

Harnessing Natural Modularity of Metabolism with Goal Attainment Optimization to Design a Modular Chassis Cell for Production of Diverse Chemicals

Sergio Garcia and Cong T. Trinh*

Cite This: *ACS Synth. Biol.* 2020, 9, 1665–1681

Read Online

ACCESS |



Metrics & More



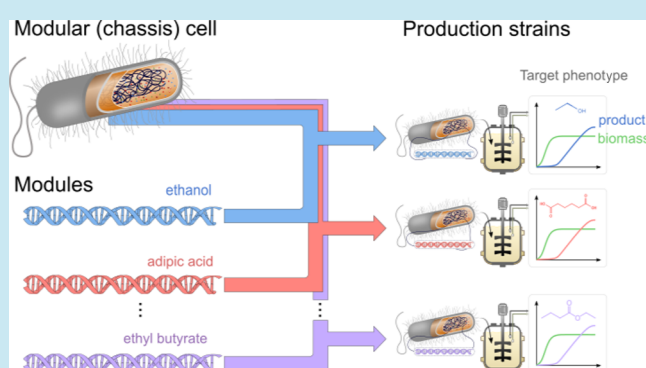
Article Recommendations



Supporting Information

ABSTRACT: Modular design is key to achieve efficient and robust systems across engineering disciplines. Modular design potentially offers advantages to engineer microbial systems for biocatalysis, bioremediation, and biosensing, overcoming the slow and costly design–build–test–learn cycles in the conventional cell engineering approach. These systems consist of a modular (chassis) cell compatible with exchangeable modules that enable programmed functions such as overproduction of a desirable chemical. We previously proposed a multiobjective optimization framework coupled with metabolic flux models to design modular cells and solved it using multiobjective evolutionary algorithms. However, such approach might not achieve solution optimality and hence limits design options for experimental implementation. In this study, we developed the goal attainment formulation compatible with optimization algorithms that guarantee solution optimality. We applied goal attainment to design an *Escherichia coli* modular cell capable of synthesizing all molecules in a biochemically diverse library at high yields and rates with only a few genetic manipulations. To elucidate modular organization of the designed cells, we developed a flux variance clustering (FVC) method by identifying reactions with high flux variance and clustering them to identify metabolic modules. Using FVC, we identified reaction usage patterns for different modules in the modular cell, revealing that its broad pathway compatibility is enabled by the natural modularity and flexible flux capacity of endogenous core metabolism. Overall, this study not only sheds light on modularity in metabolic networks from their topology and metabolic functions but also presents a useful synthetic biology toolbox to design modular cells with broad applications.

KEYWORDS: biocatalysis, modular design, modular cell, metabolic network modeling, multiobjective optimization, goal programming



Microbial metabolism can be engineered to produce a large space of molecules from renewable and sustainable feedstocks.¹ Currently, only a handful of fuels and chemicals out of the many possible molecules offered by nature are industrially produced by microbial conversion, mainly because the strain engineering process is too laborious and expensive.² To overcome this roadblock and produce a more diverse range of molecules requires innovative technologies for rapid and economically competitive strain engineering.^{1–3} The principles of modular design that have shown great success in traditional engineering disciplines can be adapted to construct modular cell biocatalysts in a plug-and-play fashion with minimal strain design–build–test–learn cycles.⁴

Multiobjective optimization is a powerful mathematical framework widely applied in engineering disciplines to tackle the optimal design of a complex system with multiple conflicting objectives.^{5,6} This framework has recently been used to design modular systems in conventional engineering⁷ and to explain the modularity of natural biological systems that enable cellular robustness and adaptability.^{8–12} Using multi-

objective optimization, microbial metabolism can be redirected to generate modular production strains that are systematically assembled from an engineered modular cell and exchangeable production modules, where each module synthesizes a target molecule.¹³ This modular cell design approach, known as ModCell2, uses the principles of mass balance and thermodynamics of biochemical reaction networks to predict metabolic fluxes upon genetic manipulations.^{13,14} Based on such flux predictions, a multiobjective optimization problem is then formulated and solved with a multiobjective evolutionary algorithm (MOEA)^{15,16} to yield a sample of the Pareto front (i.e., the set of optimal solutions to the problem with minimal

Received: December 18, 2019

Published: May 29, 2020



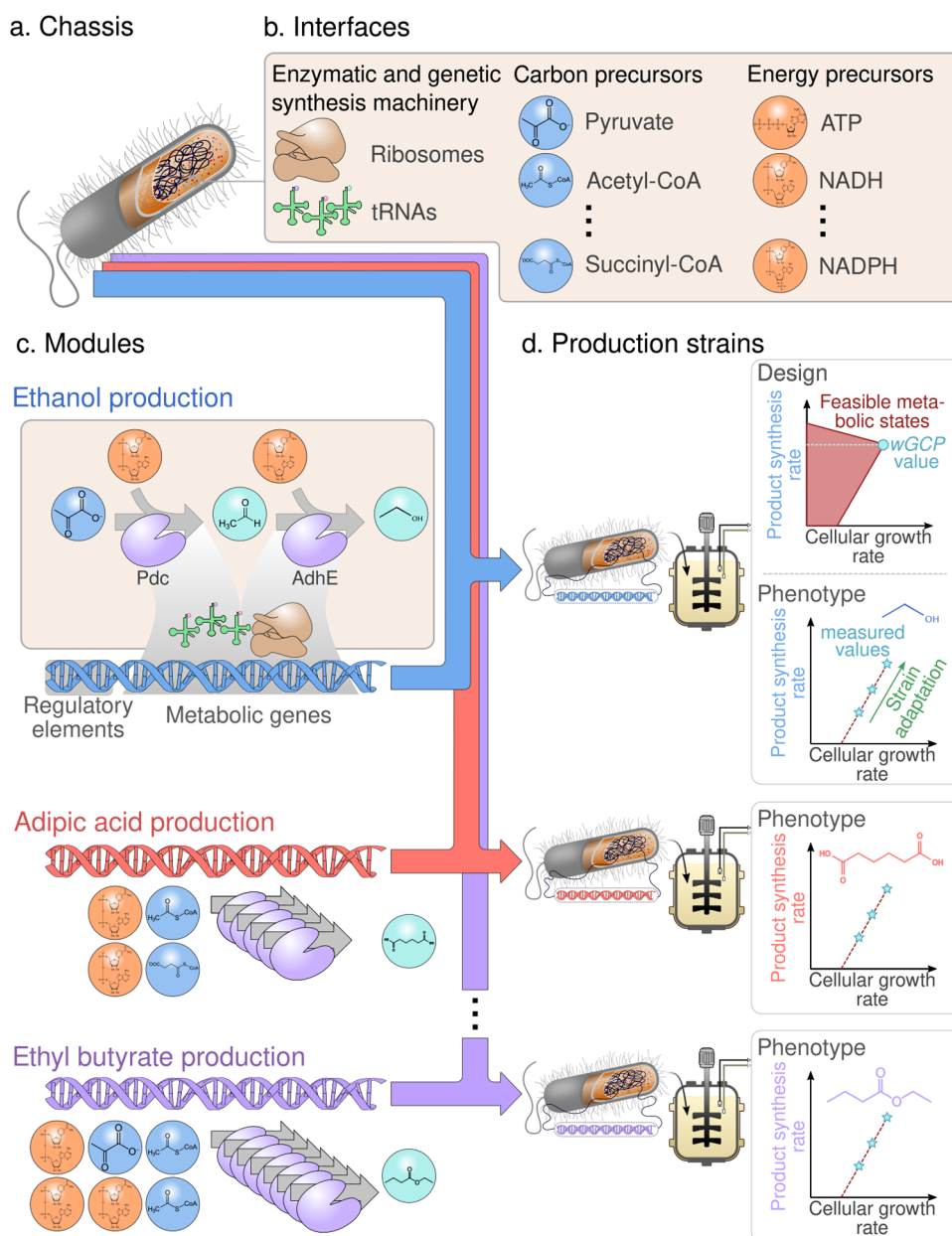


Figure 1. Principles of modular cell design. (a) Modular chassis cell. (b) Interfaces. (c) Exchangeable production modules. (d) Production strains. A modular cell is designed to provide the necessary precursors for biosynthesis pathway modules that are independently assembled with the modular cell to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

trade-offs among objectives) so that a designer can explore genetic manipulation targets for modular cell engineering.

In this study, we developed ModCell2-MILP, a ModCell2-based formulation to be compatible with mixed integer linear programming (MILP) algorithms. This framework presents a significant advancement from ModCell2 in solving the multiobjective strain design problem for modular cell engineering. Specifically, ModCell2-MILP is developed to (i) guarantee optimal solutions, (ii) completely enumerate alternative solutions of a target design, and (iii) describe practical engineering goals more directly (e.g., design of a modular cell where all production modules lead to a product yield above 50% of the theoretical maximum). By applying ModCell2-MILP to analyze the genome-scale metabolic network of *Escherichia coli*, we could identify a universal

modular cell that is compatible with a diverse class of production modules. To gain a mechanistic view into the modular organization of metabolic networks, we developed a flux variance clustering (FVC) method by identifying reactions with high flux variance and clustering them to identify metabolic modules. Using FVC, we found that broad pathway compatibility of the modular cell is facilitated by its natural modularity and flexible flux capacity of endogenous core metabolism. We anticipate ModCell2-MILP and FVC can serve as powerful tools for not only elucidating natural and synthetic metabolic modularity but also rationally designing modular cells for broad biotechnological applications in biocatalysis, bioremediation, and biosensing.

MATERIALS AND METHODS

Modular Cell Design. *Design Principles.* ModCell design enables rapid assembly of production strains with desirable phenotypes from a modular (chassis) cell.^{4,13,17} More specifically, a modular cell contains core metabolic pathways shared among exchangeable production modules (Figure 1a). The chassis interfaces with the modules through enzymatic and genetic synthesis machinery and precursor metabolites (Figure 1b). Modules contain auxiliary regulatory and metabolic pathways (Figure 1c) that enable a desired phenotype for optimal biosynthesis of a target molecule, for example, *weak growth coupled to product formation* (wGCP), where a positive correlation between growth and product synthesis rates is enforced (Figure 1d).^{13,18,19} The wGCP phenotype is useful because it enables rapid pathway optimization by adaptive laboratory evolution^{20,21} or high-throughput genetic library selection.²² The design objective phenotypes are determined from cellular growth and product synthesis rates based on steady-state stoichiometric metabolic models.²³ A modular cell is said to be *compatible* with a module if the design objective of the resulting production strain is above a specified threshold. The different biochemical natures of production modules to synthesize target metabolites can make the design objectives compete with each other and also the cellular objectives (e.g., biomass formation) compete with the engineering objectives (e.g., product formation), turning the ModCell design problem into a multiobjective and multilevel optimization problem.

Multiobjective Optimization Formulation. The modular cell design problem is stated as a general multiobjective optimization problem of the form

$$\max_{\mathbf{x}} \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_k(\mathbf{x}))^T \quad \text{s. t. } \mathbf{x} \in \mathbf{X} \quad (1)$$

where f_k is the desirable phenotype for production module k , \mathbf{x} are the problem variables including binary design variables corresponding to genetic manipulations, and \mathbf{X} is the set of constraints including mass balance of metabolism. Optimal solutions for the multiobjective optimization problem 1 are defined using the concept of domination: A vector $\mathbf{a} = (a_1, \dots, a_K)^T$ dominates another vector $\mathbf{b} = (b_1, \dots, b_K)^T$, denoted as $\mathbf{a} < \mathbf{b}$, if and only if $a_i \geq b_i \forall i \in \{1, 2, \dots, K\}$ and $a_i \neq b_i$ for at least one i . A feasible solution $\mathbf{x}^* \in \mathbf{X}$ of the multiobjective optimization problem is called a Pareto optimal solution if and only if there does not exist a vector $\mathbf{x}' \in \mathbf{X}$ such that $\mathbf{F}(\mathbf{x}') < \mathbf{F}(\mathbf{x}^*)$. The set of all Pareto optimal solutions is called the Pareto set:

$$\mathbf{PS} := \{\mathbf{x} \in \mathbf{X}: \nexists \mathbf{x}' \in \mathbf{X}, \mathbf{F}(\mathbf{x}') < \mathbf{F}(\mathbf{x})\} \quad (2)$$

The projection of the Pareto set in the objective space is denoted as the Pareto front:

$$\mathbf{PF} := \{\mathbf{F}(\mathbf{x}): \mathbf{x} \in \mathbf{PS}\} \quad (3)$$

Different feasible points in \mathbf{PS} (i.e., different genetic manipulations) that map to a single point in \mathbf{PF} (i.e., the same phenotype) are denoted *alternative solutions*.

The design variables \mathbf{x} in ModCell2 correspond to chassis reaction deletions, which remove undesired metabolic functions, and module reaction insertions, which allow identification of optimal module configurations without extensive prior knowledge of the product synthesis pathway. The constraint set \mathbf{X} is comprised of two types: (i) flux simulation constraints (e.g., mass balance, reaction reversibility, and flux bound) that allow prediction of fluxes in the

design objectives upon genetic manipulations and (ii) implementation constraints that involve the maximum number of reaction deletions in the chassis (denoted by α) and the maximum number of module reaction insertions per module (denoted by β). The following sections describe the problem formulation in detail using the definitions compiled in the **Definitions** section.

Design Objectives. Design objectives, f_k , that correspond to specific metabolic phenotypes within the space of feasible steady-state reaction fluxes, Π_{km} , of production network k (i.e., the combination of the chassis network with the production module k) and metabolic state m are defined as follows:

$$\Pi_{km}(e_{jk}) := \{v_{jkm} \in \mathbb{R}: \quad (4)$$

$$\sum_{j \in \mathcal{J}_k} S_{ijk} v_{jkm} = 0 \quad \forall i \in \mathcal{I}_k \quad (5)$$

$$l_{jkm} e_{jk} \leq v_{jkm} \leq e_{jk} u_{jkm} \quad \forall j \in \mathcal{J}_k \quad (6)$$

Here, v_{jkm} is the rate (mmol/(g CDW/h)) of reaction j in production network k under metabolic state m . Constraint 5 enforces mass balance for all metabolites according to reaction stoichiometry given by the stoichiometric coefficients S_{ijk} and constraint 6 imposes bounds, l_{jkm} and u_{jkm} , for the metabolic fluxes according to reaction reversibility, experimentally measured values, and specified metabolic state. The binary variable, e_{jk} , is used in the overall optimization problem to indicate whether reaction j in production network k is removed and thus cannot carry any flux. Two metabolic states m are considered, growth and nongrowth, denoted μ and $\bar{\mu}$, respectively. These states are differentiated by their flux bounds, l_{jkm} and u_{jkm} . For growth state, the lower bound of the biomass formation reaction that represents cell division, $v_{Xk\mu}$ is set to a minimum value of γ , that is, $l_{Xk\mu} = \gamma$ ($\forall k \in \mathcal{K}$), while there is no upper limit to growth, that is, $u_{Xk\mu} = \infty$ ($\forall k \in \mathcal{K}$). On the other hand, for the nongrowth state both bounds are set to 0, that is, $l_{Xk\bar{\mu}} = 0$ and $u_{Xk\bar{\mu}} = 0$ ($\forall k \in \mathcal{K}$).

Given the feasible metabolic flux space, Π_{km} , the following design objectives, based on the product synthesis rate reaction, $v_{Pk\mu}$, are of interest:

$$f_k^{\text{wGCP}} = \frac{v_{Pk\mu}}{v_{Pk\mu}^{\max}} \in [0, 1], \quad \forall k \in \mathcal{K} \quad (7)$$

$$f_k^{\text{lsGCP}} = b_{\mu} \frac{v_{Pk\mu}}{v_{Pk\mu}^{\max}} + b_{\bar{\mu}} \frac{v_{Pk\bar{\mu}}}{v_{Pk\bar{\mu}}^{\max}} \in [0, b_{\mu} + b_{\bar{\mu}}], \quad \forall k \in \mathcal{K} \quad (8)$$

$$f_k^{\text{NGP}} = \frac{v_{Pk\bar{\mu}}}{v_{Pk\bar{\mu}}^{\max}} \in [0, 1], \quad \forall k \in \mathcal{K} \quad (9)$$

The product synthesis fluxes, including $v_{Pk\mu}$, $v_{Pk\mu}^{\max}$, $v_{Pk\bar{\mu}}$, and $v_{Pk\bar{\mu}}^{\max}$, are computed by solving the following linear programming problems:

$$v_{Pk\mu} \in \arg \max \{v_{Xk\mu} - \epsilon v_{Pk\mu}: v_{k\mu} \in \Pi_{k\mu}(e_{jk})\} \quad (10)$$

$$v_{Pk\mu}^{\max} \in \arg \max \{v_{Pk\mu}: v_{k\mu} \in \Pi_{k\mu}(e_{jk} = 1, \forall j \in \mathcal{J}_k)\} \quad (11)$$

$$v_{Pk\bar{\mu}} \in \arg \min \{v_{Pk\bar{\mu}}: v_{k\bar{\mu}} \in \Pi_{k\bar{\mu}}(e_{jk})\} \quad (12)$$

$$v_{Pk\bar{\mu}}^{\max} \in \arg \max \{v_{Pk\bar{\mu}}: v_{k\bar{\mu}} \in \Pi_{k\bar{\mu}}(e_{jk} = 1, \forall j \in \mathcal{J}_k)\} \quad (13)$$

The maximum product synthesis fluxes 11 and 13 used for objective scaling are only calculated once by not using any deleted reactions ($e_{jk} = 1$), while the target phenotype fluxes 10 and 12 are functions of the deleted reactions e_{jk} . The design objectives, wGCP (eq 7), lsGCP (eq 8), and NGP (eq 9), were previously proposed¹³ and briefly described here. The weak growth coupled to product formation objective (wGCP, eq 7) seeks to maximize the minimum product rate at the maximum cellular growth, which is accomplished by a titled objective function²⁴ (eq 10). The linearized strong growth coupled to product formation (lsGCP, eq 8) objective seeks to maximize the minimum product synthesis rate during the nongrowth state, $v_{pk\bar{m}}$, in addition to the goal of wGCP. In f_k^{lsGCP} , b_μ and $b_{\bar{\mu}}$ are the weights on the growth and nongrowth objectives, respectively. Finally, the nongrowth production (NGP, eq 9) objective seeks to optimize the minimum product synthesis rate during the nongrowth state.

Design Constraints. All the constraints of the modular cell design problem are gathered as follows:

$$\Omega := \{f'_k \in \mathbb{R}, y_j, z_{jk}, d_{jk}, w_k, e_{jk} \in \{0, 1\}\} \quad (14)$$

$$\sum_{j \in C} (1 - y_j) \leq \alpha \quad (15)$$

$$\sum_{j \in C - N_k} z_{jk} \leq \beta_k \quad \forall k \in \mathcal{K} \quad (16)$$

$$z_{jk} \leq 1 - y_j \quad \forall j \in C - N_k, k \in \mathcal{K} \quad (17)$$

$$d_{jk} = y_j \vee z_{jk} \quad \forall j \in C, k \in \mathcal{K} \quad (18)$$

$$f'_k = f_k w_k \quad \forall k \in \mathcal{K} \quad (19)$$

$$e_{jk} = (d_{jk} \wedge w_k) \vee \neg w_k \quad \forall j \in C, k \in \mathcal{K} \quad (20)$$

$$w_k \leq M^w f_k \quad \forall k \in \mathcal{K} \quad (21)$$

$$v_{pkm} \in \Psi_{km}(e_{jk}) \quad \forall k \in \mathcal{K}, m \in \mathcal{M} \quad (22)$$

Constraints 15–18 are formulated for practical limitations and features of the modular cell. Specifically, the two variables that represent design choices for genetic manipulations include (i) y_j , which takes a value of 0 if reaction j is deleted in the chassis (and consequently in all production networks) and 1 otherwise, and (ii) z_{jk} , which takes a value of 1 if reaction j is inserted in production network k . The maximum number of reaction deletions is limited by α through constraint 15, while the maximum number of module reactions in each module β_k is imposed by 16. Constraint 16 excludes noncandidate reactions N_k (since $j \in C - N_k$) so that module reactions can be fixed (i.e., $z_{jk} = 1$), according to problem-specific knowledge. Constraint 17 ensures that only reactions deleted in the chassis can be inserted back to the modules. Constraint 18 indicates that reaction j is deleted in production network k if the reaction is deleted in the chassis and not added as a module reaction. The designer can gradually increase α and β_k to obtain solutions with higher performance.

Constraints 19–21 are introduced for modeling purposes. The indicator variable, w_k , is introduced to allow for certain production networks to be ignored from the final solution. Without w_k , the whole multiobjective problem becomes infeasible if a set of deletions renders one of the production networks infeasible (e.g., its minimum growth rate cannot be

accomplished). However, in practice it is acceptable for some modules not to work with the chassis cell. If $w_k = 0$, the objective value $f'_k = 0$ (eq 19) and reaction deletions do not apply to network k since $e_{jk} = 1$ 20; if $w_k = 1$, $f'_k = f_k$, and $e_{jk} = d_{jk}$, where f_k is any of the design objectives presented earlier, eqs 7–9. The use of w_k is likely to introduce symmetry (i.e., alternative integer solutions with no practical meaning) due to cases where $f_k = 0$ for a given k while the associated production network remains feasible, allowing w_k to take a value of 0 or 1. This symmetry is removed by enforcing w_k to be 0 if $f_k = 0$, eq 21.

Finally, constraint 22 indicates that the fluxes featured in the design objectives, v_{pkm} , are contained in the polytope Ψ_{km} . The space of v_{pkm} is originally defined as an optimization problem, eqs 10–13, thus representing a nonlinear constraint and turning the ModCell design problem into a bilevel optimization problem. These inner optimization problems are linearized, leading to Ψ_{km} as described in the following sections.

Linearization of Logical Expressions. The logical expressions in Ω are replaced by the following linear constraints in the final problem formulation.

$d_{jk} = y_j \vee z_{jk}$ corresponds to

$$d_{jk} \leq y_j + z_{jk} \quad (23)$$

$$d_{jk} \geq y_j \quad (24)$$

$$d_{jk} \geq z_{jk} \quad (25)$$

$$0 \leq d_{jk} \leq 1 \quad (26)$$

$f'_k = f_k w_k$ corresponds to

$$f'_k \leq w_k M^{\text{obj}} \quad (27)$$

$$f'_k \leq f_k - (1 - w_k) M^{\text{obj}} \quad (28)$$

$$f'_k \leq f_k \quad (29)$$

$$0 \leq f'_k \leq M^{\text{obj}} \quad (30)$$

$e_{jk} = (d_{jk} \wedge w_k) \vee \neg w_k$, given $r_{jk} = d_{jk} \wedge w_k$, corresponds to

$$e_{jk} = r_{jk} + 1 - w_k \quad (31)$$

$$r_{jk} \leq w_k \quad (32)$$

$$r_{jk} \leq d_{jk} \quad (33)$$

$$r_{jk} \geq w_k + d_{jk} - 1 \quad (34)$$

$$0 \leq r_{jk} \leq 1 \quad (35)$$

Linearization of Inner Optimization Problems. Nonlinear constraints expressed as linear programming problems can be linearized using basic mathematical programming theory. Consider the following canonical linear program, with primal variables $\mathbf{x} \in \mathbb{R}^n$ and its dual variables $\mathbf{u} \in \mathbb{R}^m$:

$$\max\{\mathbf{c}^T \mathbf{x} : \mathbf{A} \mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\} \quad (36)$$

$$\min\{\mathbf{b}^T \mathbf{u} : \mathbf{A}^T \mathbf{u} \geq \mathbf{c}, \mathbf{u} \geq \mathbf{0}\} \quad (37)$$

The strong duality theorem states that the objective functions of primal 36 and dual 37 are equal at their optima, $\mathbf{c}^T \mathbf{x}^* =$

$\mathbf{b}^T \mathbf{y}^*$. Thus, the optimal solution to the primal problem is described by the following linear constraints:

$$\mathbf{x}^* \in \{\mathbf{x} \in \mathbb{R}^n\} \quad (38)$$

$$\mathbf{A}\mathbf{x} \leq \mathbf{b} \quad (39)$$

$$\mathbf{A}^T \mathbf{u} \geq \mathbf{c} \quad (40)$$

$$\mathbf{c}^T \mathbf{x} = \mathbf{b}^T \mathbf{u} \quad (41)$$

$$\mathbf{x}, \mathbf{u} \geq \mathbf{0} \quad (42)$$

Using the strong duality theorem as presented by Maranas and Zomorodi,²⁵ the inner optimization problems 22 are linearized as follows:

$$\Psi_{km}(e_{jk}) := \{v_{jkm} \in \mathbb{R}\} \quad (43)$$

$$\sum_{j \in \mathcal{I}_k} S_{ijk} v_{jkm} = 0 \quad \forall i \in \mathcal{I}_k \quad (44)$$

$$l_{jkm} e_{jk} \leq v_{jkm} \leq e_{jk} u_{jkm} \quad \forall j \in \mathcal{J}_k \quad (45)$$

$$\sum_{i \in \mathcal{I}_k} \lambda_{ikm} S_{ijk} - \mu_{jkm}^l + \mu_{jkm}^u = c_{jkm} \quad \forall j \in \mathcal{J}_k \quad (46)$$

$$\lambda_{ikm} \in \mathbb{R} \quad \forall i \in \mathcal{I}_k \quad (47)$$

$$0 \leq \mu_{jkm}^l \leq M \quad \forall j \in \mathcal{J}_k \quad (48)$$

$$0 \leq \mu_{jkm}^u \leq M \quad \forall j \in \mathcal{J}_k \quad (49)$$

$$\sum_{j \in \mathcal{I}_k} c_{jkm} v_{jkm} = - \sum_{j \in \mathcal{J}_k - C} (l_{jkm} \mu_{jkm}^l) + \sum_{j \in \mathcal{J}_k - C} (u_{jkm} \mu_{jkm}^u) - \sum_{j \in C} (l_{jkm} p_{jkm}^l) + \sum_{j \in C} (u_{jkm} p_{jkm}^u) \quad (50)$$

$$p_{jkm}^l \leq e_{jk} M \quad \forall j \in C \quad (51)$$

$$\mu_{jkm}^l - (1 - e_{jk}) M \leq p_{jkm}^l \leq \mu_{jkm}^l \quad \forall j \in C \quad (52)$$

$$0 \leq p_{jkm}^l \leq M \quad \forall j \in C \quad (53)$$

$$p_{jkm}^u \leq e_{jk} M \quad \forall j \in C \quad (54)$$

$$\mu_{jkm}^u - (1 - e_{jk}) M \leq p_{jkm}^u \leq \mu_{jkm}^u \quad \forall j \in C \quad (55)$$

$$0 \leq p_{jkm}^u \leq M \quad \forall j \in C \quad (56)$$

Constraints 44 and 45 correspond to the primal metabolic network problem and were introduced earlier in Π_{km} . Constraints 46–49 correspond to the dual problem. We use the dual variables, λ_{ikm} , for the primal mass balance constraints 44, together with μ_{jkm}^l and μ_{jkm}^u for the primal flux bound inequalities 45 involving lower and upper reaction bounds, respectively. Constraints 47–49 emphasize the domain of the dual variables, with M being a large value above the expected value of any dual variable. Constraints 50–56 correspond to the strong duality equality. The left-hand side of the strong duality equality 50 features the objectives presented in eq 10 for $m = \mu$ and eq 12 for $m = \bar{\mu}$. On the right-hand side, products of binary and continuous variables appear, thus

requiring linearization variables p_{jkm}^l and p_{jkm}^u . Constraints 51–56 ensure that $p_{jkm}^l = e_{jk} \mu_{jkm}^l$ and $p_{jkm}^u = e_{jk} \mu_{jkm}^u$.

Conversion of a Multiobjective Problem into a Single-Objective Problem. The multiobjective optimization problem 1 is now described entirely in terms of linear constraints through Ω . However, to make the formulation compatible with MILP solver algorithms, the objective function vector, \mathbf{f}' , must be expressed as a scalar. To accomplish this without loss of relevant information, we employed blended and goal attainment formulations.²⁶

Blended Formulation. In the blended formulation, all objectives are summed as follows:

$$\max \sum_{k \in \mathcal{K}} a_k f'_k \quad \text{s. t. } f'_k \in \Omega \quad (57)$$

where 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. The blended formulation always provides Pareto optimal solutions as long as $a_k > 0$ ($\forall k \in \mathcal{K}$). 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. In the goal attainment problem, a target value is defined for each objective:

$$\min \sum_{k \in \mathcal{K}} (a_k^+ \delta_k^+ + a_k^- \delta_k^-) \quad (58)$$

such that

$$f'_k + \delta_k^+ - \delta_k^- = g_k \quad \forall k \in \mathcal{K} \quad (59)$$

$$\delta_k^+, \delta_k^- \geq 0 \quad \forall k \in \mathcal{K} \quad (60)$$

$$f'_k \in \Omega \quad (61)$$

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 general, when it is important to meet the target value without exceeding it, we set $a_k^+ = a_k^- = 1$; however, when the design objective is required to be greater or equal than the target value, we set $a_k^+ = 1$ and $a_k^- = 0$, effectively converting eq 59 into $f'_k + \delta_k^+ \geq g_k$. Solutions to the goal attainment problem are not guaranteed to be Pareto optimal, even if all demands g_k are met. To address this issue, the blended problem 57 can be solved where the objectives are constrained to be equal or greater than the values found by solving the goal attainment problem. In practice, the goal attainment formulation corresponds to the identification of the modular cell *compatible* with the largest number of modules. Here, a module k is said to be *compatible* if $f'_k \geq g_k$.

Implementation. Metabolic Models. We used two parent models from which production networks were built, including (i) a core metabolic model of *E. coli*¹⁷ to develop the ModCell2-MILP algorithm and compare with previous ModCell2 results¹³ and (ii) the iML1515 genome-scale metabolic model of *E. coli*²⁷ for biosynthesis of a library of endogenous and heterologous metabolites, including 4 organic acids, 6 alcohols, and 10 esters (Figure S1).^{28–37} These models were configured as in the previous ModCell2 study.¹³ Briefly, 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/(g CDW/h). When using the genome-

scale model iML1515 to simulate wGCP designs, only the commonly observed fermentative products (acetate, CO₂, ethanol, formate, lactate, succinate) were allowed for secretion as described elsewhere.³⁸

ModCell2-MILP Simulations. ModCell2-MILP was implemented using Pyomo,³⁹ an algebraic modeling language embedded in the Python programming language. All simulations were performed on a computer with an Intel Core i7-3770 processor, 32 GB of random access memory, and the Arch Linux operative system. The implementation and scripts used to generate the results of this manuscript are available as part of the ModCell2 package via [Supporting Information 2](#) and <https://github.com/trinhlab/modcell2>.

Optimization Solver Configuration. The Pyomo³⁹ implementation of ModCell2-MILP was solved with IBM Ilog Cplex 12.8.0. To avoid incorrect solutions associated with numerical issues the following Cplex parameters were changed from their default values: (i) *numerical emphasis* was set to “true”, (ii) *integrality tolerance* was lowered to 10⁻⁷, and (iii) the *MIP pool relative gap* was increased to 10⁻⁴ for enumerating alternative solutions. Alternative solutions were enumerated using the Cplex “populate” procedure.

Analysis methods. Reference Flux Distribution. The reference flux distribution, $\frac{v_{jk}^*}{|v_{jk}^*|}$, is determined by solving the following quadratic program based on the parsimonious enzyme usage hypothesis.^{40,41}

$$\min_{v_{jk}} \sum_{j \in \mathcal{J}_k} v_{jk}^2 \quad (62)$$

s.t.

$$\sum_{j \in \mathcal{J}_k} S_{ijk} v_{jk} = 0 \quad \forall i \in \mathcal{I}_k \quad (63)$$

$$l_{jk} \leq v_{jk} \leq u_{jk} \quad \forall j \in \mathcal{J}_k \quad (64)$$

$$v_{Xk} = \text{MaxDesignBio} \quad (65)$$

Constraint 63 corresponds to mass balance for the metabolic network. Constraint 64 corresponds to reaction bounds, including reaction deletions found in the modular cell design problem. Constraint 65 fixes the biomass formation rate, v_{Xk} , to the maximum reachable by the design. This value (MaxDesignBio) is obtained by maximizing v_{Xk} subject to eqs 63 and 64. The reference flux distribution $\frac{v_{jk}^*}{|v_{jk}^*|}$ represents the desired metabolic state of a wGCP designed production network. This distribution, if feasible, is unique because the convex optimization problem is formulated with a positive definite quadratic objective function (see Theorem 16.4 in Nocedal and Wright⁴²).

Flux Variance Clustering. Flux variance clustering (FVC) seeks to identify and group reactions that exhibit high flux changes under different conditions. Reactions with high flux variance can be considered as metabolic interfaces between core metabolic processes and pathway modules. In our study, each condition corresponds to a different product synthesis phenotype. FVC implementation entails three steps. First, flux distributions are simulated for each condition and standard deviations for each reaction are calculated. Then, a standard deviation threshold is set to select the reactions with highest flux changes across conditions. Finally, these selected reactions are clustered to identify patterns that repeat under specific

conditions. The filtering threshold is set to capture the top reactions with the most change while maintaining a sufficiently small list that is biologically meaningful. In our study, we chose an *ad hoc* value that captures well-known reactions in *E. coli* central metabolism. To cluster the selected reactions, only flux magnitudes, not directionality, were considered in our study. Further, each reaction flux was normalized by the maximum value of that same reaction across all production networks. Clustering was performed using the method *clustermap()* with default clustering-related parameters from the Python library Seaborn 0.9.

Flux Sampling. To determine an ensemble of flux distributions for a production network, we used the ACHR algorithm⁴³ in the COBRA toolbox.⁴⁴ Constraints for flux sampling simulation include the reaction deletions and module reactions found in the ModCell design problem solution, a fixed substrate uptake rate of -10 mmol glucose/(g CDW/h), and a minimum product synthesis flux of 50% of its maximum value.

Metabolic Map Drawing. Drawings of metabolic map were performed using the Escher⁴⁵ tool (<https://escher.github.io>) that produces *svg* files. Coloring, highlighting candidate reactions, and other systematic adjustments of metabolic maps were done with the Python-based *lxml* module. Additional editing for visual enhancement was done with the Inkscape software.

RESULTS AND DISCUSSION

Performance and Solution Time Optimization of ModCell2-MILP. *ModCell2-MILP Not Only Can Reproduce the Results of the Original ModCell2 Formulation but Also Can Find More Alternative Solutions.* To evaluate ModCell2-MILP, we compared its performance with the previously developed ModCell2 platform¹³ that solves the optimization problem with multiobjective evolutionary algorithms (MOEAs). As a basis of comparison, we used the same *E. coli* core metabolic model, maximum number of deletion reactions α , and maximum number of module reactions β_k for both ModCell2 and ModCell2-MILP. Due to fundamental differences in problem formulations for MOEA and MILP, we used the lsGCP design objective for ModCell2-MILP with multiple weighting factors, a_k , specifically selected to reproduce previous results, in the blended formulation and the sGCP design objective for ModCell2 ([Supporting Information 3](#)). The results showed that ModCell2-MILP could generate the same Pareto optimal designs like ModCell2. In addition, ModCell2-MILP enumerated a larger number of alternative solutions than ModCell2. For example, the design named *sGCP-5-0-6* generated by ModCell2 had 3 alternative solutions, while ModCell2-MILP found 8 alternative solutions. By increasing α to 8 and β to 2, we could identify a utopia design (i.e., one solution with the maximum value for all objectives) with 192 alternative solutions, expanding the possibilities for experimental implementation.

Tuning MILP Formulations Significantly Improves Solution Times. We considered three techniques that can improve solution times of ModCell2-MILP:

- Fixing the network feasibility indicator w_k . If all modules are expected to be compatible with a final ModCell design (i.e., $f_k > 0, \forall k \in \mathcal{K}$), w_k is set to be 1 for all $k \in \mathcal{K}$ to avoid computational efforts in finding nonoptimal feasible solutions.

- (ii) Flux bound tightening. Constraints of the form $e_{jkm}l_{jkm} \leq v_{jkm} \leq e_{jkm}u_{jkm}$ are known to result in weak linear relaxations, that is, feasible values of v_{jkm} are far from their bounds l_{jkm} and u_{jkm} . To tighten the formulation by making continuous relaxations closer to the feasible integer solution, smaller values of u_{jkm} and l_{jkm} are determined by solving a series of linear programs that maximize and minimize each flux v_{jkm} in the parent production networks $\Pi_{km}(e_{jk} = 1, \forall j \in \mathcal{J}_k)$.
- (iii) Benders decomposition. ModCell2-MILP has a separable structure compatible with Benders decomposition^{46,47} that creates a master problem, using binary variables and associated constraints 15–21, and subproblems for each production network $\Psi_{km}(e_{jk})$ with fixed binary variables. This decomposition implementation is automatically done by Cplex 12.8.

We evaluated these three techniques for tuning MILP formulations and used the core *E. coli* model¹³ for the benchmark study. The results showed that flux bound tightening, fixed w_k , and Benders decomposition could reduce the solution time to find solutions by 50%, 80%, and 95%, respectively (Table 1). By combining these techniques, the

Table 1. Solution Time Reduction by Tuning the ModCell2-MILP Formulation^a

feasibility indicator, w_k , fixed	Benders decomposition	bounds tightened	solution time (s)
no	no	no	63.3 ± 16.9
no	no	yes	32.5 ± 10.2
no	yes	no	3.6 ± 0.1
no	yes	yes	3.4 ± 0.4
yes	no	no	13.8 ± 2.7
yes	no	yes	11.9 ± 1.7
yes	yes	no	2.7 ± 0.3
yes	yes	yes	2.8 ± 0.1

^aFixed network indicator means $w_k = 1, \forall k \in \mathcal{K}$. The simulations were performed in triplicate.

solution time was shortened by 96% from 63.3 to 2.8 s. In subsequent studies, we used these three tuning techniques to solve the ModCell design problem unless otherwise noted.

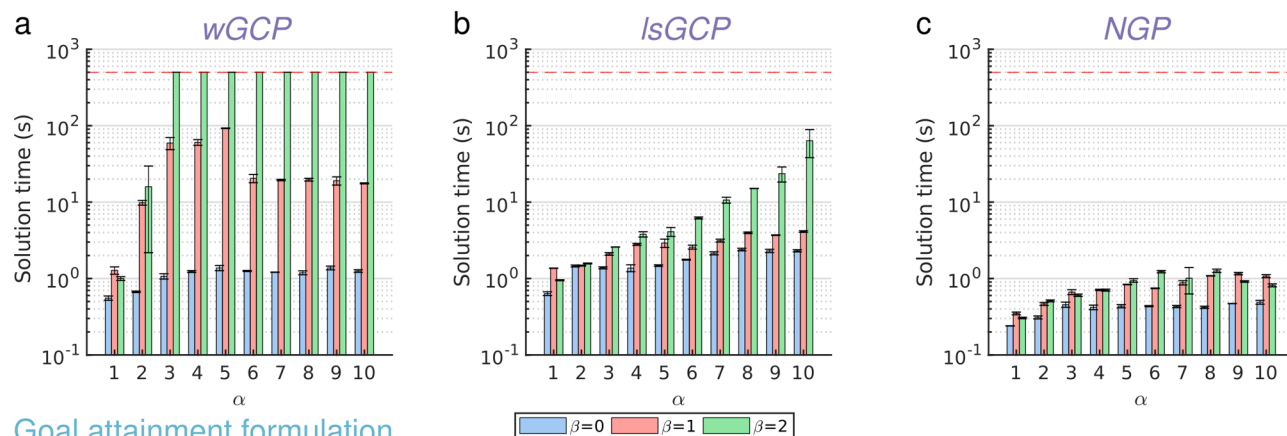
Choice of Design Parameters Affect Solution Time. In designing a modular cell with ModCell2-MILP, the designer needs to specify the formulation type (i.e., blended or goal attainment formulation), the target phenotype (e.g., wGCP, lsGCP, and NGP), and the limits of deletion reactions (α) and module-specific reactions (β_k). We evaluated the impact of these parameters on solution time using the *E. coli* core model (Figure 2). Regardless of the formulation type, increasing α and β led to harder problems and hence required more solution time due to the exponentially increasing number of feasible solutions as expected. The goal attainment formulation took longer time to solve for the lsGCP and NGP design objectives, but about the same time for the wGCP design objective. Interestingly, the overall difficulty of wGCP is higher than that of lsGCP in both the blended and goal attainment formulations, despite lsGCP having approximately twice the number of constraints. Furthermore, the NGP design objective could be solved most quickly, likely due to the narrower design space associated with the no-growth associated production of target metabolites.

Design of a Universal Modular Cell for a Genome-Scale Metabolic Model of *E. coli*. Reduction of the Candidate Reaction Deletion Set Enables ModCell2-MILP to Find Modular Cell Designs for a Large-Scale Metabolic Network. Finding genetic modifications toward a desired phenotype using mathematical optimization for large-scale metabolic networks is a computationally expensive task, due to the combinatorial search space spanned by a large number of reaction deletion candidates in the network.^{24,48} Preprocessing of metabolic networks to reduce reaction candidates is not only critical but also practical for experimental implementation. The set of reaction candidates in the iML1515 *E. coli* model²⁷ was reduced from 2712 to 276 by ModCell2.¹³ Using this model and the wGCP objective, an *E. coli* modular cell was then identified to be compatible with 17 out of 20 products with requirement of only 4 reaction deletions.¹³ Since MOEA implemented in ModCell2 does not guarantee optimality, here we aimed to evaluate the capability of ModCell2-MILP for handling a large-scale metabolic network and identifying the Pareto optimality and potential alternative solutions.

We applied ModCell2-MILP to analyze the same iML1515 model with a set of 20 products using the same design parameters (i.e., α and β_k) and the blended formulation with all objective weights $a_k = 1$. The simulation shows that ModCell2-MILP could not solve the ModCell design problem to optimality over 2 days of run time, likely due to the large number of candidate deletion reactions still present in the genome-scale model. Currently the best MILP solution algorithms do not scale well with parallel computing. In order to obtain solutions within an acceptable time, the set of candidate reactions must be further reduced. Since only a small subset of all metabolic reactions in genome-scale models tend to be deleted by strain design algorithms,^{13,24,49} we used a pool of wGCP designs with $\alpha = 4, 5, 6$ and $\beta = 0, 1$ reported with ModCell2¹³ to identify relevant deletion candidates. From a set of 601 designs found by ModCell2, only 33 out of 276 reaction deletion candidates were used at least once. Hence, these 33 reactions were used to create a new, computationally tractable set of reaction candidates. This new set contains reactions mostly from the well-characterized central metabolic pathways (Figure 3a), while the original set includes reactions in peripheral pathways that lead to biomass synthesis. Interestingly, within these 33 reaction candidates, only a few are used in most designs (Figure 3b), highlighting the importance of their removal in growth-coupled production phenotypes. Reactions with high deletion frequencies mainly occur in high-flux central metabolic pathways (Figure 3c), closely associated with cellular energetics and carbon precursors that interface with the production modules (Figure 3d).

Using the reduced reaction deletion candidate set, ModCell2-MILP could find an optimal solution in ~30 min and enumerated all optimal solutions in ~8 h. All the optimal solutions found by ModCell2-MILP in this case were in agreement with those previously found by ModCell2.¹³ It should be noted that a reduced deletion candidate set can be identified for any metabolic model and target products by using previous designs identified with either ModCell2 or single-phenotype strain design algorithms.⁵⁰ This heuristic for reaction deletion candidate selection might affect design optimality to the extent that it omits relevant reactions. However, using a sufficiently informative pool of designs from

Blended formulation



Goal attainment formulation

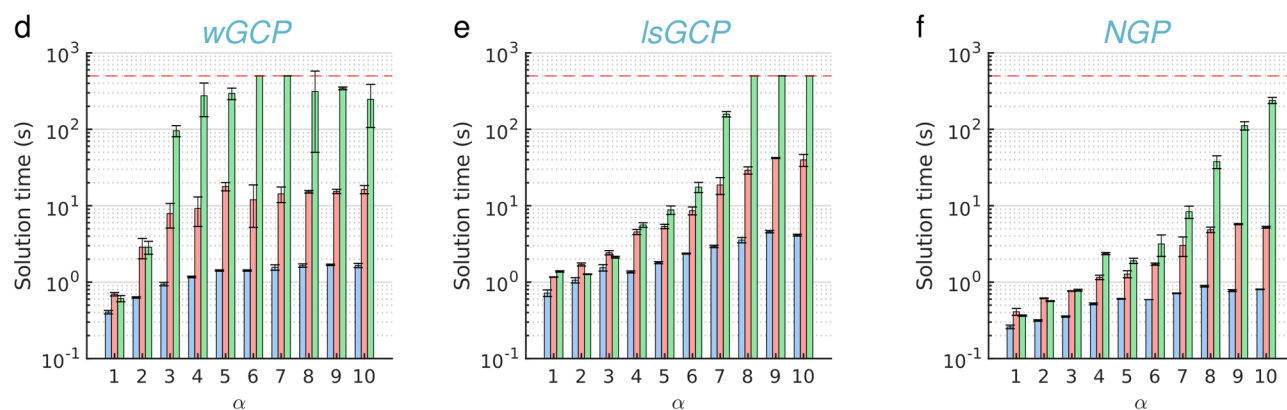


Figure 2. Effect of design parameters on solution time. We examined the target design objective (i.e., wGCP, lsGCP, and NGP) and the limits of deletion reactions, α , and module reactions, β_k , on computation time for solving the ModCell2-MILP problem with the blended (a–c) and goal attainment (d–f) formulations. A time limit of 500 s, indicated by a red dashed line, was used in all cases, but only reached by certain wGCP and lsGCP cases with $\beta \geq 2$. The simulations were performed in duplicate.

other strain design algorithms helps minimize the chances of missing relevant candidates.

ModCell2-MILP Can Identify a Universal Modular Cell Compatible with All Exchangeable Production Modules. Based on the computationally tractable reaction deletion candidate set, we next evaluated whether the goal programming formulation could help identify a universal ModCell design that is compatible with all modules. By screening for increasing α and β_k , we found that a design with $\alpha = 6$ and $\beta_k = 1$ could overcome the performance trade-offs between modules and hence constitute a universal modular cell that is compatible with all production networks considered (Figure 4a). Once coupled with a module, each resulting production strain displays the engineered phenotype with the defined minimum objective goal of 0.5 (i.e., 50% of the theoretical maximum product yield attained at the maximum growth rate). Remarkably, most products surpassed this minimum goal with yields above 90% of the theoretical maximum values (Figure 4b). All production networks displayed a feasible metabolic space where an increase in product synthesis rate is needed to attain faster growth rates (Figure 4c). This designed phenotype is useful for optimal pathway selection using adaptive laboratory evolution^{20,21} or pathway libraries.²²

Flexible Metabolic Flux Capacity of *E. coli* Core Metabolism Enables the Design of a Universal Modular Cell. Interface Reactions Responsible for Metabolic Flexibility of a Universal Modular Cell Are Identified by

Comparing Flux Distributions of Production Networks. The designed universal modular cell (Figure 4) can theoretically adapt to the conflicting metabolic requirements of all production modules (Table 2). To gain further insight into this unique metabolic capability of the modular cell, we analyzed the simulated *reference flux distributions* of each production network using the flux variance clustering (FVC) analysis (Materials and Methods). Reactions with the highest flux changes across the production networks are likely critical for the proper operation of the universal modular cell and might present potential bottlenecks. Such reactions were identified by filtering their reference flux standard deviation calculated across production networks with an *ad hoc* threshold of 0.2 (mol/substrate mol). Over 90% of the 535 active reactions, each of which carries a nonzero flux in at least one production network, had standard deviation values below the threshold, indicating highly conserved metabolic core pathways among production networks. Only 9.5% of the active reactions presented a standard deviation magnitude above the threshold (Figure 5a).

In our case study of designing a universal modular cell compatible with all 20 production modules, unbiased hierarchical clustering (Figure 5b) revealed the presence of four interface reaction types in the core metabolism of *E. coli* that are activated to fit specific production modules (Figure 5c). In the context of chassis metabolism, an endogenous module corresponds to a reaction or group of highly coupled

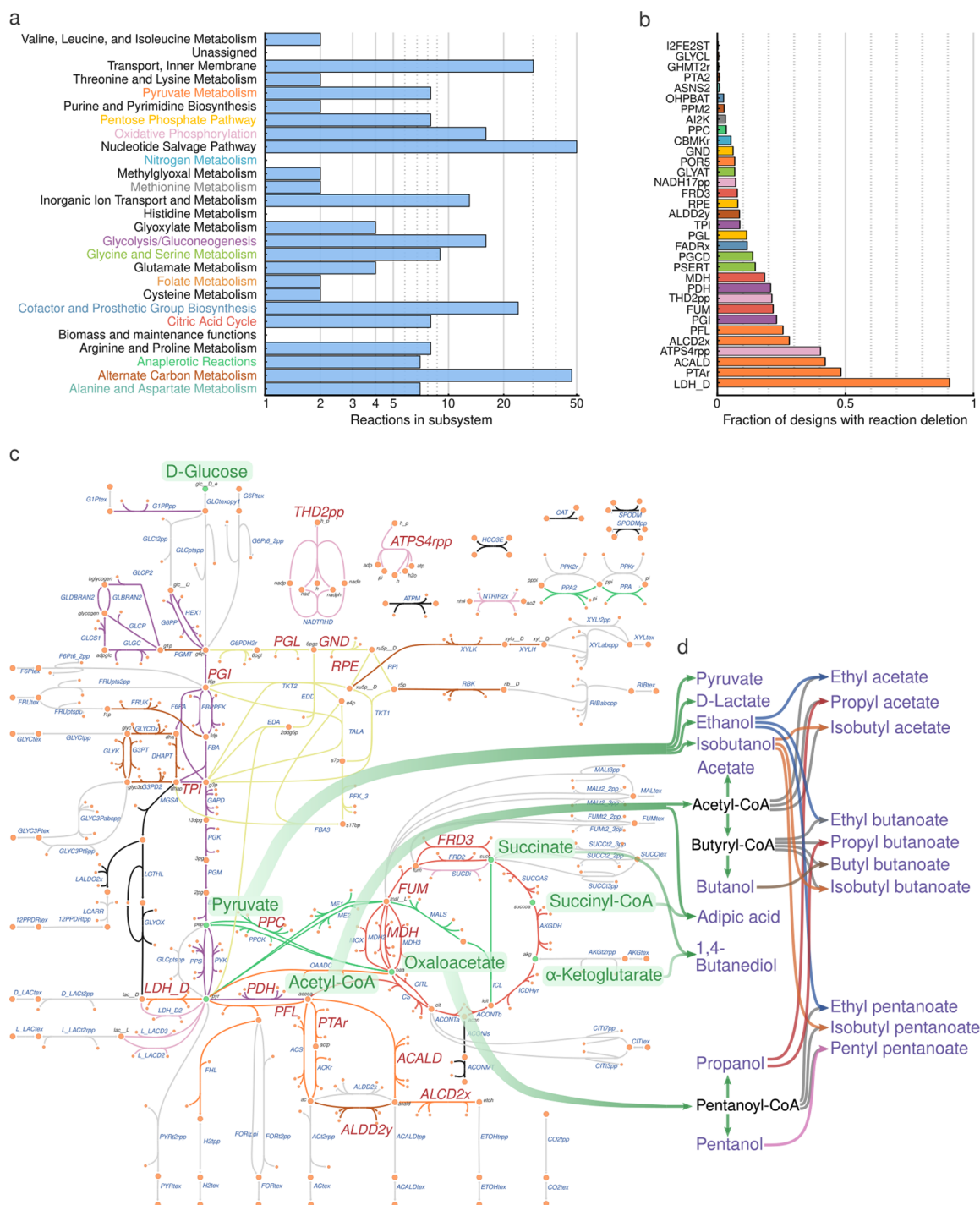


Figure 3. Metabolic functions of deletion candidate reactions. (a) Subsystem distribution for the original set of 276 candidate reactions in the iML1515 model. Those subsystems that contain a reaction used in at least one design are colored. (b) Deletion frequency for the reduced set of 33 candidate reactions. The analysis is based on a pool of 601 wGCP designs from different α and β parameters whose Pareto fronts were previously determined by ModCell2.¹³ Bar colors indicate membership of these reactions in the subsystems. (c) Metabolic map of core metabolism. Key metabolites, including precursors for the 20 product modules (i.e., pyruvate, acetyl-CoA, succinyl-CoA, succinate, and α -ketoglutarate), are highlighted in green. Reactions are colored according to subsystem labels indicated in panel a, reactions colored in light gray do not appear in any of the subsystems of panel a, and reactions that are candidates for deletion, listed in panel b, are labeled in red. (d) Link between major precursors and target products where colors are only used to facilitate visualization. Reaction and metabolite abbreviations correspond to BiGG⁵¹ identifiers (<http://bigg.ucsd.edu/>).

reactions that become active to accomplish a certain metabolic function. The interface reaction classification can be understood in terms of location (i.e., proximity in the metabolic network) and three metabolic functions. The first function is the direction of carbon toward general precursor metabolites

including (i) pyruvate and acetyl-CoA captured by acetyl-CoA-associated modules and (ii) oxaloacetate, succinate, succinyl-CoA, and α -ketoglutarate captured by TCA-associated modules. The second function is the direction of carbon from the precursor metabolites toward secretable molecules,

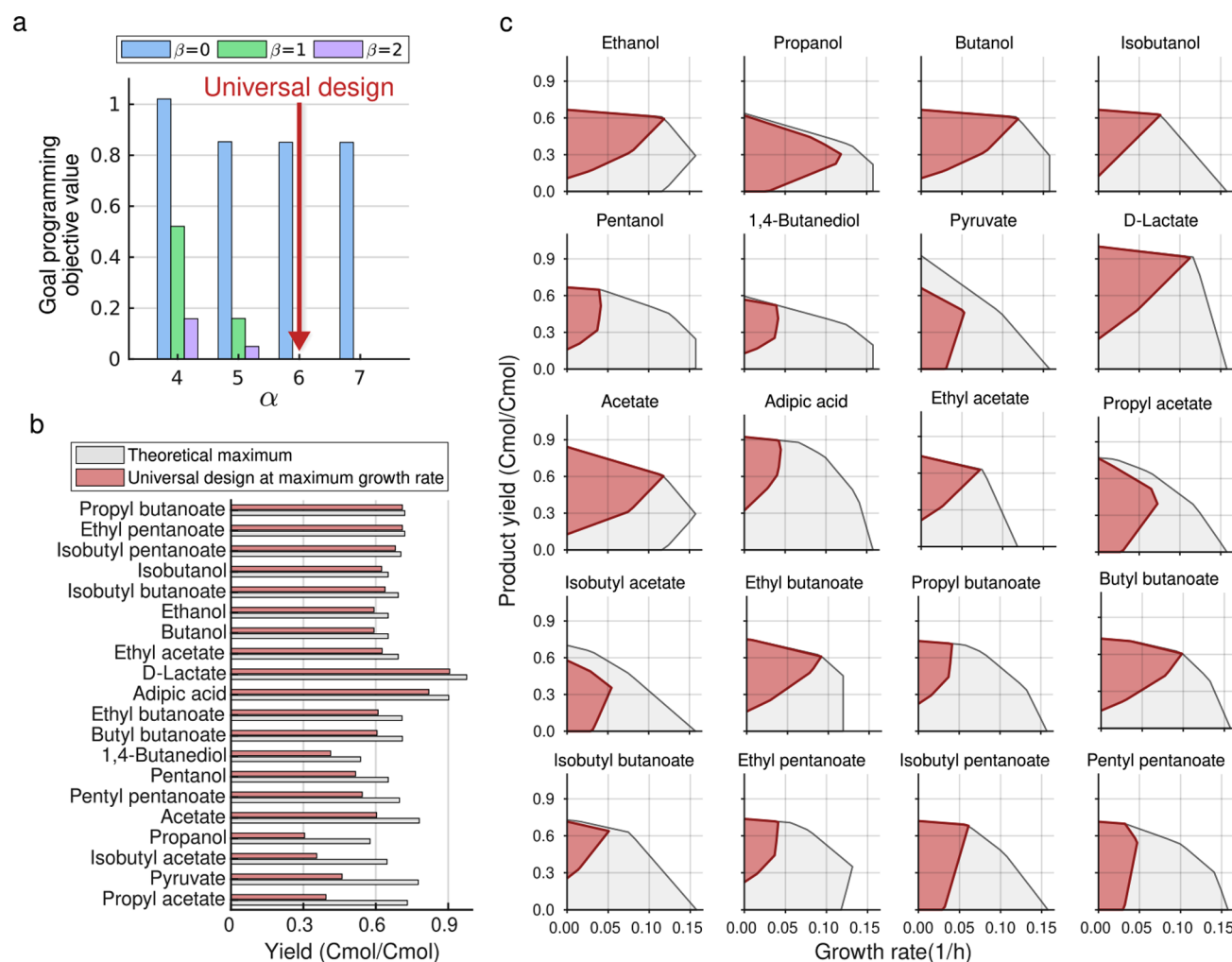


Figure 4. Identification of a universal modular cell compatible with all production modules under the wGCP design phenotype. (a) Goal programming solutions with increasing α and β values. The goal programming objective value, eq 58, in the y -axis measures the difference between the performance of production strains and the target goal, that is, $\sum_{k \in \{k \in \mathcal{K}: f'_k < g_k\}} (f'_k - g_k)$ where the target goal is set to be $g_k = 0.5$. The parameters $\alpha = 6$ and $\beta = 1$ are sufficient to identify a universal modular cell 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 5c.

captured by the upstream and TCA-associated modules. The third function is the use of ATP- and NADP(H)-dependent pathways required to maintain homeostasis, captured by the acetyl-CoA-associated and energetic modules. While these functions are conceptually separable, their biochemical manifestation overlaps, that is, specific metabolic reactions or pathways can simultaneously fulfill several functions.

Interface reactions enable the exchangeable production modules to properly operate with the universal modular cell. These interface reactions might become potential metabolic bottlenecks in practice if they cannot satisfy the predicted fluxes and thus might be critical engineering targets when the associated production modules are used.

Acetyl-CoA-Associated Interface Reactions. This interface type contains pyruvate formate lyase (PFL) and pyruvate dehydrogenase enzyme complex (PDH) reactions that convert pyruvate to acetyl-CoA. Intuitively, products derived from pyruvate, such as isobutanol, require a low flux through PFL and PDH while those derived from acetyl-CoA require a high

flux. Remarkably, the redox states of production strains determine the ratios of PFL to PDH fluxes. For example, the ethanol production network has a relatively high flux through PDH and a low flux through PFL; however, for ethyl acetate, which has a lower degree of reduction than ethanol (Table 2), PFL with formate secretion is prioritized over PDH with NADH generation. Note that our model did not include the regulatory restriction that PDH is inhibited in *E. coli* anaerobically because the function of PDH is equivalent with the coupling of PFL and heterologous NADH-dependent formate dehydrogenase (FDH) demonstrated experimentally for increased butanol^{33,52} and pentanol³⁵ production.

Upstream Interface Reactions. This module type is formed by reactions located directly upstream of a secretable metabolite, often associated with the target production module, and thus provides the necessary precursor metabolite(s). Such reactions are commonly overexpressed in practice, for example, the ECOAH1-HACD1-ACACT1r interface reaction group (comprising 3-hydroxyacyl-CoA dehydratase,

Table 2. Overall Production Module Pathway Stoichiometries and Associated Simulated Secretion Fluxes of the Universal Modular Cell Design^a

overall reaction ^b	DoR	r_p	r_{ac}	r_{CO_2}	r_{for}	r_{succ}
Pyr + NADH → ethanol/AcCoA + 2NADH → ethanol (native)	7.0	0.58	0.01	0.27	0.04	
OAA + Glu + 2ATP + 2 NADPH + NADH → AKG + propanol	6.7	0.31	0.36	0.07	0.18	
2AcCoA + 4NADH → butanol	6.5	0.59	0.01	0.28	0.04	
2Pyr + NADPH + NADH → isobutanol	6.5	0.62		0.31		
OAA + Glu + AcCoA + 3NADH + 2ATP + 2NADPH → 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 + 2NADH → 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 + 2ATP + 2NADPH + NADH → AKG + propyl acetate	5.2	0.41	0.30		0.24	
AcCoA + 2Pyr + NADPH + NADH → isobutyl acetate	5.3	0.36		0.02	0.06	0.52
2AcCoA + 3NADH + Pyr → ethyl butanoate	5.3	0.61		0.09	0.23	
2AcCoA + 3NADH + OAA + Glu + 2ATP + 2NADPH → AKG + propyl butanoate	5.4	0.68	0.03	0.23	0.04	
4AcCoA + 6NADH → butyl butanoate	5.5	0.61		0.14	0.18	
2AcCoA + 3NADH + 2pyr + NADPH → isobutyl butanoate	5.5	0.64		0.16	0.16	
OAA + Glu + AcCoA + 2NADH + 2ATP + 2NADPH + Pyr → AKG + ethyl pentanoate	5.4	0.68	0.03	0.23	0.04	
OAA + Glu + AcCoA + 2NADH + 2ATP + 3NADPH + 2Pyr → AKG + isobutyl pentanoate	5.6	0.67	0.01	0.25	0.03	
2OAA + 2Glu + 2AcCoA + 4NADH + 4ATP + 4NADPH → 2AKG + pentyl pentanoate	5.6	0.53	0.22	0.20	0.02	

^aDoR is the degree of reduction of the final product (mol e⁻/mol C). Metabolite secretion profiles are determined from the simulated reference flux distributions (mol C/mol C) of the universal modular cell design. Flux abbreviations: r_p , product; r_{ac} , acetate; r_{CO_2} , CO₂; r_{for} , formate; r_{succ} , succinate. Note that the negative CO₂ fluxes in pyruvate and acetate production networks indicate overall CO₂ uptake enabled by phosphoenolpyruvate carboxylase (PPC). ^bAbbreviations: Pyr, pyruvate; NADH, nicotinamide adenine dinucleotide, reduced; AcCoA, acetyl-coenzyme A; OAA, oxaloacetic acid; ATP, adenosine 5'-triphosphate; NADPH, nicotinamide adenine dinucleotide phosphate, reduced; AKG, α -ketoglutaric acid; Succ, succinate; SucCoA, succinyl-coenzyme A.

3-hydroxyacyl-CoA dehydrogenase, and acetyl-CoA acetyl transferase) responsible for generating butyryl-CoA and the ACLS-DHAD1-KARA1 interface reaction group (comprising of acetolactate synthase, dihydroxy-acid dehydratase, and keto-acid reductoisomerase) responsible for generating isobutyryl-CoA. These interface reactions can also become active to form byproducts in certain production networks, for example, the PTAr-ACKr-ACT2rpp-Actex interface reaction group (comprising phosphate acetyl transferase, acetate kinase, and cytosolic and periplasmic acetate transport) that not only carries the highest flux in the acetate production network but also becomes active in the propanol-associated modules.

TCA-Associated Interface Reactions. This interface type has the same function as the upstream interface reactions, but it is localized in the TCA (Krebs) cycle. Several products, including adipic acid, 1,4-butanediol, propanol, pentanol, and their associated esters, are derived from the TCA intermediates and interface with the universal modular cell via the TCA-derived interface reactions. The SUCOAS-MMM-MMCD interface reaction group (comprising succinyl-CoA synthetase, methylmalonyl-CoA mutase, and methylmalonyl-CoA decarboxylase) must be activated to convert succinate into succinyl-CoA and then propanoyl-CoA. Remarkably, two routes are present to synthesize fumarate from oxaloacetate, including the conventional MDH-FUM interface reaction group (comprising malate dehydrogenase and fumarase) that consumes NADH and the cyclic ASPTA-GLUDY-ASPT interface reaction group (comprising aspartate transaminase, glutamate dehydrogenase, and L-aspartase) that consumes NADPH. These NADH/NADPH cofactors are not interchangeable due to the deletion of the transhydrogenase THD2pp in the universal modular cell, so the isobutyl pentanoate and pentyl pentanoate

modules, which are derived from the ASPTA-GLUDY-ASPT interface reaction group, also have a high NADPH requirement. Some production networks, such as pyruvate and isobutyl acetate, that are not based on the TCA-derived interface reactions secrete succinate instead of ethanol or lactate to balance redox by using the PPC-MDH-FUM-SUCCTex interface reactions (comprising phosphoenolpyruvate carboxylase, malate dehydrogenase, fumarase, and succinate transport).

Energetic Interface Reactions. This interface type primarily involves NAD(P)-dependent transhydrogenase (THD2pp) and ATP synthase (ATPS4rpp). Other reactions that allow coupling of phosphate and electron-transfer cofactors are also included. The reactions in this module help buffer the diverse electron and ATP requirements of production networks. THD2pp is deleted in the chassis but used as a module reaction in the isobutanol and acetate production networks. In the case of isobutanol production, transhydrogenase expression has been demonstrated to increase the synthesis of NADPH and thus isobutanol.⁵³ Acetate has the smallest degree of reduction after pyruvate, which results in redox imbalance that is compensated via formate secretion. In conjunction with these mechanisms, ATP synthase works in the reverse direction by hydrolyzing excess ATP. Other production networks also use ATPS4rpp to eliminate excess ATP as observed, for example, in the ethyl acetate production network. This strategy is consistent with ATP wasting approaches recently demonstrated.⁵⁴

Comparison between Simulated and Measured Intracellular Fluxes Reveals Flexible Metabolic Flux Capacity of E. coli to Accommodate the Required Wide Flux Ranges. Flux analysis of the production networks suggests that the core

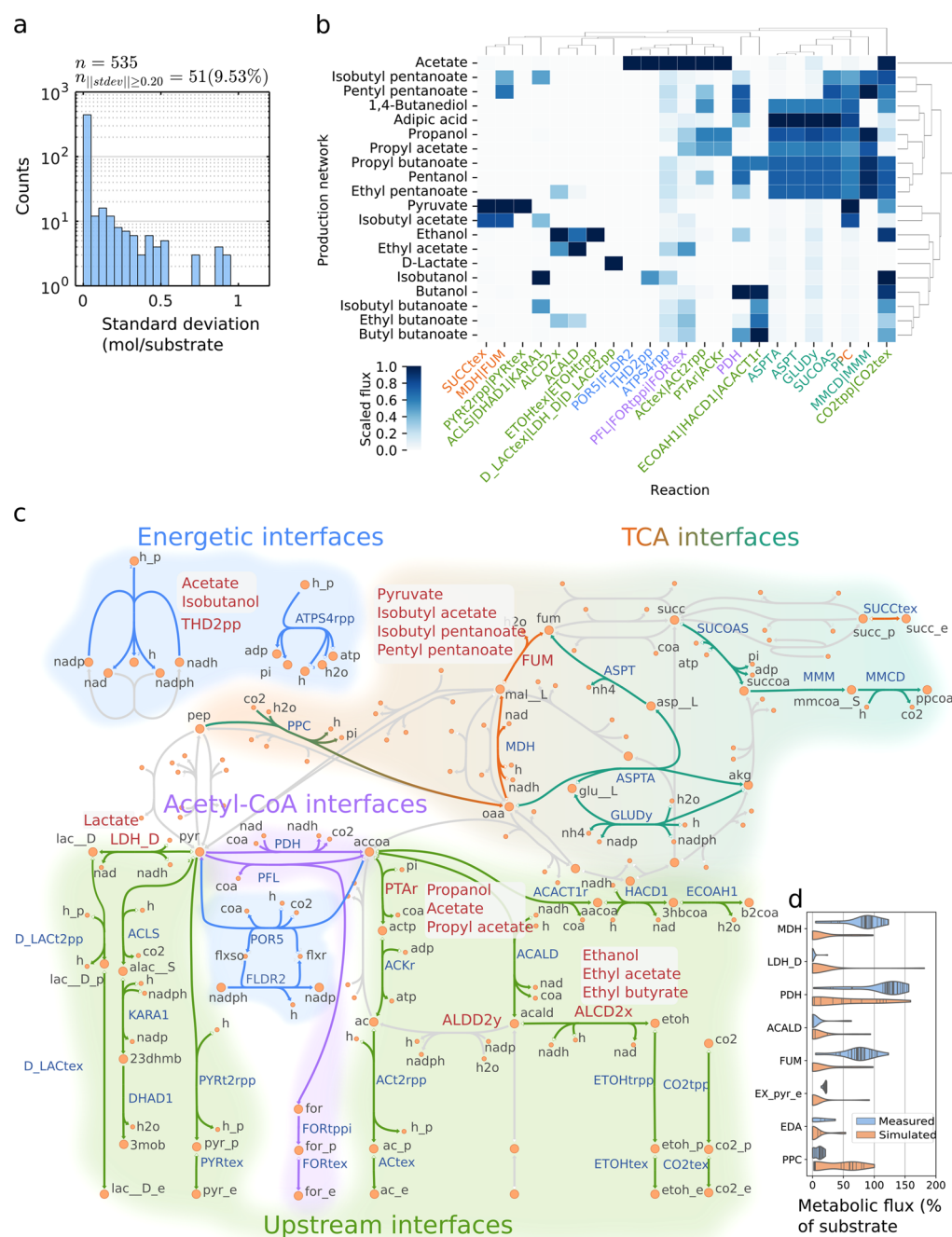


Figure 5. Flexible metabolic flux capacity of *E. coli* metabolism enables the universal modular cell design. (a) Standard deviation of each reaction flux across production networks. (b) Scaled fluxes of the 51 reactions with standard deviation magnitude above 0.2, excluding proton, water transport, and exchange reactions. A scaled flux for a reaction is determined by the reference flux distribution value divided by the maximum value of that reaction across all production networks. Hence, a scaled flux of 0 indicates a given reaction does not carry any flux, and a scaled flux of 1 indicates that this reaction carries the highest flux across production networks. Several columns have multiple reactions, separated by |, since they carry exactly the same flux. (c) Interface reactions of the universal modular cell. The reactions colored in red are deleted in the chassis but are used as module reactions in the production networks shown in the adjacent gray boxes. Metabolites in periplasmic and extracellular compartments have “_p” and “_e”, respectively, suffixed to their abbreviations. Metabolite and reaction abbreviations follow BiGG⁵¹ notation. (d) Comparison between simulated and measured fluxes. The solid lines within the “violins” correspond to samples. The simulated fluxes for the reversible reactions, including FUM, LDH, MDH, and ACALD, were multiplied by -1 to reflect their most common direction.

metabolic reactions (Figure 5b) require a wide range of fluxes when coupled with different production modules. To successfully implement this modular design in practice, we need to evaluate whether the metabolism of *E. coli* has the inherent metabolic flux capacity to accommodate these required fluxes. We compared the simulated reference flux

distributions with a recent collection of 45 measured metabolic fluxes⁵⁵ that are collected from multiple studies across various conditions (e.g., growth under aerobic and anaerobic conditions, use of glucose or acetate or pyruvate as a carbon source) and genotypes (e.g., wild-type *E. coli* and mutants with single gene deletions).^{56–59} Note that while the experimental

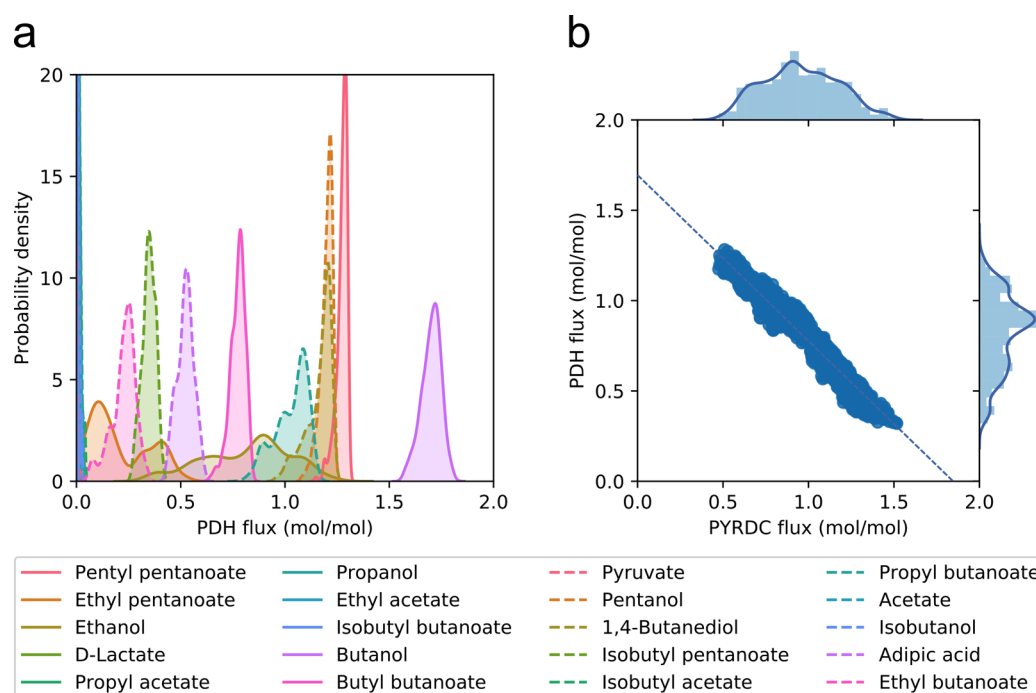


Figure 6. Sampled flux distributions of the ethanol biosynthesis pathways. (a) Probability density function of sampled fluxes for pyruvate dehydrogenase (PDH) across all production networks. Note that ethanol and ethyl pentanoate have the widest operation range. (b) Sampled metabolic fluxes of pyruvate decarboxylase (PYRDC) and PDH in the ethanol production network.

data set provides a flux distribution baseline for wild-type and relatively small deviations for single gene deletion mutants, we anticipate that highly engineered strains with more gene deletions are likely to exhibit wider flux distributions.

Within the reactions present in the 23 groups that constitute interface reactions (Figure 5b), 8 reactions also appeared in the experimental data set (Figure 5d). Remarkably, a highly consistent overlap of flux ranges was observed between the simulated and measured fluxes for malate dehydrogenase (MDH), pyruvate dehydrogenase (PDH), acetaldehyde dehydrogenase (ACALD), fumarase (FUM), and 2-dehydro-3-deoxy-phosphogluconate aldolase (EDA). For the cases of D-lactate dehydrogenase (LDH_D) and pyruvate secretion (EX_pyr_e) that are directly coupled with the biosynthesis of lactate and pyruvate, respectively, we observed that the maximum simulated fluxes surpass the measured values, suggesting that further engineering of wild-type and single-gene deletion *E. coli* is needed to attain the required fluxes. Indeed, previous studies^{60,61} have been able to redirect metabolic fluxes in *E. coli* for yields of lactate and pyruvate above 75% of the theoretical maximum values by simultaneous elimination of competing fermentative pathways for biosynthesis of acetate ($\Delta ackA$), formate ($\Delta pflB$), and ethanol ($\Delta adhE$). The only remaining discrepancy between the simulated and measured fluxes is PPC. Studies not included in the comparison data set have reported up to 50% more PPC flux observed under aerobic conditions,^{62,63} which is still considerably below several of the simulated fluxes. This result suggests that PPC can be a potential metabolic bottleneck in certain production modules. One possible solution is to include in the affected production modules the heterologous PPC from *Actinobacillus succinogenes*, which has been successfully overexpressed in *E. coli* for increased succinate production.⁶⁴ Additionally, bacterial PPC activity can be increased by elevating the acetyl-CoA pool.⁶⁵

Random Sampling of Metabolic Fluxes Confirms the Narrow Operation Range of Interface Reactions. The reference flux distributions analyzed so far represent the ideal metabolic states for each production strain. However, other metabolic states might also exist. To address this uncertainty, we performed randomized flux sampling^{43,44} for each production network under the constraint that product synthesis rate must be above 50% of the maximum value. The results show that the metabolic flux distributions (Figure S2) for most reactions involved in the interface reactions are very narrow, except for the two alternative pathways of ethanol biosynthesis, that is, the endogenous PDH-ACALD-ALCD2x route comprising pyruvate dehydrogenase, acetaldehyde dehydrogenase, and alcohol dehydrogenase (Figure 6a) and the heterologous PYRDC-ALCD2x route comprising pyruvate decarboxylase and alcohol dehydrogenase. These two pathways can be used interchangeably in the ethanol production network, where there is a linear correlation between PYRDC and PDH fluxes (Figure 6b). Notably, the sampled fluxes cannot indicate preferential use of the ethanol synthesis route because the model does not take into account kinetic (k_{cat} , k_M) and regulatory constraints. In summary, even though interface reactions must have flexible metabolic flux capacities to enable a universal modular cell to be compatible with various exchangeable production modules, they must also operate within a narrow flux range when interfacing with a specific production module.

CONCLUSIONS

In this study, we formulated multiobjective modular strain design as blended and goal attainment optimization problems. These problems can be solved by MILP algorithms that guarantee Pareto optimal solutions, exhaustively search the space of alternative solutions, and specify design requirements such as module prioritization or universal compatibility. This

multiobjective strain design approach can be extended with additional design variables (e.g., up- and down-regulation⁶⁶ or reaction insertion from database⁶⁷) or alternative flux prediction models^{68,69} to expand its applications, including use of exchangeable metabolic modules for bioremediation and biosensing. In terms of biological significance, the ModCell2-MILP and FVC methods developed could identify a universal modular cell that harnesses the inherent modularity and flexibility of native *E. coli* metabolism to properly interface with a variety of biochemically diverse pathways. This universal design was predicted to display a growth-coupled-to-product-formation phenotype for all pathways, enabling its use as a platform for pathway optimization through high-throughput library selection or adaptation. The feasibility of this universal design strategy is found to be consistent with experimental evidence of isolated metabolic engineering strategies toward target products and measured intracellular flux ranges. Furthermore, analysis of the metabolic fluxes in this universal design revealed clusters of reactions in central metabolism, named interface reactions, that become activated to interact with specific pathways, providing a mechanistic view into the modularity of metabolic pathways. In this study, the universal design developed was limited to a library of 20 molecules in *E. coli* because one primary aim was to compare the MILP solution with the MOEA solution previously presented. Future studies will use the developed methods to design and understand modularity for different organisms and a larger library of product synthesis modules.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acssynbio.9b00518>.

Biochemical and metabolic diversity of the 20 production modules and violin plots of sampled reaction flux distributions (PDF)

Computer programs used to generate the results of this study (ZIP)

Spreadsheet with modular cell designs for *E. coli* core model (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Cong T. Trinh – Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, Tennessee 37996, United States; Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, United States; orcid.org/0000-0002-8362-725X; Phone: +1 865-974-8121; Email: ctrinh@utk.edu

Author

Sergio Garcia – Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, Tennessee 37996, United States; Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acssynbio.9b00518>

Author Contributions

Conceptualization: S.G., C.T. Data Curation: S.G. Formal Analysis: S.G., C.T. Funding Acquisition: C.T. Investigation:

S.G. Methodology: S.G. Project Administration: C.T. Resources: C.T. Software: S.G. Supervision: C.T. Validation: S.G., C.T. Visualization: S.G., C.T. Writing, Review, and Editing: S.G., C.T.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was funded by the NSF CAREER Award (NSF no. 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, decision to publish, or preparation of the manuscript.

■ DEFINITIONS

Sets

\mathcal{I}_k metabolites in production network k .

\mathcal{J}_k reactions in production network k .

\mathcal{K} production networks that are derived from a combination of the parent metabolic network with the metabolic pathways associated with production modules; the parent metabolic network is the network of the host strain that is genetically manipulated to build a modular cell chassis.

\mathcal{M} metabolic states that correspond to the growth phase, denoted μ , and the nongrowth or stationary phase, denoted $\bar{\mu}$.

\mathcal{C} candidate deletion reaction set; the removal of these reactions are applied to all production networks, $\mathcal{C} \subseteq \mathcal{J}^{\text{parent}} \subseteq \mathcal{J}_k, \forall k \in \mathcal{K}$.

\mathcal{N}_k nontargeted deletion reaction set in production network k ; this set arises from the use of fixed module reactions z_{jk} in certain production networks.

Binary Variables

y_j reaction deletion indicator that takes a value of 0 if reaction j is deleted in the chassis and 1 otherwise.

z_{jk} module reaction indicator that takes a value of 1 if reaction j is added back as module reaction in production network k and 0 otherwise.

d_{jk} reaction activity indicator that takes a value of 0 if reaction j in production network k might not carry a flux and 0 otherwise; thus $d_{jk} = y_j \vee z_{jk}$. This variable is declared as a continuous and linear constraints enforce the OR relation and thus makes the variable binary.

w_k production network feasibility indicator that takes a value of 0 if reaction deletions are ignored and the objective value is set to 0 for production network k and a value of 1 otherwise.

e_{jk} Reaction activity indicator adjusted to w_k that takes the value of d_{jk} if $w_k = 1$ and a value of 1 if $w_k = 0$; thus $e_{jk} = (d_{jk} \wedge w_k) \vee \neg w_k$.

r_{jk} Linearization variable, $r_{jk} = d_{jk} \vee w_k$.

Continuous Variables

v_{jkm} flux (mmol/(g CDW/h)) of reaction j from network k at metabolic state m .

v_{pkm} flux (mmol/(g CDW/h)) of product synthesis reaction from network k at metabolic state m .

v_{xkm} flux (mmol/(g CDW/h)) of biomass synthesis reaction from network k at metabolic state m .

- f_k general objective function for production network k that can be represented by f_k^{wGCP} , f_k^{sGCP} , or f_k^{NGP} .
- f_k' objective function adjusted by w_k such that $f_k' = f_k$ if $w_k = 1$ and $f_k' = 0$ otherwise.
- δ_k^+ amount required by the objective value f_k' to attain the target goal g_k , that is, $\delta_k^+ = g_k - f_k$ if $f_k' < g_k$.
- δ_k^- amount that the objective value f_k' surpasses the target goal g_k , that is, $\delta_k^- = f_k' - g_k$ if $f_k' > g_k$.
- λ_{ikm} dual variable associated with mass balance constraint of metabolite i from production network k at growth state m .
- μ_{jkm}^l dual variable associated with the lower bound of reaction j from production network k at growth state m .
- μ_{jkm}^u dual variable associated with the upper bound of reaction j from production network k at growth state m .
- p_{jkm}^l linearization variable, $p_{jkm}^l = e_{jk} \mu_{jkm}^l$.
- p_{jkm}^u linearization variable, $p_{jkm}^u = e_{jk} \mu_{jkm}^u$.

Parameters

- S_{ijk} stoichiometric coefficient of metabolite i in reaction j of production network k .
- l_{jkm} lower bound for reaction j of production network k at metabolic state m .
- u_{jkm} upper bound for reaction j of production network k at metabolic state m .
- γ minimum biomass synthesis rate required for growth states. Note that in this study a conservative value of 20% of the maximum predicted growth rate of the wild-type strain was used to generate all results.
- α maximum number of deleted reactions in the modular cell chassis.
- β_k maximum number of module reactions in production network k .
- ϵ small scalar used for tilting the biomass objective function, leading to the minimum product rate available at the maximum growth rate; note that in our study $\epsilon = 0.0001$ was used to generate all results.
- $b_\mu, b_{\bar{\mu}}$ weights on the growth and nongrowth objectives of f_k^{sGCP} , respectively; note that in our study $b_\mu = 1$ and $b_{\bar{\mu}} = 10$ were used to generate all results.
- a_k weighting factor applied to the objective function for production network k in the blended formulation; note that in our study $a_k = 1, \forall k \in \mathcal{K}$ was used unless otherwise noted.
- g_k target value for objective f_k' in the goal programming formulation.
- a_k^+ weighting factor applied to δ_k^+ , which emphasizes the importance of objective value f_k' to avoid falling below the target value g_k ; note that in our study $a_k^+ = 1, \forall k \in \mathcal{K}$ was used in all cases.
- a_k^- weighting factor applied to δ_k^- , which emphasizes the importance of the objective f_k' to avoid surpassing the target value g_k ; note that in our study $a_k^- = 1, \forall k \in \mathcal{K}$ was chosen everywhere except to determine the universal modular cell design, where $a_k^- = 0, \forall k \in \mathcal{K}$ was used.
- M^w determines the minimum value of f_k that allows w_k to not be 0; a value of 10, corresponding to $f_k \geq 0.01$ for $w_k \neq 0$, was used in all cases.
- M^{obj} upper bound for each objective value; note that in our study a value of 20 was set for all cases.
- M upper bound for dual variables; note that in our study a value of 100 was set for all cases.

REFERENCES

- (1) 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. *Nat. Catal.* 2, 18.
- (2) Nielsen, J., and Keasling, J. D. (2016) Engineering Cellular Metabolism. *Cell* 164, 1185–1197.
- (3) Trinh, C. T., and Mendoza, B. (2016) Modular cell design for rapid, efficient strain engineering toward industrialization of biology. *Curr. Opin. Chem. Eng.* 14, 18–25.
- (4) Garcia, S., and Trinh, C. T. (2019) Modular design: Implementing proven engineering principles in biotechnology. *Biotechnol. Adv.* 37, 107403.
- (5) Coello, C. A. C., and Lamont, G. B. (2004) *Applications of multi-objective evolutionary algorithms*, Vol. 1, World Scientific, Singapore.
- (6) Rangaiah, G. P. (2009) *Multi-objective optimization: techniques and applications in chemical engineering*, Vol. 1, World Scientific, Singapore.
- (7) Helmer, R., Yassine, A., and Meier, C. (2010) Systematic module and interface definition using component design structure matrix. *J. Eng. Des.* 21, 647–675.
- (8) Kitano, H. (2004) Biological robustness. *Nat. Rev. Genet.* 5, 826.
- (9) Kashtan, N., Noor, E., and Alon, U. (2007) Varying environments can speed up evolution. *Proc. Natl. Acad. Sci. U. S. A.* 104, 13711–13716.
- (10) Clune, J., Mouret, J.-B., and Lipson, H. (2013) The evolutionary origins of modularity. *Proc. R. Soc. London, Ser. B* 280, 20122863.
- (11) Shoal, O., Sheftel, H., Shinar, G., Hart, Y., Ramote, O., Mayo, A., Dekel, E., Kavanagh, K., and Alon, U. (2012) Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* 336, 1157.
- (12) Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., and Sauer, U. (2012) Multidimensional optimality of microbial metabolism. *Science* 336, 601–604.
- (13) Garcia, S., and Trinh, C. T. (2019) Multiobjective strain design: A framework for modular cell engineering. *Metab. Eng.* 51, 110.
- (14) Garcia, S., and Trinh, C. T. (2019) Comparison of Multi-Objective Evolutionary Algorithms to Solve the Modular Cell Design Problem for Novel Biocatalysis. *Processes* 7, 361.
- (15) Zhou, A., Qu, B.-Y., Li, H., Zhao, S.-Z., Suganthan, P. N., and Zhang, Q. (2011) Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm Evol. Comput.* 1, 32–49.
- (16) Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* 6, 182–197.
- (17) Trinh, C. T., Liu, Y., and Conner, D. J. (2015) Rational design of efficient modular cells. *Metab. Eng.* 32, 220–231.
- (18) Burgard, A. P., Pharkya, P., and Maranas, C. D. (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84, 647–657.
- (19) Klamt, S., and Mahadevan, R. (2015) On the feasibility of growth-coupled product synthesis in microbial strains. *Metab. Eng.* 30, 166–178.
- (20) Fong, S. S., Burgard, A. P., Herring, C. D., Knight, E. M., Blattner, F. R., Maranas, C. D., and Palsson, B. O. (2005) In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol. Bioeng.* 91, 643–648.
- (21) Trinh, C., and Srien, F. (2009) Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol. *Appl. Environ. Microbiol.* 75, 6696–6705.
- (22) Garst, A. D., Bassalo, M. C., Pines, G., Lynch, S. A., Halweg-Edwards, A. L., Liu, R., Liang, L., Wang, Z., Zeitoun, R., Alexander, W. G., et al. (2017) Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.* 35, 48.

- (23) Palsson, B. Ø. (2015) *Systems biology: constraint-based reconstruction and analysis*, Cambridge University Press, United Kingdom.
- (24) Feist, A. M., Zielinski, D. C., Orth, J. D., Schellenberger, J., Herrgard, M. J., and Palsson, B. Ø. (2010) Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. *Metab. Eng.* 12, 173–186.
- (25) Maranas, C. D., and Zomorodi, A. R. (2016) *Optimization Methods in Metabolic Networks*, John Wiley & Sons, Hoboken, NJ.
- (26) Marler, R. T., and Arora, J. S. (2004) Survey of multi-objective optimization methods for engineering. *Struct. Multidiscip. Optim.* 26, 369–395.
- (27) Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., et al. (2017) iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* 35, 904.
- (28) Akita, H., Nakashima, N., and Hoshino, T. (2016) Pyruvate production using engineered *Escherichia coli*. *AMB Express* 6, 94.
- (29) Atsumi, S., Hanai, T., and Liao, J. C. (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* 451, 86.
- (30) Layton, D. S., and Trinh, C. T. (2014) Engineering modular ester fermentative pathways in *Escherichia coli*. *Metab. Eng.* 26, 77–88.
- (31) Niu, D., Tian, K., Prior, B. A., Wang, M., Wang, Z., Lu, F., and Singh, S. (2014) Highly efficient L-lactate production using engineered *Escherichia coli* with dissimilar temperature optima for L-lactate formation and cell growth. *Microb. Cell Fact.* 13, 78.
- (32) Rodriguez, G. M., Tashiro, Y., and Atsumi, S. (2014) Expanding ester biosynthesis in *Escherichia coli*. *Nat. Chem. Biol.* 10, 259–265.
- (33) Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., Cho, K. M., and Liao, J. C. (2011) Driving Forces Enable High-Titer Anaerobic 1-Butanol Synthesis in *Escherichia coli*. *Appl. Environ. Microbiol.* 77, 2905–2915.
- (34) Trinh, C. T., Unrean, P., and Sreenc, F. (2008) Minimal *Escherichia coli* Cell for the Most Efficient Production of Ethanol from Hexoses and Pentoses. *Appl. Environ. Microbiol.* 74, 3634–3643.
- (35) Tseng, H.-C., and Prather, K. L. (2012) Controlled biosynthesis of odd-chain fuels and chemicals via engineered modular metabolic pathways. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17925.
- (36) Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., et al. (2011) Metabolic engineering of *Escherichia coli* for direct production of 1, 4-butanediol. *Nat. Chem. Biol.* 7, 445–452.
- (37) Yu, J., Xia, X., Zhong, J., and Qian, Z. (2014) Direct biosynthesis of adipic acid from a synthetic pathway in recombinant *Escherichia coli*. *Biotechnol. Bioeng.* 111, 2580–2586.
- (38) von Kamp, A., and Klamt, S. (2017) Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nat. Commun.* 8, 15956.
- (39) Hart, W. E., Laird, C. D., Watson, J.-P., Woodruff, D. L., Hackett, G. A., Nicholson, B. L., and Sirola, J. D. (2017) *Pyomo - Optimization Modeling in Python*, Springer Optimization and Its Applications, Vol. 67, Springer International Publishing, Cham.
- (40) Machado, D., and Herrgård, M. (2014) Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* 10, No. e1003580.
- (41) Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* 6, 390.
- (42) Nocedal, J., and Wright, S. (2006) *Numerical optimization*, Springer Science & Business Media, New York.
- (43) Kaufman, D. E., and Smith, R. L. (1998) Direction choice for accelerated convergence in hit-and-run sampling. *Oper. Res.* 46, 84–95.
- (44) Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., et al. (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0. *Nat. Protoc.* 14, 639–702.
- (45) King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., and Palsson, B. O. (2015) Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Comput. Biol.* 11, No. e1004321.
- (46) Geoffrion, A. M. (1972) Generalized benders decomposition. *Journal of optimization theory and applications* 10, 237–260.
- (47) Fischetti, M., Ljubić, I., and Sinnl, M. (2016) Benders decomposition without separability: A computational study for capacitated facility location problems. *Eur. J. Oper. Res.* 253, 557–569.
- (48) von Kamp, A., and Klamt, S. (2014) Enumeration of Smallest Intervention Strategies in Genome-Scale Metabolic Networks. *PLoS Comput. Biol.* 10, e1003378.
- (49) King, Z. A., O'Brien, E. J., Feist, A. M., and Palsson, B. O. (2017) Literature mining supports a next-generation modeling approach to predict cellular byproduct secretion. *Metab. Eng.* 39, 220–227.
- (50) Long, M. R., Ong, W. K., and Reed, J. L. (2015) Computational methods in metabolic engineering for strain design. *Curr. Opin. Biotechnol.* 34, 135–141.
- (51) King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., and Lewis, N. E. (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* 44, D515–D522.
- (52) Nielsen, D. R., Leonard, E., Yoon, S.-H., Tseng, H.-C., Yuan, C., and Prather, K. L. J. (2009) Engineering alternative butanol production platforms in heterologous bacteria. *Metab. Eng.* 11, 262–273.
- (53) Shi, A., Zhu, X., Lu, J., Zhang, X., and Ma, Y. (2013) Activating transhydrogenase and NAD kinase in combination for improving isobutanol production. *Metab. Eng.* 16, 1–10.
- (54) Hädicke, O., Bettenbrock, K., and Klamt, S. (2015) Enforced ATP futile cycling increases specific productivity and yield of anaerobic lactate production in *Escherichia coli*. *Biotechnol. Bioeng.* 112, 2195–2199.
- (55) Khodayari, A., and Maranas, C. D. (2016) A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nat. Commun.* 7, 13806.
- (56) Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., et al. (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* 316, 593–597.
- (57) Kabir, M. M., Ho, P. Y., and Shimizu, K. (2005) Effect of *ldhA* gene deletion on the metabolism of *Escherichia coli* based on gene expression, enzyme activities, intracellular metabolite concentrations, and metabolic flux distribution. *Biochem. Eng. J.* 26, 1–11.
- (58) Zhao, J., Baba, T., Mori, H., and Shimizu, K. (2004) Global metabolic response of *Escherichia coli* to *gnd* or *zwf* gene-knockout, based on 13 C-labeling experiments and the measurement of enzyme activities. *Appl. Microbiol. Biotechnol.* 64, 91–98.
- (59) Zhao, J., and Shimizu, K. (2003) Metabolic flux analysis of *Escherichia coli* K12 grown on 13C-labeled acetate and glucose using GC-MS and powerful flux calculation method. *J. Biotechnol.* 101, 101–117.
- (60) Zhou, S., Causey, T., Hasona, A., Shanmugam, K., and Ingram, L. (2003) Production of optically pure D-lactic acid in mineral salts medium by metabolically engineered *Escherichia coli* W3110. *Appl. Environ. Microbiol.* 69, 399–407.
- (61) Causey, T., Shanmugam, K., Yomano, L., and Ingram, L. (2004) Engineering *Escherichia coli* for efficient conversion of glucose to pyruvate. *Proc. Natl. Acad. Sci. U. S. A.* 101, 2235–2240.
- (62) Peng, L., Arauzo-Bravo, M. J., and Shimizu, K. (2004) Metabolic flux analysis for a *ppc* mutant *Escherichia coli* based on 13C-labelling experiments together with enzyme activity assays and

intracellular metabolite measurements. *FEMS Microbiol. Lett.* 235, 17–23.

(63) Al Zaid Siddiquee, K., Arauzo-Bravo, M., and Shimizu, K. (2004) Metabolic flux analysis of pykF gene knockout *Escherichia coli* based on 13 C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. *Appl. Microbiol. Biotechnol.* 63, 407–417.

(64) Kim, P., Laivenieks, M., Vieille, C., and Zeikus, J. G. (2004) Effect of overexpression of *Actinobacillus succinogenes* phosphoenolpyruvate carboxykinase on succinate production in *Escherichia coli*. *Appl. Environ. Microbiol.* 70, 1238–1241.

(65) Lin, H., Vadali, R. V., Bennett, G. N., and San, K.-Y. (2004) Increasing the acetyl-CoA pool in the presence of overexpressed phosphoenolpyruvate carboxylase or pyruvate carboxylase enhances succinate production in *Escherichia coli*. *Biotechnol. Prog.* 20, 1599–1604.

(66) Pharkya, P., Burgard, A. P., and Maranas, C. D. (2004) OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.* 14, 2367–2376.

(67) Pharkya, P., and Maranas, C. D. (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab. Eng.* 8, 1–13.

(68) Chowdhury, A., Zomorodi, A. R., and Maranas, C. D. (2014) k-OptForce: integrating kinetics with flux balance analysis for strain design. *PLoS Comput. Biol.* 10, No. e1003487.

(69) Dinh, H. V., King, Z. A., Palsson, B. O., and Feist, A. M. (2018) Identification of growth-coupled production strains considering protein costs and kinetic variability. *Metab. Eng. Commun.* 7, No. e00080.