

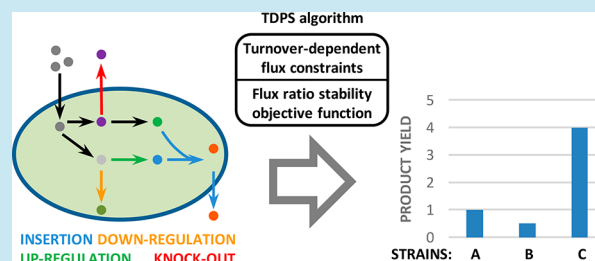
Turnover Dependent Phenotypic Simulation: A Quantitative Constraint-Based Simulation Method That Accommodates All Main Strain Design Strategies

Rui Pereira,^{†,‡} Paulo Vilaça,^{†,§} Paulo Maia,^{†,§} Jens Nielsen,^{‡,||} and Isabel Rocha^{*,†,⊥}[†]CEB – Centre of Biological Engineering, University of Minho, Campus de Gualtar, Braga 4710-057, Portugal[‡]Department of Biology and Biological Engineering, Chalmers University of Technology, SE412 96 Gothenburg, Sweden[§]SilicoLife Lda., Rua do Canastreiro 15, 4715-387 Braga, Portugal^{||}Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark[⊥]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa (ITQB-NOVA), 2775-412 Oeiras, Portugal

Supporting Information

ABSTRACT: The uncertain relationship between genotype and phenotype can make strain engineering an arduous trial and error process. To identify promising gene targets faster, constraint-based modeling methodologies are often used, although they remain limited in their predictive power. Even though the search for gene knockouts is fairly established in constraint-based modeling, most strain design methods still model gene up/down-regulations by forcing the corresponding flux values to fixed levels without taking in consideration the availability of resources. Here, we present a constraint-based algorithm, the turnover dependent phenotypic simulation (TDPS) that quantitatively simulates phenotypes in a resource conscious manner. Unlike other available algorithms, TDPS does not force flux values and considers resource availability, using metabolite production turnovers as an indicator of metabolite abundance. TDPS can simulate up-regulation of metabolic reactions as well as the introduction of heterologous genes, alongside gene deletion and down-regulation scenarios. TDPS simulations were validated using engineered *Saccharomyces cerevisiae* strains available in the literature by comparing the simulated and experimental production yields of the target metabolite. For many of the strains evaluated, the experimental production yields were within the simulated intervals and the relative strain performance could be predicted with TDPS. However, the algorithm failed to predict some of the production changes observed experimentally, suggesting that further improvements are necessary. The results also showed that TDPS may be helpful in finding metabolic bottlenecks, but further experiments would be required to confirm these findings.

KEYWORDS: genome-scale models, *Saccharomyces cerevisiae*, phenotype simulation, metabolite turnovers, network rigidity, metabolic engineering



In strain engineering, the impact of genetic alterations on cell phenotype is rarely straightforward, typically resulting in several cycles of trial and error during cell factory development. To test the effects of genetic alterations quickly and at genome scale, constraint-based modeling of metabolic fluxes can be used with several methods already having been developed for this purpose (reviewed^{1,2}). Currently, however, most computational methods developed for strain design focus only on gene knockout based strategies.^{1,3} Moreover, many of these methods were only designed for use in an optimization context; that is, they can identify possible metabolic engineering targets, but cannot predict effects that a set of genetic modifications may have on the organism's phenotype *via* quantitative simulations. Therefore, despite recent advances in constraint-based modeling, the simulation of strain engineering designs involving up and down-regulation of a few genes remains challenging.

Among the options available from literature, the under/overexpression plugin⁴ in OptFlux⁵ is the most oriented method for simulation purposes, allowing simulation of deletions as well as up- or down-regulations of genes and reactions, according to user specific preferences. The main disadvantage of this tool, however, is its inability to regulate metabolic reactions that are inactive under reference conditions. As a consequence, this algorithm is not able to simulate the activation of a given gene or reaction, nor can it simulate the introduction of heterologous genes. Although not developed for strain simulation purposes, other constraint-based methods also exist which can predict the effects of up/down-regulations based on different strategies. For example,

Received: June 10, 2018

Published: March 29, 2019

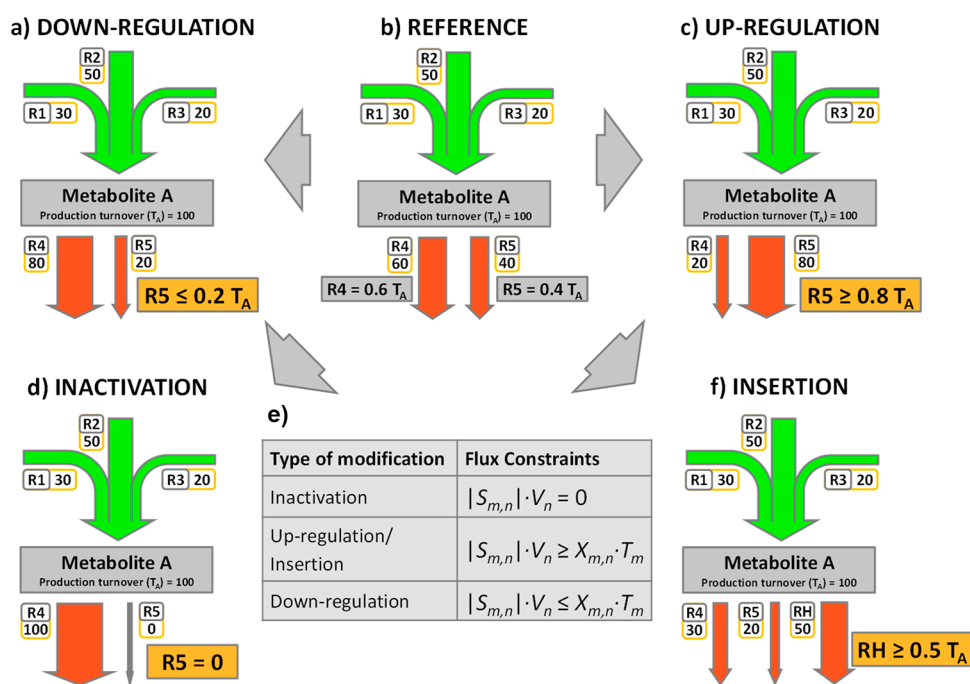


Figure 1. Schematic representation of flux constraints applied by TDPS to model different types of genetic modifications. An illustrative metabolic node is shown in its reference state (b) and after four different genetic modifications: the down-regulation of reaction R5 to half of its reference turnover fraction (a); the up-regulation of reaction R5 to double the reference turnover fraction (c); the inactivation of reaction R5 (d); and the insertion of a heterologous reaction RH by forcing it to consume 50% of the available precursors (f). Additionally, a summary table (e) is also shown with an overview of the types of genetic modifications allowed by TDPS with the corresponding flux constraints applied in each case (for further details see the [Methods](#) section and the [Supplementary Text S2](#)). T_A , production turnover of metabolite A; T_m , production turnover of metabolite m ; V_n , flux in reaction n ; $X_{m,n}$, flux partition ratios.

OptReg⁶ constrains the regulated flux to be considerably higher or lower than the reference value (steady-state range) by forcing the flux to go above or below the threshold computed from the maximum/minimum fluxes possible in that reaction. Another method, EMILiO⁷ can manipulate target fluxes to any feasible value, while another approach by Redirector⁸ enables addition of the flux variable to the objective function linear problem. All these strategies were formulated with the aim of finding new genetic targets; therefore, their use for simulating phenotypes is neither direct nor recommended.

Besides the lack of suitability for simulation purposes, all the above methods for simulating up/down-regulations also share a common limitation: they force flux to increase in up-regulated/activated reactions to fixed values, a feature that can lead to the increase of additional fluxes in adjacent reactions to supply the required precursors. One extreme example of this is to consider the last step of an inactive linear pathway being up-regulated, as this would lead to flux in all other reactions of that pathway also increasing (see [Supplementary Text S1](#) for examples). In biological terms, however, it would make more sense for this flux to be constrained by the fluxes that produce the required precursors under these conditions, which would mean that, considering only the effect of the up-regulation, the reaction would remain with no flux.

The framework for activating reactions proposed by Ip *et al.*, called proportional flux forcing (PFF),⁹ offers an alternative for modeling precursor availability. PFF makes the flux through an activated reaction dependent on the availability of its substrate, by assuming that the flux is proportional to the production turnover of the metabolite (*i.e.*, the sum of the fluxes producing the precursor metabolite).⁹ After applying the turnover-

dependent flux constraints, the phenotype is simulated using flux balance analysis (FBA)^{10,11} by maximizing biomass production. PFF can then be used to find gene deletions that increase the turnover (availability) of the precursor metabolite and/or that decrease the flux in reactions competing for the same precursor. However, PFF is still limited in functionality with regard to its applicability to the simulation of strains that have different types of modifications: the simulations are restricted to the activation of a single reaction, only gene knockouts are achievable with PFF (up/down-regulations are only possible when PFF is used together with other algorithms), and the authors do not explain how to handle the activation of reactions with multiple reactants.

Here, we present the turnover dependent phenotypic simulation (TDPS) algorithm. TDPS is a simulation-oriented constraint-based method developed to generate more biologically meaningful simulations for strains including up/down-regulations. Similar to PFF,⁹ TDPS uses metabolite turnover to integrate knowledge of precursor availability for the regulated reactions. Besides gene deletions, TDPS can also simulate up/down-regulations, as well as the activation of inactive reactions. Here, heterologous reactions are included in the category of reaction activations, allowing TDPS to simulate, in quantitative terms, the performance of heterologous genes, based on the availability of required precursor metabolites in their new host. Furthermore, a newly developed objective function, based on Stephanopoulos and Vallino's¹² work on the rigidity of metabolic networks, is also implemented in TDPS, which reinforces network rigidity by minimizing the change in the split ratios between the reference and disturbed network.

To validate TDPS, we collected production yields from *Saccharomyces cerevisiae* strains already characterized in the

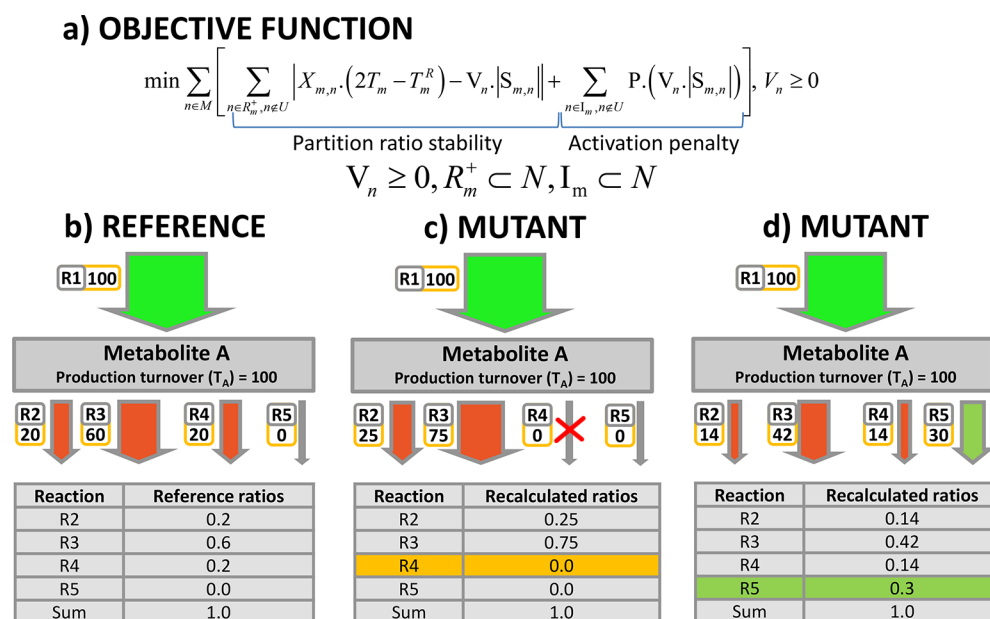


Figure 2. Objective function implemented in TDPS (a) and the schematic representation of a metabolic node in the reference state (b), after the inactivation of R4 (c) or the activation of reaction R5 (d). The flux changes in response to the inactivation of reaction R4 and the activation of reaction R5 are used to illustrate how the objective function in TDPS promotes the stability of the flux partition ratios (see the text and [Methods](#) section for more details about the variables and the objective function). I_m , set of inactive reactions in the reference flux distribution that can consume metabolite m ; M , set of all metabolites; N , set of all reactions; P , penalty constant for activated reactions; R_m^+ , set containing all the active reactions consuming metabolite m in the reference flux distribution; $S_{m,n}$, stoichiometric coefficient of metabolite m in reaction n ; T_m , production turnover of metabolite m in the simulation; T_m^R , production turnover of metabolite m in the reference flux distribution; V_n , flux value of reaction n in the simulation; $X_{m,n}$, flux partition ratios.

literature and challenged the algorithm to predict, quantitatively, the effect of genetic changes in strain performance. Using three case studies, our validation showed that TDPS can indeed predict many of the phenotypes tested, with many experimental production yields being within simulated intervals. Overall, this simulation-oriented method offers a promising approach for more effective strain development, serving as a solid platform for future optimization.

RESULTS AND DISCUSSION

Turnover Dependent Phenotypic Simulation. To account for the availability of a metabolite in constraint-based simulations one can calculate the magnitude of the fluxes producing it (turnover).^{13,14} The production turnover (rate) of a metabolite (T_m) (which equals its consumption under the steady state assumption) is hereby defined as the sum of all fluxes producing metabolite m multiplied by the corresponding stoichiometric coefficients.

In TDPS, the T_m is used to model all types of genetic modifications (Figure 1). The insertion or activation of a reaction is modeled by assuming that a fraction of the total T_m will be consumed by the activated reaction (as suggested by Ip *et al.*⁹). In this way, flux through the new or activated reaction becomes proportional to T_m and can even be zero if T_m is equal to zero, *i.e.*, if no reaction in the network is producing metabolite m . This heuristic respects the availability of precursors when imposing constraints because the regulated fluxes are not fixed to a certain value or interval, but made dependent on the T_m of the precursors. Figure 1f shows an example of how TDPS would model the insertion of a heterologous reaction *RH* into the toy network depicted. In this case, it was assumed that 50% of the turnover of metabolite A (T_A) would be consumed by the inserted reaction

RH. It is important to note that the T_m used in these constraints is not the reference T_m value, but a variable in the simulation, meaning that the flux is not forced to be above a certain value, but at least 50% of whatever value T_m will have in the final simulation.

Besides reaction insertion and activation, TDPS also models up and down-regulations in a turnover dependent manner. By computing the fraction of T_m that each reaction consumes in a reference state, it is assumed that an increase in that fraction value will correspond to an up-regulation and similarly, that a decrease will correspond to a down-regulation. Using the example shown in Figure 1, the fraction of precursor turnover that R5 consumes in the reference state is about 0.4/40%. Therefore, a possible down-regulation performed with TDPS would involve imposing that the flux through R5 must be smaller than or equal to 20% of T_A (Figure 1a). Using the same logic, R5 can be up-regulated with TDPS by increasing the minimum fraction of T_A that R5 must consume. In the up-regulation example shown in Figure 1c, the fraction of T_A consumed by R5 was therefore increased to 0.8 (double that of the reference value). It is important to note here that the changes in turnover fractions depicted in Figure 1 were chosen for illustrative purposes. Full details of the procedure used by the TDPS algorithm to perform each type of genetic modification can be found in the [Methods](#) section and the [Supplementary Text S2](#).

The genetic modifications shown in Figure 1 were all exemplified using reactions with a single precursor metabolite; however, most biochemical reactions typically have more than one reactant. Since TDPS models genetic modifications using a precursor dependent methodology, when a reaction has several precursors it is crucial to determine the precursor that will limit the flux in the modulated reaction. Therefore, TDPS handles

the regulation of reactions with more than one precursor using a strategy based on the most limiting precursor. Concerning down-regulated reactions, TDPS formulates multiple turnover-dependent constraints according to the number of precursors, and imposes that the flux in the regulated reaction must be lower or equal to the constraint that imposes the lowest flux. Likewise, the flux in an up-regulated/inserted/activated reaction should be higher than or equal to the lowest flux imposed by the turnover-dependent constraints (see [Supplementary Text S2](#) and [Methods](#) for the mathematical formulation).

The turnover-dependent constraints applied by TDPS ([Figure 1](#)) can be related to *in vivo* enzyme activities at two levels: first, the abundance of the substrate is modeled using the T_m values; second, the turnover fraction values can be interpreted as the capacity of a certain enzyme to compete with other enzymes for a precursor (mimicking the effects of enzyme kinetic parameters and enzyme concentration). Although TDPS should not be used as a replacement for kinetic modeling, its simple formulation is less computationally intensive, allowing its use for modeling metabolism at the genome scale.

In order to simulate mutant phenotypes, besides the turnover-dependent constraints applied in TDPS, it is also necessary to define a cellular goal (objective function) that represents how the network reacts to the disturbances applied. Since the constraints applied by TDPS manipulate the flux consumption ratios for certain metabolites, it would be appropriate to define an objective function that represents how a cell controls the partition of flux at important nodes. Stephanopoulos and Vallino¹² have introduced the idea that some of the nodes in biochemical networks have evolved to maintain rigid flux ratios, with the objective of sustaining biomass growth with a relatively stable supply of building blocks. Therefore, the objective function included in TDPS was developed to reinforce network rigidity at the level of flux ratios observed for the wild-type organism. Moreover, similar to other mutant simulation strategies,¹⁵ the objective function also minimizes the flux in reactions that are inactive in the reference flux distribution, which have not been subject to direct manipulations.

The rigidity-based objective function was thus formulated as the minimization of the sum of two different terms ([Figure 2a](#)). The first term was formulated to keep the flux partition ratios ($X_{m,n}$) of the mutant as close as possible to the reference, while penalizing changes in the metabolite turnovers (T_m). In a node where the turnover has not changed ($T_m = T_m^R$), the first term ensures that the consumption of metabolite m in reaction n predicted by the simulation ($V_n \cdot |S_{m,n}|$) is as close as possible to the amount of metabolite m that reaction n should be using according to the reference flux distribution and the effect of the reaction regulations imposed ($X_{m,n} \cdot T_m$). The formulation of the second term is simply the sum of all fluxes (V_n) passing in reactions that are inactive in the reference flux distribution (I_m) multiplied by the corresponding stoichiometric coefficient ($|S_{m,n}|$) and by a penalty constant (P). A more detailed description of the implementation of the objective function and the calculation of the updated partition ratios ($X_{m,n}$) is provided in the [Supplementary Text S2](#) and the [Methods](#) section.

As shown in [Figure 2b](#), the first requirement for simulating a phenotype with TDPS is the computation of a reference flux distribution for the wild-type organism and the flux partition

ratios (turnover fraction values) for each metabolite. In [Figure 2c](#), we show an example of a metabolic node consisting of three reactions that consume metabolite A. When one of the reactions is inactivated (R4 in this case), the objective function promotes the stability of the flux partition ratios for the remaining reactions by adjusting their fluxes. Furthermore, the reaction activation penalty ensures that reactions that were inactive in reference flux distributions are less likely to carry any flux. In another example shown in [Figure 2d](#), the same metabolic node is shown for when a reaction is activated. In this case, the flux partition ratios were recalculated to accommodate activation, and the objective function ensures that relative fluxes in the other reactions are kept as close as possible to the wild-type state.

Similar to the objective functions of previously described algorithms, such as MOMA¹⁶ or LMOMA,¹⁷ the objective function of TDPS also promotes the stability of a mutant network relative to a reference flux distribution ([Figure 2](#)). However, in TDPS, the objective function promotes stability at the level of the flux partition ratios. This contrasts with objective functions used in MOMA and LMOMA, which penalize any changes in the fluxes of the mutant organism compared to the reference state. Taking the knockout example shown in [Figure 2c](#), simulating this with MOMA or LMOMA would result in the nonmodified fluxes maintaining their reference values (R2 = 20 and R3 = 60). Therefore, while many simulation methods which promote flux stability always try to keep the network identical to the reference, the objective function of TDPS allows for flux flexibility, an important feature for simulating reaction up- and down-regulations and/or insertions.

The flux changes depicted in [Figure 1](#) and [Figure 2](#) assume that the turnover of metabolite A remained constant when the network was disturbed. In more realistic scenarios, changes in the turnover will affect the fluxes in regulated reactions ([Figure 1](#)) and in the flux distribution computed by the objective function ([Figure 2](#)). The proportionality between the production turnover and consumption fluxes is present in TDPS at two levels: in regulated fluxes, a rigid relationship between precursor turnover and flux is imposed and the two will always be proportional; in the objective function, the proportionality is promoted but not enforced, and it can happen that there are nodes with fluxes changing independently of the precursor turnovers. By making the flux in regulated reactions dependent on the precursor turnover, TDPS assumes that changes in the abundance of the precursor must be reflected in the flux through the regulated reaction. For example, assuming that the production turnover of a precursor doubles, mass balance constraints dictate that at least a few of the consuming fluxes will have to increase accordingly. For the fluxes computed by the objective function, the flux partition ratios are not enforced, which can result in cases where fluxes change independently of the changes in the precursor turnovers. For example, deletions of certain reactions may cause fluxes to become zero in nodes where the turnover has increased.

Validation Using a Toy Model. The TDPS algorithm was developed to simulate mutant phenotypes of engineered strains using genome-scale metabolic models (GEMs). However, GEMs are difficult to visualize due to their size, making it challenging to validate some aspects of TDPS using such models that incorporate hundreds of metabolic reactions. Subsequently, a small toy model was created to demonstrate

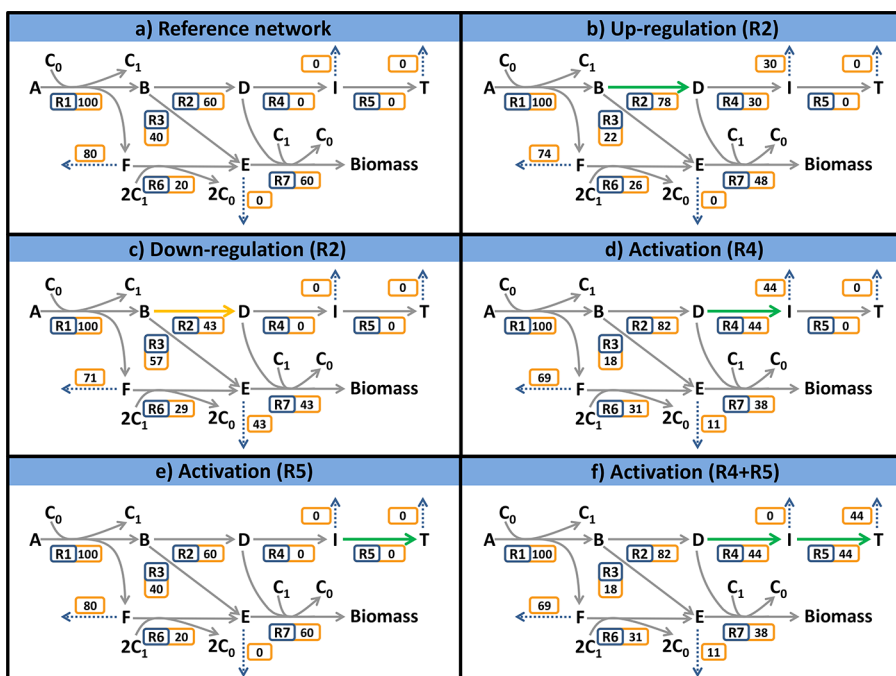


Figure 3. TDPS validation using a toy model. The reference flux distribution (a) was computed for the toy model using pFBA by maximizing the biomass flux (reaction R7) and assuming a substrate (metabolite A) uptake rate of 100 flux units. The up-regulation (b) and down-regulation panels (c) illustrate how TDPS can manipulate the flux partition ratios in certain metabolic nodes and divert fluxes accordingly. The activation of reaction R4 (d) or R5 (e) demonstrate that precursor availability has a significant impact in the flux of activated reactions; *i.e.*, the flux in an activated reaction is only induced if the precursors are available in the network. The simultaneous activation of reaction R4 and R5 (f), exemplifies how TDPS allows genetic modifications to interact with each other and have a synergistic effect. All mutant simulations were obtained with TDPS assuming a C parameter of 2.0 for up-regulations and 0.50 for down-regulation (see [Supplementary Text S2](#) and the [Methods](#) section for more details about the C parameter). The reaction activation penalty constant used was 1.0.

that each component of TDPS functions correctly. The toy model was designed to be as small as possible while also allowing different tests to be performed with it, including (i) up-regulation of an active reaction, (ii) down-regulation of an active reaction, (iii) activation of a reaction with available precursors, (iv) activation of a reaction without available precursors, and (v) the simultaneous activation of two consecutive reactions. In order to simulate mutant phenotypes, TDPS requires a reference flux distribution to serve as a basis for calculations performed by the algorithm (see [Supplementary Text S2](#) and the [Methods](#) section). [Figure 3a](#) shows the reference flux distribution obtained for the toy model using parsimonious flux balance analysis (pFBA)¹⁸ by maximizing the biomass production flux (reaction R7) and fixing the substrate (metabolite A) uptake rate to at 100 flux units. Furthermore, [Figure 3](#) also shows TDPS simulations of selected types of genetic modifications.

To validate the application of the TDPS algorithm to up-regulated reactions, reaction R2 was selected as the target. As shown in [Figure 3b](#), the up-regulation of R2 resulted in an increase of 17 flux units causing, in response, an activation of reaction R4, to dispose of excess metabolite D. There was also a drop in biomass flux, consistent with the diversion of metabolite B from reaction R3. Looking at the flux partition ratios, up-regulation of R2 also caused a decrease in the partition ratio of metabolite B to reaction R3. In contrast, when reaction R2 is down-regulated, [Figure 3c](#) shows that the flux in R3 rises while flux in R2 decreases. Furthermore, down-regulation of R2 also limits the amount of metabolite D available in the network, leading to a reduction in the rate of biomass formation. The ability to simulate both up- and down-

regulations confirms that TDPS may be used to both manipulate partition ratios of selected metabolic nodes and divert fluxes, without constraining fluxes to fixed values.

One main feature of TDPS is that it makes flux through an up-regulated/activated reaction be dependent on the availability of the relevant precursors. To demonstrate this, two inactive reactions with precursors that were available in different amounts were selected as targets for activation (R4 and R5). [Figure 3d](#) shows that a significant flux is predicted when reaction R4 is activated because of the high availability of its precursor (metabolite D). In contrast, when reaction R5 is activated, there are no changes in flux distribution in comparison to the reference network ([Figure 3e](#)) due to a lack of precursor (metabolite I). This lack of flux, when low precursor availability is recognized by the TDPS algorithm, highlights that the activation of a reaction does not always imply that there will be changes in the network. This stresses the advantage of using TDPS to simulate reaction activations, over methods that only manipulate either the flux values directly^{6,7} or the objective function.⁸

Although the activation of R5 on its own resulted in a phenotype with zero flux on that reaction, TDPS can also be used to model multiple genetic modifications in an integrated approach. For example, when R4 is activated simultaneously with R5, the production turnover for metabolite I becomes higher than zero and R5 can then also be active as a consequence ([Figure 3f](#)). This highlights another relevant feature of TDPS: since turnover-dependent constraints are not imposed based on the reference turnover values, genetic modifications can interact with each other. This can subsequently lead to either antagonistic or synergistic effects

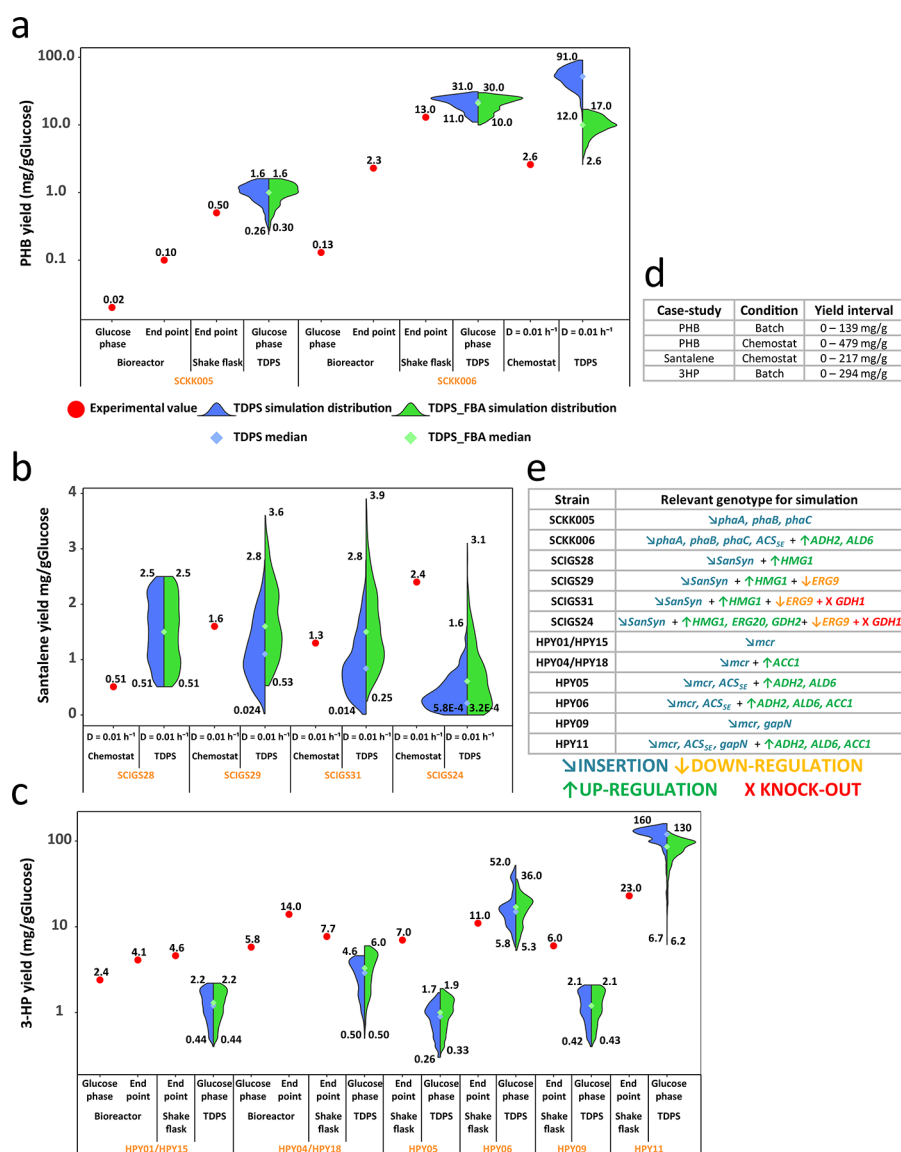


Figure 4. TDPS validation using experimental results from three different case studies. (a) The polyhydroxybutyrate (PHB) yields from *S. cerevisiae* strains characterized by Kocharin *et al.*²⁰ (red dots) are compared to the violin plots of simulated yields from TDPS (blue) and TDPS_FBA (green). (b) The product yield of four different santalene producing *S. cerevisiae* strains characterized by Scalcinati *et al.*²⁴ (red dots) are compared to the violin plots of simulated yields from TDPS (blue) and TDPS_FBA (green). (c) A series of 3-hydroxypropionic acid producing *S. cerevisiae* strains, characterized by Chen *et al.* and Shi *et al.*^{25,26} (red dots) are compared to the violin plots of simulated yields from TDPS (blue) and TDPS_FBA (green). (d) Minimum and maximum product yields for each case study computed with flux variability analysis. (e) Sets of genetic modifications, relevant for the simulations, present in each strain analyzed. Gene abbreviations: all genes endogenous to *S. cerevisiae* follow the accepted nomenclature. ACS_{SE}, acetyl-CoA synthetase (L641P) from *Salmonella enterica*; gapN, glyceraldehyde-3-phosphate dehydrogenase from *Streptococcus mutans*; mcr, malonyl-CoA reductase from *Chloroflexus aurantiacus*; phaA/phaB/phaC, genes of the PHB production operon from *Ralstonia eutropha*; SanSyn, santalene synthase (*SanSyn*) from *Clausaena lansium*.

depending on which genes are modified and how. This feature may be of significant relevance when simulations include several genetic modifications that are associated with closely related precursors.

From the toy model simulations we could see that the strategy used to model genetic modifications using metabolite turnovers results in more meaningful simulations in biological terms. Nevertheless, this can also lead to changes in fluxes that seem counterintuitive. For example, in the simulation shown in Figure 3e the flux in an up-regulated reaction does not increase in comparison to the reference strain. Because the flux constrains are made dependent on the turnover of the precursor metabolites, up-regulating a certain reaction may

result in the flux staying the same or even decreasing in comparison to the nondisturbed state, assuming that manipulations in other parts of the network may affect the turnover. The same is valid for down-regulations, which can lead to an increased flux in the event of the precursor turnovers increasing a lot between the reference and disturbed state.

TDPS Validation Using Experimental Results. The major motivation behind the development of TDPS was to create an algorithm capable of simulating complex strain designs in quantitative terms. Therefore, the ultimate test to evaluate this algorithm consisted of comparing phenotypes of experimentally characterized strains with their simulated counterparts. *S. cerevisiae* was chosen as a case study here as

the flux distributions obtained with the genome-scale models for this organism are well established, and have been validated and curated previously.¹⁹ Besides the availability of curated flux distributions, *S. cerevisiae* was also selected due to its status as a model organism, with abundant physiological data being available, as well as many strain engineering strategies accessible in the literature. Since the reference flux distribution plays an important role in TDPS simulations, we restricted validation efforts to strains of *S. cerevisiae* grown aerobically on minimal media with glucose as the carbon source.

Ideally, the validation of steady-state simulations obtained with TDPS should be done using data obtained from steady-state conditions, such as glucose limited chemostat experiments. However, as chemostat fermentations are labor intensive, the majority of metabolic engineering studies available in the literature are predominantly batch cultures in bioreactors or shake-flasks for strain characterization. To maximize the amount of experimental data used in the validation, we included both batch and chemostat experiments. Two different reference flux distributions were created for TDPS simulations: the first describing glucose-limited conditions characteristic of chemostat cultures growing at a dilution rate of 0.1 h⁻¹ and the second for simulating respiro-fermentative conditions typical of the glucose consumption phase in batch fermentations.

Besides a reference flux distribution, to simulate a certain set of genetic modifications in TDPS it is also necessary to select the strength of each modification. In TDPS this is achieved by assigning appropriate values to the *C* parameter (see the [Methods](#) section and [Supplementary Text S2](#)) for each genetic modification simulated. Experimentally, the strength of an up/down-regulation will depend on several variables, from the specific strategy used to alter the abundance to the kinetic characteristics of the enzymes involved. The translation of these experimental modifications therefore, to a set of *C* parameters can be challenging, when simulating mutant strains with TDPS. In order to get an overview of the phenotype range that TDPS could predict for the same set of genetic modifications, each strain was simulated by attributing 1000 uniformly distributed random values to each *C* parameter within a defined range (see [Methods](#) section for range values). For each case study, we therefore present the distribution of production yields obtained from these 1000 simulations.

To simulate the phenotypes of strains obtained from literature, two different versions of TDPS were used: the regular TDPS formulation and a relaxed version (TDPS_FBA) where the optimal value of the objective function was allowed to relax slightly while biomass was maximized (for a detailed description see [Methods](#)). TDPS_FBA was therefore used to determine how robust the phenotypes obtained with TDPS are, that is, to check if the production yields obtained with TDPS can vary significantly when the objective function is allowed to vary slightly. As explained in the introduction, most existing constrained-based simulation methods only allow to account for reaction knockouts, while other methods cannot be used for simulation. Thus, the results obtained with TDPS were not benchmarked against any other *in silico* methods, but rather only with the experimental results.

Case Study 1: PHB Production in *S. cerevisiae*. The first case study used for validating TDPS involves the production of polyhydroxybutyrate (PHB) in *S. cerevisiae* using genes from the *phaCAB* operon in *Ralstonia eutropha*.^{20,21} This case study uses experimental PHB production yields obtained by

Kocharin *et al.* in batch cultures (bioreactor and shake-flask) for strains expressing three heterologous genes from *R. eutropha* (SCKK005 and SCKK006).^{20,21} The PHB production yields calculated from the results of Kocharin *et al.*^{20,21} along with the corresponding yields simulated with TDPS are shown in [Figure 4a](#). The detailed strain genotypes are described in [Figure 4e](#) and interval of possible yields are shown in [Figure 4d](#) to show how TDPS simulations restrict product excretion. The simulated yields shown for batch conditions in [Figure 4a](#) were obtained specifically for the respiro-fermentative metabolism on glucose, making their value comparable with the glucose consumption phase of bioreactor cultivations. Since the shake-flask yields were calculated at the end of the fermentation, the values shown are the sum of both the respiro-fermentative phase on glucose and the growth on fermentation products phase. Therefore, the shake-flask yields should not be directly comparable to the simulated yields obtained with TDPS but are still shown for a qualitative comparison. Not all strains in refs 20, 21 were used, however, as some included the inactivation of genes in the glyoxylate cycle, which should not be active during the glucose consumption phase.^{22,23} Therefore, for this case study these strains were not included. *In vivo*, these modifications were intended to have an impact only during ethanol assimilation and indeed, *in silico* simulations with TDPS confirmed that deletion of either of the two glyoxylate cycle's genes (*MLS1* and *CIT2*) had no impact on PHB accumulation in the glucose consumption phase (data not shown).

Regarding strain SCKK005 ([Figure 4a](#)), the minimum yield simulated with TDPS (0.26 mg/g) is 1 order of magnitude higher than the experimental value obtained in the glucose phase of the bioreactor fermentation (0.020 mg/g). However, it is interesting to note that the shake-flask final yield (0.50 mg/g) is 5 times higher than the bioreactor. This shows how the same strain can perform quite differently in different conditions and that the yield simulated with TDPS is much closer to the PHB yield obtained in shake-flask for SCKK005. The analysis of the flux distribution in the conditions tested showed that the turnover of acetoacetyl-CoA (a precursor to PHB production) is 0.011 mmol/gCDW·h. Assuming that all precursors would be converted to PHB, the maximum theoretical yield would be 0.46 mg/g. Since the minimum PHB yield simulated with TDPS represents more than 50% of this value, it is likely that TDPS is overestimating the basal performance of the heterologous PHB producing enzymes.

In comparison with SCKK005, the strain SCKK006 includes the additional expression of *acsSE*^{L641P} from *S. enterica* and the up-regulation of *ALD6* and *ADH2* with the aim of improving the availability of acetyl-CoA in the cytosol of *S. cerevisiae*.^{20,21} Since the up-regulation of *ADH2* was implemented to improve ethanol utilization after glucose is exhausted, this gene was excluded in the simulations performed with TDPS for the glucose phase. Looking at [Figure 4a](#), the improvement in the PHB production yield observed for the bioreactor cultivation of SCKK006 was 6.5 times higher in the glucose phase and 23 times higher overall in comparison to SCKK005. TDPS predicts an improvement of 19-fold in the maximum simulated value for the PHB yield, while the improvement in the minimum simulated value is 42 fold. Although TDPS overestimates the improvement in the glucose phase, the simulated improvement for the maximum PHB yield is not too different from the overall improvement observed for the bioreactor cultivation (19 vs 23 fold increase). As seen for

strain SCKK005, the minimum production yield predicted by TDPS for strain SCKK006 (11.0 mg/g) is also considerably above the experimental value determined in bioreactor cultivations (0.13 mg/g). Again, only the shake-flask yields can approximate the values simulated with TDPS.

Besides the batch cultivation, strain SCKK006 was also characterized in glucose limited chemostat cultures by Kocharin *et al.*²¹ (Figure 4a). As seen for the batch cultures, the minimum yield simulated with TDPS (12.0 mg/g) is higher than the experimentally determined value (2.6 mg/g) but in this case, the difference is lower than in batch cultures. Furthermore, the minimum yield predicted with TDPS_FBA is actually coincident with the experimentally determined value (2.6 mg/g). Although the relative amount of data is limited for chemostat cultures, the better performance of TDPS in computing PHB yields here may be a consequence of the superior activity of the PHB producing enzymes under this type of cultivation. Alternatively, since genome-scale models are most suitable for simulating steady-state conditions, the simulation of chemostat cultures may also benefit from the increased accuracy of these flux distributions.

Case Study 2: Santalene Production in *S. cerevisiae*.

Santalene is an isoprenoid that belongs to the sesquiterpene class (15 carbon atoms) and can be synthesized from farnesyl pyrophosphate (FPP), a precursor for sterol biosynthesis in *S. cerevisiae*. Scalcinati *et al.* have characterized a series of *S. cerevisiae* strains engineered for improved production of santalene, which included among other modifications, the expression of santalene synthase (*SanSyn*) from *Clausena lansium*.²⁴ The product yields reported by Scalcinati *et al.*²⁴ for the santalene producing strains are shown in Figure 4b alongside the simulation values obtained with TDPS. It is important to note that the strains constructed by Scalcinati *et al.*,²⁴ which included the expression of a mutated transcription factor (*UPC2*), were excluded from this analysis due to TDPS being unable to simulate this type of genetic modification.

Strain SCIGS28 was built by introducing the *SanSyn* into *S. cerevisiae* together with the expression of a truncated version of HMG-CoA reductase (*tHMG1*), which lacks feedback regulation.²⁴ Figure 4b shows that the santalene yield interval predicted by TDPS for SCIGS28 (0.51–2.5 mg/g) encompasses the yield obtained experimentally (0.51 mg/g). Further optimization of the santalene production yield in strain SCIGS29 was achieved by down-regulating squalene synthase (*ERG9*) and deleting the lipid phosphate phosphatase (*LPP1*) to avoid unspecific hydrolysis of FPP by this enzyme.²⁴ Since the Yeast 6 model lacks the *LPP1* gene, its deletion had to be excluded from the simulations; however, it is noteworthy that even if the model did include *LPP1*, it would be unlikely that the unspecific dephosphorylation of farnesyl-pyrophosphate would be associated with it. Figure 4b shows that in comparison to SCIGS28, the experimental santalene yield in the strain SCIGS29 is 3.1 times higher (1.6 mg/g). However, the simulations performed with TDPS do not show such an improvement; in fact, the minimum simulated yield decreased, while the maximum yield increased by only 12% (Figure 4b). An analysis of the simulation results revealed that the lack of improvement in the minimum simulated yield was a consequence of poor growth rates for simulations where *ERG9* down-regulation was more severe, in turn decreasing the availability of FPP precursor.

Given the essentiality of ergosterol for growth, it would be expectable that down-regulating *ERG9* *in silico* would promote

an increase in the turnover of FPP, to compensate for limiting *ERG9* activity. However, the simulations performed with TDPS failed to reproduce the phenotype observed *in vivo*, suggesting that the objective function used in TDPS is not suited for this type of simulation, wherein growth is anticipated to be severely impacted. This finding is further supported by the improved performance of TDPS_FBA in simulating strain SCIGS29 as a result of partially allowing the maximization of biomass production (Figure 4b). We then simulated both SCIGS28 and SCIGS29 using a modified TDPS algorithm that uses biomass maximization as the objective function. Here, the results showed that the santalene yield for SCIGS28 remained very similar to the TDPS simulation (0.52–2.5 mg/g), but the simulated yield interval for strain SCIGS29 increased significantly (0.56–43 mg/g). The increase in the maximum simulated yield for SCIGS29, using only biomass maximization as the objective, suggests that the rigidity based objective function is not enough to reproduce some cellular behaviors.

Since santalene production in *S. cerevisiae* also requires NADPH in its synthesis, this led Scalcinati *et al.* to manipulate the abundance of this cofactor by deleting the NADP⁺-dependent glutamate dehydrogenase (*GDH1*), which is the main reaction consuming NADPH.²⁴ Strain SCIGS31 also included the knockout of *DPP1*, but similarly as was the case for the *LPP1* deletion, this knockout was not included in the simulations. Figure 4b shows that the santalene yield for SCIGS31 dropped slightly in comparison to SCIGS29, and accordingly we also saw a decrease in the minimum yields simulated with TDPS and TDPS_FBA. The lack of improvement in the santalene yield for SCIGS31 subsequently indicates that the limiting factor *in silico* is not related to NADPH abundance. We can therefore infer from this (the lack of improvement in the experimental santalene yield), that TDPS was able to correctly predict NADPH metabolism as not being the limiting factor for santalene production.

Additional improvement in the santalene production yield (strain SCIGS24) was achieved by Scalcinati *et al.* up-regulating *GDH2* and *ERG20*.²⁴ As shown in Figure 4b, the santalene production yield in SCIGS24 improved to 2.4 mg/g, which is 1.5 times higher than strain SCIGS29. In contrast, the simulations performed with TDPS for this strain resulted in lower minimum and maximum yields when compared to the simulations of SCIGS29. One of the main reasons for the lack of improvement observed in the simulations of strain SCIGS24 could be attributed to the absence of any effect of up-regulating *ERG20* using TDPS. *ERG20p* catalyzes the conversion of three isopentenyl diphosphate (IPP) molecules into a single FPP molecule, but in the reference strain simulation, all of the available IPP was already being channeled to the production of FPP. In other words, when a reaction is already consuming 100% of the precursor, TDPS is not able to further increase that value. In this particular case, *ERG20p* is already consuming all the available IPP; therefore, TDPS could not increase it any further. Although this should be expected from the TDPS formulation, it contrasts with some situations observed *in vivo*; for example, when a rate-limiting enzyme is the only consumer for a certain precursor, its up-regulation can still result in increased flux.

Case Study 3: 3-Hydroxypropionic Acid Production in *S. cerevisiae*. The biosynthesis of 3-hydroxypropionic acid (3-HP) is not native to *S. cerevisiae*, but it can be achieved from malonyl-CoA by a double step reduction. In the work of Chen *et al.* and Shi *et al.*,^{25,26} the authors introduced into *S. cerevisiae*

the bifunctional malonyl-CoA reductase (*mcr*) from *Chloroflexus aurantiacus* and created a series of strains with improved 3-HP production yields that we used here as case studies to validate the simulation accuracy of TDPS (Figure 4c).

The strains HPY01 and HPY15 include the heterologous expression of the *mcr* gene from *C. aurantiacus* and, as shown in Figure 4c, the 3-HP production yield in the glucose phase of the bioreactor cultivation was very close to the maximum yield obtained with TDPS (2.4 vs 2.2 mg/g). Although the experimental yield is slightly above the maximum simulated yield, it is important to point out that three different combinations of enzymes were tested by the authors to catalyze the conversion of malonyl-CoA to 3-HP.²⁵ In shake-flasks, the yields for the three candidates ranged from 0.23 mg/g to 4.6 mg/g, showing that the range of yields obtained with TDPS approximate well the variability observed experimentally. In comparison to HPY01/HPY15, strains HPY04/HPY18 also include the up-regulation of the acetyl-CoA carboxylase gene (*ACC1*)^{25,26} for improved availability of malonyl-CoA. The difference between HPY04 (shake-flask yield) and HPY18 (bioreactor yield) is the extent of *ACC1* up-regulation: in HPY04, *ACC1* was simply overexpressed;²⁵ and in HPY18 the *ACC1* gene was modified to improve its enzymatic activity²⁶ (Figure 4c). Here, the 3-HP yield of strain HPY18 improved 2.4 fold in the glucose consumption phase and 3.4 times overall in comparison with strain HPY15 (Figure 4c). The 2.4 fold improvement observed in the glucose phase is remarkably similar to the improvement of 2.1 fold obtained in the maximum value simulated with TDPS. In regards to the production yield value in the glucose phase for HPY18 (5.8 mg/g), it falls outside the interval predicted by TDPS (0.42–4.5 mg/g) but is included in the interval predicted by TDPS_FBA (0.50–6.0 mg/L). Regarding the strain HPY04, the shake-flask final production yield showed an improvement of 1.7 fold in comparison to strain HPY01 (Figure 4c), which is similar to the 2.1 fold improvement simulated with TDPS.

The remaining strains presented in Figure 4c for the 3-HP case study were only characterized by the authors in shake-flask cultures, which hampered quantitative comparison of the final yields against the simulations. However, the overall improvement in comparison to HPY01 can still indicate if TDPS captures the differences between strains in qualitative terms. Strain HPY05 has an improved acetyl-CoA availability when compared to HPY01 because of the additional expression of *acs_{SE}^{L641P}* from *S. enterica* and the up-regulation of *ALD6* and *ADH2* (which, as explained for the PHB case study, was similarly excluded). The performance of the strain HPY05 in shake flask revealed an improvement of 1.5 times compared to HPY01 (Figure 4c). However, the yields predicted by TDPS actually decreased in comparison with the control strain. Since the up-regulation of the acetyl-CoA boosting genes was shown to be functional *in silico* for the PHB producing strain SCKK006, the mismatch between the simulation and experimental result of HPY05 may lie in the limited conversion of acetyl-CoA to malonyl-CoA. This is supported by the considerable increase in 3-HP production in the simulation of strain HPY06, which is similar to HPY05 but also includes the up-regulation of *ACC1*.

Strain HPY09 was engineered for improved NADPH availability by introducing in *S. cerevisiae* a NADP⁺-dependent glyceraldehyde-3-phosphate dehydrogenase (*gapN*) from *Streptococcus mutans*.²⁵ Although the 3-HP yield is 1.3 times higher in strain HPY09 that was cultivated in shake-flasks, the

TDPS simulations failed to predict the advantage of the increased availability of NADPH (Figure 4c). However, when the expression of *gapN* was combined with all the other genetic modifications (strain HPY11 in Figure 4c), we observed a synergistic effect in the TDPS simulations. Specifically, the production yield obtained experimentally (23 mg/g) for HPY11 is included in the interval predicted by TDPS (6.7 to 160 mg/g).

The simulation of the 3-HP productions yields with TDPS also showed that this algorithm was successful in predicting the relative performance between metabolically engineered strains. Among all the strains shown in Figure 4c, only two of them failed to be ranked correctly among the others (HPY05 and HPY09). Although the simulations of strains HPY05 and HPY09 did not show a direct improvement in the 3-HP production yield, when combined with additional genetic modifications, there was a clear positive effect of the underlying strategy.

Besides predicting production yields, TDPS can also be used to analyze internal fluxes of engineered strains to help troubleshooting strategies that are not working as expected. For example, the authors in ref 27 used ¹³C metabolic flux analysis to conclude that part of a strategy to improve 3-HP production did not work as expected. The issue was pinpointed to the insertion of a NADPH-dependent glyceraldehyde phosphate dehydrogenase (*gapC*), which proved insufficient to provide enough NADPH for 3-HP production.²⁷ By using TDPS to simulate the fluxes of the 3-HP producing strain with and without the insertion of the *gapC* gene from *Clostridium acetobutylicum*, we could also conclude that without *gapC* the flux in the oxidative branch of the pentose phosphate pathway has to increase to compensate for the additional NADPH sink. However, if *gapC* is significantly up-regulated, then the flux in the oxidative branch of the pentose phosphate pathway decreases and 3-HP production increases (see Supplementary Figure S4 for the detailed flux distributions).

Overall, the results obtained for the three case studies were satisfactory in confirming the predictive accuracy of the TDPS algorithm. The production yields simulated with TDPS were very close to the values reported experimentally with most of the engineering strategies implemented *in vivo* showing a similar trend to what was simulated with TDPS. However, problems were identified with genetic modifications that significantly affect biomass production and that include up-regulation of reactions that are already fully using their precursors. Nonetheless, these limitations can be surmountable by using different objective functions or by integrating enzyme kinetic parameter information into genome-scale models (such as with the enzyme constrained genome scale metabolic model GECKO²⁸). Although not addressed here, TDPS can also be used together with optimization algorithms to find strain designs with improved production of target metabolites. Therefore, any validation efforts to fine-tune the simulation accuracy of TDPS should also improve the quality of strains designs obtained with it.

METHODS

Implementation of the TDPS Algorithm. The TDPS algorithm was implemented in the JAVA programming language within the OptFlux software.⁵ OptFlux is an *in silico* metabolic engineering framework that allows the user to import and manipulate GEMs and perform strain optimization/simulation tasks. With the integration of TDPS in

OptFlux it is possible to access additional functionalities such as strain optimization algorithms and flux visualization tools, which expands the possible applications of TDPS. TDPS is available in OptFlux from version 3.3.3 and can be accessed from the Simulation/Under-Overexpression menu.

Precalculations. The reference flux distributions were calculated with pFBA using biomass maximization as the objective function.¹⁸ Using the reference flux distribution, the fraction values ($F_{m,n}^R$) were calculated for all reactions (n) consuming metabolite m ($n \in R_m$):

$$F_{m,n}^R = \frac{V_n^R \cdot |S_{m,n}|}{T_m^R} \quad \forall m \in M, n \in R_m, V_n^R \geq 0 \quad (1)$$

where R_m is the set of reactions that consume metabolite m , V_n^R is the reference flux value in reaction n , $S_{m,n}$ is the stoichiometric coefficient of metabolite m in reaction n and M is the set of all the metabolites in the network. The reference production turnovers (T_m^R) were calculated for all metabolites in the network ($m \in M$) using the equation:

$$T_m^R = \sum_{n \in P_m} V_n^R \cdot |S_{m,n}| \quad \forall m \in M, n \in P_m, V_n^R \geq 0 \quad (2)$$

where P_m is the set of reactions that can produce metabolite m .

Splitting Reversible Reactions in Two Half-Reactions.

The formulation of TDPS requires the creation of variables for the production turnovers to compute mutant flux distributions. These variables are required in the formulation of the turnover-dependent flux constraints and for the formulation of the objective function. However, the existence of reversible reactions in metabolic models makes the creation of production turnover variables complex when using linear programming. Since the flux variable can be either positive or negative it was necessary to split all reversible reactions in two positive half-reactions in order to have flux variables that are specific for the forward or reverse direction:

$$V_n = V_n^{\text{Pos}} - V_n^{\text{Neg}} \quad \forall n \in U \quad (3)$$

$$0 \leq V_n^{\text{Pos}} \leq V_n^{\text{UB}} \cdot B_n^{\text{Pos}} \quad B_n^{\text{Pos}} \in \{0, 1\} \quad (4)$$

$$V_n^{\text{LB}} \cdot B_n^{\text{Neg}} \leq -V_n^{\text{Neg}} \leq 0 \quad B_n^{\text{Neg}} \in \{0, 1\} \quad (5)$$

$$B_n^{\text{Pos}} + B_n^{\text{Neg}} \leq 1 \quad (6)$$

where V_n^{Pos} is the flux value in the forward direction of reaction n during the simulation, V_n^{Neg} is the flux value in the reverse direction of reaction n during the simulation, U is the set containing all reversible reactions in the network, V_n^{LB} is the lower flux bound of reaction n , V_n^{UB} is the upper flux bound of reaction n , and B_n^{Pos} and B_n^{Neg} are the binary variables used to prevent both half-reactions from being active simultaneously. As a consequence of using binary variables, the simulations performed with TDPS become a mixed integer linear programming (MILP) problem.

TDPS Objective Function and Formulation. The mathematical implementation of the TDPS objective function, which is shown in the Figure 2 in a simplified version, also includes the reference turnover in the split ratio stability term to address the possibility of the simulation returning a flux distribution filled with zeros. This formulation ensures that the turnovers in the mutant flux distribution remain close to the reference value. TDPS was therefore mathematically formulated as follows (eqs 7–21):

$$\begin{aligned} \min \sum_{m \in M} [& \sum_{n \in R_m^+, n \notin U} |X_{m,n}| \cdot (2T_m - T_m^R) - V_n \cdot |S_{m,n}| \\ & + \sum_{n \in I_m, n \notin U} P \cdot (V_n \cdot |S_{m,n}|) + \\ & \sum_{n \in R_m^+, n \in U, S_{m,n} < 0} |X_{m,n}| \cdot (2T_m - T_m^R) - V_n^{\text{Pos}} \cdot |S_{m,n}| \\ & + \sum_{n \in I_m, n \in U, S_{m,n} < 0} P \cdot (V_n^{\text{Pos}} \cdot |S_{m,n}|) + \\ & \sum_{n \in R_m^+, n \in U, S_{m,n} > 0} |X_{m,n}| \cdot (2T_m - T_m^R) - V_n^{\text{Neg}} \cdot |S_{m,n}| \\ & + \sum_{n \in I_m, n \in U, S_{m,n} > 0} P \cdot (V_n^{\text{Neg}} \cdot |S_{m,n}|)], \quad V_n \geq 0 \end{aligned} \quad (7)$$

$$\begin{aligned} T_m = & \sum_{n \in P_m, n \notin U} V_n \cdot |S_{m,n}| + \sum_{n \in P_m, n \in U, S_{m,n} > 0} V_n^{\text{Pos}} \cdot |S_{m,n}| + \\ & \sum_{n \in P_m, n \in U, S_{m,n} < 0} V_n^{\text{Neg}} \cdot |S_{m,n}| \quad \forall m \in M, V_n \geq 0 \end{aligned} \quad (8)$$

$$V_n^{\text{LB}} \leq V_n \leq V_n^{\text{UB}} \quad (9)$$

$$V_n = V_n^{\text{Pos}} - V_n^{\text{Neg}} \quad \forall n \in U \quad (10)$$

$$0 \leq V_n^{\text{Pos}} \leq V_n^{\text{UB}} \cdot B_n^{\text{Pos}} \quad B_n^{\text{Pos}} \in \{0, 1\} \quad (11)$$

$$V_n^{\text{LB}} \cdot B_n^{\text{Neg}} \leq -V_n^{\text{Neg}} \leq 0 \quad B_n^{\text{Neg}} \in \{0, 1\} \quad (12)$$

$$B_n^{\text{Pos}} + B_n^{\text{Neg}} \leq 1 \quad (13)$$

$$\begin{aligned} V_n > \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), \quad m \in M_n, M_n \\ = \{m_0, \dots, m_i\}, \quad n \notin U, V_n \geq 0, n \in \text{GM}_{\text{UR}} \end{aligned} \quad (14)$$

$$\begin{aligned} V_n^{\text{Pos}} > \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), \quad m \in M_n, M_n \\ = \{m_0, \dots, m_i\}, \quad n \in U, n \in \text{GM}_{\text{UR}}, D_n = 1 \end{aligned} \quad (15)$$

$$\begin{aligned} V_n^{\text{Neg}} > \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), \quad m \in M_n, M_n \\ = \{m_0, \dots, m_i\}, \quad n \in U, n \in \text{GM}_{\text{UR}}, D_n = -1 \end{aligned} \quad (16)$$

$$\begin{aligned} V_n < \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), \quad m \in M_n, M_n \\ = \{m_0, \dots, m_i\}, \quad n \notin U, V_n \geq 0, n \in \text{GM}_{\text{DR}} \end{aligned} \quad (17)$$

$$\begin{aligned} V_n^{\text{Pos}} < \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), \quad m \in M_n, M_n \\ = \{m_0, \dots, m_i\}, \quad n \in U, n \in \text{GM}_{\text{DR}}, D_n = 1 \end{aligned} \quad (18)$$

$$\begin{aligned} V_n^{\text{Neg}} < \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), \quad m \in M_n, M_n \\ = \{m_0, \dots, m_i\}, \quad n \in U, n \in \text{GM}_{\text{DR}}, D_n = -1 \end{aligned} \quad (19)$$

$$V_n = 0 \quad V_n \in \text{GM}_{\text{KO}} \quad (20)$$

$$\sum_{n \in N} S_{m,n} \cdot V_n = 0 \quad \forall m \in M \quad (21)$$

where GM_{UR} is the set of fluxes that should be up-regulated, GM_{DR} is the set of fluxes that should be down-regulated, GM_{KO} is the set of fluxes that should be knocked-out and D_n is the direction of reaction n that should be modified if a reversible reaction is targeted for up- or down-regulation. As shown in eqs 7 and 8, the number of terms present in these equations had to be increased to adjust for the existence of reversible reactions in the metabolic model. This was achieved by replacing V_n by the appropriate half-reaction reaction (V_n^{Pos} or V_n^{Neg}) for all reversible reactions ($n \in U$). The flux constraints formulated in eqs 14–19 also take into consideration which direction of reaction n should be modified by checking a directionality parameter (D_n) provided by the user. For modifying the forward reaction of n , D_n must be set to 1 and for the reverse reaction to -1 . Further details regarding the calculation of the partition ratios ($X_{m,n}$) according to a set of genetic modifications are given in Supplementary Text S2.

TDPS_FBA Objective Function. TDPS_FBA was implemented to test the robustness of the predictions obtained with the TDPS algorithm. Initially, a normal TDPS simulation was performed and the optimal value of the objective function (OF) was stored. Subsequently, a new TDPS problem was formulated with all the initial constraints, while also including eq 22 as well:

$$\begin{aligned} & \sum_{m \in M} [\sum_{n \in R_m^+, n \notin U} |X_{m,n} \cdot (2T_m - T_m^R) - V_n \cdot |S_{m,n}|| \\ & + \sum_{n \in I_m, n \notin U} P \cdot (V_n \cdot |S_{m,n}|) + \\ & \sum_{n \in R_m^+, n \in U, S_{m,n} < 0} |X_{m,n} \cdot (2T_m - T_m^R) - V_n^{\text{Pos}} \cdot |S_{m,n}|| \\ & + \sum_{n \in I_m, n \in U, S_{m,n} < 0} P \cdot (V_n^{\text{Pos}} \cdot |S_{m,n}|) + \\ & \sum_{n \in R_m^+, n \in U, S_{m,n} > 0} |X_{m,n} \cdot (2T_m - T_m^R) - V_n^{\text{Neg}} \cdot |S_{m,n}|| \\ & + \sum_{n \in I_m, n \in U, S_{m,n} > 0} P \cdot (V_n^{\text{Neg}} \cdot |S_{m,n}|)] \leq T \cdot \text{OF}, \quad V_n \geq 0 \end{aligned} \quad (22)$$

where T is the tolerance constant used to relax the initial objective function value. The objective function used in TDPS was then replaced in TDPS_FBA by the maximization of the biomass production flux.

TDPS_FBA was also used to investigate the impact of the rigidity enforcing objective function in the predictions performed with TDPS as shown in Supplementary Text S3.

Mathematical Solver. The academic version of IBM ILOG CPLEX optimization studio V12.5.1 64bit was used as the LP and MILP solver in all simulations performed with TDPS. Given the size and complexity of the problems generated by TDPS, the CPLEX parameter “NumericalEmphasis” was activated in all simulations performed to avoid occasional numerical instability issues. All other CPLEX parameters were used as predefined by the manufacturer.

CPLEX is also required for performing TDPS simulations when using OptFlux.

Toy Model Simulations. TDPS was used to simulate the effects of different types of genetic modifications on a toy model using a penalty constant for activated reactions (P) equal to 1.0, the C parameter for up-regulation was assumed to be 2 and for down-regulation it was assumed to be 0.5 (for further details about the C parameter see Supplementary Text S2). The reference flux distribution used in the simulations was calculated using pFBA by maximizing the biomass reaction R7.

Genome Scale Models and Reference Flux Distributions. The consensus GSMM for *S. cerevisiae* version 6.06²⁹ was downloaded in the SBML format from the project’s Web site: <http://sourceforge.net/projects/yeast/files/> and modified according to Supplementary Table S1. The models including the modifications for each case study were then imported into OptFlux 3.3.3⁵ and all stoichiometric coefficients were multiplied by 1000 to help solve occasional numerical instability issues reported by CPLEX when handling reactions with low fluxes.

The case studies collected from literature to validate TDPS contained data from two types of cultivation methods, chemostat and batch cultures. In order to simulate glucose limited chemostat conditions with a dilution rate of 0.1 h^{-1} , the glucose uptake rate was set to $1.15 \text{ mmol/gCDW}\cdot\text{h}$ and the uptake rates of ammonia, phosphate, sulfate and oxygen were unconstrained. The cofactor (NADH/NADPH) modifications applied by Pereira *et al.*,¹⁹ to improve the reference flux distribution, were also applied here during the computation of the reference flux distribution.

With the purpose of simulating batch cultures, an approximate glucose uptake rate was estimated from batch cultivations of *S. cerevisiae* (initial glucose concentration of 20 g/L) and set to $11.4 \text{ mmol/gCDW}\cdot\text{h}$. To mimic respiro-fermentative metabolism, the oxygen uptake rate was limited to $1.3 \text{ mmol/gCDW}\cdot\text{h}$ in order to obtain a final biomass yield on glucose characteristic of *S. cerevisiae* (0.12 g/g).³⁰ During computation of this flux distribution, the constraints described in¹⁹ were also applied, although only those related to NADPH metabolism were used. All models used here and reference flux distributions are provided in Supplementary File S1 in formats that can be imported into OptFlux.

Parametrization of the TDPS Simulation of Mutant Phenotypes. The simulation of mutant phenotypes was performed with TDPS using a penalty constant (P) of 50 for activated reactions. This value ($P = 50$) provided a good balance between the maintenance of the split ratio values (first term of the objective function in Figure 2) and the inhibition of reaction activation (second term of the objective function in Figure 2) under the conditions that Yeast 6.06 was tested. The penalty constant can be changed in OptFlux by navigating to Help → Preferences → Simulation → Advanced → TDPS Penalty.

The drain reactions were excluded from the objective function formulation as they do not represent any metabolic entity and are only present in the model to allow the exchange of metabolites with the medium. Some metabolites were also excluded from the TDPS constraints and objective function as they are available in excess, making it not logical to include them in the resource based formulation used to model genetic modifications, (as their availability will not change from being in excess). In chemostat conditions the metabolites excluded were as follows: H_2O , H^+ , SO_4^{2-} , NH_4^+ , HPO_4^{2-} , Fe^{2+} and O_2 .

Coenzyme A was also excluded from the algorithm calculations because it has a passive role in metabolism; *i.e.*, it is always recycled to its original form; therefore, its availability should not be a limiting factor for metabolic fluxes at steady-state. In batch conditions, we used the same conditions with the exception of O₂, which was included in the algorithm calculations since its consumption flux has been limited in the simulations. The sets of excluded metabolites are provided as [Supporting Information](#) in a format that can be imported in OptFlux (files can be imported using the menu: File → Import → Reactions/Metabolites extra info).

The strength of the genetic modifications simulated with TDPS can be adjusted using the *C* parameter (see [Supplementary Text S2](#)). Given the lack of knowledge regarding which values for the *C* parameter would be more appropriate to simulate the genetic modifications extracted from the literature, a random number generator was used to create virtual mutant populations of at least 1000 individuals with diverse phenotypes for the same set of genetic modifications. For each up-regulation, 1000 random numbers between 1 and 5 with a uniform distribution were attributed to the *C* parameter ($1 < C \leq 5$), while for each down-regulation 1000 random numbers between 0 and 1 with a uniform distribution ($0 < C < 1$) were generated. Finally, for reaction inactivation the *C* parameter was set to zero. For each virtual strain, the distribution of production yields obtained for each randomly generated population is shown. TDPS_FBA simulations were carried by relaxing 10% ($T = 1.1$) the optimal value of the objective function obtained with TDPS and maximizing the biomass formation with the added constraint.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acssynbio.8b00248](https://doi.org/10.1021/acssynbio.8b00248).

Text S1–S3, Figures S1–S4 (PDF)

Table S1 (PDF)

File S1: archive containing the genome-scale models, reference flux distributions and list of metabolites excluded from the simulations (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: irocha@itqb.unl.pt.

ORCID

Rui Pereira: [0000-0002-0572-875X](https://orcid.org/0000-0002-0572-875X)

Author Contributions

Conceptualization: R.P., J.N., I.R.; Software: R.P., P.V., P.M.; Investigation: R.P.; Writing, Original Draft: R.P.; Writing, Review and Editing: R.P., J.N., I.R.; Supervision: J.N., I.R.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was supported by the Portuguese Foundation for Science and Technology (FCT) under the scope of the strategic funding of UID/BIO/04469/2019 unit and BioTecNorte operation (NORTE-01-0145-FEDER-000004) funded by the European Regional Development Fund under

the scope of Norte2020 – Programa Operacional Regional do Norte. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 686070 and from the Novo Nordisk Foundation. The work of Rui Pereira was supported by a PhD grant from FCT (ref SFRH/BD/51111/2010). We thank Kate Campbell for proof reading this manuscript.

■ REFERENCES

- (1) Lewis, N. E., Nagarajan, H., and Palsson, B. O. (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* *10*, 291–305.
- (2) Long, M. R., Ong, W. K., and Reed, J. L. (2015) Computational methods in metabolic engineering for strain design. *Curr. Opin. Biotechnol.* *34*, 135–141.
- (3) Maia, P., Rocha, M., and Rocha, I. (2016) In Silico Constraint-Based Strain Optimization Methods: the Quest for Optimal Cell Factories. *Microbiol. Mol. Biol. Rev.* *80*, 45–67.
- (4) Gonçalves, E., Pereira, R., Rocha, I., and Rocha, M. (2012) Optimization approaches for the in silico discovery of optimal targets for gene over/underexpression. *J. Comput. Biol.* *19*, 102–14.
- (5) Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J. P., Nielsen, J., Patil, K. R., Ferreira, E. C., and Rocha, M. (2010) OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Syst. Biol.* *4*, 45.
- (6) Pharkya, P., and Maranas, C. D. (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab. Eng.* *8*, 1–13.
- (7) Yang, L., Cluett, W. R., and Mahadevan, R. (2011) EMILiO: a fast algorithm for genome-scale strain design. *Metab. Eng.* *13*, 272–81.
- (8) Rockwell, G., Guido, N. J., and Church, G. M. (2013) Redirector: designing cell factories by reconstructing the metabolic objective. *PLoS Comput. Biol.*, DOI: [10.1371/journal.pcbi.1002882](https://doi.org/10.1371/journal.pcbi.1002882).
- (9) Ip, K., Donoghue, N., Kim, M. K., and Lun, D. S. (2014) Constraint-based modeling of heterologous pathways: Application and experimental demonstration for overproduction of fatty acids in *Escherichia coli*. *Biotechnol. Bioeng.* *111*, 2056–66.
- (10) Fell, D. A., and Small, J. R. (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem. J.* *238*, 781–6.
- (11) Varma, A., and Palsson, B. O. (1994) Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology* *12*, 994–998.
- (12) Stephanopoulos, G., and Vallino, J. (1991) Network rigidity and metabolic engineering in metabolite overproduction. *Science (Washington, DC, U. S.)* *252*, 1675–1681.
- (13) Brochado, A. R., Andrejev, S., Maranas, C. D., and Patil, K. R. (2012) Impact of stoichiometry representation on simulation of genotype-phenotype relationships in metabolic networks. *PLoS Comput. Biol.*, DOI: [10.1371/journal.pcbi.1002758](https://doi.org/10.1371/journal.pcbi.1002758).
- (14) Chung, B. K. S., and Lee, D.-Y. (2009) Flux-sum analysis: a metabolite-centric approach for understanding the metabolic network. *BMC Syst. Biol.* *3*, 117.
- (15) Shlomi, T., Berkman, O., and Ruppin, E. (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 7695–700.
- (16) Segrè, D., Vitkup, D., and Church, G. M. (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 15112–7.
- (17) Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. Ø., and Herrgard, M. J. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protoc.* *2*, 727–38.
- (18) Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., Weitz, K. K., Eils, R., König, R., Smith, R. D., and Palsson, B. Ø. (2010) Omic data from evolved *E. coli* are

consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* 6, 390.

(19) Pereira, R., Nielsen, J., and Rocha, I. (2016) Improving the flux distributions simulated with genome-scale metabolic models of *Saccharomyces cerevisiae*. *Metab. Eng. Commun.* 3, 153–163.

(20) Kocharin, K., Chen, Y., Siewers, V., and Nielsen, J. (2012) Engineering of acetyl-CoA metabolism for the improved production of polyhydroxybutyrate in *Saccharomyces cerevisiae*. *AMB Express* 2, 52.

(21) Kocharin, K., and Nielsen, J. (2013) Specific growth rate and substrate dependent polyhydroxybutyrate production in *Saccharomyces cerevisiae*. *AMB Express* 3, 18.

(22) Caspary, F., Hartig, A., and Schüller, H. J. (1997) Constitutive and carbon source-responsive promoter elements are involved in the regulated expression of the *Saccharomyces cerevisiae* malate synthase gene *MLS1*. *Mol. Gen. Genet.* 255, 619–27.

(23) Daran-Lapujade, P., Jansen, M. L. A., Daran, J.-M., van Gulik, W., de Winde, J. H., and Pronk, J. T. (2004) Role of transcriptional regulation in controlling fluxes in central carbon metabolism of *Saccharomyces cerevisiae*. A chemostat culture study. *J. Biol. Chem.* 279, 9125–38.

(24) Scalcinati, G., Partow, S., Siewers, V., Schalk, M., Daviet, L., and Nielsen, J. (2012) Combined metabolic engineering of precursor and co-factor supply to increase alpha-santalene production by *Saccharomyces cerevisiae*. *Microb. Cell Fact.* 11, 117.

(25) Chen, Y., Bao, J., Kim, I.-K., Siewers, V., and Nielsen, J. (2014) Coupled incremental precursor and co-factor supply improves 3-hydroxypropionic acid production in *Saccharomyces cerevisiae*. *Metab. Eng.* 22, 104–9.

(26) Shi, S., Chen, Y., Siewers, V., and Nielsen, J. (2014) Improving production of malonyl coenzyme A-derived metabolites by abolishing *Snf1*-dependent regulation of *Accl1*. *mBio* 5, e01130–14.

(27) Kildegaard, K. R., Jensen, N. B., Schneider, K., Czarnotta, E., Özdemir, E., Klein, T., Maury, J., Ebert, B. E., Christensen, H. B., Chen, Y., Kim, I.-K., Herrgård, M. J., Blank, L. M., Forster, J., Nielsen, J., and Borodina, I. (2016) Engineering and systems-level analysis of *Saccharomyces cerevisiae* for production of 3-hydroxypropionic acid via malonyl-CoA reductase-dependent pathway. *Microb. Cell Fact.* 15, 53.

(28) Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P., Kerkhoven, E. J., and Nielsen, J. (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.* 13, 935.

(29) Heavner, B. D., Smallbone, K., Price, N. D., and Walker, L. P. (2013) Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database*, DOI: 10.1093/database/bat059.

(30) van Dijken, J., Bauer, J., Brambilla, L., Duboc, P., Francois, J., Gancedo, C., Giuseppin, M. L., Heijnen, J., Hoare, M., Lange, H., Madden, E., Niederberger, P., Nielsen, J., Parrou, J., Petit, T., Porro, D., Reuss, M., van Riel, N., Rizzi, M., Steensma, H., Verrips, C., Vindeløv, J., and Pronk, J. (2000) An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* 26, 706–714.