



Comprehensive evaluation of the capacities of microbial cell factories

Received: 23 February 2025

Accepted: 17 March 2025

Published online: 24 March 2025

Gi Bae Kim ^{1,2}, Ha Rim Kim ^{1,2} & Sang Yup Lee ^{1,2,3,4,5,6} 

Systems metabolic engineering is facilitating the development of high-performing microbial cell factories for producing chemicals and materials. However, constructing an efficient microbial cell factory still requires exploring and selecting various host strains, as well as identifying the best-suited metabolic engineering strategies, which demand significant time, effort, and costs. Here, we comprehensively evaluate the capacities of various microbial cell factories and propose strategies for systems metabolic engineering steps, including host strain selection, metabolic pathway reconstruction, and metabolic flux optimization. We analyze the metabolic capacities of five representative industrial microorganisms as cell factories for the production of 235 different bio-based chemicals and suggest the most suitable host strain for the corresponding chemical production. To improve the innate metabolic capacity by constructing more efficient metabolic pathways, heterologous metabolic reactions, and cofactor exchanges are systematically analyzed. Additionally, we present metabolic engineering strategies, which include up- and down-regulation target reactions, for the improved production of chemicals. Altogether, this study will serve as a comprehensive resource for the systems metabolic engineering of microorganisms in the bio-based production of chemicals.

Systems metabolic engineering^{1,2}, which integrates the strategies and tools of synthetic biology, systems biology, and evolutionary engineering with traditional metabolic engineering, allows more efficient development of microorganisms for the sustainable production of various chemicals, including bulk chemicals^{3–6}, fine chemicals^{7,8}, fuels^{9–11}, polymers^{12–14}, and natural products^{15–19} from renewable resources instead of fossil resources. Starting with project design, systems metabolic engineering aims to optimize host strain selection, metabolic pathway construction, and metabolic fluxes, while considering fermentation and downstream processes²⁰. However, exploring the vast metabolic space, represented by the combinations of the metabolic networks of different

host strains and strain engineering strategies, still demands significant time, effort, and costs.

Model microorganisms such as *Escherichia coli* and *Saccharomyces cerevisiae* have been the primary workhorses for metabolic engineering due to the availability of the most abundant knowledge on their genetic and metabolic characteristics and also the gene manipulation tools. However, selecting a host strain requires consideration of the most suitable metabolic characteristics for the production of a target chemical. These include the presence of a native biosynthetic pathway for the target chemical, or the potential to produce it effectively when a heterologous or new biosynthetic pathway is introduced, capacity to produce the target chemical, the safety of the microorganism, and the

¹Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 four), Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea. ²Systems Metabolic Engineering and Systems Healthcare Cross-Generation Collaborative Laboratory, KAIST, Daejeon, Republic of Korea. ³KAIST Institute for the BioCentury, KAIST, Daejeon, Republic of Korea. ⁴BioProcess Engineering Research Center, KAIST, Daejeon, Republic of Korea. ⁵Graduate School of Engineering Biology, KAIST, Daejeon, Republic of Korea. ⁶Center for Synthetic Biology, KAIST, Daejeon, Republic of Korea. ✉ e-mail: leesy@kaist.ac.kr

environmental conditions in which the microorganism can thrive²¹. Recent advancements in bioengineering tools, such as clustered regularly interspaced short palindromic repeats (CRISPR)²² and serine recombinase-assisted genome engineering (SAGE)²³, have enabled the metabolic engineering of non-model organisms that naturally produce target chemicals more amenable. Obviously, performing metabolic engineering on a host strain that possesses the highest biosynthetic capacity toward the target product is a promising strategy, as the strain has a potential to more efficiently produce chemicals compared to the other strains with lower biosynthetic capacity. The production performance is defined by three key metrics: titer (the amount of product per volume), productivity (specific productivity, the rate of production per unit of biomass, or volumetric productivity, the rate of production per volume), and yield (the amount or mole of product per amount or mole of consumed substrate)²⁴. Among these key metrics, yield determines the required raw material costs, significantly affecting the overall bioprocess costs. Thus, selecting a host strain with a biosynthetic pathway that maximizes the yield of chemical production is crucial.

Genome-scale metabolic models (GEMs), which represent gene-protein-reaction associations in organisms through mathematical models, have been used to analyze the biosynthetic capacities and engineering strategies for developing microbial cell factories^{20,25,26}. For example, gene knockout targets for the improved production of L-valine in *E. coli* were identified at the systems level by performing in silico knockout simulations for each gene in the strain, which would otherwise require considerable time, effort, cost for real experiments²⁷. GEM-based approaches have not only identified gene targets for engineering but also characterized strain variations²⁸, constructed biosynthetic pathways toward desired chemicals^{29,30}, analyzed metabolic resource allocations in host strains³¹, and predicted metabolic interactions between microbial communities³². Although GEMs have been utilized to optimize host strain selection, metabolic pathway construction, and metabolic fluxes, a comprehensive exploration of the processes at the systems level still demands significant effort.

In this study, we aim to provide resources for host strain selection, metabolic pathway construction, and metabolic flux optimization. To support host strain selection, we provide the metabolic capacities for 235 chemicals that have been produced, even if only minimally, in representative industrial microorganisms by calculating the maximum theoretical yield (Y_T), the maximum production of the target chemical per given carbon source when resources are fully used for the target chemical production, and maximum achievable yield (Y_A), the maximum production of the target chemical per given carbon source, considering cell growth and maintenance. For further improvement of metabolic pathway reconstruction, we have also systematically analyzed the expansion of innate metabolic capacity through the addition of heterologous reactions and cofactor exchanges in native metabolic reactions, and rewiring of innate metabolism to improve target chemical production. Furthermore, metabolic engineering strategies, which include the target reactions to be up- and down-regulated, are suggested for the improved production of chemicals. To demonstrate the versatility and applicability of these resources, we selected various products, including amino acids (L-lysine and L-glutamate) and ornithine used as nutritional supplements; precursors for biopolymers (sebacic acid and putrescine); a bulk chemical (propan-1-ol); and a key precursor for various natural products (mevalonic acid) as case studies. The resources presented in this study can also be employed for analyzing the other 229 chemicals (Supplementary Data 1–23) and also for other chemicals not described here using similar approaches.

Results

Selection of a suitable host strain having the high metabolic capacity

Bacillus subtilis, *Corynebacterium glutamicum*, *E. coli*, *Pseudomonas putida*, and *S. cerevisiae* are the five most frequently employed and

preferred microbial strains in industrial biomanufacturing and academic research. Here, we analyzed the metabolic capacities - the potential of metabolic networks to produce chemicals - of five representative industrial host strains for the production of 235 chemicals (Supplementary Figs. 1–7). To calculate metabolic capacity, two types of yields of chemical production are used: maximum theoretical yield (Y_T) and maximum achievable yield (Y_A). Ignoring metabolic fluxes toward cell growth and maintenance makes the Y_T to be determined solely by the stoichiometry of reactions in the given metabolic network. However, unlike chemical processes, bioprocesses require resources and energy for the generation and maintenance of cells, which serve as biocatalysts, making it impossible to achieve the Y_T . To more realistically describe the metabolic capacity of strains for chemical production, we calculated Y_A , which accounts for non-growth-associated maintenance energy (NGAM) and setting the lower bound of the specific growth rate to 10% of the maximum biomass production rate to ensure minimum growth requirements, as suggested by Monk et al. (Supplementary Note 1)³³.

We calculated both Y_T and Y_A for 235 chemicals when produced in five microorganisms using nine key carbon sources (i.e., L-arabinose, D-fructose, D-galactose, D-glucose, D-xylose, glycerol, sucrose, formate, and methanol) under different aeration conditions (aerobic, micro-aerobic, and anaerobic conditions) (Supplementary Data 1–5). To calculate the yields of chemical production, we constructed GEMs that incorporate the biosynthetic pathways for each chemical, using metabolic reactions that have been previously reported to function properly for target chemical production. For the construction of this GEM, we selected 235 target chemicals from a metabolic map previously compiled³⁴. We organized all metabolic reactions associated with these target chemicals into mass- and charge-balanced equations using the Rhea database³⁵. For reactions not found in the Rhea database, we manually constructed the corresponding equations. Overall, we developed 272 metabolic pathways leading to the biosynthesis of 235 chemicals, including multiple pathways for a single target chemical when available. We constructed a separate GEM for each chemical biosynthesis pathway in each host, resulting in a total of 1360 GEMs. Out of these, 1092 GEMs were supplemented with heterologous reactions not present in the host strain's GEM to establish functional biosynthetic pathways. The remaining 268 GEMs utilized native biosynthetic pathways for the production of the target chemicals. For more than 80% of the target chemicals, fewer than five heterologous reactions were required to construct biosynthetic pathways in the host strains, with percentages of 88.24%, 84.56%, 88.97%, 85.29%, and 90.81% for *B. subtilis*, *C. glutamicum*, *E. coli*, *P. putida*, and *S. cerevisiae*, respectively (Supplementary Fig. 8). These results indicate that the majority of bio-based chemicals can be synthesized with minimal expansion of metabolic networks. Furthermore, the length of biosynthetic pathways exhibited a weak negative correlation with maximum yields (Spearman correlations of -0.3005 and -0.3032 for Y_T and Y_A under aerobic conditions with D-glucose as the carbon source, respectively; p -values of $8.991e-30$ and $2.601e-30$ for Y_T and Y_A , respectively. $n = 1360$ for both cases), suggesting that maximum yields should be analyzed at the systems level for more comprehensive insights.

Based on metabolic capacities, it is possible to identify the most potent strain for producing a specific chemical. To explore the variability in host performance across chemicals, we performed hierarchical clustering of host ranks based on maximum yields (Supplementary Figs. 9 and 10). Under aerobic conditions with D-glucose as the carbon source, the clustering shows that while most chemicals achieve their highest yields in *S. cerevisiae*, a few chemicals display clear host-specific superiority (e.g., pimelic acid in *B. subtilis*; see Discussion for other selection criteria). Notably, these chemicals do not group according to conventional biosynthetic pathways or chemical categories, highlighting the necessity of evaluating each

chemical individually rather than applying a universal rule. For instance, the metabolic capacities of host strains for producing L-lysine, an essential amino acid used in animal feed and as a human nutritional supplement, were compared under aerobic conditions with D-glucose as the sole carbon source. Among the strains, *S. cerevisiae* showed the highest yield (Y_T) of 0.8571 mol/mol D-glucose, followed by *B. subtilis* (0.8214 mol/mol D-glucose), *C. glutamicum* (0.8098 mol/mol D-glucose), *E. coli* (0.7985 mol/mol D-glucose), and *P. putida* (0.7680 mol/mol D-glucose). Except for *S. cerevisiae*, which synthesizes L-lysine via the L-2-aminoadipate pathway, the other strains utilize the diaminopimelate pathway, albeit with differing metabolic capacities (Supplementary Fig. 11). While metabolic capacities are crucial for selecting host strains for chemical production, other factors, such as actual in vivo metabolic fluxes toward the target chemical and chemical tolerance, play important roles in industrial applications. For example, *C. glutamicum* is widely utilized as an industrial strain for L-glutamate production due to its high metabolic fluxes in the L-glutamate biosynthetic pathway, capability for high cell density cultivation, and the GRAS status of its products. However, *C. glutamicum* has lower Y_T and Y_A under aerobic conditions with D-glucose as the sole carbon source (1.0000 mol/mol D-glucose and 0.9290 mol/mol D-glucose, respectively) compared to the other four strains (*B. subtilis*: 1.1729 mol/mol D-glucose, 0.9225 mol/mol D-glucose; *E. coli*: 1.1917 mol/mol D-glucose, 1.0652 mol/mol D-glucose; *P. putida*: 1.1915 mol/mol D-glucose, 1.0708 mol/mol D-glucose; *S. cerevisiae*: 1.2000 mol/mol D-glucose, 1.0868 mol/mol D-glucose). It should be noted that maximum yields represent upper-bound estimates based on in silico modeling and do not capture factors such as enzyme kinetics, product tolerance, and GRAS status of strains/products. Thus, our resource is intended to serve as an initial screening tool that provides a quantitative basis for narrowing down candidate strains. Although *C. glutamicum* is successfully used in industry, strains with higher maximum yields suggest a higher capacity for chemical production. Therefore, if strains with higher maximum yields are fully explored and engineered, they could potentially be developed into even more efficient cell factories. This also highlights that optimizing metabolic fluxes, a key objective of metabolic engineering, remains crucial for strain development. To further elucidate host-specific differences in maximum yields, we calculated the coefficient of variation (CV) for both Y_T and Y_A across the host strains for each target chemical. Our analysis shows that the mean CV for Y_T is 0.3684, whereas for Y_A it is 0.5172. A one-sided Wilcoxon signed-rank test confirmed that the CV for Y_A is significantly higher than that for Y_T (p -value of 0.000, $n = 6292$), indicating that Y_A captures additional host strain-specific variability. These results suggest that Y_A provides more discriminatory power for identifying hosts with a higher latent capacity for chemical production, which is not evident from Y_T alone. Importantly, even small differences in maximum yield can result in meaningful improvements in large-scale production, emphasizing the value of these metrics in guiding host selection for metabolic engineering.

The emission of greenhouse gases accelerates global warming, spurring the development of technologies to reduce one-carbon gases (e.g., carbon dioxide, carbon monoxide, and methane). Consequently, one-carbon compounds have emerged as promising carbon sources for chemical production. Although the development of microbial cell factories for producing value-added chemicals exclusively from one-carbon compounds is still in its early stages, we also analyzed the potential to convert one-carbon compounds (i.e., methanol, carbon dioxide, and formate) into chemicals, aiming to provide a basis for future research. For instance, in the production of sebacic acid, a precursor of nylon-6,10, using methanol, the Y_T and Y_A were higher in *E. coli* (0.1091 and 0.0969 mol/mol methanol) and *P. putida* (0.1082 and 0.0970 mol/mol methanol), followed by *S. cerevisiae* (0.1000 and 0.0900 mol/mol methanol), *B. subtilis* (0.0944 and 0.0778 mol/mol

methanol), and *C. glutamicum* (0.0667 and 0.0600 mol/mol methanol) when methanol was used as the sole carbon source through the ribulose monophosphate (RuMP) cycle. Maximum yields varied depending on the one-carbon compounds and their assimilation pathways. In *E. coli*, when using carbon dioxide as the carbon source and formate as both a carbon source and reducing power, a strain utilizing the reductive glycine cleavage (rGly) pathway showed higher maximum yields (Y_T and Y_A of 0.0321 mol/mol formate and 0.0270 mol/mol formate, respectively) than a strain using the Calvin-Benson-Bassham (CBB) cycle (Y_T and Y_A of 0.0243 mol/mol formate and 0.0204 mol/mol formate, respectively) (Fig. 1). Methanol exhibits a more negative standard enthalpy of combustion (−638.2 kJ/mol) compared to carbon dioxide (0 kJ/mol) and formate (−211.5 kJ/mol)³⁶. This lower enthalpy of combustion indicates a higher intrinsic energy content, which provides increased availability of reducing power and ATP during metabolism. As a result, strains employing the RuMP cycle achieve higher maximum yields for sebacic acid production compared to those using the CBB cycle and the rGly pathway. Taking into account the toxicity of intermediate metabolites, the catalytic efficiency of enzymes in the one-carbon assimilation pathway, and other economic factors including storage and transport costs of the sources, these maximum yields of each pathway can help identify the most viable assimilation pathway and carbon sources.

Overall, this comprehensive analysis of metabolic capacities for bio-based chemical production highlights which microbial cell factories offer the most efficient biosynthetic pathways under targeted bioprocess conditions, providing a valuable resource for systems metabolic engineering.

Improving the innate metabolic capacity for chemical production

Metabolic engineering enhances the chemical production abilities of microbial cell factories by optimizing cellular characteristics within their metabolic capacity or by improving their innate metabolic capacity³⁷. We systematically analyzed metabolic reactions expected to improve the metabolic capacity of a host strain for target chemical production. To improve the innate metabolic capacity of a host strain, we performed two approaches: expanding the native metabolic network by introducing heterologous reactions and replacing cofactors used in the native metabolic network to non-native cofactors.

To expand the native metabolic network, we explored heterologous metabolic reactions that could build more efficient biosynthetic pathways when coordinated with native metabolic reactions. These heterologous reactions were collected by constructing a universal model, an assembly of all metabolic reactions. The universal model was curated from the universal model provided by the BiGG database³⁸ and contains 3814 metabolites and 6846 metabolic reactions. We identified yield-improving heterologous reactions if the Y_T of a target chemical was improved by at least 1% when these reactions were added to the GEM of a production strain. The simulation was performed for all 1360 constructed GEMs (representing 1360 different microorganisms), limiting the number of added heterologous reactions to three. We also provide the source code for the simulation to analyze more diverse conditions of interest to researchers.

The simulation identified candidate reactions that resulted in carbon- or energy-efficient biosynthetic pathways for the chemicals. For example, phosphoketolase, an enzyme that converts xylulose 5-phosphate (or fructose 5-phosphate) into acetyl phosphate and glyceraldehyde 3-phosphate (or D-erythrose 4-phosphate), was most frequently predicted to improve Y_T of chemicals across host strains. This enzyme facilitates a non-oxidative glycolysis pathway that conserves all carbons from the consumed sugar to acetyl-CoA (Fig. 2a). It has been demonstrated that this non-oxidative glycolysis pathway enhances carbon assimilation and improves the innate metabolic capacity of *E. coli* for acetyl-CoA derived chemicals including acetate³⁹

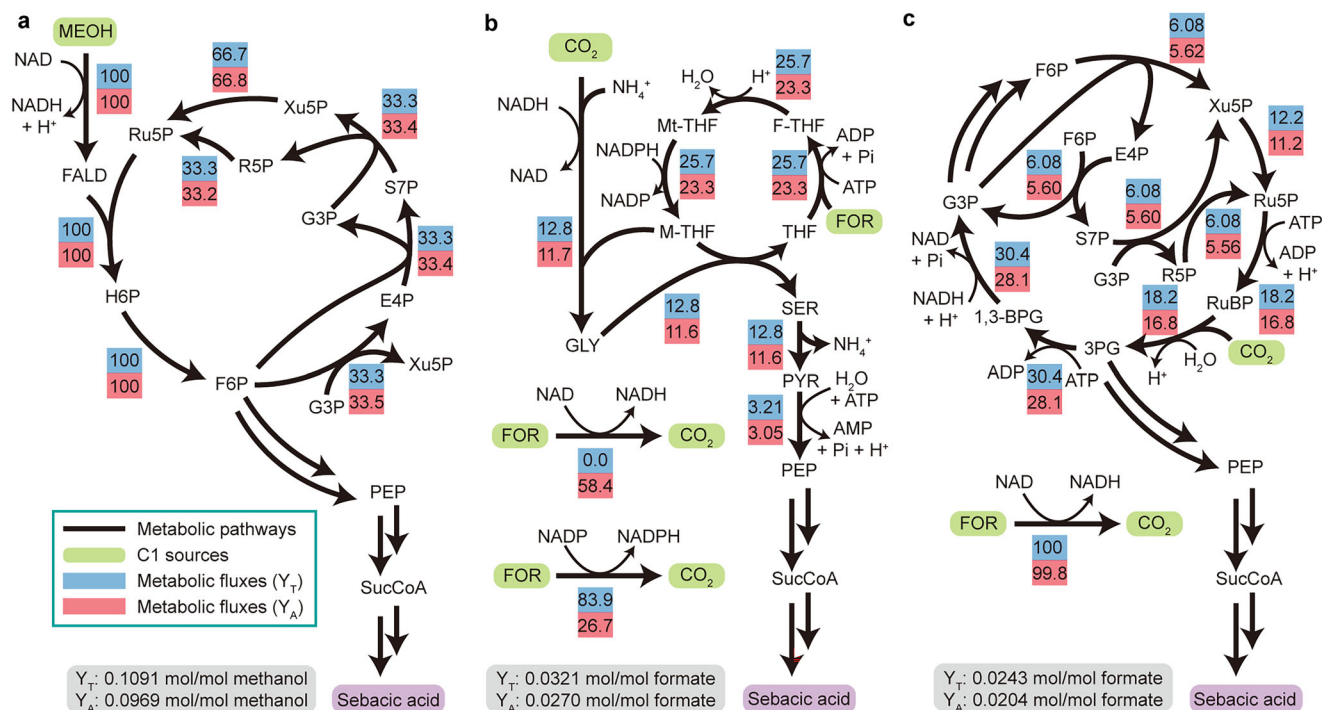


Fig. 1 | Comparison of one-carbon assimilation pathways for sebamic acid production. a Methanol assimilation via RuMP cycle for sebamic acid production. **b** Formate and CO₂ assimilation via CBB cycle for sebamic acid production.

c Formate and CO₂ assimilation via rGLY pathway for sebamic acid production. Metabolic fluxes normalized by carbon source uptake rate to achieve Y_T and Y_A are shown in blue and red boxes, respectively.

and mevalonic acid production⁴⁰, as predicted in this study. While several heterologous reactions demonstrated conserved improvements in maximum yields across host strains, hierarchical clustering of the heterologous reactions showed that the sets of target reactions differ among host strains (Supplementary Fig. 12). For instance, *meso*-diaminopimelate dehydrogenase was commonly predicted for L-lysine-derived chemicals (i.e., 7-hydroxyheptanoic acid, pimelic acid, cadaverine, L-2-ammoniohexano-6-lactam, 3,6-diammoniohexanoate, L-lysine, glutaric acid, 5-oxopentanoic acid, 5-ammoniopentanamide, 5-aminopentanoic acid, and valerolactam) under aerobic conditions using D-glucose as the sole carbon source in *E. coli* (Supplementary Fig. 13). It should be noted that detailed engineering strategies for introducing heterologous reactions vary with the chemical of interest. For example, ornithine transacetylase, which catalyzes the conversion of L-glutamate to ornithine, was predicted to be a candidate enzyme for improving Y_T of ornithine production in *E. coli*, increasing it from 0.9817 mol/mol D-glucose to 1.0142 mol/mol D-glucose under aerobic conditions—a 3.31% increase. Ornithine transacetylase replaces two enzymatic steps—specifically, N-acetylglutamate synthase and acetylornithine deacetylase—in the ornithine biosynthetic pathway (Fig. 2b). Flux-sum analysis, which calculates the sum of the incoming and outgoing fluxes from a metabolite (see Methods), was performed to identify major differences in flux distribution between the two pathways (Fig. 2d). Acetate, acetyl phosphate, and acetyl-CoA showed the largest decreases in flux-sum from the native ornithine biosynthesis pathway to the pathway with the introduced ornithine transacetylase. The native pathway requires acetyl-CoA for N-acetylglutamate synthase and produces acetate via acetylornithine deacetylase. To supply the required acetyl-CoA, the metabolic network must convert acetate back to acetyl-CoA, consuming an ATP via acetate kinase. We performed flux variability analysis (FVA) and confirmed that introducing ornithine transacetylase reduces the metabolic fluxes of acetate kinase and phosphotransacetylase (Fig. 2d). This reduction decreases the energy required to convert acetate into acetyl-CoA. The yield of ornithine production in the pathway with the newly introduced

ornithine transacetylase is consequently increased (Fig. 2d). Although *C. glutamicum* and *S. cerevisiae* natively possess ornithine transacetylase, experimental studies have demonstrated that further overexpression of ornithine transacetylase-encoding genes (i.e., *argI*)—thereby reinforcing the flux through this pathway—can enhance ornithine production in microorganisms^{41,42}.

Exploring heterologous reactions also suggests biosynthetic pathways that offer advantages in redox balance. For propan-1-ol production, NADH-dependent homoserine dehydrogenase, homoserine deaminase, and pyruvate carboxylase were predicted to improve the Y_T in *E. coli* from 0.7059 mol/mol D-glucose to 1.0909 mol/mol D-glucose under anaerobic conditions—a 54.55% increase (Fig. 2c). The propan-1-ol biosynthetic pathway without these heterologous reactions requires an ATP and two NADPH from glyceraldehyde 3-phosphate. Furthermore, producing the intermediate L-aspartate from L-glutamate and oxaloacetate indirectly requires NADPH, necessitating a sufficient NADPH supply for propan-1-ol production. The introduced enzymes reduce the dependency on the NADPH pool. The native NADPH-dependent homoserine dehydrogenase consumes one NADPH, while the introduced NADH-dependent homoserine dehydrogenase consumes an NADH to convert L-aspartate 4-semialdehyde to L-homoserine, thus reducing the total NADPH requirement for propan-1-ol production. Additionally, homoserine deaminase bypasses the ATP-requiring native pathway to convert L-homoserine to 2-oxobutanoate, and pyruvate carboxylase converts pyruvate to oxaloacetate, generating ATP in the reaction. Therefore, the heterologous reactions make the metabolic network more ATP efficient, redirecting metabolic fluxes from acetate and formate production pathways to the pentose phosphate pathway (Supplementary Fig. 14). The addition of heterologous reactions increased the flux-sum of metabolites in the pentose phosphate pathway, demonstrating that the engineered metabolic network can redirect metabolic fluxes toward the pentose phosphate pathway to meet the high NADPH demand rather than producing ATP from fermentation pathways (Fig. 2e).

We further assessed how the number of added heterologous reactions affects increases in maximum yields. Among the 784,774

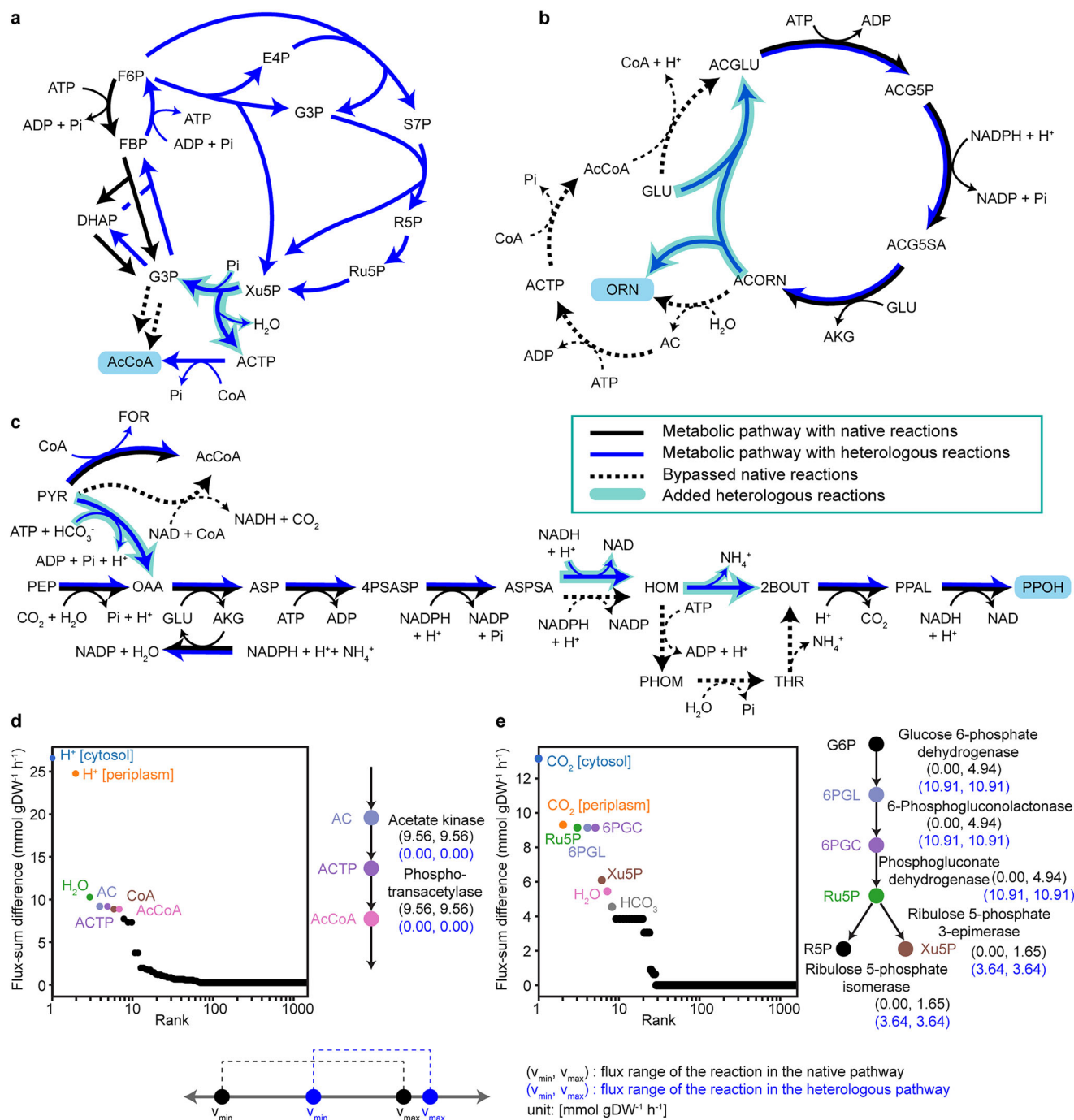


Fig. 2 | Improved Y_T of chemical production via introduction of heterogeneous reactions. **a** Non-oxidative glycolysis for improving Y_T of acetate production. **b** Use of ornithine transacetylase for improving Y_T of ornithine production. **c** Use of NADH-dependent homoserine dehydrogenase, homoserine deaminase, and pyruvate carboxylase for improving Y_T of propan-1-ol production. Black arrows indicate native metabolic pathways, while blue arrows indicate metabolic pathways with heterogeneous reactions. Dotted arrows show bypassed reactions due to the addition of heterogeneous reactions. The added heterogeneous reactions are shaded in cyan. **d** Flux-sum differences of metabolites between the native pathway and the pathway with introduced ornithine transacetylase for Y_T calculation. Metabolites

are sorted in descending order of flux-sum decrease. **e** Flux-sum differences of metabolites between the native pathway and the pathway with introduced NADH-dependent homoserine dehydrogenase, homoserine deaminase, and pyruvate carboxylase for Y_T calculation. Metabolites are sorted in descending order of flux-sum increase. Flux values in parentheses are the lower and upper bounds of the FVA results when maximizing the target chemical production flux. FVA results for native pathways and pathways with introduced heterogeneous reactions are in black and blue, respectively. Abbreviations for metabolites and reactions are available in Supplementary Data 24. Source data are provided as a Source Data file.

heterologous reaction targets that improve Y_T of chemicals, 57,045 involved a single reaction, 191,901 involved two reactions, and 535,828 involved three reactions. The distribution of the percentage increases in Y_T differed significantly across these groups (Kruskal-Wallis H-test; p -value of 0.000, $n = 784,774$) with median increases of

6.296%, 8.800%, and 10.37% for the addition of one, two, and three reactions, respectively (Supplementary Fig. 15). However, introducing multiple heterologous reactions should be carefully optimized to balance yield improvements with the potential metabolic burden on the host.

Similarly, we identified metabolically efficient biosynthetic pathways for chemical production in industrial microorganisms (Supplementary Data 6–10). However, the predicted heterologous reaction sets should be carefully selected for validation experiments. Although all metabolic reactions used in the analysis are mass- and charge-balanced, flux balance analysis (FBA) often fails to exclude thermodynamically infeasible solutions due to insufficient knowledge of the reaction directions. To address this issue, we employed loopless FBA to reduce thermodynamically infeasible solutions⁴³. Thermodynamics-based FBA, which utilizes metabolite concentration ranges and Gibbs free energy of reactions to predict feasible flux profiles, could further reduce infeasible solutions. However, incomplete knowledge of genome-scale metabolome data and reaction thermodynamics hinders the application of such thermodynamics-based network analysis. Additionally, the search space for non-native metabolic reactions is constrained by the reactions in the universal model. Therefore, further improvement of the universal model in both quality and quantity is necessary for better construction of metabolically efficient biosynthetic pathways. Community efforts to elucidate genome-scale metabolome, kinome, and other relevant omics data will facilitate the construction of efficient biosynthetic pathways for chemical production.

Next, we analyzed how to improve metabolic capacity by modifying the use of cofactors in native metabolic reactions. Addressing the burden of cofactor usage is crucial for the biosynthesis of chemicals, particularly for those that are highly oxidized or reduced. To alleviate this burden in biosynthetic pathways, metabolic engineering studies have employed strategies such as changing enzymes⁴⁴ or engineering the cofactor specificities⁴⁵. For example, GEMs have been utilized to identify cofactors for exchange in constructing microbial cell factories^{46,47}. Furthermore, the development of computational designs⁴⁸ or adaptive evolution-based approaches⁴⁹ for altering enzyme specificity makes cofactor swapping a more viable strategy for metabolic engineering. To propose potential strategies for improving metabolic capacity through changes in cofactor usage in metabolic reactions, we analyzed the effect of cofactor exchanges in the 1360 constructed GEMs to improve the Y_T of target chemicals. Metabolic reactions involving NADPH or NADH were considered candidates for cofactor exchange. Cofactor exchanges were identified if the Y_T of the target chemical improved when NADP and NADPH in the candidate reactions were exchanged to NAD and NADH, or vice versa (Supplementary Data 11–13).

For example, under aerobic conditions using D-glucose as the sole carbon source in *E. coli*, 80.6% (29/36) of the biosynthetic pathways for chemicals derived from acetyl-CoA showed improved Y_T 's when the cofactor (NADH) of the glyceraldehyde 3-phosphate dehydrogenase was exchanged for NADPH. This cofactor exchange enables the metabolic network to produce NADPH from glycolysis, reducing the high demand for NADPH in the production of acetyl-CoA-derived fatty acids and isoprenoids. Mevalonic acid, a key precursor in the mevalonate pathway that produces isoprenoids, showed improved Y_T from 0.8000 mol/mol D-glucose to 0.8229 mol/mol D-glucose when the cofactor (NAD) of the glyceraldehyde 3-phosphate dehydrogenase was swapped for NADP. Introducing NADPH-producing reaction reduces metabolic fluxes toward NADP transdehydrogenase and NADH dehydrogenase, providing NADPH required for 3-hydroxy-3-methylglutaryl-CoA reductase from the glyceraldehyde 3-phosphate dehydrogenase (Fig. 3a). Similarly, 61.2% (30/49) of the 2-oxoglutarate-derived chemicals showed improved Y_T 's when the cofactor (NADPH) of the glutamate dehydrogenase was exchanged for NADH, under aerobic conditions using D-glucose as the sole carbon source. Putrescine, a four-carbon diamine used in manufacturing engineering plastics, showed improved Y_T from 0.9907 mol/mol D-glucose to 1.0221 mol/mol D-glucose when the cofactor (NADPH) of the glutamate dehydrogenase was exchanged for NADH (Fig. 3b). Exchanging the required NADPH for NADH in glutamate dehydrogenase allows the metabolic network to provide cofactors from glycolysis, without diverting additional metabolic fluxes toward NADPH-producing

pathways. Experimental studies have demonstrated that employing enzymes with different cofactor specificities can substantially improve chemical production. For example, engineering glyceraldehyde 3-phosphate dehydrogenase to favor NADP⁺ increased L-lysine production in *C. glutamicum*⁵⁰, and overexpression of an NADP⁺-dependent glyceraldehyde 3-phosphate dehydrogenase gene enhanced 3-hydroxypropionic acid production in *E. coli*⁵¹. Furthermore, introducing an NADH-dependent aspartate-semialdehyde dehydrogenase improved L-homoserine production in *E. coli*⁵². These experimental findings support our simulation predictions on cofactor exchange strategies.

These cofactor preferences are largely conserved across different carbon sources but exhibit distinct patterns under varying aeration conditions (Supplementary Fig. 16). For instance, in *E. coli* under aerobic conditions, the predicted cofactor exchange targets predominantly favor NADPH for isoprenoids (e.g., bisabolene, carotene, farnesene, farnesol, farnesyl diphosphate, geraniol, geranyl diphosphate, geranylgeraniol, geranylgeranyl diphosphate, limonene, myrcene, pinene, sabinene, and santalene) and for aromatic compounds (e.g., 2-phenylethanol, 4-amino-L-phenylalanine, 4-aminocinnamic acid, 4-aminophenyl ethylamine, 4-hydroxyphenyl acetaldehyde, 4-hydroxyphenylacetate, and resveratrol). In contrast, under micro-aerobic and anaerobic conditions, both NADH- and NADPH-dependent reactions are predicted with similar frequencies for these chemical groups. These findings suggest that while the inherent cofactor demands of biosynthetic pathways remain relatively consistent, the optimal cofactor utilization strategy is modulated by other factors (e.g., oxygen availability), reflecting shifts in redox balance and energy requirements. Consequently, tailoring cofactor exchange to the specific condition may be important for optimizing production performance. Although engineering or designing enzymes with specific cofactor preferences is challenging, recent advances in machine learning are expected to facilitate the enzyme engineering and design^{53–55}. In this context, we present metabolic engineering strategies that employ cofactor swapping for the production of 235 bio-based chemicals (Supplementary Data 11–13).

While our *in silico* simulations were designed to predict engineering strategies (i.e., the addition of heterologous reactions and cofactor exchanges) to improve Y_T of chemical production, we also found these predicted strategies enhanced Y_A of chemical production (Supplementary Note 2). This indicates that the resource presented in this study have practical applicability, potentially leading to more efficient microbial production processes.

Rewiring metabolic fluxes toward target chemicals

Although a high theoretical yield of a biosynthetic pathway can indicate its maximum potential efficiency, the actual metabolic flux toward the target chemical does not necessarily correlate with the theoretical yield. Therefore, constructing an effective microbial cell factory requires rewiring metabolic fluxes toward the target chemical, rather than solely relying on biosynthetic pathways with high theoretical yields. To suggest metabolic rewiring strategies for chemicals, we conducted flux variability scanning based on enforced objective function (FVSEOF)^{56,57} and iBridge^{56,57} to predict metabolic reactions to be up-regulated or down-regulated to enhance the production fluxes of target chemicals. Initially, FVSEOF analysis was performed for each chemical production using each GEM of the host strains (Supplementary Data 14–18). A metabolic reaction was considered an up-regulation candidate if its flux showed a positive Pearson correlation with the target chemical production flux (see Online Methods). To provide general engineering strategies for chemical production, we analyzed target reaction profiles, which are sets of metabolic reactions predicted to be up-regulation targets, for 272 metabolic pathways leading to the biosynthesis of 235 chemicals (Fig. 4). Hierarchical clustering of these reaction profiles suggests that target chemicals within the same cluster share common engineering strategies. For instance, isoprenoids and metabolites in the isoprenoid biosynthetic

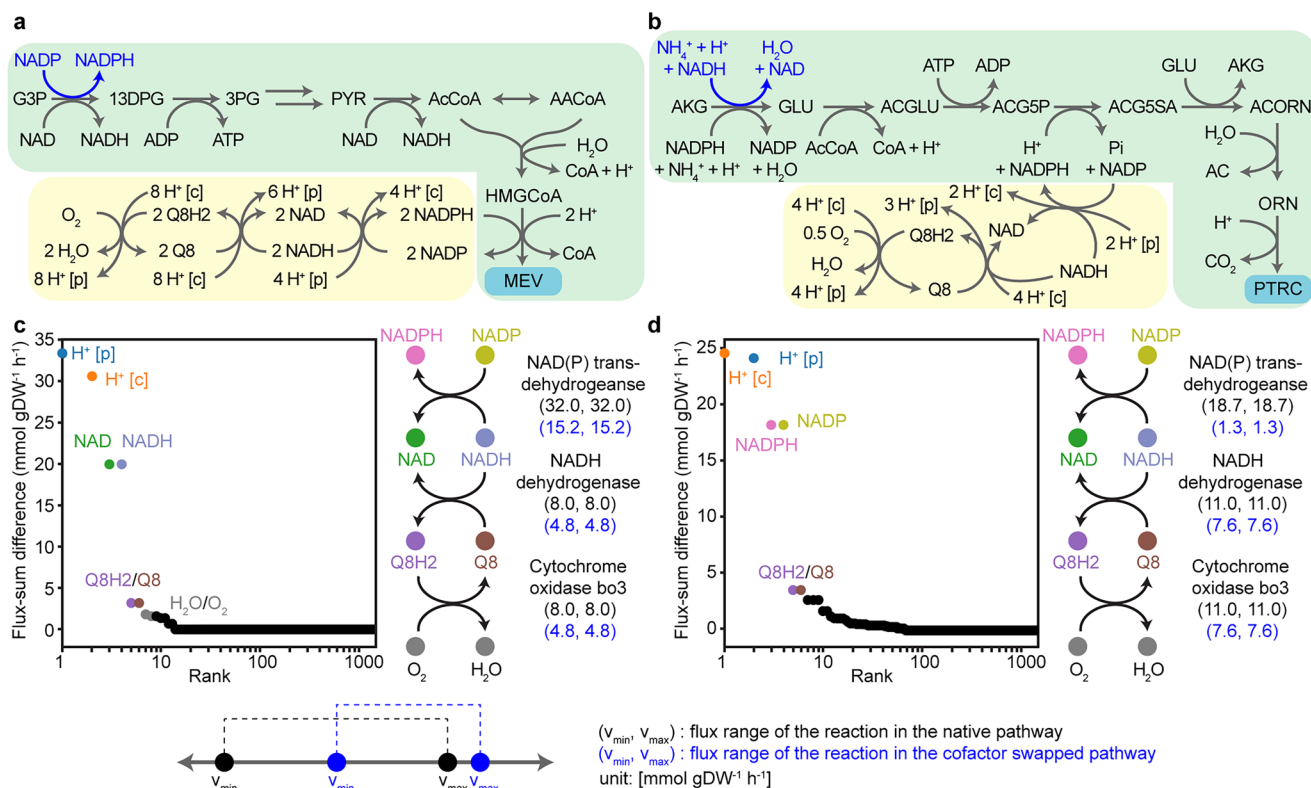


Fig. 3 | Improved Y_T of chemical production via cofactor exchange.

a Improvement of Y_T of mevalonic acid production by exchanging the native cofactor (NAD) of glyceraldehyde 3-phosphate dehydrogenase with NADP.

b Improvement of Y_T of putrescine production by exchanging the native cofactor (NADPH) of glyceraldehyde 3-phosphate dehydrogenase with NADH. Blue arrows represent cofactor swapped reactions. **c** Flux-sum differences of metabolites between the native pathway and the pathway with cofactor-swapped glyceraldehyde 3-phosphate dehydrogenase for Y_T calculation. Metabolites are sorted in

descending order of flux-sum decrease. **d** Flux-sum differences of metabolites between the native pathway and the pathway with cofactor-swapped glutamate dehydrogenase for Y_T calculation. Metabolites are sorted in descending order of flux-sum decrease. Flux values in parentheses are the lower and upper bounds of the FVA results when maximizing the target chemical production flux. FVA results for native pathways and pathways with cofactor swapped reactions are in black and blue, respectively. Abbreviations of metabolites and reactions are available in Supplementary Data 24. Source data are provided as a Source Data file.

pathways formed a distinct cluster (cluster A in Fig. 4). In this cluster, metabolic reactions within the isopentenyl diphosphate biosynthetic pathways, as well as those in the oxidative pentose phosphate pathway (glucose 6-phosphate dehydrogenase, 6-phosphogluconolactonase, and phosphogluconate dehydrogenase) and lower glycolysis (glyceraldehyde-3-phosphate dehydrogenase, enolase, and pyruvate dehydrogenase), were predicted to be up-regulation targets for 20 biosynthetic pathways in cluster A. This highlights the importance of enhancing metabolic fluxes toward key precursor production (acetyl-CoA and isopentenyl diphosphate) and NADPH production to meet the high cofactor demand for isoprenoid production.

Metabolic flux optimization strategies can vary depending on the target chemicals, even if they share common precursors. For example, aromatic compounds derived from D-erythrose 4-phosphate were grouped into two clusters, cluster I and J (Supplementary Fig. 17). While the chorismate biosynthetic pathway was predicted to be an up-regulation target for both clusters, the oxidative pentose phosphate pathway was exclusively predicted for 30 biosynthetic pathways in cluster I, and the TCA cycle (citrate synthase, aconitase, 2-oxoglutarate dehydrogenase, succinate dehydrogenase, and fumarase) was exclusively predicted for 13 biosynthetic pathways in cluster J. This suggests that, despite sharing the same precursor, the engineering strategies for different chemicals can vary considerably, emphasizing the need for tailored metabolic engineering approaches to optimize metabolic fluxes for each target chemical.

Similarly, iBridge analysis was conducted for each chemical production in each GEM of the host strains to predict reactions to be

regulated, providing metabolic engineering strategies complementary to those from FVSEOF (Supplementary Data 19–23). It is important to note that these two algorithms capture different aspects of metabolic networks. Specifically, FVSEOF identifies reactions whose fluxes exhibit a strong positive correlation with the target production flux, thereby robustly predicting up-regulation targets. In contrast, iBridge leverages a metabolite-centric analysis based on flux covariances, enabling the identification of both up- and down-regulation targets that may be missed by FVSEOF. For example, hierarchical clustering of reaction profiles from iBridge showed that pyruvate kinase, not predicted by FVSEOF, was predicted as an up-regulation target for isoprenoids and metabolites in the isoprenoid biosynthetic pathways in cluster A (Fig. 5). As another example, aromatic compounds derived from D-erythrose 4-phosphate formed a distinct clade, cluster B (Fig. 5). In this cluster, 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase and transaldolase were predicted as up-regulation targets, where 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase was also predicted by FVSEOF while transaldolase was unique to iBridge. iBridge also provided down-regulation targets not predicted by FVSEOF. For instance, phosphoglycerate dehydrogenase was predicted as a down-regulation target for cluster D, which includes pyruvate-derived chemicals (e.g., L-leucine, L-valine, isobutanol, isovalerate) and acetyl-CoA-derived chemicals (e.g., butyrate, hexanoate, octanoate, propan-2-ol). Similarly, phosphoenolpyruvate carboxylase was predicted as a down-regulation target for cluster E, which includes pyruvate-derived chemicals (e.g., acetoin, 2,3-butaneol, propionate, isopentane) (Fig. 5). These differences highlight that while some

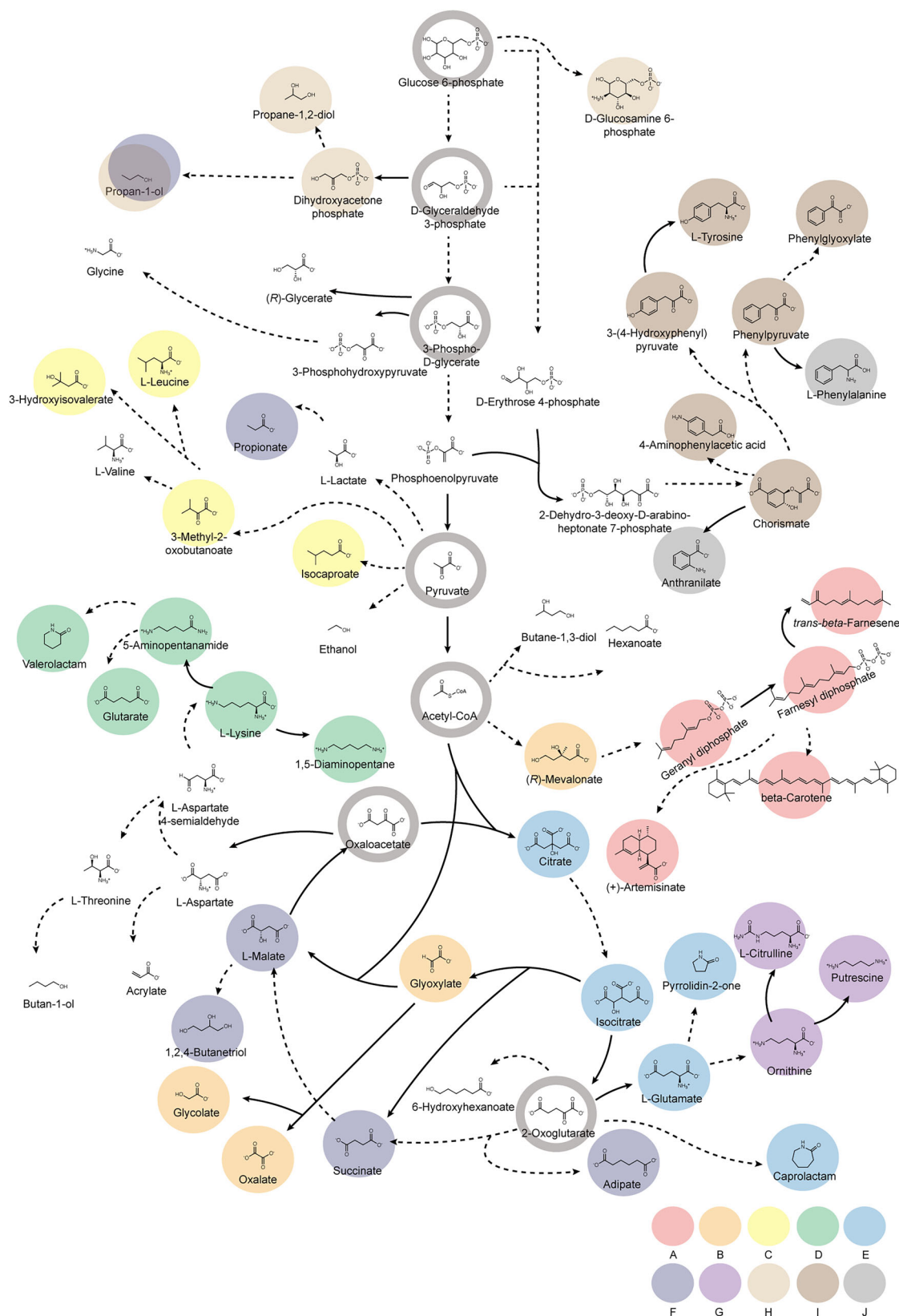


Fig. 4 | Metabolic pathways for bio-based chemicals with cluster groups predicted by metabolic reaction profiles using FVSEOF. Metabolic reaction profiles, representing the types of predicted metabolic reactions to be up-regulated for improved chemical production, were analyzed across 272 metabolic pathways which lead to the production of 235 chemicals. Pairwise similarities of metabolic reaction profiles were calculated using cosine similarity. Chemicals with similar

metabolic profiles were grouped into distinct clusters (Supplementary Fig. 18), and each chemical was colored according to the cluster it belongs to. When multiple pathways for a chemical were available and their clusters differ, both clusters are denoted. The simulation was performed for *E. coli* under aerobic conditions with D-glucose as the sole carbon source. Metabolic pathways for chemicals in each cluster are available in Supplementary Figs. 19–28.

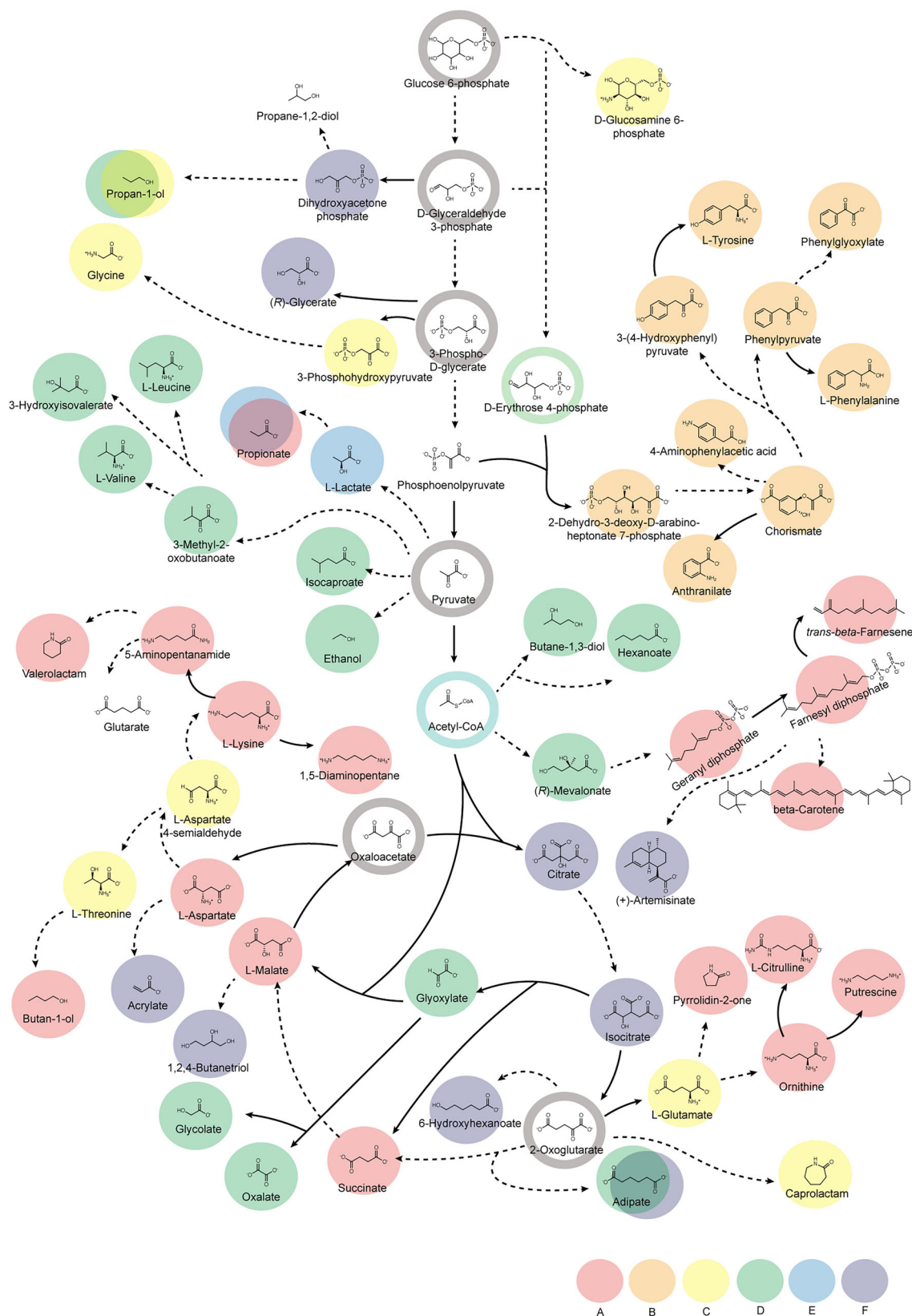


Fig. 5 | Metabolic pathways for bio-based chemicals with cluster groups predicted by metabolic reaction profiles using iBridge. Metabolic reaction profiles, representing the types of predicted metabolic reactions to be up- and down-regulated for improved chemical production, were analyzed across 272 metabolic pathways which lead to the production of 235 chemicals. Pairwise similarities of metabolic reaction profiles were calculated using cosine similarity. Chemicals with

similar metabolic reaction profiles were grouped into distinct clusters (Supplementary Fig. 29), and each chemical was colored according to the cluster it belongs to. When multiple pathways for a chemical were available and their clusters differ, both clusters are denoted. The hierarchical clustering heatmap of metabolic reaction profiles and metabolic pathways for chemicals in each cluster are available in Supplementary Figs. 30–35.

engineering targets were commonly predicted by both approaches, each algorithm captures distinct metabolic features. Consequently, the complementary use of FVSEOF and iBridge enriches our resource by providing a broader set of potential engineering strategies tailored to specific chemicals (Supplementary Data 14–23).

Discussion

Planning a metabolic engineering project necessitates an extensive search through the entire decision-making process, including the selection of target chemicals, host strains, and pathways to be engineered. Finding an optimal strategy for the project is challenging without systematically exploring the vast metabolic space. To aid the initiation of a metabolic engineering project, we provide a comprehensive evaluation of the capabilities of microbial cell factories. The maximum yield of a chemical indicates how efficiently a biosynthesis pathway can transform a carbon source into the target chemical, thereby guiding metabolic engineers in selecting the most optimal host strain by comparing the maximum yields across different strains.

In this study, we calculated the Y_T and Y_A of 235 bio-based chemicals in five representative host strains for metabolic engineering (i.e., *B. subtilis*, *C. glutamicum*, *E. coli*, *P. putida*, and *S. cerevisiae*). It should be noted that the development of genome engineering tools for non-model organisms increasingly enables the use of non-traditional strains for metabolic engineering^{58–60}. Therefore, expanding the analysis of this study to encompass all available genome data could inspire the use of less-explored organisms for metabolic engineering as well.

Conventional constraint-based modeling approaches (e.g., FBA) do not account for gene expression, regulatory network, or allocation of macromolecules within an organism^{61–63}. Although our study is limited to calculating the maximum yields of bio-based chemicals using FBA, integrating multi-scale mechanisms will allow for a more accurate calculation of the biotechnologically achievable maximum yields of these chemicals. Moreover, since we neglected the target chemical-specific transport reactions in our analysis, further characterization and inclusion of these exporters will enable more accurate yield calculations. Recent advancements in metabolic modeling have highlighted the importance of integrating enzyme kinetics and proteome constraints to understand the metabolism of microorganisms. Although proteome-integrated GEMs offer valuable insights under specific experimental conditions, their applicability is limited when exploring diverse environmental and substrate conditions. To assess the potential impact of such constraints on maximum yields, we compared the maximum yields obtained from iML1515, a GEM of *E. coli*, and its enzyme-constrained counterpart, ecML1515 (Supplementary Note 1)⁶⁴. While the magnitude of maximum yields from the enzyme-constrained model can vary, the overall trends and distributions were statistically indistinguishable. These findings demonstrate that while enzyme constraints could provide a more realistic value of maximum yields, the general trends across conditions remain consistent. Thus, the current GEMs provide robust results for exploring metabolic capacities over a wide range of conditions, although further incorporation of enzymatic information continues to be a valuable tool when more precise, condition-specific yield calculations are required.

We also proposed engineering strategies to enhance the innate metabolic capacities of microbial cell factories or to rewire their metabolism toward target chemical production. To identify heterologous reactions for introduction into a microbial cell factory, we utilized and curated a universal model that accounts for all reported metabolic reactions. It is evident that the search of heterologous reactions is limited by the quality and quantity of the universal model. As most currently available GEMs have been reconstructed using highly curated reference GEMs or established reaction databases, the universal model is constrained by limited knowledge of biological reactions rather than reflecting the extensive metabolic space of nature. Developing GEM reconstruction pipelines that directly extract

specific metabolic reactions from genomes would enable the exploration of more diverse and plausible metabolic engineering strategies. Altogether, the resources showcase 42,976 cases detailing the capacities of host strains for 235 bio-based chemicals under different aeration conditions using different carbon sources (5440 cases for *B. subtilis*, 9792 cases for *C. glutamicum*, 11,424 cases for *E. coli*, 3264 cases for *P. putida*, and 13,056 cases for *S. cerevisiae*), alongside 1,925,500 cases detailing engineering strategies (784,774 heterologous reaction targets + 32,867 cofactor exchange targets + 613,863 targets identified by FVSEOF + 493,996 targets identified by iBridge). These resources provided in this study will be useful for selecting a host strain, improving innate metabolic capacity by constructing more efficient metabolic pathways through the introduction of heterologous metabolic reactions and cofactor exchanges, and identifying target reactions for up- and down-regulation to enhance the bio-based production of chemicals.

Selecting the production strain needs to consider various factors, such as growth rate, maximum achievable or optimal/desirable cell concentration, culture condition (e.g., temperature, pH, nutritional requirement, and medium cost), ease of product purification, GRAS status, and others, in addition to the maximum theoretical and achievable yields presented in this study. While the presented maximum yields alone do not capture all dynamic aspects such as growth kinetics and process-specific conditions, they provide a valuable approximation for assessing the inherent metabolic capacity of different strains. As such, this resource serves as an essential reference for narrowing down candidate strains for further experimental validation. Moreover, when combined with additional criteria that reflect the conditions of interest, such as the high productivity achieved in fed-batch fermentations driven by rapid cell growth, the resource can guide strain selection and further cell factory design. Although our approach does not offer a complete solution, it will play an essential role in advancing towards the development of high-performing microbial cell factories.

Methods

Model construction

GEMs containing metabolic pathways for bio-based chemicals were constructed by introducing metabolic reactions of the biosynthetic pathways into previously reported GEMs of representative industrial hosts: *B. subtilis* iY0844⁶⁵, *C. glutamicum* iCW773⁶⁶, *E. coli* iML1515⁶⁷, *P. putida* iJN1463⁶⁸, and *S. cerevisiae* Yeast8⁶⁹. To obtain basic information about the heterologous reactions to introduce, the metabolic reactions, which were compiled in a previously published metabolic map³⁴, were obtained from the Rhea database³⁵. The mass- and charge-balances of the reactions were manually curated. For 235 bio-based chemicals in the metabolic map, overall reactions and pathways toward the biosynthesis of the chemical from key metabolites (i.e., 2-oxoglutarate, 3-phospho-D-glycerate, acetyl-CoA, D-erythrose 4-phosphate, D-glucose 6-phosphate, D-glyceraldehyde 3-phosphate, oxaloacetate, and pyruvate) were identified. Based on the identified biosynthetic pathways, GEMs were constructed by introducing required heterologous reactions for each chemical production into the aforementioned five template GEMs. All reactions and metabolites were curated to have consistent BiGG IDs when possible³⁸. To calculate the metabolic capacity for each target chemical production, an exchange reaction, which exports the target chemical from the cytosol to extracellular space, was added in the GEMs. If multiple pathways have been reported for a single target chemical production, a GEM was constructed separately for each biosynthetic pathway. For the construction GEMs with one-carbon assimilation pathways, the following reactions are added if not included in the original GEMs: alcohol dehydrogenase (methanol), hexulose-6-phosphate synthase, and phosphohexulose isomerase for RuMP cycle; formate dehydrogenase (NAD), formate dehydrogenase (NADP), formate-tetrahydrofolate ligase, and glycine cleavage system (bidirectional) for

rGly pathway; ribulose-bisphosphate carboxylase, phosphoribulokinase and formate dehydrogenase (NAD) for CBB cycle.

Calculation of maximum theoretical yields and maximum achievable yields

To analyze the metabolic capacity for target chemical production, Y_T and Y_A of each bio-based chemical in the five representative industrial host strains (i.e., *B. subtilis*, *C. glutamicum*, *E. coli*, *P. putida*, and *S. cerevisiae*) were calculated. Using the default minimal media composition of each GEM constructed in this study, Y_T and Y_A were calculated using FBA⁷⁰. Each of nine carbon substrates (i.e., L-arabinose, D-fructose, D-galactose, D-glucose, D-xylose, glycerol, sucrose, formate, and methanol) was used as a single carbon source in the simulation. To simulate other carbon sources, we scaled the carbon source uptake rate proportionally based on the number of carbons in the carbon source relative to D-glucose, the default carbon source in the GEMs. For anaerobic and microaerobic conditions, the oxygen uptake rates were constrained to be less than or equal to 0 mmol gDCW⁻¹ h⁻¹ and 0.5 mmol gDCW⁻¹ h⁻¹, respectively⁷¹. The biomass production equation, NGAM requirement, and reaction directions of Yeast8 were curated to enable simulation under anaerobic and microaerobic conditions^{69,72}. The flux of the target exchange reaction, which extracts the target chemical from the cytosol, was maximized as an objective function using loopless FBA simulation. The exchange reaction flux was divided by the carbon source uptake rate to calculate the yields. Because Y_T does not account for the requirements of NGAM, the lower bound for NGAM was set to zero. On the other hand, to calculate Y_A the lower bound for NGAM was set to the default value in each GEM. Additionally, 10% of the maximum biomass production rate was used as the lower bound of biomass production to account for the minimum cell growth of the microbial cell factories when calculating Y_A . As iCW773 has no requirement for NGAM, 3.2 mmol gDCW⁻¹ h⁻¹ was used as the lower bound for ATP maintenance requirement reactions in *C. glutamicum* GEMs, according to a previous *C. glutamicum* GEM⁷³. For bio-based chemicals with multiple biosynthetic pathways, the pathway with the highest yield value was used for comparison of the maximum yields between carbon sources or host strains. Yield calculations were performed for carbon sources that a host strain can metabolize in native pathways. Since *B. subtilis* subsp. *subtilis* str. 168 and *P. putida* KT2440 are aerobic strains, the yields of iY0844 and iJN1463 derivative GEMs were calculated only for aerobic conditions. One-carbon assimilation pathways were also analyzed only for aerobic conditions. Construction of the linear programming problem and solving the problem were performed using cobrapy package⁷⁴ with Gurobi Optimizer (Gurobi Optimization Inc., Houston TX) in Python environment (Python Software Foundation, Delaware, United States).

Flux-sum analysis and flux variability analysis

Flux distributions were calculated using parsimonious FBA before conducting the flux-sum analysis⁷⁵. Flux-sum analysis was performed by summing the incoming or outgoing fluxes of each metabolite in a GEM of interest⁷⁶. Differences between the flux-sum of each metabolite in the flux distributions of the native pathway and the pathway with introduced heterologous reactions were calculated. FVA was performed using loopless solutions^{77,78}.

Universal model construction

The metabolic capacity of a strain is constrained by its overall metabolic reactions. To systematically identify candidate reactions that could improve metabolic capacity for target chemical production, a universal model was constructed. This universal model, derived from the BiGG database³⁸, was curated to contain only mass- and charge-balanced reactions. Reactions in BiGG universal model that involved metabolites without annotation for mass and charge annotations were removed. Annotation data were retrieved from the CarveMe universal

model and the five template microorganism GEMs used in this study⁷⁹. Reactions from the CarveMe universal model and the template microorganism GEMs were added to the universal model if the reactions had the necessary annotation data. Finally, reactions involving metabolites in compartments other than the cytosol, periplasm, and extracellular space were removed, resulting in a universal model comprising 3814 metabolites and 6846 reactions. The universal model was manually curated to correct incorrect reaction directions.

Identification of yield-improving heterologous reactions for target chemical production

Heterologous reactions that improve maximum yields of target chemical production were identified. For a GEM that produces a given target chemical, the following procedures were performed sequentially. To analyze the improvement of Y_T , NGAM was neglected (i.e., the lower bound of ATPM reaction was set to zero). Among metabolic reactions in the universal model that are not in the target GEM, those occurring in the cytosol were added. The target GEM was then used to solve mixed-integer linear programming (MILP), which identifies metabolic reactions required for improving Y_T .

$$\min \sum_{u \in U} y_u$$

$$\text{s.t.} \sum_{j \in M} S_{i,j} \cdot v_j = 0, \forall i \in N \quad (1)$$

$$v_j \leq v_{j,ub}, \forall j \in M \quad (2)$$

$$v_j \geq v_{j,lb}, \forall j \in M \quad (3)$$

$$v_u \leq 1000 \cdot y_u, \forall u \in U \quad (4)$$

$$v_u \geq -1000 \cdot y_u, \forall u \in U \quad (5)$$

$$y_u = y_u^+ + y_u^-, \forall u \in U \quad (6)$$

$$v_u = v_u^+ - v_u^-, \forall u \in U \quad (7)$$

$$v_u^+ \geq 0, \forall u \in U \quad (8)$$

$$v_u^- \geq 0, \forall u \in U \quad (9)$$

$$v_u^+ \leq 1000 \cdot y_u^+, \forall u \in U \quad (10)$$

$$v_u^- \leq 1000 \cdot y_u^-, \forall u \in U \quad (11)$$

$$v_u^+ \geq \epsilon \cdot y_u^+, \forall u \in U \quad (12)$$

$$v_u^- \geq \epsilon \cdot y_u^-, \forall u \in U \quad (13)$$

$$\sum_{u \in U} y_u \leq L \quad (14)$$

$$v_{\text{target}} \geq J_{\text{max}} \quad (15)$$

In the above MILP, the meaning of the variables are as follows: v , a flux of a metabolic reaction; y , a binary variable whether the

corresponding reaction from the universal model should be active or not; $S_{i,j}$, a coefficient of metabolite i which participates in metabolic reaction j ; subscript ub, upper bound of a metabolic reaction; subscript lb, lower bound of a metabolic reaction; ϵ , a minimum threshold which ensures the activity of the candidate metabolic reaction; L , a limit of the number of metabolic reactions to be added in the target GEM; N , a set of metabolites in the target GEM; M , a set of metabolic reactions in the target GEM; U , a set of metabolic reactions from the universal model; J_{\max} , the flux of target chemical producing reaction when the Y_T was achieved. The constraint (1) makes mass balance be validated. Constraints (2) and (3) are for the innate constraints for metabolic reactions such as the thermodynamic, the limit of carbon substrate uptake. Constraints (4–13) identify candidate reactions from the universal model. If a binary variable is on, the corresponding metabolic reaction should be added to the target GEM to improve the Y_T . Here we used 0.001 as a value for ϵ . Constraint (14) limits the number of reactions to be added to the target GEM. We set the number of reactions to be added to three in this study. Constraint (15) ensures that the added metabolic reaction should make the target production flux be improved, which subsequently makes improved Y_T . To analyze improved Y_T , J_{\max} was sequentially decreased from 1.5-fold of the v_{target} at Y_T to the native v_{target} at Y_T in steps of 0.01-fold. Reactions from the universal model that formed infeasible cycles in the target GEMs were manually curated. The simulation was performed until no additional reaction sets were predicted.

Identification of yield-improving cofactor exchange for target chemical production

The identification of cofactor exchange was also performed for all of the constructed chemical-producing models. The default setting for the GEM simulation is the same as that used for identifying heterologous reactions. NADH (or NADPH) participating reactions were duplicated with the cofactor changed to NADPH (or NADH). The target GEM was used to solve a MILP which identifies sets of metabolic reactions that improve the Y_T when the participating cofactors are exchanged.

$$\min \sum_{c \in C} y_c \quad (16)$$

$$s.t. \sum_{j \in M} S_{i,j} \cdot v_j = 0, \forall i \in N \quad (17)$$

$$v_j \leq v_{j,ub}, \forall j \in M \quad (18)$$

$$v_j \geq v_{j,lb}, \forall j \in M \quad (19)$$

$$v_c \leq 1000 \cdot y_c, \forall c \in C \quad (20)$$

$$v_c \geq -1000 \cdot y_c, \forall c \in C \quad (21)$$

$$s_c + t_c = 1, \forall c \in C \quad (22)$$

$$t_c \leq y_c, \forall c \in C \quad (23)$$

$$v_{\text{target}} \geq J_{\max} \quad (24)$$

$$\sum_{c \in C} y_c \leq L \quad (25)$$

In the above MILP, the meaning of the variables are as follow: v , a flux of a metabolic reaction; y , a binary variable whether the corresponding reaction should be active or not; s , a binary variable for the original cofactor participating reaction; t , a binary variable for the cofactor exchanged reaction; $S_{i,j}$, a coefficient of metabolite i which participates in metabolic reaction j ; subscript ub, upper bound of a metabolic reaction; subscript lb, lower bound of a metabolic reaction; L , the upper limit for the number of cofactor swapping; N , a set of metabolites in the target GEM; M , a set of metabolic reactions in the target GEM; C , a set of metabolic reactions which NADH or NADPH participate in; J_{\max} , the flux of target chemical producing reaction when the Y_T was achieved. The constraint (16) makes mass balance be validated. Constraints (17) and (18) are for the innate constraints for metabolic reactions such as the thermodynamic, the limit of carbon substrate uptake. Constraints (19), (20), (21), and (22) restrict only a single reaction to being active between the original metabolic reaction and the cofactor exchanged reaction. If a binary variable is on, the cofactor of the corresponding metabolic reaction should be exchanged to improve the theoretical yield. Constraint (23) ensures that the cofactor exchange reaction should make the target production rate be improved, which subsequently makes improved Y_T . Constraint (24) limits the number of cofactor exchanges. We limited the number of reactions to be exchanged to one in this study.

In silico simulation for identification of up- and down-regulation targets

To achieve a high production yield from a bioprocess, the flux toward the target chemical production should be sufficiently high. To enhance the flux of target chemical producing reaction, iBridge analysis was performed for all constructed GEMs³⁷. iBridge analysis was performed to predict up- and down-regulation target reactions to enhance the production flux of a target chemical. First, parsimonious FBA was used to calculate a reference metabolic flux distribution. Based on the reference flux distribution, ten metabolic flux distributions were calculated with linear minimization of metabolic adjustment, varying the target chemical production flux from zero to its maximum. The covariance between each intracellular reaction flux and the target chemical production flux was calculated from these ten metabolic flux distributions. Metabolites were annotated as positive or negative based on whether the sum of covariances (SoCs) of their outgoing reactions was positive or negative, respectively. Bridge reactions that convert negative metabolites to positive metabolites were identified, and their scores were calculated by subtracting the SoC of the negative metabolites from that of the positive metabolites participating in the reaction, followed by min-max normalization. The reactions with scores equal to or greater than 0.5 are regarded as up-regulation targets, while reactions with scores equal to or smaller than −0.5 are regarded as down-regulation targets³⁷. FVSEOF analysis was also performed for all constructed GEMs under aerobic conditions³⁶. FVSEOF analyzes variations in fluxes in response to the enforced target chemical production rate. For FVSEOF analysis, ten metabolic flux distributions were calculated varying the target chemical production flux from zero to its maximum. From each metabolic flux distribution, FVA was performed for each metabolic reaction. To select reactions that should be up-regulated, the Pearson correlation between the enforced target production rate and the minimal flux of an analyzed reaction was used as a criterion. If the Pearson correlation was positive, the reaction was selected for up-regulation. Metabolic reactions that have corresponding gene information were analyzed using iBridge and FVSEOF.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Source data are provided with this paper and also available from Figshare: <https://doi.org/10.6084/m9.figshare.27874275> (ref. 80). Source data are provided with this paper.

Code availability

The computational pipeline for constructing the resources is available at <https://github.com/kaistsystemsbiology/MEResource> (ref. 81). The source code for FVSEOF simulation is available at <https://github.com/kaistsystemsbiology/FVSEOF> (ref. 82).

References

- Lee, S. Y. & Kim, H. U. Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.* **33**, 1061–1072 (2015).
- Kim, G. B., Choi, S. Y., Cho, I. J., Ahn, D. H. & Lee, S. Y. Metabolic engineering for sustainability and health. *Trends Biotechnol.* **41**, 425–451 (2023).
- Li, Z. et al. Systems metabolic engineering of *Corynebacterium glutamicum* for high-level production of 1,3-propanediol from glucose and xylose. *Metab. Eng.* **70**, 79–88 (2022).
- Liew, F. E. et al. Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nat. Biotechnol.* **40**, 335–344 (2022).
- Tran, V. G. et al. An end-to-end pipeline for succinic acid production at an industrially relevant scale using *Issatchenkia orientalis*. *Nat. Commun.* **14**, 6152 (2023).
- Huang, J. et al. Artificial intelligence system for enhanced automated 1,3-propanediol green biosynthesis. *Green. Chem.* **25**, 9175–9186 (2023).
- Ro, D. K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
- Choi, K. R., Luo, Z. W., Kim, G. B., Xu, H. & Lee, S. Y. A microbial process for the production of benzyl acetate. *Nat. Chem. Eng.* **1**, 216–228 (2024).
- Malaviya, A., Jang, Y. S. & Lee, S. Y. Continuous butanol production with reduced byproducts formation from glycerol by a hyper-producing mutant of *Clostridium pasteurianum*. *Appl. Microbiol. Biotechnol.* **93**, 1485–1494 (2012).
- Kim, H. M., Chae, T. U., Choi, S. Y., Kim, W. J. & Lee, S. Y. Engineering of an oleaginous bacterium for the production of fatty acids and fuels. *Nat. Chem. Biol.* **15**, 721–729 (2019).
- Fang, L. et al. Genome-scale target identification in *Escherichia coli* for high-titer production of free fatty acids. *Nat. Commun.* **12**, 4976 (2021).
- Choi, S. Y. et al. One-step fermentative production of poly(lactate-co-glycolate) from carbohydrates in *Escherichia coli*. *Nat. Biotechnol.* **34**, 435–440 (2016).
- Yan, X. et al. Biosynthesis of diverse alpha,omega-diol-derived polyhydroxyalkanoates by engineered *Halomonas bluephagenesis*. *Metab. Eng.* **72**, 275–288 (2022).
- Ji, M. et al. PHB production from food waste hydrolysates by *Halomonas bluephagenesis* harboring PHB operon linked with an essential gene. *Metab. Eng.* **77**, 12–20 (2023).
- Park, S. Y., Eun, H., Lee, M. H. & Lee, S. Y. Metabolic engineering of *Escherichia coli* with electron channelling for the production of natural products. *Nat. Catal.* **5**, 726–737 (2022).
- Ma, Y., Zu, Y., Huang, S. & Stephanopoulos, G. Engineering a universal and efficient platform for terpenoid synthesis in yeast. *Proc. Natl. Acad. Sci. USA* **120**, e2207680120 (2023).
- Zhang, J. et al. A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* **609**, 341–347 (2022).
- Zhan, C. et al. Improved polyketide production in *C. glutamicum* by preventing propionate-induced growth inhibition. *Nat. Metab.* **5**, 1127–1140 (2023).
- Jiang, B. et al. Characterization and heterologous reconstitution of *Taxus* biosynthetic enzymes leading to baccatin III. *Science* **383**, 622–629 (2024).
- Ko, Y. S. et al. Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. *Chem. Soc. Rev.* **49**, 4615–4636 (2020).
- Nielsen, J. & Keasling, J. D. Engineering cellular metabolism. *Cell* **164**, 1185–1197 (2016).
- Gao, J. et al. Biosynthesis of catharanthine in engineered *Pichia pastoris*. *Nat. Synth.* **2**, 231–242 (2023).
- Elmore, J. R. et al. High-throughput genetic engineering of non-model and undomesticated bacteria via iterative site-specific genome integration. *Sci. Adv.* **9**, eade1285 (2023).
- Konzock, O. & Nielsen, J. TRYing to evaluate production costs in microbial biotechnology. *Trends Biotechnol.* **42**, 1339–1347 (2024).
- Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* **10**, 291–305 (2012).
- Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol.* **20**, 121 (2019).
- Park, J. H., Lee, K. H., Kim, T. Y. & Lee, S. Y. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc. Natl. Acad. Sci. USA* **104**, 7797–7802 (2007).
- Brunk, E. et al. Characterizing strain variation in engineered *E. coli* using a multi-omics-based workflow. *Cell Syst.* **2**, 335–346 (2016).
- Campodonico, M. A., Andrews, B. A., Asenjo, J. A., Palsson, B. O. & Feist, A. M. Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path. *Metab. Eng.* **25**, 140–158 (2014).
- Kumar, A., Wang, L., Ng, C. Y. & Maranas, C. D. Pathway design using de novo steps through uncharted biochemical spaces. *Nat. Commun.* **9**, 184 (2018).
- Chen, Y. et al. Reconstruction, simulation and analysis of enzyme-constrained metabolic models using GECKO Toolbox 3.0. *Nat. Protoc.* **19**, 629–667 (2024).
- Schafer, M. et al. Metabolic interaction models recapitulate leaf microbiota ecology. *Science* **381**, eadf5121 (2023).
- Monk, J. M. et al. Multi-omics quantification of species variation of *Escherichia coli* links molecular features with strain phenotypes. *Cell Syst.* **3**, 238–251.e212 (2016).
- Lee, S. Y. et al. A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* **2**, 18–33 (2019).
- Lombardot, T. et al. Updates in Rhea: SPARQLing biochemical reaction data. *Nucleic Acids Res.* **47**, D596–D600 (2019).
- Green, D. W. & Southard, M. Z. *Perry's chemical engineers' handbook*, 9th Edn, (McGraw Hill Education, 2019).
- Zheng, Y., Yuan, Q., Luo, H., Yang, X. & Ma, H. Engineering NOG-pathway in *Escherichia coli* for poly-(3-hydroxybutyrate) production from low cost carbon sources. *Bioengineered* **9**, 209–213 (2018).
- Norsigian, C. J. et al. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* **48**, D402–D406 (2020).
- Bogorad, I. W., Lin, T. S. & Liao, J. C. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature* **502**, 693–697 (2013).
- Wang, Y. et al. Reassessing acetyl-CoA supply and NADPH availability for mevalonate biosynthesis from glycerol in *Escherichia coli*. *Biotechnol. Bioeng.* **119**, 2868–2877 (2022).
- Zhang, B., Yu, M., Zhou, Y., Li, Y. & Ye, B. C. Systematic pathway engineering of *Corynebacterium glutamicum* S9114 for L-ornithine production. *Micro. Cell Fact.* **16**, 158 (2017).

42. Qin, J. et al. Modular pathway rewiring of *Saccharomyces cerevisiae* enables high-level production of L-ornithine. *Nat. Commun.* **6**, 8224 (2015).
43. Schellenberger, J., Lewis, N. E. & Palsson, B. O. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J.* **100**, 544–553 (2011).
44. Papapetridis, I. et al. Improving ethanol yield in acetate-reducing *Saccharomyces cerevisiae* by cofactor engineering of 6-phosphogluconate dehydrogenase and deletion of ALD6. *Micro. Cell Fact.* **15**, 67 (2016).
45. Zheng, T. et al. Cofactor specificity of the bifunctional alcohol and aldehyde dehydrogenase (AdhE) in wild-type and mutant *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*. *J. Bacteriol.* **197**, 2610–2619 (2015).
46. King, Z. A. & Feist, A. M. Optimizing cofactor specificity of oxidoreductase enzymes for the generation of microbial production strains—OptSwap. *Ind. Biotechnol.* **9**, 236–246 (2013).
47. King, Z. A. & Feist, A. M. Optimal cofactor swapping can increase the theoretical yield for chemical production in *Escherichia coli* and *Saccharomyces cerevisiae*. *Metab. Eng.* **24**, 117–128 (2014).
48. Khoury, G. A. et al. Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity. *Protein Sci.* **18**, 2125–2138 (2009).
49. Bouzon, M. et al. Change in cofactor specificity of oxidoreductases by adaptive evolution of an *Escherichia coli* NADPH-auxotrophic strain. *mBio* **12**, e0032921 (2021).
50. Bommarreddy, R. R., Chen, Z., Rappert, S. & Zeng, A. P. A de novo NADPH generation pathway for improving lysine production of *Corynebacterium glutamicum* by rational design of the coenzyme specificity of glyceraldehyde 3-phosphate dehydrogenase. *Metab. Eng.* **25**, 30–37 (2014).
51. Batista, R. S. et al. Glycerol as substrate and NADP(+)-dependent glyceraldehyde-3-phosphate dehydrogenase enable higher production of 3-hydroxypropionic acid through the beta-alanine pathway in *E. coli*. *Bioresour. Technol.* **393**, 130142 (2024).
52. Cai, M. et al. Development of a nonauxotrophic L-homoserine hyperproducer in *Escherichia coli* by systems metabolic engineering. *Metab. Eng.* **73**, 270–279 (2022).
53. Sugiki, S., Niide, T., Toya, Y. & Shimizu, H. Logistic regression-guided identification of cofactor specificity-contributing residues in enzyme with sequence datasets partitioned by catalytic properties. *ACS Synth. Biol.* **11**, 3973–3985 (2022).
54. Ingraham, J. B. et al. Illuminating protein space with a programmable generative model. *Nature* **623**, 1070–1078 (2023).
55. Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
56. Park, J. M. et al. Flux variability scanning based on enforced objective flux for identifying gene amplification targets. *BMC Syst. Biol.* **6**, 106 (2012).
57. Kim, W. J. et al. Genome-wide identification of overexpression and downregulation gene targets based on the sum of covariances of the outgoing reaction fluxes. *Cell Syst.* **14**, 990–1001 e1005 (2023).
58. Brophy, J. A. N. et al. Engineered integrative and conjugative elements for efficient and inducible DNA transfer to undomesticated bacteria. *Nat. Microbiol.* **3**, 1043–1053 (2018).
59. Peters, J. M. et al. Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. *Nat. Microbiol.* **4**, 244–250 (2019).
60. Wang, G. et al. CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. *Nat. Microbiol.* **4**, 2498–2510 (2019).
61. Yang, L., Yurkovich, J. T., King, Z. A. & Palsson, B. O. Modeling the multi-scale mechanisms of macromolecular resource allocation. *Curr. Opin. Microbiol.* **45**, 8–15 (2018).
62. Fang, X., Lloyd, C. J. & Palsson, B. O. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat. Rev. Microbiol.* **18**, 731–743 (2020).
63. Ma, X., Ma, L. & Huo, Y. X. Reconstructing the transcription regulatory network to optimize resource allocation for robust biosynthesis. *Trends Biotechnol.* **40**, 735–751 (2022).
64. Domenzain, I. et al. Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *Nat. Commun.* **13**, 3766 (2022).
65. Oh, Y. K., Palsson, B. O., Park, S. M., Schilling, C. H. & Mahadevan, R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.* **282**, 28791–28799 (2007).
66. Zhang, Y. et al. A new genome-scale metabolic model of *Corynebacterium glutamicum* and its application. *Biotechnol. Biofuels* **10**, 169 (2017).
67. Monk, J. M. et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **35**, 904–908 (2017).
68. Nogales, J. et al. High-quality genome-scale metabolic modelling of *Pseudomonas putida* highlights its broad metabolic capabilities. *Environ. Microbiol.* **22**, 255–269 (2020).
69. Lu, H. et al. A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun.* **10**, 3586 (2019).
70. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
71. Shinfuku, Y. et al. Development and experimental verification of a genome-scale metabolic model for *Corynebacterium glutamicum*. *Micro. Cell Fact.* **8**, 43 (2009).
72. Osterlund, T., Nookaew, I., Bordel, S. & Nielsen, J. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst. Biol.* **7**, 36 (2013).
73. Mei, J., Xu, N., Ye, C., Liu, L. & Wu, J. Reconstruction and analysis of a genome-scale metabolic network of *Corynebacterium glutamicum* S9114. *Gene* **575**, 615–622 (2016).
74. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COntstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).
75. Lewis, N. E. et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* **6**, 390 (2010).
76. Kim, P. J. et al. Metabolite essentiality elucidates robustness of *Escherichia coli* metabolism. *Proc. Natl. Acad. Sci. USA* **104**, 13638–13642 (2007).
77. Mahadevan, R. & Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5**, 264–276 (2003).
78. Desouki, A. A., Jarre, F., Gelius-Dietrich, G. & Lercher, M. J. Cycle-FreeFlux: efficient removal of thermodynamically infeasible loops from flux distributions. *Bioinformatics* **31**, 2159–2165 (2015).
79. Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).
80. Kim, G. B. et al. Comprehensive evaluation of the capacities of microbial cell factories. *Figshare* <https://doi.org/10.6084/m9.figshare.27874275> (2025).
81. Kim, G. B. et al. Comprehensive evaluation of the capacities of microbial cell factories. *Zenodo* <https://doi.org/10.5281/zenodo.15008089> (2025).
82. Kim, G. B. et al. Comprehensive evaluation of the capacities of microbial cell factories. *Zenodo*, <https://doi.org/10.5281/zenodo.15008103> (2025).

Acknowledgements

This work is supported by the Development of platform technologies of microbial cell factories for the next-generation biorefineries project (2022M3J5A1056117), and Development of advanced synthetic biology source technologies for leading the biomanufacturing industry project (RS-2024-00399424) from National Research Foundation supported by the Korean Ministry of Science and ICT.

Author contributions

S.Y.L. conceived and designed the study. G.B.K., H.R.K., and S.Y.L. analyzed the data. G.B.K. contributed analysis tools. G.B.K., H.R.K., and S.Y.L. wrote the paper.

Competing interests

The authors declare no competing interests.

Declaration of generative AI and AI-assisted technologies in the writing process

We initially wrote the complete paper and then utilized ChatGPT for English grammar checking. Following the grammar editing, the authors reviewed and revised the content to ensure its accuracy.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-58227-1>.

Correspondence and requests for materials should be addressed to Sang Yup Lee.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025