Multiobjective strain design: A framework for modular cell engineering

Sergio Garcia[1,2], and Cong T. Trinh[1,2]

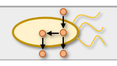
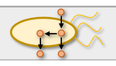
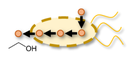
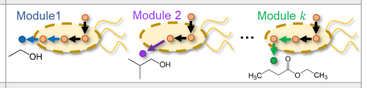
[1]Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, TN, USA ,
[2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge.

ABSTRACT

Diversity of cellular metabolism can be harnessed to produce a large space of molecules. However, development of optimal strains with high product titers, rates, and yields required for industrial production is laborious and expensive. To accelerate the strain engineering process, we have recently introduced a modular cell design concept that enables rapid generation of optimal production strains by systematically assembling a modular cell with an exchangeable production module(s) to produce target molecules efficiently. In this study, we formulated the modular cell design concept as a general multiobjective optimization problem with flexible design objectives derived from mass balance. We developed algorithms and an associated software package, named ModCell2, to implement the design. We demonstrated that ModCell2 can systematically identify genetic modifications to design modular cells that can couple with a variety of production modules and exhibit a minimal tradeoff among modularity, performance, and robustness. Analysis of the modular cell designs revealed both intuitive and complex metabolic architectures enabling modular production of these molecules. We envision ModCell2 provides a powerful tool to guide modular cell engineering and sheds light on modular design principles of biological systems.

1. ModCell2 method

Concept

Features	Conventional strain engineering	Modular cell engineering
Parent strain		
Modular cell	Absent	 Optimized common production phenotypes
Exchangeable modules	1	k
Optimal production strains		
Design-build-test cycle	Repeated for every new product	One time for many products

Comparison between the conventional single-product strain design and modular cell engineering.

- In the conventional approach, each target product requires to go through the iterative optimization cycle.
- The modular cell engineering approach exploits common phenotypes associated with high product titers, rates, and yields; and hence, the strain optimization cycle only needs to be performed once for multiple products, which helps reduce the cost and time of strain development.

Mathematical formulation

$$\underset{over\ y_j, z_{jk}}{\text{maximize}}\ \boldsymbol{F} = \left(f_1, f_2, \ldots, f_{|K|}\right)^T \quad [1]$$

subject to

$$f_k \in \arg max\left\{\sum_{j\in J_k} c_{jk} v_{jk}\right\}$$

subject to

$$\sum_{j\in J_k} S_{ij} v_{jk} = 0, for\ all\ i \in I_k \quad [2]$$

$$l_{jk} \le v_{jk} \le u_{jk}, for\ all\ j \in J_k \quad [3]$$

$$l_{jk} d_{jk} \le v_{jk} \le u_{jk} d_{jk}, for\ all\ j \in C \text{ and } C \subseteq J_k \quad [4]$$

$$\text{where } d_{jk} = y_j\ V\ z_{jk}\}, for\ all\ k \epsilon K$$

$$z_{jk} \le (1 - y_j), for\ all\ j \in C, k \in K \quad [5]$$

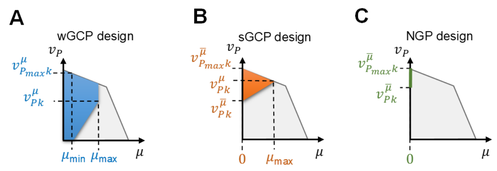
$$\sum_{j\epsilon C}(1 - y_j) \le \alpha \quad [6]$$

$$\sum_{j\epsilon C} z_{jk} \le \beta_k, for\ all\ k \epsilon K \quad [7]$$

← **The framework of modular cell engineering is formulated as a multiobjective optimization problem, named ModCell2.**

- For modular cell engineering, we seek to design a modular cell compatible with as many production modules as possible to achieve only desirable production phenotypes while requiring minimal genetic modifications.
- Since all production modules must leverage cellular resources of the modular cell (e.g. precursor metabolites, cofactors, and energy), they form competing objectives.

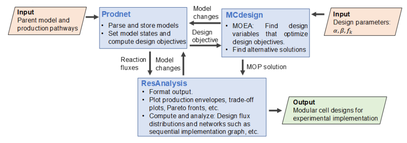
Target phenotypes



Graphical representation of phenotypic spaces for different strain design objectives

- **(A)** weak growth coupling (*wGCP*), **(B)** strong growth coupling (*sGCP*), and **(C)** no-growth production (*NGP*).
- Any point within the delimited polygon represents a metabolic flux distribution attainable by the organism. v_{Pk}^{μ} is the minimum product formation rate at the maximum growth rate for production network *k*, and $v_{P_{max}k}^{\mu}$ is the maximum product secretion rate attainable. $v_{Pk}^{\bar{\mu}}$ and $v_{P_{max}k}^{\bar{\mu}}$ are the minimum and maximum product formation rates for production network *k* during the stationary phase, respectively.
- The objectives take the following scalar form: $f_k^{wGCP} = \frac{v_{Pk}^{\mu}}{v_{P_{max}k}^{\mu}}$, $f_k^{sGCP} = \frac{v_{Pk}^{\mu}}{v_{P_{max}k}^{\mu}} \cdot \frac{v_{Pk}^{\bar{\mu}}}{v_{P_{max}k}^{\bar{\mu}}}$, and $f_k^{NGP} = \frac{v_{Pk}^{\bar{\mu}}}{v_{P_{max}k}^{\bar{\mu}}}$.

Implementation



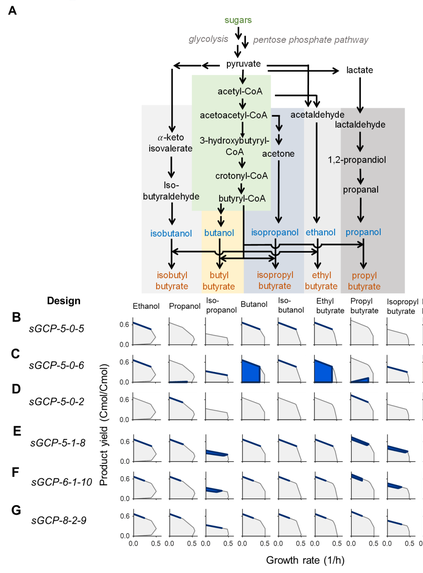
Software architecture of ModCell2.

- The Prodnet class preprocesses production network models and computes design objectives.
- The MCdesign class serves as an interface between the Multiobjective evolutionary algorithm (MOEA) optimization method and metabolic models.
- The ResAnalysis class loads the Pareto set computed by MCdesign and performs analyses to identify the most promising designs.

ModCell2 is user-friendly and available on GitHub!
https://github.com/trinhlab/modcell2

2. Comparing ModCell2 designs with first-generation MODCELL and single-product designs

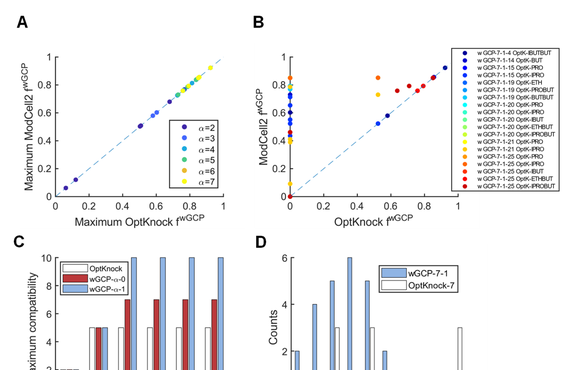
ModCell2 can generate more and better designs than the first-generation modular cell design method MODCELL



The 2-D metabolic phenotypic spaces of different *sGCP* designs using the core *E. coli* metabolic model.

- (**A**) Metabolic map, **(B)** *sGCP-5-0-5* design, **(C)** *sGCP-5-0-6* design, **(D)** *sGCP-5-0-2* design, **(E)** *sGCP-5-1-8* design, **(F)** *sGCP-6-1-10* design, and **(G)** *sGCP-8-2-9* design (See DEFINITIONS for notation details).
- For each panel, the gray and blue areas correspond to the phenotypic spaces of the wildtype and the designed mutant strain, respectively.
- In Figures 4B-4D, the phenotypic spaces are the same between ModCell2 and MODCELL; however, ModCell2 finds those same designs with fewer deletions.
- Figures 4E-4G are results only found by ModCell2 that perform better than those presented in Figure 4B-4D.
- MODCELL identifies common patterns in single-product designs to build a chassis, but this approach omits relevant information that ModCell2 is able to account for by simultaneously considering all production modules.

ModCell2 can identify designs with more compatibility than the conventional single-product design method OptKnock



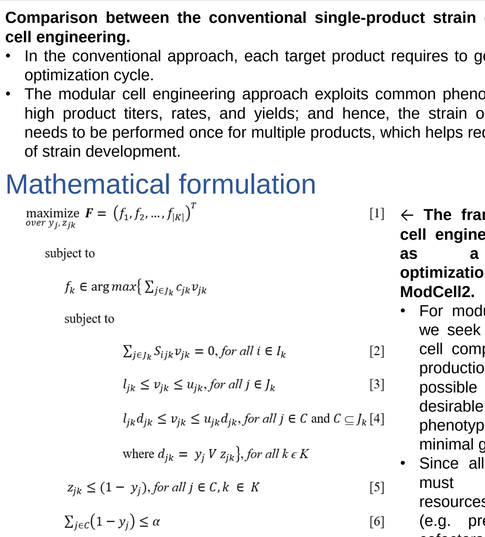
Comparison of strain design by OptKnock and Modcell2.

A. Correlation between the maximum objective values for each product generated by OptKnock and the equivalent values attained by ModCell2. For a given maximum number of reaction deletions, each circle corresponds to a product.
B. A comparison between the Optknock objective vectors with at most 7 reaction deletions and the representative ModCell2 objective vector, *wGCP-7-1* which dominates them. Each color circle represents a pair of dominating *wGCP* design and dominated OptKnock solution (Supplementary File S3).
C. Maximum compatibility of OptKnock designs and *wGCP* designs with and without module-specific reactions.
D. Compatibility distribution of Optknock and *wGCP-7-1*.

- The *E. coli* core model with 10 alcohols and esters was also used for this analysis.
- These results demonstrate that while single-product strategies have certain re-usability, as commonly done in practice, they can be far from the Pareto optimality attained by modular cell strains.

3. Exploring emergent features of modular cell design using an *E. coli* genome-scale network

20 experimentally verified product synthesis pathways with diverse biochemistry



Metabolic map and properties of the 20 production modules.

A. Distribution of precursor metabolites.
B. Distribution of degrees of reduction of target products.
C. Correlation between the number of product carbons and the number of reactions in production modules.

- The 6 alcohols, 4 organic acids, and 10 esters are biochemically diverse in terms of precursors and degree of reduction.

ModCell2 can identify highly compatible modular cells with negligible tradeoffs

ModCell2 growth-coupled chassis designs with at most 4 reaction deletions for 20 different products.

A. Pareto front of *wGCP-4-0-d*. The columns correspond to different designs labeled by their design index, *d*, where the rows correspond to different products.
B. Frequency of the top deletion reactions.
C. Product compatibility distribution across designs.
D. Compatibility distribution of Pareto optimal designs.
E. Tradeoff between modularity and performance. The bars correspond to the maximum objective values attainable for each product whereas the blue line represent the objective values of the *wGCP-4-0-48alternative* design.

- These results highlight the combinatorial and core metabolic phenotypes shared among products (e.g. deletion of lactatate dehydrogenease and phosphotransacetylase), and that it is feasible to identify highly compatible modular cell designs without a significant tradeoff between performance and modularity.

4. Metabolic switch among different design objectives

Phenotypic spaces of highly compatible wGCP design and its derived sGCP and NGP designs

← **Feasible production phenotypes of ModCell2 designs.**

- (**A**) *wGCP-4-0-48-alternative* (blue), (**B**) *wGCP→NGP* design (red), and (**C**) the *wGCP→sGCP* design (green).
- In all of the panels, the feasible production phenotypes of the parent strain are shown in gray; specifically, the gray areas are referred to the wildtype in (**A**) and wGCP design in (**B**) and (**C**).
- The deletions of the designs in (**B**) and (**C**) are supersets of the design presented in (**A**).
- These results demonstrate the desirable growth-coupled and non-growth production phenotypes of all production modules and also the possibility of dynamic phenotype switch throughout fermentation to further optimize titer, yields, and rates within the context of modular cells.

5. Experimental demonstration

In silico predicted phenotype is shown *in vivo* by regulating production pathway flux

Adjustment of coupling degree between the modular cell TCS095 and ethanol production modules.

- TCS095 is a 9 knockout strain predicted by MODCELL which is also a superset of the highly compatible design wGCP-4-0-48-alternative predicted by ModCell2.
- (**A**) Coupled cells EcDL110, EcDL111, and EcDL112 assembled from the modular cell TCS095 and the one-operon, two-gene ethanol production modules, pCT24, pAY1, and pAY3, using promoters of different strengths. (**B**) Cell growth. (**C**) Glucose consumption. (**D**) Ethanol production. (**E**) Correlation of growth and ethanol production rates.
- The linear correlation between growth and product rate directly verify the growth-coupled-to-product-synthesis phenotype as predicted *in silico*.

CONCLUSIONS

1. Formulate modular cell design as a multiobjective optimization problem.
2. Develop algorithm and computational platform, ModCell2, to streamline modular cell design.
3. Demonstrate ModCell2 to design modular cells using genome-scale *E. coli* metabolic model.
4. Show modular cells exhibit minimal tradeoff among modularity, performance, and robustness.
5. Reveal intuitive and complex metabolic architectures enabling modular design.
6. Software is available at *https://github.com/trinhlab/modcell2*

DEFINITIONS

- **Core model**: *E. coli* core model (*Trinh et al. Metab. Eng. 2015).*
- **Genome-scale model**: *E. coli* iML1515 (*Monk et al. Nat. Biotech. 2017*).
- **Design objective**: One of the three phenotypes presented in 1.ModCell Method - Target phenotypes.
- ***α***: Maximum number of reaction deletions in the chassis.
- ***β*** : Maximum number of endogenous **module-specific reactions**. These are endogenous reactions deleted in the chassis but added back to specific modules to enhance compatibility.
- **Design notation, e.g. "sGCP-5-10"**: A Pareto front is determined by the design parameters, thus the notation: <design objective>-<*α*>-<*β*>. When referring to a specific design its index is appended at the end: <design objective>-<*α*>-<*β*>-<design index>.
- **Compatibility**: Number of products (i.e. modules) that work with a chassis above an specified objective (0.6 for wGCP

REFERENCES

- Garcia, S., and Trinh, C.T., "Multiobjective Strain Design: A Framework for Modular Cell Engineering", Metab. Engineering. In press.
- Wilbanks, B., Layton, D. S., Garcia, S., & Trinh, C.T. (2017). A Prototype for Modular Cell Engineering. ACS synthetic biology, 7(1), 187-199.

AKNOWLEDGEMENTS

This research was financially supported in part by the NSF CAREER Award (NSF#1553250) and the DOE subcontract Grant (DE-AC05-000R22725) by the Center of Bioenergy Innovation (CBI), the U.S. Department of Energy Bioenergy Research Center funded by the Office of Biological and Environmental Research in the DOE Office of Science