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, University of Tennessee, Knoxville, TN

[2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN

[*]Corresponding author: Tel: 865-974-2181. Email: ctrinh@utk.edu.

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 action. 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.

Keywords

Modular cell engineering; modular cell; production modules; modular design; modularity; multiobjective optimization; multiobjective evolutionary algorithms.

INTRODUCTION

Engineering microbial cells to produce bulk and specialty chemicals from renewable and sustainable feedstocks is becoming a feasible alternative to traditional chemical methods that rely on petroleum feedstocks (1). However, only a handful of chemicals, out of the many possible molecules offered by nature, are industrially produced by microbial conversion, mainly because the current strain engineering process is laborious and expensive for profitable biochemical production (2). Thus, innovative technologies to enable rapid and economical strain engineering are needed to harness a large space of industrially-relevant molecules (3).

The modular organization of biological systems has been a source of inspiration for synthetic biology and metabolic engineering (4, 5). Modular pathway engineering breaks down target pathways into tractable pathway modules that can be finely tuned for optimal production of desirable chemicals (6, 7). Harnessing combinatorial pathways (e.g., fatty acid biosynthesis, reverse beta oxidation, polyketide or isoprenoid biosynthesis) is one excellent example of modular metabolic engineering. These pathways contain metabolic similarity (or combinatorial characteristics) such as a group of common specific enzymes capable of catalyzing linear reaction steps (8) and/or elongation cycles (9-11) and hence are capable of producing a large library of unique molecules (12). Since these molecules are derived from a common precursor metabolite(s), the optimal production strains often share common genotypes and phenotypes, and hence, the costly strain optimization process is only performed once for these molecules. Remarkably, this advantageous strain optimization strategy can be applied even for production of molecules derived from different precursors, using the concept of modular cell (ModCell) design (2, 13, 14).

With the arrival of steady-state, constraint-based stoichiometric models of cellular metabolism, various computational algorithms have been developed to guide strain engineering (15-17). These methods have featured the design of strains capable of growth-coupled product synthesis (*GCP*), enabling adaptive laboratory evolution of these designed strains to enhance product titers, rates, and yields (14, 18-20). Two approaches on growth-coupled production have been formulated ─ one based on the coexistence of maximum growth and product synthesis rates during the exponential growth phase (21) and the other based on the obligate requirement of optimal product synthesis in any growth phase (22). The distinction between these two types of growth coupling are also referred to weak coupling (*wGCP*) and strong coupling (*sGCP*) (16, 23).

Development of most strain design algorithms has been focused on overproduction of only one target molecule. The first algorithm proposed for modular cell design compatible for overproduction of multiple target molecules is MODCELL (13), which guided several experimental studies (14, 24-27). It works by generating *sGCP* strain designs for each target product based on elementary mode analysis (28), and then comparing the design strategies of different products to identify common genetic modifications among them. A similar approach was adapted in a subsequent work (29). For MODCELL to find optimal solutions for multiple target products, it requires: 1) enumerating all possible designs above a predefined minimum product yield and with minimal reaction deletion sets for each production network, which might lead to a large number of solutions for each network and hence make the problem computationally intractable, and 2) the resulting designs for all products must be compared to identify common interventions, which is a computationally-hard, set-covering problem. Thus, the current enumerative approach of MODCELL might become intractable very quickly, especially

for large-scale metabolic networks and potentially generate non-optimal designs, i.e., requiring more knock-outs than necessary or including fewer products than possible.

In this study, we generalized the concept of modular cell design and addressed the computational limitation of implementing it. We developed a novel computational platform (ModCell2), based on multiobjective optimization and analysis of mass action of cellular metabolism, to guide the design of modular cells for large-scale metabolic networks. 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. By analyzing these designs, we further revealed both intuitive and complex metabolic architectures enabling modularity in modular cell and production modules required for efficient biosynthesis of target molecules.

METHODS

Design principles of modular cell engineering

In the conventional strain engineering approach, a parent strain is genetically modified to yield an optimal production strain to make only a target product. To produce each new molecule, the design-build-test cycles of strain engineering must be repeated, which is laborious and expensive (Figure 1). To minimize the cycles, modular cell engineering is formulated by genetically transforming a parent strain into a modular (chassis) cell that must be assembled with exchangeable modules to create optimal production strains (13). A modular cell is designed to contain core metabolic pathways shared across designed optimal production strains. Exchangeable modules are production pathways designed to synthesize desirable chemicals. A combination of a modular cell and a production module(s) is required to balance redox, energy,

and precursor metabolites for sustaining cellular metabolism during growth and/or stationary phases and exhibiting only desirable phenotypes. Practically, modular cell engineering can be applied to monocultures and polycultures, where a production module(s) can be embedded in a modular cell and activated by intracellular and/or extracellular cues such as light and/or signaling molecules.

Multiobjective strain design framework for modular cell engineering

For modular cell engineering, we seek to design a chassis 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. Therefore, the framework of modular cell engineering can be formulated as a multiobjective optimization problem, named ModCell2, as described below.

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

subject to

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

subject to

$$\textstyle\sum_{j \in J_k} S_{ijk} v_{jk} = 0, \textit{for all } i \in I_k \qquad [2]$$

$$l_{jk} \leq v_{jk} \leq u_{jk}, \textit{for all } j \in J_k \qquad [3]$$

$$l_{jk} d_{jk} \leq v_{jk} \leq u_{jk} d_{jk}, \textit{for all } j \in C \text{ and } C \square J_k \qquad [4]$$

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

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

$$\textstyle\sum_{j \in C} \left(1 - y_j\right) \leq \alpha \qquad [6]$$

$$\sum_{j \in C} z_{jk} \leq \beta_k, \text{ for all } k \square K \quad [7]$$

where i, j, and k are indices of metabolite i, reaction j, and production network k, respectively; f_k is a design objective for network k; c_{jk} represents the cellular objective for reaction j in network k associated with a design objective defined in eqns. 8-10; v_{jk} (mmol/g DCW/h) is metabolic flux of reaction j bounded by l_{jk} and u_{jk} in network k, respectively; y_j and z_{jk} are binary design variables for deletion reaction j and module reaction j in network k, respectively; α and β_k are design parameters for deletion and module reactions, respectively; S_{ijk} is a stoichiometric coefficient of metabolite in reaction j of network k; and C (eqn. 4) is the candidate reaction set (Supplementary File S1). The goal of the optimization problem is to simultaneously maximize all design objectives f_k.

Steady-state mass balance constraint of cellular metabolism. Quasi steady-state flux balance of cellular metabolism (eqn. 2) is used as metabolic constraints for eqn. 1 (30, 31). A model corresponding to each modular production strain (i.e. production network k) will be derived from a parent strain (i.e. parent network) by adding necessary reactions (e.g., a production module) to produce a target molecule. A feasible flux distribution for each production network is described by mass balance (eqn. 2) and reaction flux bounds (eqns. 3 and 4). For a given production network, the phenotypic space can be illustrated by the gray area that is projected onto the two-dimensional space spanned by product synthesis and growth rates (Figure 2).

Design variables. In our formulation for modular cell engineering, we introduced two design variables − binary reaction deletions (y_j) inherent to the modular cell and module-specific reaction insertions (z_{jk}) (eqn. 4). These variables can be experimentally manipulated to constrain the desirable phenotypes of production strains as shown in Figure 2. Specifically, y_j = 0 if

reaction j is deleted from the modular cell; otherwise, y_j = 1. Deleting metabolic reactions removes undesired functional states of the network and leaves those with high design objectives. Likewise, z_{jk} = 1 if reaction j is present in the production network k; otherwise, z_{jk} = 0. These module reactions are endogenous reactions removed from the parent network (eqn. 5), but are added back to a specific production module to enhance the compatibility of a modular cell. The maximum number of reaction deletions (α) and module-specific reaction insertions (β_k) are user-defined parameters.

Design objectives. To generalize ModCell2 design, we allow three different types of design objectives (f_k, eqn. 1) that determine production phenotypes for each production network. Depending on the application, a phenotype can be designed to be weak coupling (*wGCP*), strong coupling (*sGCP*), and/or non-growth production (*NGP*) (Figure 2). The constrained phenotypic spaces based on these design objectives are shown in color; any point within these spaces is a feasible physiological state of the cell that can be represented by a metabolic flux distribution.

The *wGCP* design seeks to achieve a high product rate at maximum growth rate (Figure 2A). The *wGCP* design objective, f_k^{wGCP} ($\in$ [0, 1]), is calculated as follows:

$$f_k^{wGCP} = \frac{v_{Pk}^{\mu}}{v_{Pmaxk}^{\mu}} \qquad [8]$$

where v_{Pk}^{μ} is the minimum synthesis rate of the target product P at the maximum growth rate for production network k and v_{Pmaxk}^{μ} is the maximum synthesis rate of P (Supplementary File S1). This *wGCP* design formulation is equivalent to RobustKnock (32) or OptKnock with a tilted objective function (21, 33, 34). In eqn. 8, f_k^{wGCP} is scaled from 0 to 1 for proper comparison among products. The *wGCP* design is appropriate for applications where growth rate is not limited by the nutrients, and the product is formed during the growth phase.

The *sGCP* design seeks to achieve a high product rate not only at optimal growth rate but also during non-growth phase (Figure 2B). The *sGCP* design objective, f_k^{sGCP} ($\in$ [0, 1]), is calculated as follows:

$$f_k^{sGCP} = \frac{v_{Pk}^{\mu}}{v_{P_{max}k}^{\mu}} \cdot \frac{v_{Pk}^{\bar{\mu}}}{v_{P_{max}k}^{\bar{\mu}}} \qquad [9]$$

where $v_{Pk}^{\bar{\mu}}$ and $v_{P_{max}k}^{\bar{\mu}}$ are the minimum and maximum product formation rates for production network k in the stationary phase, respectively (Supplementary File S1). The *sGCP* design objective is comparable to the one implemented in MODCELL (13). Different from *wGCP*, *sGCP* requires high product synthesis rate for any growth phase. However, the additional constraint of optimal product synthesis during the stationary phase requires more genetic manipulations or specific experimental conditions (e.g., anaerobic growth condition, supply of intermediate metabolites). Both *wGCP* and *sGCP* designs enable fast growth selection to attain the optimum product rates by adaptive laboratory evolution (35, 36).

The *NGP* design aims to maximize the minimum product rate during the non-growth phase by eliminating carbon fluxes directed to biomass synthesis (Figure 2C). The *NGP* design objective, f_k^{NGP} ($\in$ [0, 1]), is calculated as follows:

$$f_k^{NGP} = \frac{v_{Pk}^{\bar{\mu}}}{v_{P_{max}k}^{\bar{\mu}}} \qquad [10]$$

While the *NGP* design is not suitable for growth selection, it can be derived from a *wGCP* (or *sGCP*) design by imposing additional genetic modifications. Practically, *NGP* design strains can be activated during cell culturing using a regulatory genetic circuit to toggle switch between production phases.

Design solutions. Optimal solutions for eqn. 1 are a Pareto set (***PS***) that correspond to design variables, including reaction deletions (y_j) and module reaction insertions (z_{jk}). Each solution constitutes a design of a modular cell:

$$\boldsymbol{PS} := \{\mathbf{x} \in \Omega : \nexists \mathbf{t} \in \boldsymbol{\Omega}, \mathbf{F}(\mathbf{t}) \prec \mathbf{F}(\mathbf{x})\} \quad [11]$$

Here, $\mathbf{F}(\mathbf{t}) \prec \mathbf{F}(\mathbf{x})$ means $\mathbf{F}(\mathbf{t})$ *dominates* $\mathbf{F}(\mathbf{x})$ if and only if $\mathrm{f_i}(\mathbf{t}) \geq \mathrm{f_i}(\mathbf{x})$ for all i, and $\mathbf{F}(\mathbf{t})$ differs from $\mathbf{F}(\mathbf{x})$ in at least one entry. The feasible space of design variables, Ω, is defined by the problem constraints (eqn. 2-7, also see Supplementary File S1). Phenotypes of modular cells will be the image of the Pareto set in the objective space, known as the Pareto front (***PF***):

$$\boldsymbol{PF} := \{\mathbf{F}(\mathbf{x}) : \mathbf{x} \in \boldsymbol{PS}\} \quad [12]$$

For the multiobjective strain design framework, the input parameters include α (eqn. 6), β_k (eqn. 7), and the production networks as input metabolic models. Each model contains a production module to produce one target chemical. The output is a Pareto set (genetic modifications) and its respective Pareto front (desirable production phenotypes). For a special case with no trade-off among the design objectives, an optimal solution, named a utopia point, exists where each objective achieves its maximum value. The multiobjective strain design formulation presented is general and can be applied to design modular cells for any organism.

Algorithm and implementation

ModCell2 algorithm. To solve the multiobjective optimization problem for modular cell engineering, we used multiobjective evolutionary algorithms (MOEAs) (37). MOEAs were selected because they can efficiently handle linear and non-linear problems and do not require preferential specification of design objectives (38). MOEAs start by randomly generating a population of individuals (a vector of design variables), each of which is mapped to a design

objective vector (i.e., a fitness vector). In ModCell2 (Supplementary File S1), the objective values of an individual are calculated by solving the linear programming problems for each production network. Next, individuals are shuffled to generate an offspring, from which the most fit individuals are kept. This process was repeated until the termination criteria was reached, for instance, either the solutions cannot be further improved or the simulation time limit is reached.

ModCell2 implementation. To streamline the modular cell design, we developed the ModCell2 software package based on three core classes (Figure S1 in Supplementary File S2). The Prodnet class parses and pre-processes production network models, and computes production phenotypes. The MCdesign class serves as an interface between the MOEA optimization method and metabolic models. Finally, the ResAnalysis class loads the Pareto set computed by MCdesign and identifies the most promising modular cell designs.

The code was written in MATLAB 2017b (The Mathworks Inc.) using the function gamultiobj() from the MATLAB Optimization Toolbox that implements the NSGA-II algorithm (39) to solve the multiobjective optimization problem. The solution and analysis methods were parallelized using the MATLAB Parallel Computing Toolbox. The linear programs to calculate metabolic fluxes were solved using the GNU Linear Programming Kit (GLPK). The COBRA toolbox (40, 41) and F2C2 0.95b (42) were also used for COBRA model preprocessing and manipulation.

Metabolic models. In our study, we used three parent models including i) a small metabolic network to illustrate the modular cell design concept (13), ii) a core metabolic network of *Escherichia coli* to compare the performance of ModCell2 with respect to the conventional single-product strain design strategy and the first-generation modular cell design platform MODCELL (13), and iii) a genome-scale metabolic network of *E. coli* (i.e., iML1515 (33)) for

biosynthesis of a library of endogenous and heterologous metabolites, including 4 organic acids, 6 alcohols, and 10 esters (Figures S2 in Supplementary File S2) (8, 22, 25, 43-49).

Simulation protocols. Anaerobic conditions were imposed by setting oxygen exchange fluxes to be 0, and the glucose uptake rate was constrained to be at most 10 mmol/gCDW/h, as experimentally observed for *E. coli*. When using the genome-scale model iML1515 to simulate *wGCP* designs, the commonly observed fermentative products (acetate, CO2, ethanol, formate, lactate, succinate) were allowed for secretion as described elsewhere (50). For simulation of *sGCP* and *NGP* designs, the glucose uptake rate was fixed (i.e., -10 mmol/gCDW/h); otherwise, the flux is not active during the no-growth phases, resulting in the product synthesis rate of 0 regardless of genetic manipulations. To compare ModCell2 with Optknock, we applied the OptKnock algorithm with a titlted objective function (51) to generate *wGCP* designs for each production network, using the open-source algebraic modeling language Pyomo (52). The MILP problems were solved using CPLEX 12.8.0 with a time limit of 10,000 seconds set for each product. ModCell2 is provided as an open-source software package and is freely available for academic research. The software package and documentation can be download via either web.utk.edu/~ctrinh or Github https://github.com/TrinhLab.

Analysis methods for design solutions

Compatibility. The compatibility, C ($\square$ Z^+), of a design is defined as the number of products that are coupled with a modular cell and has objective values above a specified cutoff value θ. As a default, we set $\theta = 0.6$ for the *wGCP* and *NGP* design objectives and $\theta = 0.36$ (0.6^2) for the *sGCP* design objective. For example, a *wGCP* design for 3 products that has the

design objective values of 0.4, 0.9, and 0.6 has a compatibility of 2, given a cutoff value of $\theta \geq 0.6$.

Compatibility difference and loss. Robustness is the ability of a system to maintain its function against perturbations, and hence is very important of designing biological and engineered systems (53). To evaluate the robustness of modular cell designs, we defined two metrics, the compatibility difference (*CD*) and compatibility loss (CL □ [0,1]) as follows:

$$CD = C_{initial} - C_{final} \quad [13]$$

$$CL = \frac{C_{initial} - C_{final}}{C_{initial}} \quad [13]$$

where $C_{initial}$ and C_{final} are the compatibilities of a modular cell design before and after a single reaction deletion, respectively. The value $CD > 0$ (or $CL > 0$) means the modular gains fitness while $CD < 0$ (or $CL < 0$) means that it loses its fitness. In the analysis, we did not consider essential and blocked reactions for our single-deletion analysis; for instance, there are only 1139 potential reaction deletions in the iML1515 model.

Metabolic switch design. A metabolic switch design is a modular cell that can possess multiple production phenotypes (i.e., *wGCP*, *sGCP*, and *NPG*), activated by an environmental stimulus (e.g. metabolites, lights). The metabolic switch design is enforced to have a set of reaction (gene) deletions in one production phenotype to be a subset of the other, for instance, $\{\mathbf{y}_{\mathrm{wGCP}}\} \subseteq \{\mathbf{y}_{\mathrm{NPG}}\}$, $\{\mathbf{y}_{\mathrm{wWCP}}\} \subseteq \{\mathbf{y}_{\mathrm{sGCP}}\}$. The metabolic switch design is beneficial for multiphase fermentation configurations that enable flexible genetic modification and implementation. Specifically, the metabolic switch design can exhibit the *wGCP* phenotype during the growth phase and the *NPG* (or *sGCP*) phenotype during the stationary phase. The metabolic switches can be implemented using the genetic switchboard (54).

RESULTS AND DISCUSSION

Illustrating ModCell2 for modular cell design of a simplified network

An example parent network, adapted from (13), was used to illustrate ModCell2 (Figure 3A). Inputs for the multiobjective optimization problem include i) three production networks (Figure 3B), comprising of one endogenous production module (module 1) and two heterologous production modules (modules 2 and 3) and ii) design parameters (Figure 3C), containing design objective type, maximum number of deletion reactions (α), and maximum number of module reactions (β_k). The output of ModCell2 generated the Pareto set and the corresponding Pareto front for modular cell designs (Figure 3D). The 2-D plots of product yields versus growth rates presented the feasible phenotypic spaces of the wildtype (gray area) and the designed strain (blue area).

Using various α and β_k values, ModCell2 simulation generated three *wGCP-*α*-*β_k*-d*, four *sGCP-*α*-*β_k*-d* designs, and three *NGP-*α*-*β_k*-d* designs, where *d* is the design solution index (Figure 3D). For instance, by setting α = 3 and β_k = 0, we found three *sGCP* designs including *sGCP-3-0-1*, *sGCP-3-0-2*, and *sGCP-3-0-3*. The first design *sGCP-3-0-1* has a compatibility of 2 (given θ □ 0) with the design objective values of 0.42 and 0.97 for the products P2 and P3, respectively. In contrast, the *sGCP-3-0-2* and *sGCP-3-0-3* designs have compatibilities of only 1 with the design objectives of 0.63 for P2 and 0.45 for P1, respectively.

Based on all designs, we can clearly see the trade-offs for optimization of different products for β_k = 0. However, setting $\beta_k \geq 1$ helps increase the compatibility of a modular cell with different production modules. In addition, we found that the Pareto front collapses into a utopia point as seen in the *wGCP-1-1-1*, *sGCP-3-1-1*, and *NGP-3-1-1* designs. For instance, the modular cell, *sGCP-3-1-1*, is compatible with all three products. The three corresponding

optimal production strains can couple growth and product formation during the growth phase. During the stationary phase, these strains produce the products at maximum theoretical yields. In theory, a universal modular cell always exists, provided that enough reaction deletions and module reactions are used. It might be more tractable to construct such a modular cell from a synthetic minimal cell using the bottom-up approach. However, construction of a universal modular cell from a host organism (e.g., *E. coli*, *S. cerevisiae*) using the top-down approach will require a significantly large number of genetic modifications, that might be challenging.

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 platform. To evaluate the algorithms and performance of ModCell2, we directly compared it with MODCELL (13) in two case studies, using the same core model of *E. coli* for production of five alcohols (ethanol, propanol, isopropanol, butanol, and isobutanol) and 5 derived butyrate esters (ethyl butyrate, propyl butyrate, isopropyl butyrate, butyl butyrate, and isobutyl butyrate) from glucose (Figure 4A).

In the first case study, we fixed the reaction module, i.e. $\beta_k = 2$ for ethanol dehydrogenase (FEM5) and ethanol export reaction (TRA1), in ModCell2 to emulate the same input as MODCELL (Supplementary File S3). The results showed that ModCell2 generated all the designs with the same *sGCP* objective values like MODCELL (Figure 4B, 4C, 4D) together with other alternative solutions (Supplementary File S3). Interestingly, ModCell2 only required 5 and 6 reaction deletions as opposed to 7 and 7 for the *sGCP-5-0-5* and *sGCP-5-0-6* designs, respectively. By setting the maximum reaction deletions to $\alpha \geq 6$, ModCell2 could find better

design solutions with fewer deletion reaction requirement and higher objective values (Supplementary File S3).

In the second case study, we used the same model configuration but did not enforce the module reactions. By setting $\alpha = 5$ and $\beta_k = 1$, we found the *sGCP-5-1-8* design that is compatible with all products and achieves the same objective values for products found in *sGCP-5-0-5*, *sGCP-5-0-6*, and *sGCP-5-0-2* (Figure 4E). The desirable phenotypic spaces can be further constrained for many products if α is increased from 5 to 6 (Figure 4F). Remarkably, by setting $\alpha = 8$ and $\beta_k = 2$, we found a utopia point design, *sGCP-8-2-9*, without any trade-off among design objectives (Figure 4G). This utopia point design could not be achieved with $\alpha < 8$ regardless of any β_k value.

Overall, the results demonstrate that ModCell2 can efficiently compute the Pareto front of modular cell designs. It can find better designs with fewer reaction deletion and module reaction requirements, improve design objective values, and enhance compatibility.

ModCell2 can identify designs with more compatibility than the conventional single-product designs. To evaluate if the conventional, single-product design strategy is suitable for modular cell engineering, we first used OptKnock to generate *wGCP* designs for the same 10 target molecules independently with various allowable reaction deletions ($\alpha = 2, 3, \ldots, 7$). Likewise, we employed ModCell2 to produce *wGCP* designs using the same α and various β. To directly compare OptKnock and ModCell2 solutions, we calculated the *wGCP* design objective values for all products based on each OptKnock solution (Supplementary File S3). As expected, our result showed that ModCell2 and OptKnock designs have the same highest objective values for each product (Figure 5A). However, several OptKnock solutions were always dominated by ModCell2 solutions in all parameter configurations (Figure 5B). With $\alpha \geq 4$, ModCell2 could

identify *wGCP-α-1* designs with the maximum compatibility of 10, while the best OptKnock designs only achieved the highest compatibility of 5 (Figure 5C, 5D).

Overall, ModCell2 can generate modular cells compatible with the maximum number of modules and achieve high objective values. Single-product designs might not be compatible with a large number of products, and the solutions might be far from Pareto optimality.

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

Modcell2 can design modular cells using a large-scale metabolic network. To demonstrate that ModCell2 can be applied for a genome scale metabolic network, we tested it to generate *wGCP* designs for 20 target molecules with $\alpha = 4$ and various β_k (Supplementary File S4). The α value was chosen because with 4 deletions, OptKnock could identify single product designs with objectives above 60% of the theoretical maximum (Supplementary File S5). With $\beta_k = 0$, ModCell2 could identify modular cell designs with compatibility of 17, for example, the *wGCP-4-0-50* design featuring deletion of ACALD (acetaldehyde dehydrogenase, *adhE*), ACKr/PTAr (acetate kinase, *ack;* phosphotransacetylase, *pta*), GLYAT (glycine C-acetyltransferase, *kbl*), and LDH_D (lactate dehydrogenase, *ldhA*) (Figure 6A, 6D, Supplementary File S4). By analyzing all *wGCP-4-*β_k*-d* designs (257 total for β_k = 0, 1, 2, and 3), we found that the ethanol and D-lactate production modules are most compatible with all modular cell designs (Figure 6A, 6C, Supplementary File S4). Among reaction deletions, *LdhA* (86% of designs), *Pta* (38%), and *AdhE* (25%) are the most frequent deletion reactions (Figure 6B). This finding is consistent with a comprehensive survey of metabolic engineering publications (55) showing that these deleted reactions appeared in most of *E. coli* engineered

strains for production of fuels and chemicals. The result supports the potential use of modular cell engineering to systematically build modular platform strains.

ModCell2 designs can capture combinatorial characteristics of production modules. To evaluate whether ModCell2 could capture the combinatorial properties among production modules, we analyzed the Pareto front of *wGCP-4-0-d* that have a total of 58 designs. Hierarchical clustering of this Pareto front revealed certain products with similar objective values across solutions, such as ethyl esters and butyrate esters (Figure 6A). These products together were compatible with different modular cells and exhibited metabolic similarity in their production modules. Thus, ModCell2 could generate designs that capture the combinatorial properties useful for modular cell engineering.

ModCell2 can identify highly compatible modular cells. Analysis of compatibility shows that certain modular cells can couple with production modules that may not exhibit the combinatorial properties (Figure 6D). However, there exists a tradeoff between the number of feasible designs and degree of compatibility. Some modular cell designs are compatible with up to 17 out of 20 products, for instance, the most compatible design, *wGCP-4-0-48*, featuring deletions of ACALD (*adhE*), ACKr/PTAr (*ack*, *pta*), GND (phosphogluconate dehydrogenase, *gnd*) and LDH_D (*ldhA*) (Supplementary File S4). An alternative design *wGCP-4-0-48-alternative* also exists where deletion of G6PDH2r (gluose-6-phosphate dehydrogenase, *zwf*) is replaced by that of GND, the first step in the oxidative pentose phosphate pathway. The gene deletions in the design *wGCP-4-0-48-alternative* are a subset of the modular *E. coli* strain TCS095, whose modular properties have recently been validated experimentally (14).

To determine if modular cell design is a viable alternative to single-product design, we also analyzed a potential tradeoff between design performance and modularity by comparing the

maximum value of each objective across all solutions in the Pareto front and the single-product design optima. If production modules exhibit competing phenotypes, a modular cell will not achieve the same performance in all modules as a single-product design strain. Analysis of the most compatible design *wGCP-4-0-48-alternative* showed that it could achieve objectives within 4% of the single-product optima in 14 products and within 10% in 3 products (Figure 6E). This result indicates that it is feasible to identify highly compatible modular cell designs without a significant tradeoff between performance and modularity.

Analysis of potential tradeoff between robustness and modularity can identify conserved metabolic features. To evaluate the robustness of modular cells, we analyzed the compatibility change (*CD*) of *wGCP-4-0* designs with compatibilities of 4 or greater (Figure S3 in Supplementary File 2). Remarkably, the result shows that only 7.5% of potential reaction deletions were detrimental to the robustness of modular cells while the large remaining portion did not affect *CD* values. Out of the 85 reactions whose deletion affected compatibility, only a few appeared consistently across the designs. For instance, deletion of TPI (triose-phosphate isomerase, *tpi*) led to an average compatibility loss of 95%, inactivating most modular cell designs. Based on flux variability analysis, TPI must operate in the forward direction by converting glycerone phosphate (dhap) to glyceraldehyde-3-phosphate (g3p) to drive sufficient flux through glycolysis and hence preventing synthesis of undesired byproducts (D-lactate or 1,2-propanediol) from dhap. Likewise, deletion of carbon dioxide and water transport and exchange reactions caused compatibility loss across all designs. Pyruvate carboxylase (PPC) is an important reaction to channel carbon flux through the Krebs cycle (56), and hence, deletion of PPC reduces compatibility in most modular cell designs with an average *CL* of 43%.

While some reaction deletions are critical for modular cell robustness, others are associated with specific products. For example, deletion of PDH (pyruvate dehydrogenase complex, *lpd/aceEF*) removes compatibility in all butanol-derived designs, indicating PFL (pyruvate formate lyase, *pfl*) is not an appropriate route. To make heterologous butanol-derived molecules under anaerobic conditions, FDH (NADH-dependent formate dehydrogenase, *fdh*) is required in butanol-derived modules where enzymatic reaction pairs of PFL and FDH could substitute PDH known to be anaerobically inhibited.

Overall, analysis of tradeoff between modularity and robustness can identify not only the conserved metabolic features of modular cells but also potential bottlenecks in specific production modules.

Enabling metabolic switch among different design objectives using ModCell2

The ability to dynamically control growth and production phases can potentially enhance product titers, rates, and yields. For instance, two-phase fermentation can be employed where growth phase is optimized for biomass synthesis and stationary phase for chemical production (57). Using ModCell2, we investigated the feasibility to design optimal strains to toggle switch desirable production phenotypes.

To design a *wGCP→NPG* metabolic switch, we first used our reference *wGCP* design as a parent strain (Figure 7A) and then employed ModCell2 to identify the most compatible *wGCP→NPG* designs. With 5 additional deletions, we could find *wGCP*→NPG designs that encompass both *wGCP* and *NPG* phenotypes, for instance, the *sup-NGP-5-0-23* design featuring deletion of PGI (glucose-6-phosphate isomerase, *pgi*), MDH (malate dehydrogenase, *mdh*), ASPT (L-aspartase, *aspT*), Tkt2 (transketolase, *tktB*), and ATPS4rpp (ATP synthase, *atp*) (Figure

7B). The deletion reactions in the *wGCP*→NPG designs appear in both catabolic (PGI, ATPS4tpp) and anabolic (ASP, TKT2) processes, responsible for growth disruption and direction of carbon flow to the biosynthesis of target products.

Likewise, we used ModCell2 to design a *wGCP*→*sGCP* metabolic switch. We identified the most compatible *wGCP*→*sGCP* designs with 6 additional deletions, for instance, the *sup-sGCP-6-0-39* design featuring the deletion of MGSA (Methylglyoxal synthase, *mgsA*), ALCD2x (alcohol dehydrogenase, *adhE*), PFL, MDH, FADRx (FAD reductase, *fadI*), and GLUDy ($NADP^+$dependent glutamate dehydrogenase, *gdhA*) (Figure 7C). Different from the *wGCP*→*NGP* metabolic switch, all deletions in the *wGCP*→*sGCP* designs are involved the elimination of biosynthesis pathways of undesirable byproducts.

While it is feasible to metabolically switch among different production phenotypes, it not only requires more reaction deletions but also reduces the product compatibility. For instance, the *wGCP*→sGCP and *wGCP*→NGP designs are only compatible with 5 products while the *wGCP* parent design have a compatibility of 17 out of 20 products with 4 deletions. The main reason is that both *wGCP*→sGCP and *wGCP*→NGP designs must eliminate all possible redundant pathways that result in biosynthesis of undesirable byproducts.

CONCLUSION

In this study, we developed a multiobjective strain design platform for modular cell engineering. With a new developed algorithm and computational platform, ModCell2 enables flexible design of modular cells that can couple with production modules to exhibit desirable production phenotypes. In comparison to the first-generation strain design platform, ModCell2 can handle large-scale metabolic networks and identify better solutions that require fewer genetic

modifications and exhibit more product compatibility. Different from the conventional single-product strain design, ModCell2 can find solutions that are Pareto optimal with negligible tradeoffs among modularity, performance, and robustness. We envision ModCell2 is a useful tool to implement modular cell engineering and fundamentally study modular designs in natural and synthetic biological systems.

ACKNOWLEDGEMENTS

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. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Nielsen J & Keasling Jay D (2016) Engineering Cellular Metabolism. *Cell* 164:1185-1197.
2. Trinh CT & Mendoza B (2016) Modular cell design for rapid, efficient strain engineering toward industrialization of biology. *Current Opinion in Chemical Engineering* 14:18-25.
3. Connelly T, *et al.* (2015) Industrialization of Biology: A Roadmap to Accelerate the Advanced Manufacturing of Chemicals.
4. Sauro HM (2008) Modularity defined. *Molecular systems biology* 4.
5. Purnick PEM & Weiss R (2009) The second wave of synthetic biology: from modules to systems. *Nature reviews Molecular cell biology* 10:410-422.
6. Yadav VG, De Mey M, Giaw Lim C, Kumaran Ajikumar P, & Stephanopoulos G (2012) The future of metabolic engineering and synthetic biology: Towards a systematic practice. *Metabolic Engineering* 14:233-241.
7. Biggs BW, De Paepe B, Santos CNS, De Mey M, & Ajikumar PK (2014) Multivariate modular metabolic engineering for pathway and strain optimization. *Current opinion in biotechnology* 29:156-162.
8. Rodriguez GM, Tashiro Y, & Atsumi S (2014) Expanding ester biosynthesis in Escherichia coli. *Nat Chem Biol* 10:259-265.
9. Cheong S, Clomburg JM, & Gonzalez R (2016) Energy-and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions. *Nature biotechnology* 34:556-561.

10. Tseng H-C & Prather KL (2012) Controlled biosynthesis of odd-chain fuels and chemicals via engineered modular metabolic pathways. *Proceedings of the National Academy of Sciences* 109(44):17925-17930.

11. Xu P, *et al.* (2013) Modular optimization of multi-gene pathways for fatty acids production in *E. coli*. *Nat Commun* 4:1409.

12. Ng CY, Chowdhury A, & Maranas CD (2016) A microbial factory for diverse chemicals. *Nat Biotech* 34:513-515.

13. Trinh CT, Liu Y, & Conner DJ (2015) Rational design of efficient modular cells. *Metabolic engineering* 32:220-231.

14. Wilbanks B, Layton D, Garcia S, & Trinh C (2017) A Prototype for Modular Cell Engineering. *ACS Synthetic Biology*:acssynbio.7b00269.

15. Long MR, Ong WK, & Reed JL (2015) Computational methods in metabolic engineering for strain design. *Current opinion in biotechnology* 34:135-141.

16. Machado D & Herrgård MJ (2015) Co-evolution of strain design methods based on flux balance and elementary mode analysis. *Metabolic Engineering Communications* 2:85-92.

17. Chowdhury A, Zomorrodi AR, & Maranas CD (2015) Bilevel optimization techniques in computational strain design. *Computers & Chemical Engineering* 72:363-372.

18. Fong SS, *et al.* (2005) In silico design and adaptive evolution of Escherichia coli for production of lactic acid. *Biotechnology and Bioengineering* 91:643-648.

19. Trinh CT & Srienc F (2009) Metabolic engineering of Escherichia coli for efficient conversion of glycerol to ethanol. *Applied and Environmental Microbiology* 75(21):6696-6705.

20. Otero JM, *et al.* (2013) Industrial systems biology of Saccharomyces cerevisiae enables novel succinic acid cell factory. *PLoS One* 8:e54144.

21. Burgard AP, Pharkya P, & Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering* 84:647-657.

22. Trinh CT, Unrean P, & Srienc F (2008) Minimal Escherichia coli Cell for the Most Efficient Production of Ethanol from Hexoses and Pentoses. *Applied and Environmental Microbiology* 74(12):3634-3643.

23. Klamt S & Mahadevan R (2015) On the feasibility of growth-coupled product synthesis in microbial strains. *Metabolic engineering* 30:166-178.

24. Wierzbicki M, Niraula N, Yarrabothula A, Layton DS, & Trinh CT (2016) Engineering an Escherichia coli platform to synthesize designer biodiesels. *Journal of biotechnology* 224:27-34.

25. Layton DS & Trinh CT (2014) Engineering modular ester fermentative pathways in Escherichia coli. *Metabolic Engineering* 26:77-88.

26. Layton DS & Trinh CT (2016) Expanding the modular ester fermentative pathways for combinatorial biosynthesis of esters from volatile organic acids. *Biotechnology and bioengineering*.

27. Layton DS & Trinh CT (2016) Microbial synthesis of a branched-chain ester platform from organic waste carboxylates. *Metabolic Engineering Communications* 3:245-251.

28. Trinh CT, Wlaschin A, & Srienc F (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Applied Microbiology and Biotechnology* 81:813-826.

29. Jouhten P, *et al.* (2016) Yeast metabolic chassis designs for diverse biotechnological products. *Scientific reports* 6.

30. Price ND, Papin JA, Schilling CH, & Palsson BO (2003) Genome-scale microbial in silico models: the constraints-based approach. *Trends in Biotechnology* 21:162-169.

31. Palsson BØ (2015) Systems biology: constraint-based reconstruction and analysis. (Cambridge University Press).

32. Tepper N & Shlomi T (2010) Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics* 26:536-543.

33. Feist AM, *et al.* (2010) Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. *Metabolic engineering* 12:173-186.

34. Yang L, Cluett WR, & Mahadevan R (2011) EMILiO: a fast algorithm for genome-scale strain design. *Metabolic engineering* 13:272-281.

35. Fong SS, *et al.* (2005) In silico design and adaptive evolution of Escherichia coli for production of lactic acid. *Biotechnology and bioengineering* 91(5):643-648.

36. Trinh C & Srienc F (2009) Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol. *Appl Environ Microbiol* 75(21):6696 - 6705.

37. Coello Coello CA (2002) Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art. *Computer Methods in Applied Mechanics and Engineering* 191(11):1245-1287.

38. Marler RT & Arora JS (2004) Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization* 26:369-395.

39. Deb K, Pratap A, Agarwal S, & Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *Evolutionary Computation, IEEE Transactions on* 6:182-197.

40. Schellenberger J, *et al.* (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. *Nature protocols* 6:1290-1307.

41. Heirendt L, *et al.* (2017) Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0.

42. Larhlimi A, David L, Selbig J, & Bockmayr A (2012) F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. *BMC Bioinformatics* 13:57.

43. Shen CR, *et al.* (2011) Driving forces enable high-titer anaerobic 1-butanol synthesis in Escherichia coli. *Applied and environmental microbiology* 77:2905-2915.

44. Atsumi S, Hanai T, & Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* 451:86-89.

45. Tseng H-C & Prather KLJ (2012) Controlled biosynthesis of odd-chain fuels and chemicals via engineered modular metabolic pathways. *Proceedings of the National Academy of Sciences* 109:17925-17930.

46. Yim H, *et al.* (2011) Metabolic engineering of Escherichia coli for direct production of 1, 4-butanediol. *Nature chemical biology* 7:445-452.

47. Akita H, Nakashima N, & Hoshino T (2016) Pyruvate production using engineered Escherichia coli. *AMB Express* 6(1):94.

48. Niu D, *et al.* (2014) Highly efficient L-lactate production using engineered Escherichia coli with dissimilar temperature optima for L-lactate formation and cell growth. *Microbial cell factories* 13(1):78.

49. Yu JL, Xia XX, Zhong JJ, & Qian ZG (2014) Direct biosynthesis of adipic acid from a synthetic pathway in recombinant *Escherichia coli*. *Biotechnology and bioengineering* 111(12):2580-2586.

50. von Kamp A & Klamt S (2017) Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nature communications* 8:15956.

51. Maranas CD & Zomorrodi AR (2016) Optimization Methods in Metabolic Networks.

52. Hart WE, *et al.* (2012) *Pyomo-optimization modeling in python* (Springer).

53. Kitano H (2004) Biological robustness. *Nature Reviews Genetics* 5(11):826.

54. Callura JM, Cantor CR, & Collins JJ (2012) Genetic switchboard for synthetic biology applications. *Proceedings of the National Academy of Sciences* 109(15):5850-5855.

55. Winkler JD, Halweg-Edwards AL, & Gill RT (2015) The LASER database: Formalizing design rules for metabolic engineering. *Metabolic Engineering Communications* 2:30-38.

56. Coomes MW, Mitchell BK, Beezley A, & Smith TE (1985) Properties of an Escherichia coli mutant deficient in phosphoenolpyruvate carboxylase catalytic activity. *Journal of bacteriology* 164(2):646-652.

57. Klamt S, Mahadevan R, & Hädicke O (2018) When Do Two□Stage Processes Outperform One□Stage Processes? *Biotechnology journal* 13(2):1700539.

Figure legends

Figure 1: 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.

Figure 2: Graphical representation of 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_{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.

Figure 3: Illustration of ModCell2 workflow and analysis including (**A**) parent model, (**B**) production modules, (**C**) design parameters, and (**D**) simulation output for Pareto set and Pareto front based on design input.

Figure 4: The 2-D metabolic phenotypic spaces of different *sGCP* designs using the core 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. For each panel, the gray and blue areas correspond to the phenotypic spaces of the wildtype and the optimal production strain, respectively.

Figure 5: Comparison of strain design by OptKnock and Modcell2. (**A**) A correlation between the maximum objective values for each product generated by OptKnock and the equivalent values attained by ModCell2. Each point is colored based on the number of reaction deletions, with warmer colors corresponding to more reaction deletions. (**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 (blue), *wGCP* designs (β_k = 0, orange), wGCP designs (β_k = 1, yellow). (**D**) Compatibility distribution of Optknock (α = 7, orange) and *wGCP-7-1* (blue).

Figure 6: (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**) Design compatibility. **(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.

Figure 7: Production phenotypes of (**A**) the wild type (gray) and the representative, highly-compatible design *wGCP-4-0-48-alternative* (blue), (**B**) the *wGCP→NPG* design, *sup-NGP-5-0-23*, and (**C**) the *wGCP→sGCP* design, *sup-sGCP-6-0-39*.

Figure 1

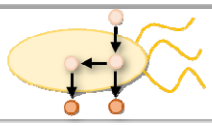
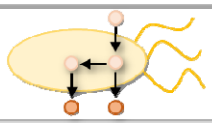
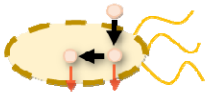
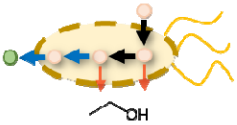
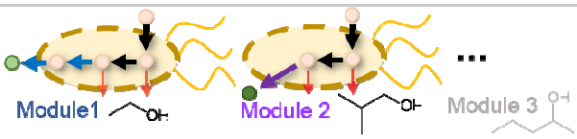
Features	Conventional strain engineering	Modular cell engineering
Parent strain		
Modular cell	Absent	 Optimized common production phenotypes
Exchangeable modules	1	Multiple
Optimal production strains		
Design-build-test cycle	Repeated for every new product	One time for many products

Figure 2

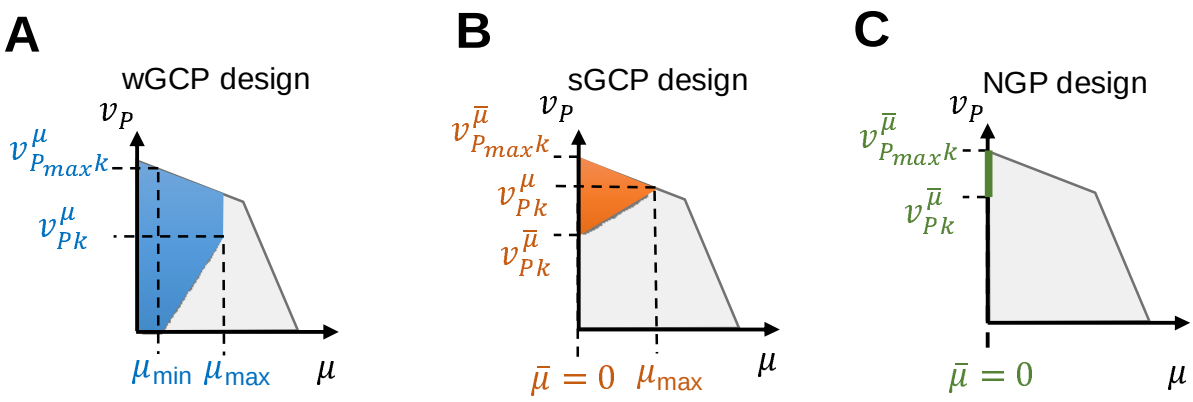
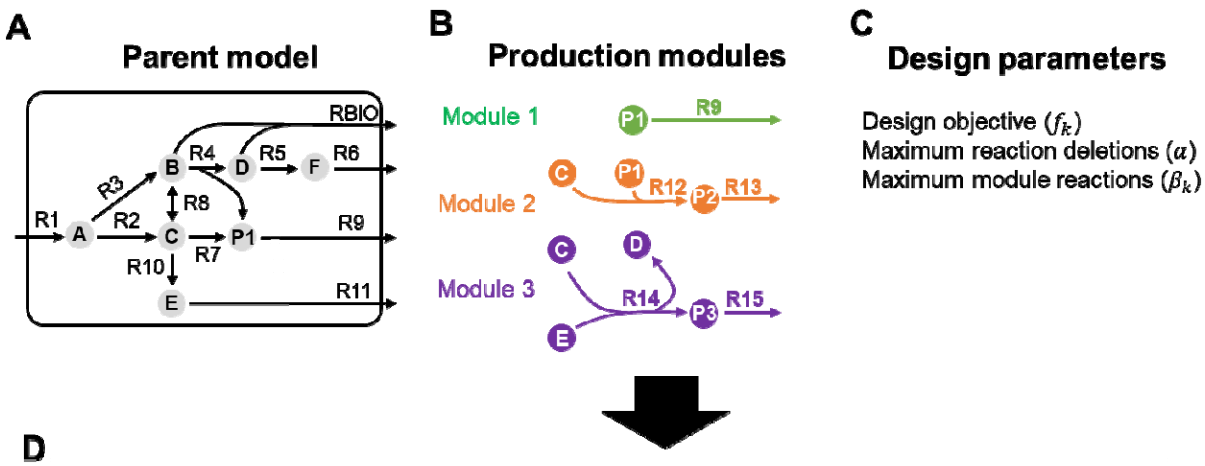

A
wGCP design
v_P
$v^{\mu}_{P_{max}k}$
v^{μ}_{Pk}
$\mu_{\min}$ $\mu_{\max}$
μ
B
sGCP design
v_P
$v^{\bar{\mu}}_{P_{max}k}$
v^{μ}_{Pk}
$v^{\bar{\mu}}_{Pk}$
$\bar{\mu} = 0$ $\mu_{\max}$
μ
C
NGP design
v_P
$v^{\bar{\mu}}_{P_{max}k}$
$v^{\bar{\mu}}_{Pk}$
$\bar{\mu} = 0$
μ

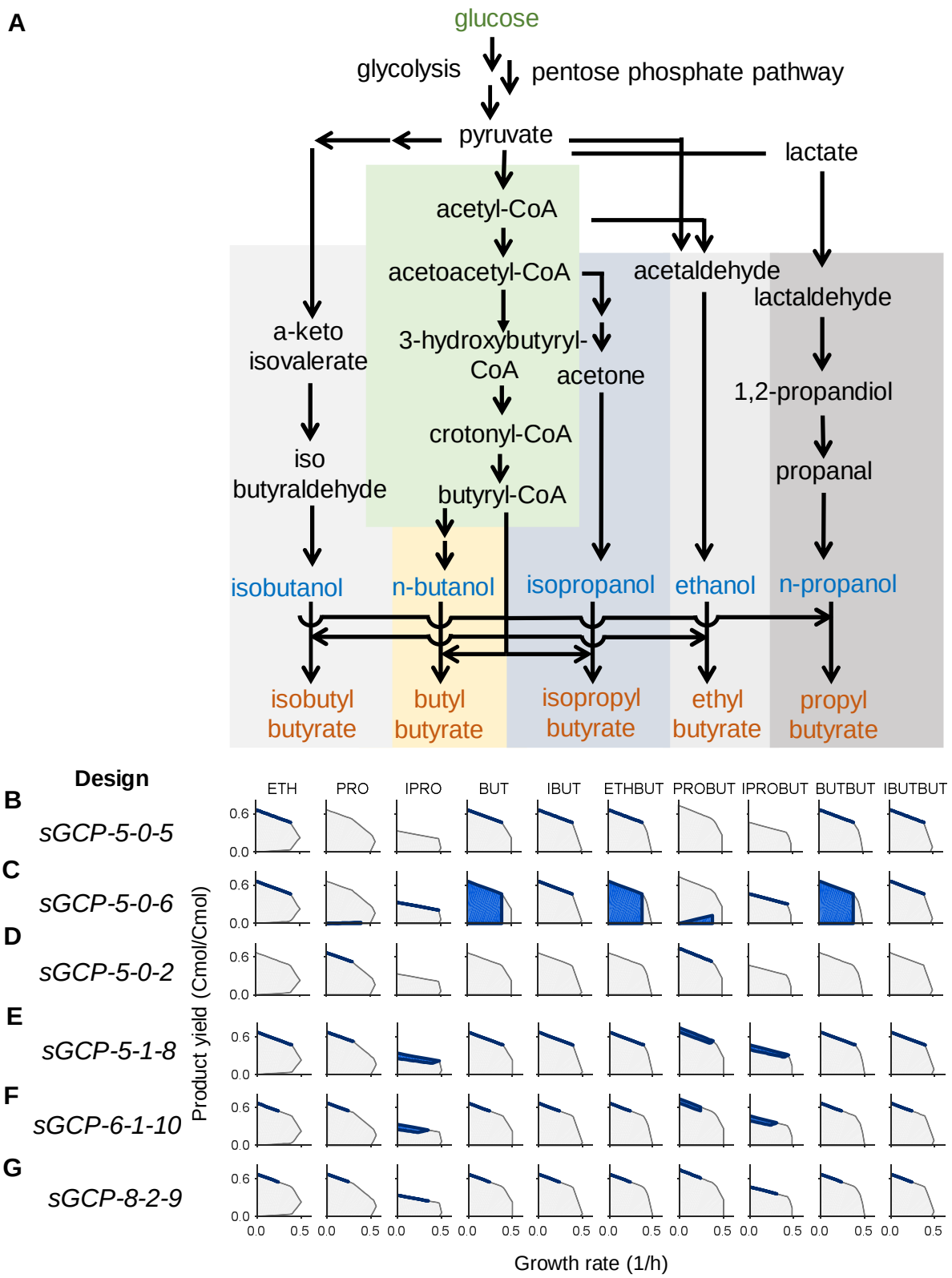
Figure 3



D

Design parameters				*Pareto set*				*Pareto front*					
Objective	α	β_k	d	**Reaction deletions**	**Module reactions**			f_k			**Prod. yield vs growth**		
					P1	P2	P3	P1	P2	P3	P1	P2	P3
wGCP	1	0	1	R10				0.45	0.00	0.00			
wGCP	1	0	2	R9				0.00	0.63	0.97			
wGCP	1	1	1	R9	R9			0.45	0.63	0.97			
sGCP	3	0	1	R8, R9, R11				0.00	0.42	0.97			
sGCP	3	0	2	R6, R9, R10				0.00	0.63	0.00			
sGCP	3	0	3	R6, R8, R11				0.45	0.00	0.00			
sGCP	3	1	1	R6, R9, R11	R9		R6	0.45	0.63	0.97			
NGP	3	0	1	R6, R11				1.00	0.00	0.00			
NGP	3	0	2	R4, R9, R11				0.00	1.00	1.00			
NGP	3	1	1	R4, R9, R11	R9			1.00	1.00	1.00			

Figure 4


A
glucose
glycolysis
pentose phosphate pathway
pyruvate
lactate
acetyl-CoA
acetoacetyl-CoA
acetaldehyde
lactaldehyde
a-keto isovalerate
3-hydroxybutyryl-CoA
acetone
1,2-propandiol
crotonyl-CoA
iso butyraldehyde
propanal
butyryl-CoA
isobutanol
n-butanol
isopropanol
ethanol
n-propanol
isobutyl butyrate
butyl butyrate
isopropyl butyrate
ethyl butyrate
propyl butyrate
Design
B
sGCP-5-0-5
C
sGCP-5-0-6
D
sGCP-5-0-2
E
sGCP-5-1-8
F
sGCP-6-1-10
G
sGCP-8-2-9
ETH
PRO
IPRO
BUT
IBUT
ETHBUT
PROBUT
IPROBUT
BUTBUT
IBUTBUT
Product yield (Cmol/Cmol)
0.6
0.0
0.5
Growth rate (1/h)

668

Figure 5

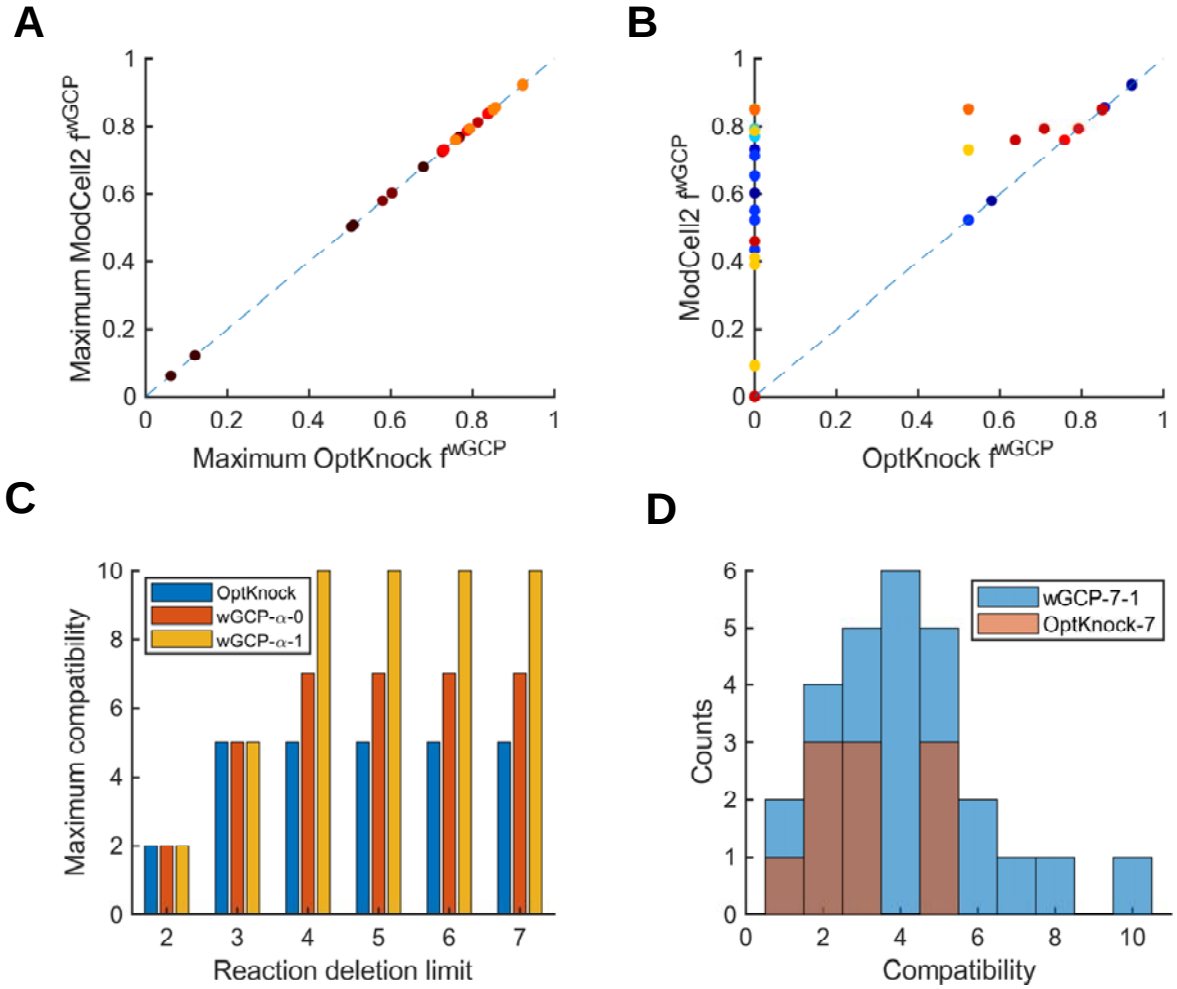

A
Maximum ModCell2 f^wGCP
Maximum OptKnock f^wGCP
B
ModCell2 f^wGCP
OptKnock f^wGCP
C
Maximum compatibility
Reaction deletion limit
OptKnock
wGCP-α-0
wGCP-α-1
D
Counts
Compatibility
wGCP-7-1
OptKnock-7

Figure 6

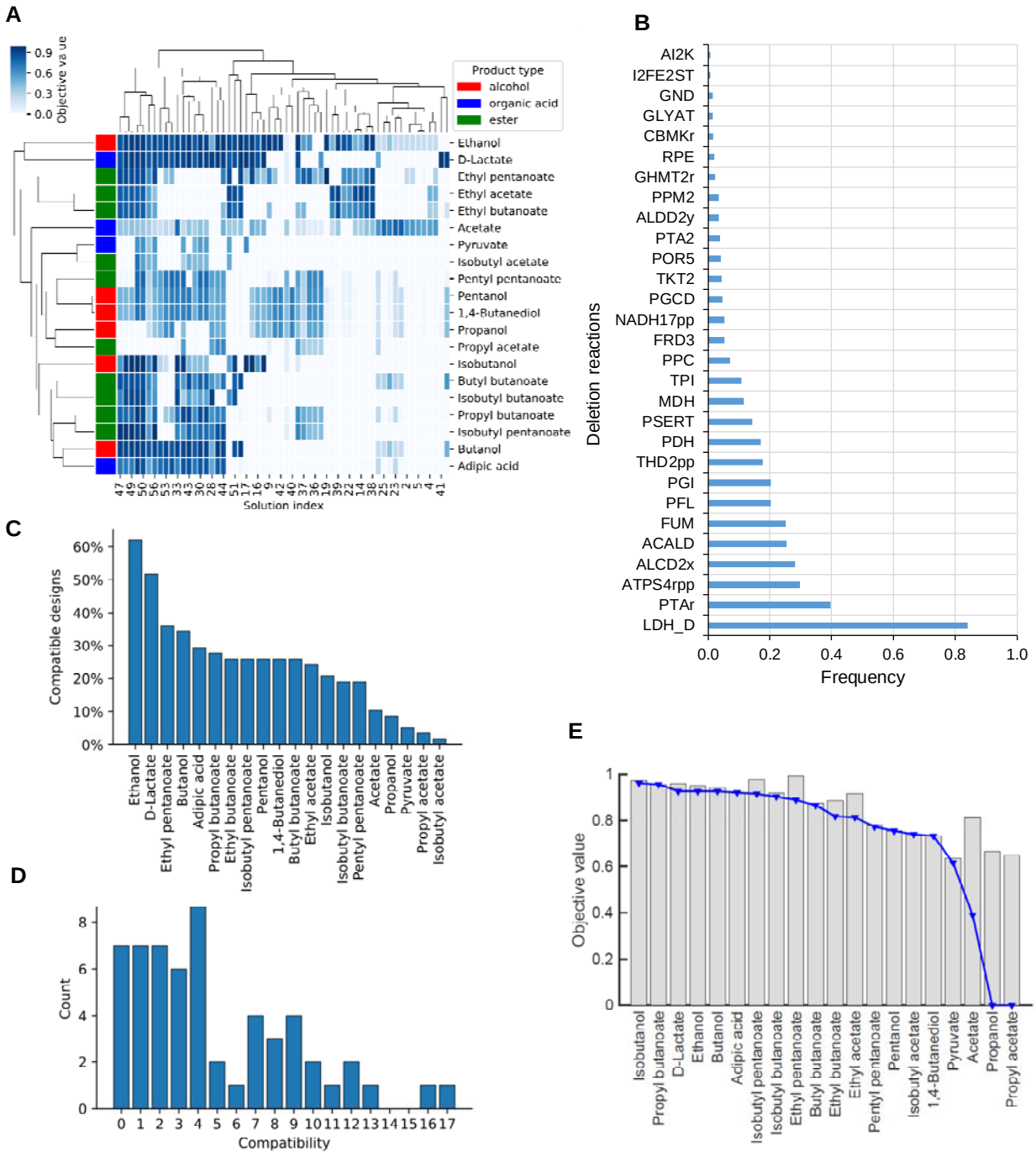
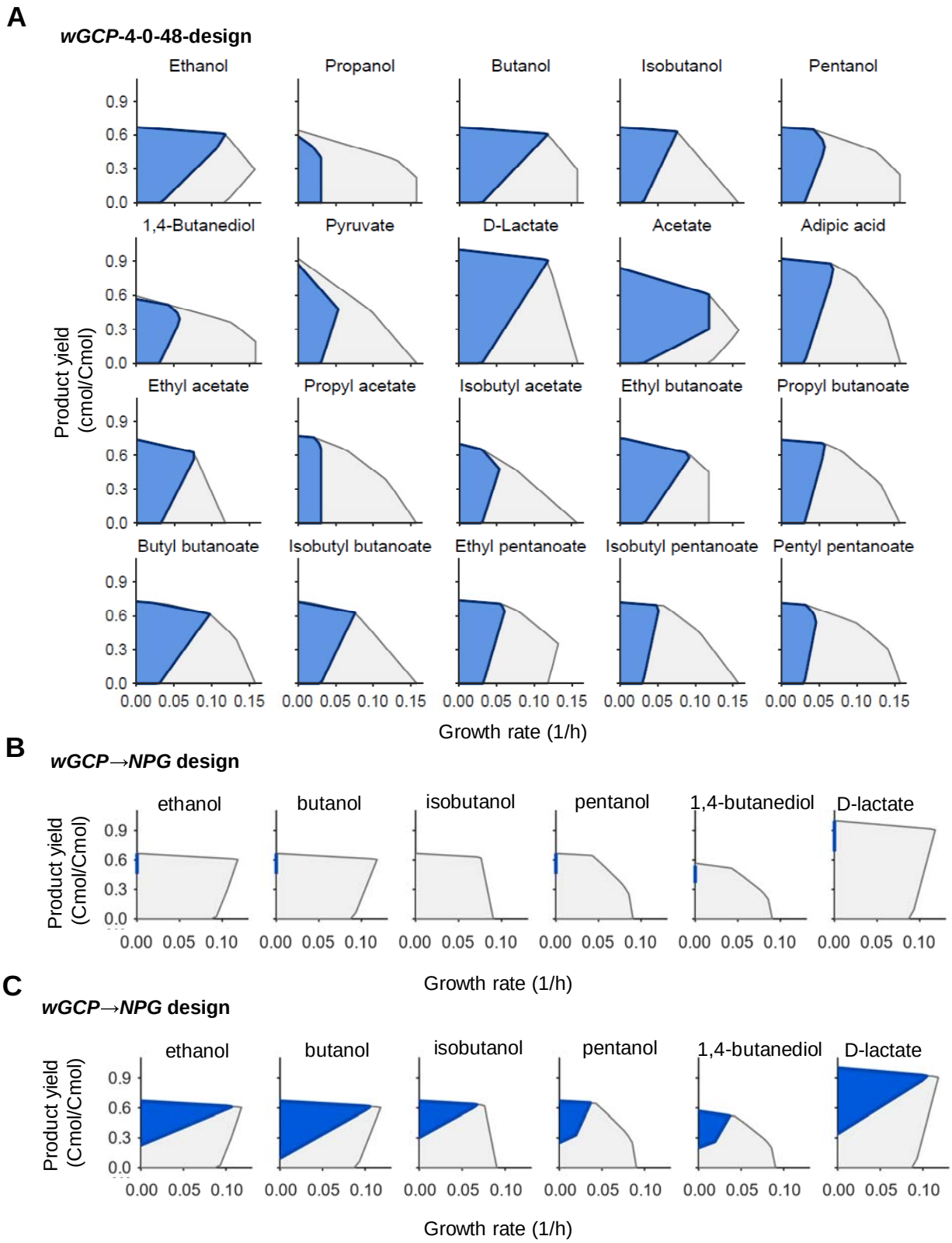

A
Objective value
0.9
0.6
0.3
0.0
Product type
alcohol
organic acid
ester
Ethanol
D-Lactate
Ethyl pentanoate
Ethyl acetate
Ethyl butanoate
Acetate
Pyruvate
Isobutyl acetate
Pentyl pentanoate
Pentanol
1,4-Butanediol
Propanol
Propyl acetate
Isobutanol
Butyl butanoate
Isobutyl butanoate
Propyl butanoate
Isobutyl pentanoate
Butanol
Adipic acid
Solution index
B
Deletion reactions
AI2K
I2FE2ST
GND
GLYAT
CBMKr
RPE
GHMT2r
PPM2
ALDD2y
PTA2
POR5
TKT2
PGCD
NADH17pp
FRD3
PPC
TPI
MDH
PSERT
PDH
THD2pp
PGI
PFL
FUM
ACALD
ALCD2x
ATPS4rpp
PTAr
LDH_D
0.0
0.2
0.4
0.6
0.8
1.0
Frequency
C
Compatible designs
60%
50%
40%
30%
20%
10%
0%
Ethanol
D-Lactate
Ethyl pentanoate
Butanol
Adipic acid
Propyl butanoate
Ethyl butanoate
Isobutyl pentanoate
Pentanol
1,4-Butanediol
Butyl butanoate
Ethyl acetate
Isobutanol
Isobutyl butanoate
Pentyl pentanoate
Acetate
Propanol
Pyruvate
Propyl acetate
Isobutyl acetate
D
Count
8
6
4
2
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Compatibility
E
Objective value
1
0.8
0.6
0.4
0.2
0
Isobutanol
Propyl butanoate
D-Lactate
Ethanol
Butanol
Adipic acid
Isobutyl pentanoate
Isobutyl butanoate
Ethyl pentanoate
Butyl butanoate
Ethyl butanoate
Ethyl acetate
Pentyl pentanoate
Pentanol
Isobutyl acetate
1,4-Butanediol
Pyruvate
Acetate
Propanol
Propyl acetate

Figure 7


A
wGCP-4-0-48-design
Ethanol
Propanol
Butanol
Isobutanol
Pentanol
1,4-Butanediol
Pyruvate
D-Lactate
Acetate
Adipic acid
Ethyl acetate
Propyl acetate
Isobutyl acetate
Ethyl butanoate
Propyl butanoate
Butyl butanoate
Isobutyl butanoate
Ethyl pentanoate
Isobutyl pentanoate
Pentyl pentanoate
Product yield (cmol/Cmol)
Growth rate (1/h)
B
wGCP→NPG design
ethanol
butanol
isobutanol
pentanol
1,4-butanediol
D-lactate
Product yield (Cmol/Cmol)
Growth rate (1/h)
C
wGCP→NPG design
ethanol
butanol
isobutanol
pentanol
1,4-butanediol
D-lactate
Product yield (Cmol/Cmol)
Growth rate (1/h)