



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

## **Computational biology predicts metabolic engineering targets for increased production of 103 valuable chemicals in yeast**

Downloaded from: <https://research.chalmers.se>, 2025-04-01 22:47 UTC

Citation for the original published paper (version of record):

Domenzain Del Castillo Cerecer, I., Lu, Y., Wang, H. et al (2025). Computational biology predicts metabolic engineering targets for increased production of 103 valuable chemicals in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 122(9). <http://dx.doi.org/10.1073/pnas.2417322122>

N.B. When citing this work, cite the original published paper.



# Computational biology predicts metabolic engineering targets for increased production of 103 valuable chemicals in yeast

Iván Domenzain<sup>a,b,1</sup> , Yao Lu<sup>c,1</sup>, Haoyu Wang<sup>d,e,f</sup> , Junling Shi<sup>g</sup>, Hongzhong Lu<sup>d,2</sup>, and Jens Nielsen<sup>a,b,h,2</sup>

Affiliations are included on p. 8.

Contributed by Jens Nielsen; received September 2, 2024; accepted January 22, 2025; reviewed by Costas D. Maranas and Isabel Rocha

Development of efficient cell factories that can compete with traditional chemical production processes is complex and generally driven by case-specific strategies, based on the product and microbial host of interest. Despite major advancements in the field of metabolic modeling in recent years, prediction of genetic modifications for increased production remains challenging. Here, we present a computational pipeline that leverages the concept of protein limitations in metabolism for prediction of optimal combinations of gene engineering targets for enhanced chemical bioproduction. We used our pipeline for prediction of engineering targets for 103 different chemicals using *Saccharomyces cerevisiae* as a host. Furthermore, we identified sets of gene targets predicted for groups of multiple chemicals, suggesting the possibility of rational model-driven design of platform strains for diversified chemical production.

metabolic engineering | yeast | genome scale modeling | synthetic biology

The accelerated rise of metabolic engineering, the rewiring of cells metabolism for enhanced production of metabolites (1), and synthetic biology, the assembly of novel synthetic biological components and their integration into cells (2), has enabled the development of microbial strains with increased production capabilities of chemicals from renewable feedstocks. These engineered microbes, also known as microbial cell factories (MCF), have been generated for production of multiple specialized compounds, such as pharmaceuticals (3, 4), biofuels (5, 6), food additives (7, 8) and platform chemicals (9). Most of these cases have relied on use of the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae* as platform cell factories. Despite success in development of many processes, complete development of MCFs usually takes several years of research and costs USD50M, on average, in order to bring a proof-of-concept strain forward for commercial production (10).

As metabolism is a complex and highly interconnected network, the time- and resource-intensive process of MCF development can be alleviated by the use of genome-scale metabolic models (GEMs) together with computational algorithms, aiming to find non-intuitive gene engineering targets for enhanced production (11). Several methods for MCF design have been developed in past years and used to drive metabolic engineering projects such as production of lycopene (12, 13) and lactate (14) in *E. coli*, and drug precursors in *S. cerevisiae* cells (15). In general, the most widely used methods for MCF design (MOMA, FSEOF, optKnock, and optForce) (12, 16–18) tend to predict extensive lists of gene target candidates, and although they are capable of generating ranked lists and prioritize suggestions, they lack integration with additional analysis performed. Therefore, modelers often find themselves in need of imposing custom criteria to delimit the number of candidate gene targets to be tested, in order to reduce the amount of experimental work. Additionally, state-of-the-art GEMs tend to overpredict metabolic capabilities of cells due to the lack of kinetic and regulatory information in their formulation, hindering their applicability for further quantitative evaluation and comparison of predicted metabolic engineering strategies. Kinetic models have also been used for the development of strain design algorithms, such as k-OptForce (19); however, the limited size of this kind of models impedes prediction of metabolic gene targets in a genome-scale (20).

Here, we present a computational method (ecFactory) for prediction of optimal metabolic engineering strategies, that circumvents the problem of arbitrary selection of the number of gene candidates by leveraging the vast amount of enzymatic capacity data, together with the improved phenotype prediction capabilities, of enzyme-constrained metabolic models (ecModels, generated by the GECKO toolbox) (21). The performance of ecFactory was systematically tested and evaluated by comparing the predictions with experimental data for multiple study cases. Using this method we identified gene targets

## Significance

Prediction of metabolic engineering targets through mathematical modeling can reduce costs of strain development. Here, we developed a computational pipeline, ecFactory, designed to predict optimal gene targets for enhancing the production of 103 valuable chemicals by *Saccharomyces cerevisiae* using enzyme-constrained metabolic models. The ecFactory helps to reduce the extensive lists of candidate gene targets for metabolic engineering, thereby simplifying the experimental validation process and accelerating the development of high-producing strains. The predictive capabilities of ecFactory were assessed based on reported experimental results, showcasing that it could successfully predict the gene targets for strain metabolic engineering. The identification of common gene targets for groups of chemicals opens up the possibility of creating platform strains with versatile applications.

Copyright © 2025 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

<sup>1</sup>I.D. and Y.L. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: hongzhonglu@sjtu.edu.cn or nielsenj@chalmers.se.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2417322122/-/DCSupplemental>.

Published February 25, 2025.

for increased production of 103 different chemicals in *S. cerevisiae*, enabling identification of gene targets common to multiple groups of products, suggesting the opportunity for development of platform strains that can be used for diverse chemical production. Moreover, our analysis quantitatively estimates the production capability of representative chemicals in ecModels, revealing the limitation of enzymes or substrates on them. Altogether, we expect this integrative computational toolbox with those product-specific reaction resources to facilitate significant advancements in development of yeast MCFs through systematic metabolic engineering in coming years.

## Results

**Modeling Production of 103 Chemical Products in Yeast.** A list of 103 industrially relevant natural products, whose metabolic production pathways are known and reported in the literature, was collected. Products were grouped into 10 different families according to their chemical characteristics: amino acids, terpenes, organic acids, aromatic compounds, fatty acids and lipids, alcohols, alkaloids, flavonoids, bioamines, and stilbenoids. From these, 50 products were found to be native metabolites in *S. cerevisiae*, while 53 products were identified as heterologous, according to an enzyme-constrained metabolic model for yeast (ecYeastGEM v8.3.4) (21). A summary of the chemical classification of products is shown in Fig. 1A and *SI Appendix, Table S1*. Production pathways were reconstructed for all these heterologous products and incorporated into ecYeastGEM, during which all the corresponding heterologous reactions and enzyme kinetic data were taken into account so that ecYeastGEM could successfully predict the formation of the target product (*SI Appendix*). All of the 53 reconstructed heterologous pathways are described in the *Dataset S1*.

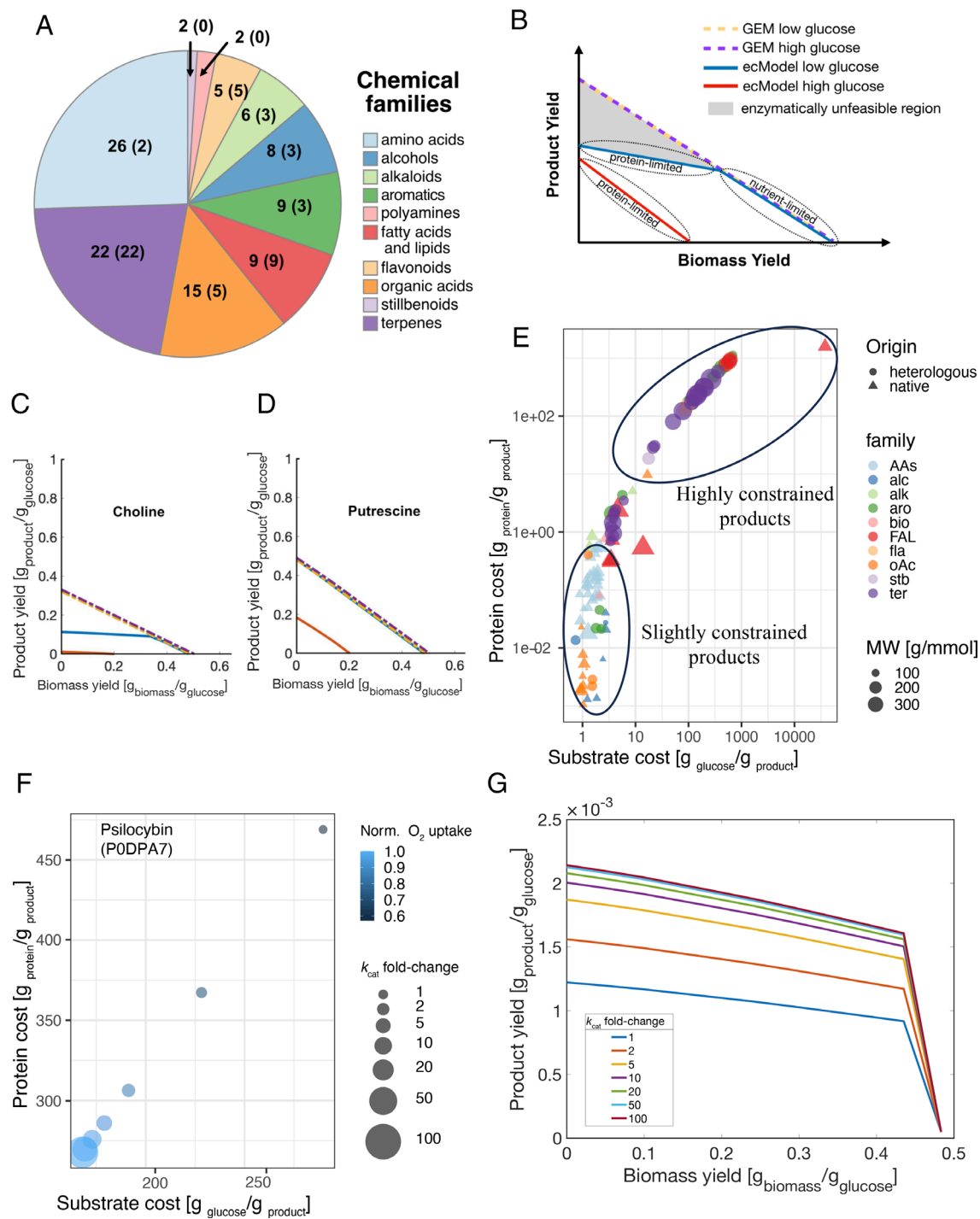
**In Silico Assessment of Production Capabilities for 103 Chemicals in Yeast.** The production envelope of *S. cerevisiae* was quantitatively explored to evaluate its maximal potential in synthesize distinct biochemicals, using both YeastGEM and ecYeastGEM (Fig. 1B), by computing optimal production yields for all of the 103 studied chemicals, constrained by low and high glucose consumption regimes (1 mmol/gDw h; and 10 mmol/gDw h) and biomass production rates spanning the range between zero and a maximum attainable value, using flux balance analysis (FBA) simulations (22).

As FBA relies on optimality principles, usually assuming maximization of cellular growth as a cellular objective (23), there is a trade-off between biomass formation and accumulation or secretion of products of interest. Yeast has evolved the ability to switch to mixed respiro-fermentative metabolic regimes when nutrients are available in excess, favoring enzymatic efficiency over biomass yield on substrate (24–27). As ecModels account for a limited enzymatic machinery in cells, different production capabilities are predicted when changing from low to high glucose uptake rate, in contrast to classic GEMs, that solely rely on stoichiometric constraints. This additional constraint results in a different production phase-plane as illustrated by Fig. 1B, i.e., instead of the standard linear trade-off between product formation and biomass formation there will be a regime where the product formation is limited by the protein constraint. Furthermore, the phase-plane becomes dependent on the glucose uptake rate, such that at high glucose consumption the ecModel predicts a protein-limited regime of production, yielding lower production levels and biomass formation per unit of glucose. Protein limitations may also arise at low glucose consumption levels, for cases in which the production pathways for the chemical of

interest involve inefficient enzymes (low specific activity). This introduces enzymatically unfeasible regions in the production space of a cell, indicated by the gray region in Fig. 1B. A typically protein-constrained production landscape with a region of difference between YeastGEM and ecYeastGEM predictions in the low glucose regime is shown for the alkaloid choline in Fig. 1C. In contrast, a production landscape solely governed by stoichiometric constraints at low glucose levels is shown in Fig. 1D for the polyamine putrescine. Additional examples of yield plots for chemicals belonging to all studied families can be found in *SI Appendix, Fig. S1*.

Highly protein-constrained products were found by identifying those chemicals whose maximum production level (or yield) demands the totality of the available enzyme mass in the model, at low levels of glucose consumption. In total, 40 out of the 53 analyzed heterologous products were found to be highly protein-constrained, in comparison to production of native metabolites, for which just 5 products were classified as part of the same group (*SI Appendix, Fig. S2A*). Furthermore, strong protein limitations arise often for groups of heterologous chemicals derived from a native pathway with high enzymatic demands, such as terpenes and flavonoids, derived from the mevalonate pathway. On the other hand, few strongly protein-limited products were found among families connected to native biosynthetic processes, such as amino acids, organic acids, and diverse alcohols (*SI Appendix, Fig. S2B*). Protein-constrained models offer the possibility of computing optimal costs of chemical production both in terms of substrate and required protein mass. Minimal protein and substrate mass costs per unit mass of product were computed for each of the 103 products, during which the respective product secretion reaction was set as the objective function for each case and the glucose uptake rate was constrained with an upper bound of 1 mmol/gDW.h (*SI Appendix*) (28). Fig. 1E shows that a positive correlation between these two production costs exists, allowing the identification of slightly and highly constrained groups of products, with an overrepresentation of native products (amino acids, organic acids, and some alcohols) in the former group, and heterologous chemicals (terpenes, flavonoids, and some aromatic compounds) in the latter. This plot shows that for heterologous products, it is usually necessary to invest on improving enzyme properties, i.e., increase their catalytic efficiency, whereas for native products it is predominantly stoichiometric constraints that should be considered for minimizing costs. Moreover, it was found that slightly constrained products tend to be lighter, in terms of molecular weight, than those in the highly constrained group. This is also suggested by substrate costs, as larger organic molecules require more carbon to be formed, notwithstanding, this also suggests that a heavier enzymatic burden is needed for assembly of large molecules, as it is likely that additional, and less efficient, enzymatic steps are involved in their synthesis.

The effect of increasing enzyme catalytic efficiency for improving production levels was explored with FBA simulations using ecYeastGEM at different activity levels of rate-limiting enzymes. For highly protein-constrained products, such as the alkaloid psilocybin, a monotonic linear decrease of the substrate cost is observed when decreasing the total protein production cost by enhancing the activity of the heterologous tryptamine 4-monooxygenase (P0DPA7). Fig. 1F shows that when the P0DPA7 catalytic efficiency is increased by 100-fold, the total oxygen consumption is predicted to increase by 75%, which suggests that reducing the protein burden of the psilocybin biosynthetic pathway releases protein mass that can be used by the cell to meet its energy demands by an increased respiratory rate. Overall, this metabolic rewiring shifts the psilocybin production space in a direction of



**Fig. 1.** Exploration of chemical production in yeast using enzyme-constrained metabolic modeling. (A) Chemical classification of 103 chemicals for in silico prediction. Numbers within parentheses indicate the number of heterologous products in the different families, those outside the parentheses indicate the total number of products in the family. (B) Production landscape predicted by a metabolic model with and without enzyme constraints at low and high glucose uptake levels. (C) Production yield plot for the highly protein-constrained product choline. (D) Production yield plot for the substrate-limited putrescine. (E) Predicted substrate and protein cost of chemical production in yeast. Product origin, chemical classification, and molecular weights are indicated by the characteristics of the 103 markers. (F) Effect of PODPA7 (tryptamine 4-monooxygenase) enzyme capacity on the production cost for psilocybin production. (G) Psilocybin production space predicted by ecYeastGEM.

higher product yields (Fig. 1G). However, the product yield is still low indicating that other enzymes in the pathway may have to be improved to further increase yield. A similar behavior was obtained for the case of valencene, a moderately protein-constrained terpene, by increasing activity levels of the sole heterologous limiting enzyme, terpene synthase (S4SC87), from 1 to a 100-fold. A positive correlation was also observed between substrate and protein costs for this product (SI Appendix, Fig. S3), however, lower

slopes in the production cost space were obtained for higher activity values of S4SC87. SI Appendix, Fig. S3 illustrates that increased activity of this limiting enzyme reduces the enzymatically unfeasible region of the valencene production space, bringing its optimal production line closer to the stoichiometrically constrained limit. Here, only the  $k_{cat}$  values of key enzymes were evaluated. However, the enzyme abundance should be also paid much attention in actual metabolic engineering as the protein expression is

controlled strictly at multiscale levels from transcription to maturation, whereby carbon and redox costs may not always be the limiting factor for enzyme abundance.

**An Integrative Constraint-Based Method for Prediction of Metabolic Engineering Strategies.** The flux scanning with enforced objective function algorithm (FSEOF) (12) has been successfully used for identification of metabolic engineering targets in yeast, due to its implicit consideration of the tradeoff between biomass and metabolite production. It is of particular interest to explore this method in the context of ecModels as it can predict engineering strategies that minimize the substrate and protein costs for optimal bioproduction. For example, it is verified that the predictions by FSEOF on top of ecModels have boosted heme accumulation in yeast cells by 70-fold (29). In order to ensure predictive robustness and minimizing the number of false positives among predictions, we revised and systematized this approach and developed ecFactory, a multistep constrained-based method for prediction of engineering gene targets for enhanced biochemical production, based on the principles of FSEOF and on the ability of ecModels to compute enzyme demands for biochemical reaction, providing systematic criteria to predict an optimal minimal set of modifications for increasing production of target metabolites.

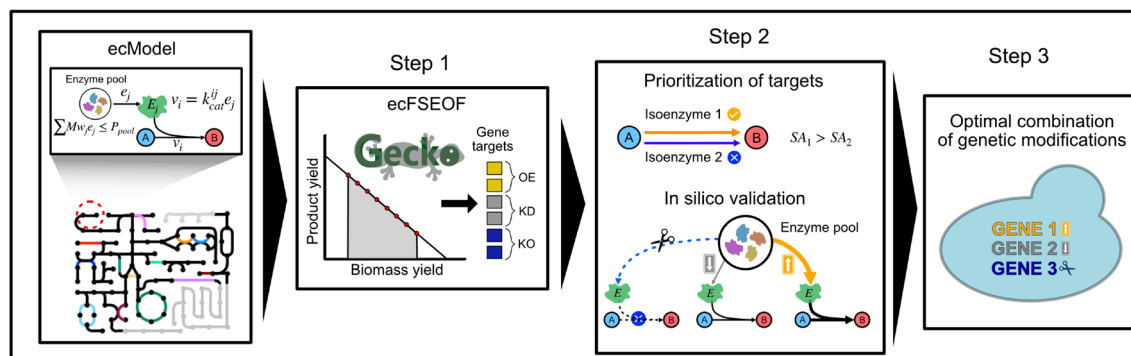
In summary, ecFactory consists of three basic steps: 1) prediction of gene expression scores, indicating intensity and directionality of genetic modifications; 2) discard gene targets encoding for unfavorable enzymes (redundant, low efficiency) and; 3) obtention of a minimal combination of modifications required for driving cells from optimal biomass formation to a metabolic production regime. The overall objective of this method is to obtain a reduced list of gene targets, focusing on the optimal strategies for enhanced production by taking enzyme allocation and connectivity into account. All the constituent steps of the ecFactory method are illustrated in Fig. 2 and explained in detail in *SI Appendix*. Note that the input ecModels for ecFactory can be in the standard SBML format. Thus, all existing ecModels from other species can be also directly fed into the pipeline of ecFactory.

Furthermore, the classification of targets according to the biological role of their respective enzymes (illustrated by *SI Appendix*, Fig. S4 and S5), facilitates a deeper understanding of the predicted optimal metabolic engineering strategies. The list of 12 gene targets for 2-phenylethanol (*SI Appendix*, Table S2) suggests that, in order to increase production of this chemical, enzymes that are optimal for providing the necessary metabolic precursors and cofactors are predicted as targets for overexpression. Knock-down and knock-out targets aim to direct the metabolic flux toward

optimal production while reducing the formation of biomass precursors in excess (glycerolipids in this case).

**Enzyme Constraints Enable Prediction of Optimal Combinations of Genetic Modifications for 103 Chemicals in Yeast.** The ecFactory method was used to predict gene targets for enhanced production of each of the 103 chemicals. The method proved to be effective at returning predictions for all cases, while reducing the number of candidate gene targets at each of its sequential steps. The distributions of the predicted number of gene targets per product (shown in Fig. 3A) show the major contribution of classifying targets according to their enzymatic characteristics (step 2) at reducing the number of predicted OEs, KDs, and KOs. On average the first step of the method (FSEOF), running on ecYeastGEM, predicted 85 gene targets per product (28 OEs, 42 KDs, and 15 KOs, on average), the number of targets is then reduced by the following steps by 73%, as only optimal candidates are returned by the ecFactory algorithm (7 OEs, 9 KDs, and 5 KDs per product, on average). Notably, predictions reveal that increasing production of protein-limited and heterologous chemicals require significantly more genetic modifications, compared to substrate-limited and native products ( $P$ -values =  $1.16 \times 10^{-5}$  and  $2.3 \times 10^{-3}$ , respectively, under a one-sided two-sample Kolmogorov–Smirnov test). These differences are caused by the large number of gene knock-downs and knock-outs that are required to reallocate the protein resource so that the limited cellular enzyme capacity can be optimally allocated to the final production reaction steps, which tend to be inefficient for these kinds of products. A more detailed presentation of results, by chemical family, method steps, and target types, is available in *SI Appendix*, Table S3.

Overall, 150 endogenous genes in yeast are predicted as OE target for at least one of the modeled products; 88 different genes are predicted as KD targets and 129 as KO targets. More than 50% of the targets predicted for OE, KD, and KO are specific to one or two of the 103 products (Fig. 3B and *SI Appendix*, Fig. S6 A and C). Nonetheless, small sets of genes are predicted as targets for a high number of products (also named as highly frequent targets), spanning almost all chemical classifications in this study. Genes encoding for reaction steps in the pentose-phosphate pathway and pyruvate metabolism, together with PFK2 in the glycolysis pathway, are predicted as the most common OE targets across products; the most common KD and KO gene targets encode for enzymes in the TCA cycle, oxidative phosphorylation, and synthesis of biomass precursors (steroids, glycerolipids, nucleotides and amino acids), as shown in Fig. 3C and *SI Appendix*, Fig. S6 B and D, suggesting it as a relevant strategy for redirecting



**Fig. 2.** Prediction of metabolic engineering targets with ecFactory. A metabolic model with enzyme constraints is used for (1) prediction of gene targets for rewiring flux toward increased production. (2) Gene targets are classified and filtered according to enzymatic efficiency and connectivity. (3) A minimal combination of targets for sustaining optimal production levels is obtained.

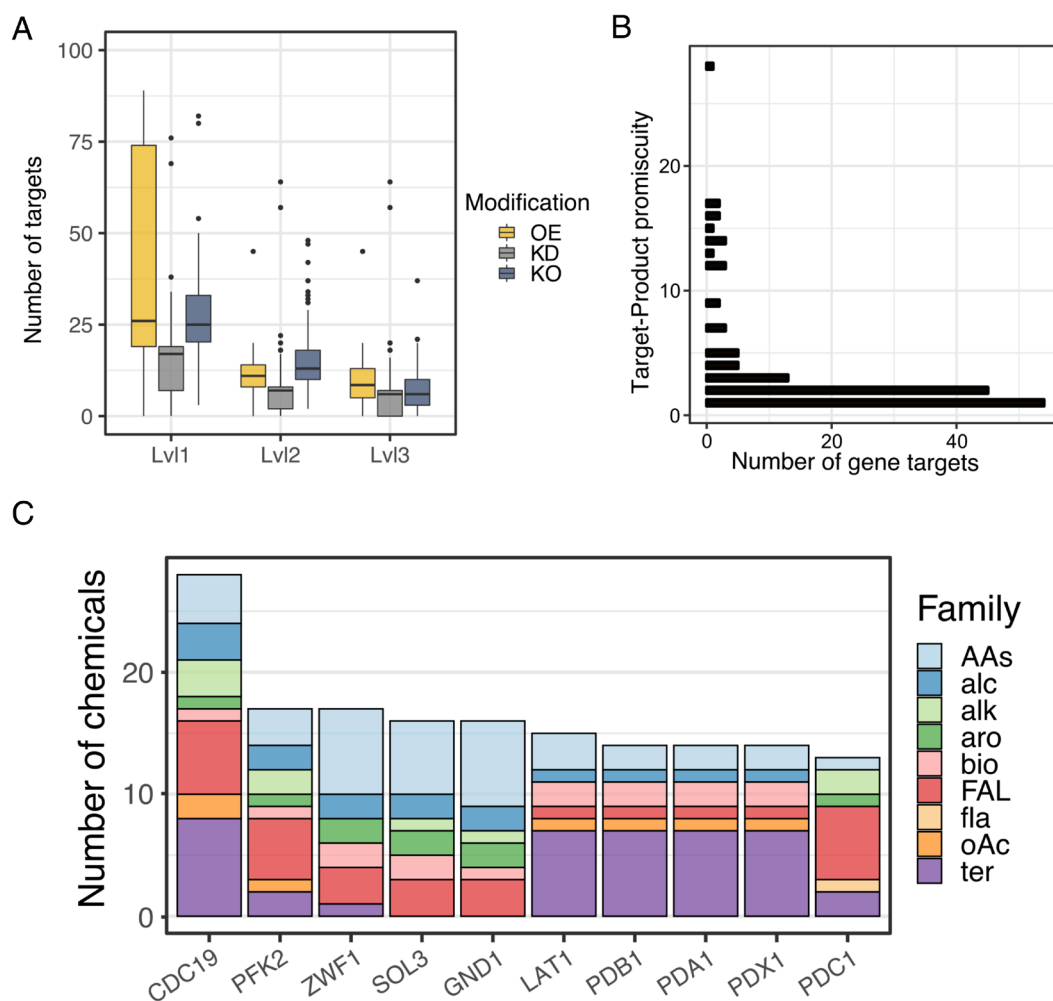
carbon flux into heterologous pathways and alternative energy production mechanisms.

**In Silico Predictions Capture Validated Metabolic Engineering Strategies in Yeast.** It was found that 7 out of the 12 predicted gene targets to increase 2-phenylethanol have been previously engineered in *S. cerevisiae*, *Yarrowia lipolytica*, and *Kluyveromyces marxianus* strains with enhanced 2-phenylethanol production levels (30–32) (*SI Appendix, Table S2*), indicating that ecFactory predictions can capture gene targets proposed by rational engineering approaches. As another example of experimentally validated predictions, the case of spermidine is of particular interest. In this case, the ecFactory method was able to capture 9 of 85 predicted gene targets (OEs: SPE1, SPE2, SPE3, MEU1, ORT1, AGC1, and GDH1; KO/KD: CAR2 and SPE4) in a successfully engineered strain for spermidine production, an immediate precursor of spermine (33). Note that ecFactory could also identify some additional gene targets besides experimentally verified targets. It was found that the experimental implementation of a heterologous cytosolic ornithine cycle was resembled by a general predicted overexpression of its native mitochondrial version, but ecFactory also identified the targets ARG8 and AGC1 (*SI Appendix, Fig. S7*).

These results suggest that the method is able to capture the underlying logic of highly complex rational engineering approaches

that require the coordination of multiple sectors at the metabolic network level. Overall, the predicted gene modifications aim to increase spermidine biosynthesis by overexpression of the whole ornithine cycle, a direct precursor, together with the Yang cycle and some steps in the pentose phosphate pathway (PPP) in order to increase S-adenosyl-L-methionine, another important precursor of polyamines. Interestingly, when focusing on the final prediction for this product from the third step of ecFactory, which could generate a minimal combination of genetic modifications for the optimal production, 6 of the 29 predicted optimal gene targets were experimentally verified for spermidine production (OEs: SPE1, SPE2, SPE3, MEU1; KO/KD: SPE4, CAR2). Interestingly, it is found that the ratio of verified gene targets per total predicted gene targets could be continuously improved when running the filtration steps within ecFactory using ecModels compared to that from FSEOF (12) and OptForce (18) using GEMs (*SI Appendix, Fig. S8*). This suggests that, according to enzyme capacity and metabolic connectivity, it is possible to reduce complex rational metabolic engineering strategies, to fewer modifications on crucial reaction steps in pathways that need to be rewired and coordinated, one of the purposes for which this method was designed.

In order to further validate the quality of the ecFactory predictions, we searched the literature for independent experimental studies in *S. cerevisiae* that have been successful at increasing production levels of chemicals included in our list. Gene modifications



**Fig. 3.** Prediction of gene engineering targets for increased production of 103 chemicals in yeast. (A) Distribution of the number of gene targets per product predicted at different steps in the ecFactory pipeline. Level 1, FSEOF-based prediction; Level 2, filtering by enzyme characteristics; Level 3, obtention of minimal set of targets for optimal production. (B) Distribution of product specificity of gene targets across 103 chemicals. (C) Representation of the presence of the top 10 most common predicted overexpression targets across products and families.

validated for diverse chemicals were found to be predicted as optimal gene targets by ecFactory, shown in *SI Appendix, Table S4*. Interestingly, several of these targets are common to multiple products. In total, 30 different gene targets from prediction were found as experimentally validated across 18 products, which are also part of different chemical classes. The most repeated genes among these targets correspond to overexpression in the ergosterol, mevalonate, shikimate, and polyamine biosynthesis pathways.

These similarities at the gene and pathway level among predictions for different chemical products, suggest the existence of metabolic engineering strategies capable of providing the necessary precursors for increasing production of groups of chemicals. This kind of strategies have been sought in experimental metabolic engineering, following rational approaches, and have proved to be successful for the development of platform yeast strains for production of different groups of molecules such as opioids (4) and other alkaloids (34, 35), polyketides (36), and terpenes (37, 38). Furthermore, cumulative combination of individual genetic modifications in a production strain is needed for achieving meaningful flux toward the desired chemical (29), therefore, it is desirable to identify multiple gene targets, encompassing multiple metabolic pathways, that constitute the chassis for robust and diversified chemical production.

Gene targets common to all products in a given chemical family were sought for all cases in this study. The only chemical family with common predicted targets was found to be flavonoids, with 9 KDS (ADO1, ATP19, IDP1, LPD1, MAE1, MDH2, MET6, PPA2, and SAH1) and 7 KOS (CAR2, FAA, FAA4, FDH1, RNR1, RNR3, and RNR4). This combination of targets reveals an engineering strategy that decreases the TCA cycle and respiratory fluxes, the amount of carbon going toward acetyl-CoA and posterior fatty acid synthesis, synthesis of amino acids derived from 2-oxoglutarate and nucleotides biosynthesis. Altogether, this shows an optimal way of allocating carbon flux and the limited enzymatic machinery of yeast for the biosynthetic pathways producing catechin, genistein, kaempferol, naringenin, and quercetin. Nevertheless, the impact of these modifications on other biological processes, such as regulatory networks, is not accounted for in the metabolic model and should be further assessed.

**Model-Driven Design of Platform Strains for Diversified Chemical Production.** As highly frequent gene targets, for all kinds of modifications, were found to be predicted for products present in most of the studied chemical families, other sets of targets common to groups of multiple products may exist among the ecFactory predictions. In order to systematize the analysis of gene target profiles across products, the 103 lists of targets were represented as mathematical vectors (Fig. 4A for further details and *SI Appendix*). Highly similar gene expression vectors were identified using the t-distributed stochastic neighbor embedding method (t-SNE), which is suited for visualization and identification of clusters in high dimensional datasets (39). Two-dimensional representation of t-SNE results facilitated identification of 8 different clusters of target vectors, representing different groups of products. Product clusters are shown in Fig. 4B. Notably, gene targets common to all products in a group were found for all clusters (*SI Appendix, Table S5*).

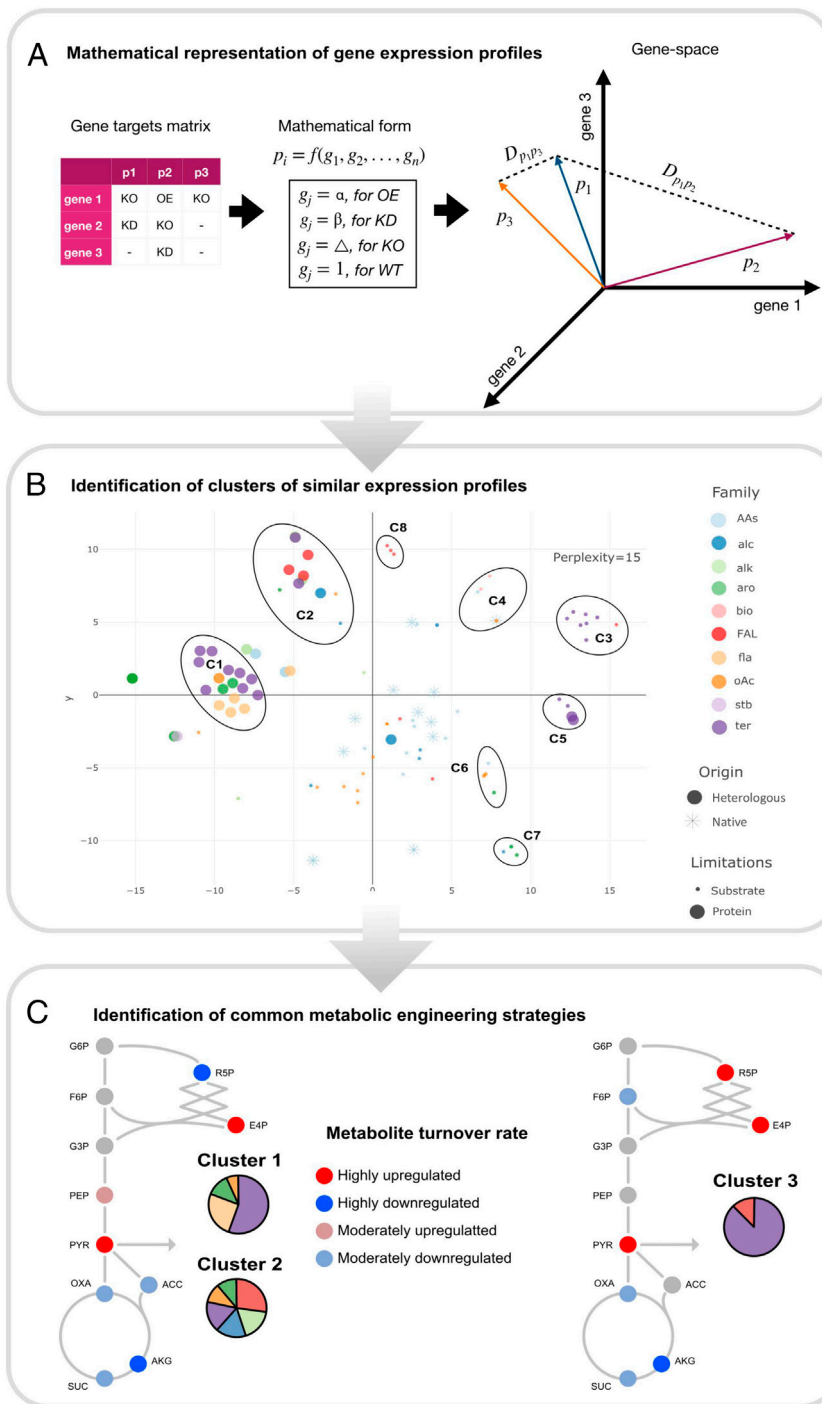
In general, these clusters are composed of products that belong to different chemical families, except for clusters 3 and 5, composed mostly by terpenes, and cluster 8, formed just by lipid compounds. Mapping product origin (native or heterologous) and protein limitations information into the clustering results showed that clusters 1 and 2 are composed by heterologous and highly protein-constrained products belonging to different compound

classes; terpenes whose production is constrained by substrate availability tend to group together, in cluster 3; and most native products, despite their protein limitations, do not fall into the identified clusters. Altogether, this shows that metabolic engineering strategies for the different product clusters are defined by gene modifications that are related with redirecting flux and energy from central metabolism to the final specific heterologous pathways. This suggests that shared molecular characteristics between products (i.e., chemical classification of products) might not be the most decisive aspect when designing genetic modification strategies for enhanced production of multiple chemicals (platform or chassis strains).

In order to understand the particular metabolic rewiring required by each platform strain designed with the aid of the cluster analysis, turnover rates were calculated for the 12 main precursor metabolites in central carbon metabolism (D-glucose-6-phosphate, D-fructose-6-phosphate, ribose-5-phosphate, erythrose-4-phosphate, glyceraldehyde-3-phosphate, 3-phosphoglycerate, phosphoenolpyruvate, pyruvate, acetyl-CoA, 2-oxoglutarate, succinyl-CoA, and oxaloacetate) (11) using FBA simulations with ecYeastGEM for different scenarios, optimal biomass formation, and optimal production of each of the studied chemicals. Fold-changes were then computed for each of the precursor turnover rates, by comparing the optimal production flux distributions to their optimal biomass formation counterpart, for all 103 production scenarios. In this way fold-changes higher than one indicate that, for increased production, the overall flux toward a precursor should be upregulated, in comparison to a wild-type metabolic state, while fold-changes lower than one imply that the flux toward a precursor needs to be down-regulated (*SI Appendix*).

Fig. 4C shows that significant upregulation of flux toward erythrose-4-phosphate (E4P) and pyruvate, moderate upregulation of phosphoenolpyruvate (PEP), a drastic decrease in ribose-5-phosphate (R5P) and  $\alpha$ -ketoglutarate (AKG) turnover rates and, a moderate downregulation of the flux toward oxaloacetate (OXO), acetyl-CoA, and succinyl-CoA should be combined to achieve optimal production levels of the products in clusters 1 and 2. Additionally, it can be seen that fluxes toward precursors located downstream from pyruvate (TCA cycle intermediates and acetyl-CoA) are needed to be downregulated for products in these clusters. This can be explained by a lower demand of building blocks, due to the decrease of biomass formation rate in a production scenario. Moreover, as all products in these clusters were found to be protein-limited, a predicted coordinated downregulation of the lower section of central carbon metabolism suggests that forcing a fermentative regime, in which most of the energy is produced by glycolysis to minimize the protein burden induced by cellular respiration, thus, leaving room for expression of inefficient heterologous enzymes, offers the optimal conditions (metabolic mode) for production of these chemicals.

For the case of products in cluster 3, predictions indicate that a metabolic rewiring that induces significant upregulation of R5P, E4P, and pyruvate production, and intense downregulation of the flux toward  $\alpha$ -ketoglutarate is needed to improve production of these terpene compounds (Fig. 4D), suggesting that an increased supply of NADPH (produced in the first steps of the pentose phosphate pathway, preceding ribose-5-phosphate) is needed for these products. The gene target profiles for the bioamines putrescine and spermine were found to cluster together with their precursor amino acids proline and glutamate, as well as itaconic acid (cluster 4). *SI Appendix, Fig. S9A* shows that genetic modifications common to all products in this cluster cause only moderate changes in the turnover rate of central carbon metabolism precursors, mostly for those in lower glycolysis, indicating that the



**Fig. 4.** Model-driven design of platform strains for diverse chemical production. (A) representation of gene targets for optimal production as mathematical vectors. (B) Identification of clusters of products with highly similar predicted engineering targets using t-SNE. Chemical families are indicated as AA, amino acids; Alc, alcohols; Alk, alkaloids; Aro, aromatics; Bio, bioamines; FAL, fatty acids and lipids; fla, flavonoids; oAc, organic acids; stb, stilbenoids; ter, terpenes. (C) FBA predicts cluster-specific metabolic demands. Fold-change in turnover rate of the main metabolic precursors, compared to wild-type, necessary for optimal production of the products in clusters 1, 2, and 3

optimal production mode for these products does not differ significantly from a wild-type optimal growing metabolic strategy. A strong requirement for increased flux toward E4P was found to be common to all terpenes, despite the protein limitations involved in their production pathways, as shown by Figs. 4 C and D and *SI Appendix, Fig. S9B*.

Production of native and heterologous products derived from the shikimate pathway, those in clusters 6 and 7, were found to require an increase of flux toward the immediate precursors E4P

and PEP, together with enhanced NADPH supply, provided by an increased flux to R5P, and a reduction of the metabolic turnover of precursors located downstream of PEP, in order to maximize carbon conversion (*SI Appendix, Fig. S9C*). Finally, significant increase of acetyl-CoA turnover, together with a moderate upregulation of the pentose-phosphate pathway for increased NADPH flux, was found to be the optimal reprogramming strategy for production of free fatty acids, oleate, and palmitoleate (*SI Appendix, Fig. S9D*), resembling previous successful work in yeast cells (40).



## Discussion

Here, we demonstrated that, by accounting for enzyme limitations, the use of metabolic models for quantitative prediction in metabolic engineering can be extended and improved. Enzyme-constrained models enabled assessment of the impact of enzyme capacity on the total protein and substrate costs of chemical production in cell factories. As with better prediction capability, several toolboxes, i.e., GECKO2 (21) and ECMpy (41), have been developed to accelerate the automatic reconstruction of ecModels for nonmodel organisms. Different from those model organisms, i.e., *E. coli* and yeast, the kinetic parameters for enzymes from nonmodel organisms are scarce, which hinders the reconstruction of high-quality ecModels and even brings potential errors in gene target prediction for metabolic engineering. To overcome this bottleneck, several novel deep learning models, like DLkcat (42), UniKP (43), and DeepEnzyme (44), were developed to predict kinetic parameters for enzymes from any metabolic network (45). Meanwhile, the enzyme parameters in the ecModel could be fine-tuned by comparing the model output with the measured dataset using Bayesian inference (46). Thus, the improved ecModels for nonmodel organisms using the above procedures will, to some extent, promote the application of ecFactory across most industrial organisms.

The model ecYeastGEM was used to predict gene engineering targets for enhanced production of 103 chemical products with yeast cells, including native and heterologous biochemicals with distinct chemical characteristics. Predictions showed to resemble complex engineering strategies that involve coordinated modulation and coordination of multiple pathways. Notably, supportive experimental evidence was found in the literature to verify the gene target predictions in 22 of the studied chemicals. Additionally, incorporation of enzyme constraints enabled drastic reduction of the number of predicted gene targets per product in contrast to purely stoichiometric constrain-based methods, aiming to prevent scenarios in which experimentalists and modelers need to arbitrarily choose targets from extensive lists of predictions. Sets of gene targets common across multiple products were identified for 8 different groups of chemicals, inferred with a clustering algorithm. FBA simulations indicate that these core genetic modifications represent the expression tuning profiles, which are needed to rewire the central carbon metabolism of yeast toward increased production of the main metabolic precursors required by groups of valuable chemicals. Implementation of these predictions in yeast cells can be used to drive the development of platform strains, specialized in providing the production scaffold for multiple chemicals. Platform strains can then be transformed into product-specific ones by introducing the necessary heterologous genetic components.

By visualizing the 8 different metabolic rewiring schemes we learned that clustering of products according to their gene target predictions complies with combinations of these three basic factors: 1) protein burden induced by the specific production pathways and its impact on energy production; 2) the specific metabolic precursor that provides the main carbon flux for final product formation; 3) products that require increased NADPH supply levels. Thus, the presented approach suggests the advantages of using enzyme-constrained models for design and understanding of platform strains optimized for diverse chemical production. Nonetheless, expanding the scope and number of chemicals and host organisms for this kind of large-scale studies might help to

unveil additional core principles for rationally engineering of metabolism.

A platform-based procedure can potentially reduce the resources and efforts involved in the development of next-generation cell factories; therefore, we envision that the tools and methodology developed in this study will contribute to accelerate development of robust and efficient microbial strains both for specialized and also versatile production of valuable chemicals, promoting the conversion from petrol to a biobased economy.

## Methods and Materials

All the *Materials and Methods* are detailed in *SI Appendix*: incorporation of heterologous metabolic pathways into ecYeastGEM; Computing production capabilities using GEMs and ecModels; Computing substrate and protein costs of biochemical production using ecModels; Prediction of metabolic engineering gene targets using ecFactory; Computation of genetic target profile vectors; Computation of optimal flux vectors; Cluster analysis of gene target profiles; Metabolic precursors analysis. The model simulations and gene target prediction for selected chemicals were carried out in MATLAB using MATLAB toolbox with the Gurobi solver (Gurobi Optimization, LLC).

**Data, Materials, and Software Availability.** The 103 lists of gene targets for increased chemical production in yeast are publicly available at: [https://github.com/SysBioChalmers/CellFactory-ecYeastGEM/tree/dev/results/production\\_targets](https://github.com/SysBioChalmers/CellFactory-ecYeastGEM/tree/dev/results/production_targets) (47). Lists of combined gene targets for the 8 proposed platform strains are available at: [https://github.com/SysBioChalmers/CellFactory-ecYeastGEM/tree/dev/results/cluster\\_strains](https://github.com/SysBioChalmers/CellFactory-ecYeastGEM/tree/dev/results/cluster_strains) (48). A detailed tutorial to run ecFactory step-by-step could be found at [https://github.com/SysBioChalmers/ecFactory/blob/main/tutorials/ecFactory\\_prediction\\_example.m](https://github.com/SysBioChalmers/ecFactory/blob/main/tutorials/ecFactory_prediction_example.m) (49). The ecFactory method is available as a standalone tool at the public repository: <https://github.com/SysBioChalmers/ecFactory> (50). All scripts for reproduction of results and figures in this article can be found at: <https://github.com/SysBioChalmers/CellFactory-ecYeastGEM> (51).

**ACKNOWLEDGMENTS.** This project has received funding from the Novo Nordisk Foundation (Grant no. NNF10CC1016517), the Knut and Alice Wallenberg Foundation, and the European Union's Horizon 2020 research and innovation program with projects DD-DeCaF and CHASSY (Grant agreements No 686070 and 720824). This project was also supported by the Shanghai Pujiang Program and the National Natural Science Foundation of China (Grants no. 22208211, 22378263, and 32302255). Portions of this work were used in the PhD thesis of the first author. We would like to thank Wenbin Liao and Xingcun Fan for their valuable comments to this paper.

Author affiliations: <sup>a</sup>Department of Life Sciences, Chalmers University of Technology, Gothenburg SE41296, Sweden; <sup>b</sup>Novo Nordisk Foundation Center for Biosustainability, Chalmers University of Technology, Gothenburg SE41296, Sweden; <sup>c</sup>College of Enology, Northwest A&F University, Yangling, Shannxi 712100, China; <sup>d</sup>State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China; <sup>e</sup>Division of Biotechnology, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China; <sup>f</sup>University of Chinese Academy of Sciences, Beijing 100049, China; <sup>g</sup>Key Laboratory for Space Bioscience and Biotechnology, School of Life Sciences, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China; and <sup>h</sup>BioInnovation Institute, Copenhagen DK2200, Denmark

Author contributions: H.L. and J.N. designed research; I.D., Y.L., and H.L. performed research; H.W. contributed new reagents/analytic tools; I.D., Y.L., H.W., and J.S. analyzed data; H.L. and J.N. supervised the work; and I.D., H.L., and J.N. wrote the paper.

Reviewers: C.D.M., Pennsylvania State University; and I.R., Universidade Nova de Lisboa Instituto de Tecnologia Química e Biológica Antonio Xavier.

Competing interest statement: Elypta (cancer diagnostic company) Melt&Marble (food tech company) Chrysea (dietary supplement company). Several patents, but none related to this work.

1. J. Nielsen, J. D. Keasling, Engineering cellular metabolism. *Cell* **164**, 1185–1197 (2016).
2. D. Julleson, F. David, B. Pfeleger, J. Nielsen, Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnol. Adv.* **33**, 1395–1402 (2015).

3. D.-K. Ro *et al.*, Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
4. S. Galanie, K. Thodey, I. J. Trenchard, M. Filsinger Interrante, C. D. Smolke, Complete biosynthesis of opioids in yeast. *Science* **349**, 1095–1100 (2015).

5. Y. J. Choi, S. Y. Lee, Microbial production of short-chain alkanes. *Nature* **502**, 571–574 (2013).
6. Y. J. Zhou *et al.*, Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. *Nat. Commun.* **7**, 1–9 (2016).
7. J. H. Park, K. H. Lee, T. Y. Kim, S. Y. Lee, Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc. Nat. Acad. Sci.* **104**, 7797–7802 (2007).
8. Y. Wei, D. Bergenholm, M. Gossing, V. Siewers, J. Nielsen, Expression of cocoa genes in *Saccharomyces cerevisiae* improves cocoa butter production. *Microbial cell factories* **17**, 1–11 (2018).
9. J. M. Otero *et al.*, Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory. *PLoS one* **8**, e54144 (2013).
10. J. Nielsen, Yeast cell factories on the horizon. *Science* **349**, 1050–1051 (2015).
11. J. Nielsen, Systems biology of metabolism. *Ann. Rev. Biochem.* **86**, 245–275 (2017).
12. H. S. Choi, S. Y. Lee, T. Y. Kim, H. M. Woo, In silico identification of gene amplification targets for improvement of lycopene production. *Appl. Environ. Microb.* **76**, 3097–3105 (2010).
13. H. Alper, Y.-S. Jin, J. Moxley, G. Stephanopoulos, Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab. Eng.* **7**, 155–164 (2005).
14. S. S. Fong *et al.*, In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol. Bioeng.* **91**, 643–648 (2005).
15. M. Suástegui *et al.*, Multilevel engineering of the upstream module of aromatic amino acid biosynthesis in *Saccharomyces cerevisiae* for high production of polymer and drug precursors. *Metabolic Eng.* **42**, 134–144 (2017).
16. D. Segre, D. Vitkup, G. M. Church, Analysis of optimality in natural and perturbed metabolic networks. *Proc. Nat. Acad. Sci.* **99**, 15112–15117 (2002).
17. A. P. Burgard, P. Pharkya, C. D. Maranas, OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84**, 647–657 (2003).
18. S. Ranganathan, P. F. Suthers, C. D. Maranas, OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.* **6**, e1000744 (2010).
19. A. Chowdhury, A. R. Zomorodi, C. D. Maranas, k-OptForce: Integrating kinetics with flux balance analysis for strain design. *PLoS Comput. Biol.* **10**, e1003487 (2014).
20. A. Khodayari, C. D. Maranas, A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nat. Commun.* **7**, 13806 (2016).
21. I. Domenzain *et al.*, Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *Nat. Commun.* **13**, 3766 (2022).
22. J. D. Orth, I. Thiele, B. Ø. Palsson, What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
23. J. M. Savinell, B. O. Palsson, Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *J. Theor. Biol.* **154**, 421–454 (1992).
24. A. Vazquez *et al.*, Impact of the solvent capacity constraint on *E. coli* metabolism. *BMC Syst. Biol.* **2**, 1–10 (2008).
25. A. Vazquez, J. Liu, Y. Zhou, Z. N. Oltvai, Catabolic efficiency of aerobic glycolysis: The Warburg effect revisited. *BMC Syst. Biol.* **4**, 1–9 (2010).
26. A. Nilsson, J. Nielsen, Metabolic trade-offs in yeast are caused by F1F0-ATP synthase. *Sci. Rep.* **6**, 22264 (2016).
27. R. Adadi, B. Volkmer, R. Milo, M. Heinemann, T. Shlomi, Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. *PLoS Comput. Biol.* **8**, e1002575 (2012).
28. Y. Chen, J. Nielsen, Yeast has evolved to minimize protein resource cost for synthesizing amino acids. *Proc. Nat. Acad. Sci.* **119**, e2114622119 (2022).
29. O. P. Ishchuk *et al.*, Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae*. *Proc. Nat. Acad. Sci.* **119**, e2108245119 (2022).
30. E.-J. Hassing, P. A. de Groot, V. R. Marquenie, J. T. Pronk, J.-M. G. Daran, Connecting central carbon and aromatic amino acid metabolisms to improve de novo 2-phenylethanol production in *Saccharomyces cerevisiae*. *Metabolic Eng.* **56**, 165–180 (2019).
31. M. Larroude, J. M. Nicaud, T. Rossignol, *Yarrowia lipolytica* chassis strains engineered to produce aromatic amino acids via the shikimate pathway. *Microbial Biotechnol.* **14**, 2420–2434 (2021).
32. A. S. Rajkumar, J. P. Morrissey, Rational engineering of *Kluyveromyces marxianus* to create a chassis for the production of aromatic products. *Microbial Cell Factories* **19**, 1–19 (2020).
33. J. Qin *et al.*, Engineering yeast metabolism for the discovery and production of polyamines and polyamine analogues. *Nat. Catal.* **4**, 498–509 (2021).
34. M. E. Pyne *et al.*, A yeast platform for high-level synthesis of tetrahydroisoquinoline alkaloids. *Nat. Commun.* **11**, 3337 (2020).
35. M. McKeague, Y.-H. Wang, A. Cravens, M. N. Win, C. D. Smolke, Engineering a microbial platform for de novo biosynthesis of diverse methylxanthines. *Metabolic Eng.* **38**, 191–203 (2016).
36. T. Jakićūnas *et al.*, Programmable polyketide biosynthesis platform for production of aromatic compounds in yeast. *Synth. Syst. Biotechnol.* **5**, 11–18 (2020).
37. M. Farhi *et al.*, Harnessing yeast subcellular compartments for the production of plant terpenoids. *Metabolic Eng.* **13**, 474–481 (2011).
38. S. Tippmann, G. Scalcinati, V. Siewers, J. Nielsen, Production of farnesene and santalene by *Saccharomyces cerevisiae* using fed-batch cultivations with RO-controlled feed. *Biotechnol. Bioeng.* **113**, 72–81 (2016).
39. L. Van der Maaten, G. Hinton, Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
40. T. Yu *et al.*, Reprogramming yeast metabolism from alcoholic fermentation to lipogenesis. *Cell* **174**, 1549–1558.e1514 (2018).
41. Z. Mao *et al.*, ECMpy 2.0: A Python package for automated construction and analysis of enzyme-constrained models. *Synth. Syst. Biotechnol.* **9**, 494–502 (2024).
42. F. Li *et al.*, Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nat. Catal.* **5**, 662–672 (2022).
43. H. Yu, H. Deng, J. He, J. D. Keasling, X. Luo, UniKP: A unified framework for the prediction of enzyme kinetic parameters. *Nat. Commun.* **14**, 8211 (2023).
44. T. Wang *et al.*, DeepEnzyme: A robust deep learning model for improved enzyme turnover number prediction by utilizing features of protein 3D-structures. *Briefings Bioinform.* **25**, bbae409 (2024).
45. F. Li, Y. Chen, M. Anton, J. Nielsen, GotEnzymes: An extensive database of enzyme parameter predictions. *Nucleic Acids Res.* **51**, D583–D586 (2022).
46. G. Li *et al.*, Bayesian genome scale modelling identifies thermal determinants of yeast metabolism. *Nat. Commun.* **12**, 190 (2021).
47. I. Domenzain, Y. Lu, H. Wang, H. Lu, SysBioChalmers/CellFactory-ecYeastGEM. GitHub. [https://github.com/SysBioChalmers/CellFactory-ecYeastGEM/tree/dev/results/production\\_targets](https://github.com/SysBioChalmers/CellFactory-ecYeastGEM/tree/dev/results/production_targets). Deposited 12 February 2025.
48. I. Domenzain, Y. Lu, H. Wang, H. Lu, SysBioChalmers/CellFactory-ecYeastGEM. GitHub. [https://github.com/SysBioChalmers/CellFactory-ecYeastGEM/tree/dev/results/cluster\\_strains](https://github.com/SysBioChalmers/CellFactory-ecYeastGEM/tree/dev/results/cluster_strains). Deposited 12 February 2025.
49. I. Domenzain, Y. Lu, H. Wang, H. Lu, SysBioChalmers/ecFactory. GitHub. [https://github.com/SysBioChalmers/ecFactory/blob/main/tutorials/ecFactory\\_prediction\\_example.m](https://github.com/SysBioChalmers/ecFactory/blob/main/tutorials/ecFactory_prediction_example.m). Deposited 12 February 2025.
50. I. Domenzain, Y. Lu, H. Wang, H. Lu, SysBioChalmers/ecFactory. GitHub. <https://github.com/SysBioChalmers/ecFactory>. Deposited 12 February 2025.
51. I. Domenzain, Y. Lu, H. Wang, H. Lu, SysBioChalmers/CellFactory-ecYeastGEM. GitHub. <https://github.com/SysBioChalmers/CellFactory-ecYeastGEM>. Deposited 12 February 2025.