass, organic waste, CO_2, etc. However, the current R&D cycles needed to develop economically feasible microbial biocatalysis processes are too costly and slow, due primarily to the repetition of tasks and lack of standardization.

Approach. To accelerate microbial biocatalyst R&D, we propose to apply modular design principles widely implemented in conventional engineering[1] to metabolic engineering.[2] We formulate these design principles into an approach named Modular Cell (ModCell),[3] and mathematically express the design problem as a multi-objective optimization problem.[4]

Results. The multi-objective modular strain design problem was previously solved with multi-objective evolutionary algorithms (MOEA).[4,5] In this study we propose a MILP-compatible formulation that leads to optimal solutions, enumerates the full space of alternative solutions, and allows to formulate the design goals in conceptual terms closer to practical goals. We used ModCell2-MILP to design a modular cell universally compatible with a wide variety of modules, and identified the metabolic features that make such design possible.

Methods

ModCell design concept

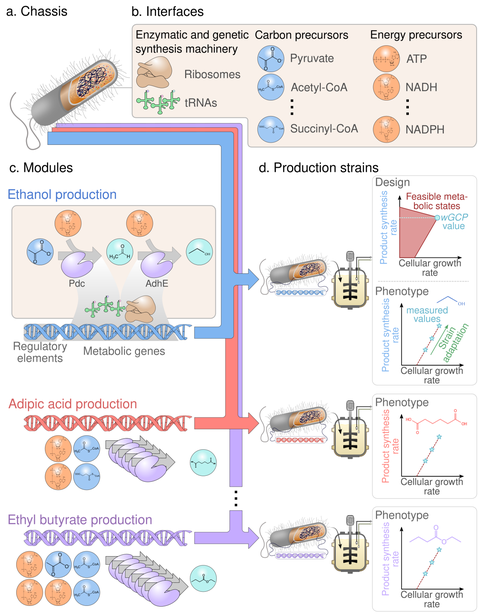

Fig. 1: Principles of modular cell design. (a) Modular (chassis) cell. (b) Interfaces. (c) Production modules. (d) Production strains. A modular cell is designed to provide the necessary precursors for biosynthesis pathway modules that are independently assembled with the modular cell to generate production strains exhibiting desirable phenotypes. The *wGCP* phenotype, one of the possible design objectives, enforces the coupling between desirable product synthesis rate and the maximum cellular growth rate.

ModCell multi-objective optimization formulation

$$\max_{y_j, z_{jk}} \quad (f_1, f_2, \ldots, f_{|\mathcal{K}|})^T \quad \text{s.t.} \tag{1}$$

$$f_k \in \arg\max \left\{ \frac{1}{f_k^{max}} \sum_{j \in \mathcal{J}_k} c_{jk} v_{jk} \quad \text{s.t.} \right. \tag{2}$$

$$\sum_{j \in \mathcal{J}_k} S_{ijk} v_{jk} = 0 \qquad \text{for all } i \in \mathcal{I}_k \tag{3}$$

$$l_{jk} \le v_{jk} \le u_{jk} \qquad \text{for all } j \in \mathcal{J}_k \tag{4}$$

$$l_{jk} d_{jk} \le v_{jk} \le u_{jk} d_{jk} \qquad \text{for all } j \in \mathcal{C} \tag{5}$$

$$\left. \text{where } d_{jk} = y_j \vee z_{jk} \right\} \qquad \text{for all } k \in \mathcal{K}$$

$$z_{jk} \le (1 - y_j) \qquad \text{for all } j \in \mathcal{C},\ k \in \mathcal{K} \tag{6}$$

$$\sum_{j \in \mathcal{C}} z_{jk} \le \beta_k \qquad \text{for all } k \in \mathcal{K} \tag{7}$$

$$\sum_{j \in \mathcal{C}} (1 - y_j) \le \alpha \tag{8}$$

The design objectives for each production module f_k (Fig. 2) are to be simultaneously optimized (1) subject to: i) modeling constraints (2-5) that predict metabolic fluxes v_{jk} (mmol/gCDW/hr) using a steady-state metabolic model, where S_{ijk} is the stoichiometric coefficient of metabolite i in reaction j of production network k; and ii) design constraints (6-8), where the binary design variables y_j and z_{jk} represent reaction deletions, limited by α and module reaction insertions, limited by β, respectively.

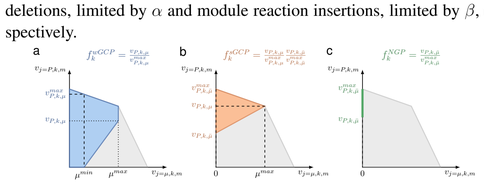

Fig. 2: Phenotypic spaces for different strain design objectives including: (a) weak growth coupling (*wGCP*), (b) strong growth coupling (*sGCP*), and (c) no-growth production (*NPG*). $v_{P,k,\hat{\mu}}$ is the minimum product formation rate at the maximum growth rate for production network k, and $v^{max}_{P,k,\hat{\mu}}$ is the maximum product secretion rate attainable. $v_{P,k,\hat{\mu}}$ and $v^{max}_{P,k,\hat{\mu}}$ are the minimum and maximum product formation rates for production network k during the stationary phase, respectively.

Single-objective reformulations

After linearizing the constraints of the multi-objective optimization problem (1-8) (not shown here), the objective function is noted as f'_k and the set of linear constraints is noted as Ω. Now, the objective function can be linearized in two ways:

Blended formulation

$$\max \quad \sum_{k \in \mathcal{K}} a_k f'_k \quad \text{s.t. } f' \in \Omega \tag{9}$$

Here a_k is a scalar weighting factor associated with the design objective of product k. Different Pareto optimal solutions can be obtained by varying these weights. In practice, the product priority, a_k, can be determined by criteria such as product market value or "pathway readiness level" (i.e., certain pathways are easier to engineer than others).

Goal attainment formulation

$$\min \quad \sum_{k \in \mathcal{K}} (a_k^+ \delta_k^+ + a_k^- \delta_k^-) \tag{10}$$

s.t.

$$f'_k + \delta_k^+ - \delta_k^- = g_k \quad \forall k \in \mathcal{K} \tag{11}$$

$$\delta_k^+, \delta_k^- \ge 0 \quad \forall k \in \mathcal{K} \tag{12}$$

$$f' \in \Omega \tag{13}$$

In the goal attainment problem (10-13), a target value g_k is defined for each objective k. The problem seeks to minimize the variables δ_k^+ and δ_k^- that represent the deficiency and excess of the objective f'_k from the target value g_k, respectively. Weighting parameters a_k^+ and a_k^- correspond to different types of discrepancy to be minimized. In practice, the goal attainment formulation corresponds to the identification of the modular cell *compatible* (i.e., a module k is said to be compatible if $f'_k \ge g_k$) with the largest number of modules.

Results

Design of universal modular cell

- We validated correctness of the MILP formulation and optimized solution times using techniques such as bound tightening and benders decomposition (not shown here).
- We applied the goal attainment formulation to design a universal modular cell that displays $g_k \ge 0.5$ (Fig. 3) for all 20 biochemically diverse products (Table 1).

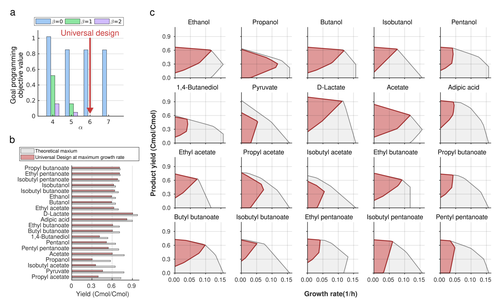

Fig. 3: Identification of a universal modular cell compatible with all production modules using the *wGCP* design objective. (a) Goal programming solutions using increasing α and β values with a target goal $g_k = 0.5$. The parameters $\alpha = 6$ and $\beta = 1$ are sufficient to identify a universal ModCell design meeting the required goal for all production networks. (b) Comparison between the yield performances of the designed modular production strains and maximum theoretical values. (c) The feasible flux spaces for the wild-type (gray) and designed modular production strains (crimson). Based on the *wGCP* design phenotype, to increase growth rate, each mutant must increase product synthesis rate. The genetic manipulations of this universal modular cell design are indicated in the metabolic map of Figure 4c.

- Despite the broad diversity of the production modules in terms of carbon and electron precursor requirements, the universal modular cell is *compatible* with all of them (Table 1).

Overall reaction	DoR	Production network yields at maximum growth rate				
		product	*ac*	*co2*	*for*	*succ*
pyr + nadh → **ethanol** | accoa + 2 nadh → **ethanol** (native)	7.0	0.58	0.01	0.27	0.04	-
oaa + glu + 2 atp + 2 nadph + nadh → akg + **propanol**	6.7	0.31	0.36	0.07	0.18	-
2 accoa + 4 nadh → **butanol**	6.5	0.59	0.01	0.28	0.04	-
2 pyr + nadph + nadh → **isobutanol**	6.5	0.62	-	0.31	-	-
oaa + glu + accoa + 3 nadh + 2 atp + 2 nadph → akg + **pentanol**	6.4	0.50	0.21	0.24	0.03	-
succ + akg + atp + 4 nadh + accoa → ac + **1,4-butanediol**	5.5	0.46	0.33	0.17	-	-
→ **pyruvate**	3.0	0.46	-	-0.16	-	0.66
pyr + nadh → **D-lactate**	3.7	0.91	-	-	-	-
accoa → atp + **acetate**	3.5	0.60	0.60	-0.30	0.61	-
accoa + succoa + 2 nadh → atp + **adipic acid**	4.0	0.82	0.05	0.04	0.06	-
accoa + pyr + nadh → **ethyl acetate**	5.0	0.63	-	-	0.32	-
accoa + oaa + glu + 2 atp + 2 nadph + nadh → akg + **propyl acetate**	5.2	0.41	0.30	-	0.24	-
accoa + 2 pyr + nadph + nadh → **isobutyl acetate**	5.3	0.36	-	0.02	0.06	0.52
2 accoa + 3 nadh + pyr → **ethyl butanoate**	5.3	0.61	-	0.09	0.23	-
2 accoa + 3 nadh + oaa + glu + 2 atp + 2 nadph → akg + **propyl butanoate**	5.4	0.68	0.03	0.23	0.04	-
4 accoa + 6 nadh → **butyl butanoate**	5.5	0.61	-	0.14	0.18	-
2 accoa + 3 nadh + 2 pyr + nadph → **isobutyl butanoate**	5.5	0.64	-	0.16	0.16	-
oaa + glu + accoa + 2 nadh + 2 atp + 2 nadph + pyr → akg + **ethyl pentanoate**	5.4	0.68	0.03	0.23	0.04	-
oaa + glu + accoa + 2 nadh + 2 atp + 3 nadph + 2 pyr → akg + **isobutyl pentanoate**	5.6	0.67	0.01	0.25	0.03	-
2 oaa + 2 glu + 2 accoa + 4 nadh + 4 atp + 4 nadph → 2 akg + **pentyl pentanoate**	5.6	0.53	0.22	0.20	0.02	-

Table 1: Overall production module stoichiometries, degree of reduction (DoR) of the final product (mol e^- / mol C), and metabolite secretion profiles (mol C / mol C) from simulated flux distributions of the universal modular cell design.

Flexible metabolic flux capacity of *E. coli* central metabolism enables the design of a universal modular cell

- We simulated metabolic fluxes in each production strain of the universal design using pFBA and flux sampling, leading to the identification of highly-variable reactions that activate to interface with specific production modules (Fig. 4).
- The flexible flux capacities and modularity of core *E. coli* metabolism, consistent with fluxomics data (Fig. 4d), enable this universal modular cell.

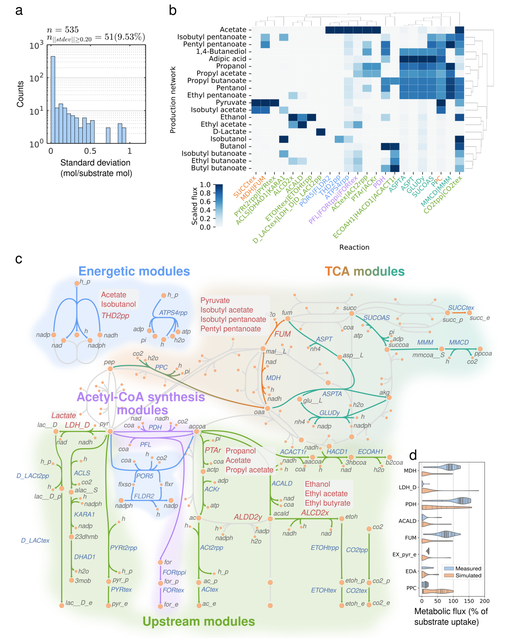

Fig. 4: (a) Standard deviation of each simulated reaction flux across production networks. (b) Scaled fluxes of the 51 reactions with standard deviation magnitude above 0.2, excluding proton, water transport, and exchange reactions. Scaled fluxes correspond to the simulated flux distribution value of each reaction divided by the maximum value of that reaction across all production networks, thus a scaled flux of 0 indicates a given reaction does not carry any flux, and a scaled flux of 1 indicates that a reaction carries the highest flux observed for that particular reaction across production networks. Reactions with the same flux are separated by a vertical line (|). (c) Reactions colored in red are deleted in the chassis, while the production networks where such reactions are used as module reaction are included in an adjacent gray box. (d) Violin plot comparing the simulated fluxes (reference flux distributions) that appear in each production network against metabolic fluxes measured in a variety of conditions and mutants.[6] The solid lines within the "violins" correspond to samples.

References

1. Garcia, S. & Trinh, C. T. Modular design: Implementing proven engineering principles in biotechnology. *Biotechnology Advances* (2019).
2. Trinh, C. T. & Mendoza, B. Modular cell design for rapid, efficient strain engineering toward industrialization of biology. *Current Opinion in Chemical Engineering* **14,** 18–25 (2016).
3. Trinh, C. T. *et al.* Rational design of efficient modular cells. *Metabolic engineering* **32,** 220–231 (2015).
4. Garcia, S. & Trinh, C. T. Multiobjective strain design: A framework for modular cell engineering. *Metabolic Engineering* **51** (2019).
5. Garcia, S. & Trinh, C. T. Comparison of Multi-Objective Evolutionary Algorithms to Solve the Modular Cell Design Problem for Novel Biocatalysis. *Processes* **7** (2019).
6. Khodayari, A. & Maranas, C. D. A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nature Communications* **7** (2016).

Acknowledgements: This research was funded by the NSF CAREER Award (NSF#1553250) and the Center of Bioenergy Innovation (CBI), U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The funders had no role in the study design, data collection and analysis, or decision to publish.

Poster presented at MPA 2019. Riga, Latvia.