

RESEARCH ARTICLE

Harnessing the optimization of enzyme catalytic rates in engineering of metabolic phenotypes

Zahra Razaghi-Moghadam^{1,2*}, Fayaz Soleymani Babadi^{1,2}, Zoran Nikoloski^{1,2*}

1 Systems Biology and Mathematical Modeling Group, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany, **2** Bioinformatics Department, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany

* razaghi@mpimp-golm.mpg.de (ZR-M); zoran.nikoloski@uni-potsdam.de (ZN)



OPEN ACCESS

Citation: Razaghi-Moghadam Z, Soleymani Babadi F, Nikoloski Z (2024) Harnessing the optimization of enzyme catalytic rates in engineering of metabolic phenotypes. PLoS Comput Biol 20(11): e1012576. <https://doi.org/10.1371/journal.pcbi.1012576>

Editor: Marija Cvijovic, University of Gothenburg: Goteborgs Universitet, SWEDEN

Received: March 28, 2024

Accepted: October 21, 2024

Published: November 4, 2024

Copyright: © 2024 Razaghi-Moghadam et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data and code are publicly available in the GitHub repository: <https://github.com/MonaRazaghi/OKO/>.

Funding: This study was financed by the European Union's Horizon 2020 research and innovation program through grant 862201 (to Z.N.). The KT Boost financially also supported this work through grant KT41 (to Z.R.M.). Fayaz Soleymani Babadi acknowledges the financial support by the German Research Foundation, project number 517516852

Abstract

The increasing availability of enzyme turnover number measurements from experiments and of turnover number predictions from deep learning models prompts the use of these enzyme parameters in precise metabolic engineering. Yet, there is no computational approach that allows the prediction of metabolic engineering strategies that rely on the modification of turnover numbers. It is also unclear if modifications of turnover numbers without alterations in the host's transcriptional regulatory machinery suffice to increase the production of chemicals of interest. Here, we present a constraint-based modeling approach, termed Overcoming Kinetic rate Obstacles (OKO), that uses enzyme-constrained metabolic models to predict *in silico* strategies to increase the production of a given chemical, while ensuring specified cell growth. We demonstrate that the application of OKO to enzyme-constrained metabolic models of *Escherichia coli* and *Saccharomyces cerevisiae* results in strategies that can at least double the production of over 40 compounds with little penalty to growth. Interestingly, we show that the overproduction of compounds of interest does not entail only an increase in the values of turnover numbers. Lastly, we demonstrate that a refinement of OKO, allowing also for manipulation of enzyme abundance, facilitates the usage of the available compendia and deep learning models of turnover numbers in the design of precise metabolic engineering strategies. Our results expand the usage of genome-scale metabolic models toward the identification of targets for protein engineering, allowing their direct usage in the generation of innovative metabolic engineering designs for various biotechnological applications.

Author summary

Enzymes play a crucial role in metabolic processes, and by selecting enzymes with optimal activity, we can enhance the production of desired compounds. However, there has been no computational approach to designing metabolic engineering strategies based on the modification of enzyme activities. In this study, we developed a new computational method called Overcoming Kinetic rate Obstacles (OKO) to increase chemical production

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

Competing interests: The authors have declared that no competing interests exist.

by modifying enzyme activities in *E. coli* and *S. cerevisiae*. Our method uses growing data on enzyme efficiency from experiments and deep learning models to suggest strategies for metabolic engineering. We applied OKO to increase the production of over 40 different compounds in models of *E. coli* and *S. cerevisiae*, and found that it can at least double their production without severely affecting cell growth. Finally, we showed that by refining OKO to account for changes in both enzyme activity and abundance, our method can more effectively use existing data and models to design precise strategies for improving chemical production. Our findings pave the way for advanced metabolic engineering techniques for various biotechnological applications.

Introduction

Microbial cell factories have emerged as a tractable and sustainable solution for the production of important chemicals from renewable biomass [1]. Building such factories necessitates the design and implementation of systems-level metabolic engineering strategies that streamline, modify, and expand the biosynthetic capabilities of microbes serving as design chassis [2]. The *in silico* design of metabolic engineering strategies has been propelled by the developments of constraint-based modeling approaches that make use of genome-scale metabolic models [3–6]. These approaches have focused on *in silico* manipulation of fluxes, achieved *in vivo* by modifying the expression of genes and abundance of downstream proteins by using knock-out, knock-down, or over-expression techniques [7–13]. The production of chemicals by cell factories is, however, often impeded by insufficient catalytic activity of enzymes, resulting in negligible gains despite manipulations in protein abundance. Yet, rational manipulation of the catalytic activity of enzymes has not yet been considered for *in silico* design of metabolic engineering strategies, despite recent advances in enzyme engineering strategies that can be readily employed to address this problem [14].

The turnover number, k_{cat} , of an enzyme denotes the maximum number of substrate molecules converted to product molecules per unit time per active site, and quantifies the enzyme catalytic activity. Recent advances in constraint-based modeling have resulted in approaches that: (i) generate enzyme-constrained genome-scale metabolic models (ecGEMs) [15,16] by integrating turnover numbers from major repositories [17,18] and deep learning predictions [19–22], (ii) combine fluxomics and proteomics data to estimate turnover numbers for several model organisms [23–29], and (iii) correct turnover number estimates to better match measured phenotypes [30,31]. The essence of ecGEMs is that the rate of a metabolic reaction is bounded by the product of the turnover number and abundance of an enzyme catalyzing the reaction. These constraints allow establishing direct relations between metabolic activity, proteome allocation, physiology, and growth for organisms with available ecGEMs, including: *Escherichia coli* [32], *Saccharomyces cerevisiae* [31], Chinese hamster ovary cells [33], *Homo sapiens* [34], *Bacillus subtilis* [35], *Corynebacterium glutamicum* [36], *Chlamydomonas reinhardtii* [28] and *Rhizophagus irregularis* [37]. Importantly, such enzyme constraints indicate a viable way to manipulate native metabolism by designing strategies focused on modification of turnover numbers of enzymes. These strategies can be implemented using advances in deep learning of turnover numbers [19–22], in predictions of protein structures that indicate key amino acid residues driving modifications of catalytic activities [38], and in new tools for directed enzyme evolution [39].

However, the existing ecGEMs have been used to design engineering strategies based only on manipulation of enzyme abundance, following the same principles underlying flux

manipulation [7–11]. For instance, an ecGEM of *E. coli* was used to improve lysine production, while ensuring fixed minimal growth, by using flux balance analysis (FBA) [40]. Overexpression of proteins with largest predicted abundance was used to validate the predicted increase in lysine production [32]. Similar approach was used with an ecGEM for *C. glutamicum* to determine enzymes with predicted differential protein abundance between low and near-optimal growth rate obtained from the *in silico* maximization of lysine production [36]. Another study used an ecGEM of *B. subtilis* to identify knockout strategies to overproduce poly- γ -glutamic acid (γ -PGA) [35]. The knock-out strategy was simulated using the well-established minimization of metabolic adjustment method (MOMA) [41], and indicated differences to strategies obtained from the application of MOMA to classical GEMs; the *in vivo* implementation of the resulting strategies indicated a two-fold increase in the production of γ -PGA. While these studies have demonstrated the use of ecGEMs in the design and investigation of strategies to overproduce a few, specific compounds, another study has shown that overproducing certain chemicals including many amino acids, cannot be achieved by manipulation of gene expression due to conflicts arising from promiscuous enzymes, captured by the participation of one gene in the gene-protein-reaction rules of multiple reactions [11]. Specifically, when considering metabolic engineering strategies compatible with gene-protein-reaction rules, present in both ecGEMs and classical GEMs, the flux of one reaction catalysed by a promiscuous enzyme should be increased, while decreasing the flux of another reaction catalysed by the same enzyme. This leads to infeasibilities of designing metabolic engineering strategies that can be resolved by targeted manipulation of turnover numbers that do not tinker with the transcriptional regulatory machinery.

Motivated by developments of ecGEMs that include constraints on how fluxes depend on enzyme abundance and catalytic rates, here we devise, implement, and test constraint-based approaches to design metabolic engineering strategies that: (1) rely on manipulation of turnover numbers, without modifying enzyme abundances compared to the wild type; this avoids the principal problem of gene conflicts arising due to the presence of promiscuous enzymes, (2) consider the simultaneous change of turnover numbers and enzyme abundances, compared to the wild type, and (3) test the resulting strategies by deploying them with a catalogue of proteins with diverse turnover numbers from 343 fungal species. Therefore, we provide a pioneering constraint-based approach applicable with large-scale ecGEMs systematically identifying which enzyme turnover numbers need to be manipulated and in what direction to enhance the capacity of cell factories with respect to a large set of chemicals native to the hosts.

Results and discussion

Targeting turnover numbers for engineering applications

The integration of enzymatic constraints into ecGEMs [15] facilitates the development of our optimization-based approach, termed Overcoming Kinetic rate Obstacles (OKO), that predicts metabolic engineering strategies targeting turnover numbers. OKO manipulates the turnover numbers of enzymes under the assumption that the abundances of enzymes in the ecGEMs of wild type and the engineered organism are not significantly changed. This setting is particularly relevant for *in vivo* metabolic engineering strategies, since it does not require tinkering with the underlying transcriptional and (post)translational regulatory machinery of cell factories. That said, practical applications of OKO may still necessitate heterologous expression of genes from other organisms whose proteins have desired catalytic properties.

OKO is formulated as a two-step approach: In the first step, we determine the maximum product yield in the wild-type model at optimal growth rate under the set of constraints imposed by ecGEMs. Next, we identify the protein allocation by minimizing the enzyme usage

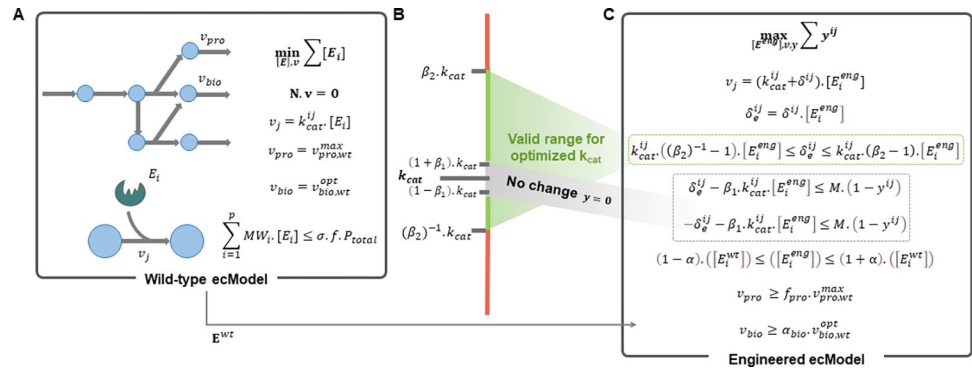


Fig 1. Schematic overview of OKO, a constraint-based approach to predict metabolic engineering strategies targeting turnover numbers. OKO is a two-step approach: In the first step (A), the protein allocation (E^{wt}) is determined by minimizing the enzyme usage at the maximum product yield ($v_{pro,wt}^{max}$) in the wild-type model at optimal growth rate ($v_{bio,wt}^{opt}$) under the set of constraints imposed by ecGEMs. (B) It is guaranteed that the changed turnover number is within the admissible range defined by a tuneable parameter β_2 . In addition, we introduce a binary variable (y) for each turnover number, that keeps track of whether a turnover number is considered to be changed, using a tuneable parameter β_1 . The second step of OKO (C) minimizes the number of significantly changed turnover numbers while ensuring that a desired level of chemical production ($f_{pro} \cdot v_{pro,wt}^{max}$) is achieved at a factor (α_{bio}) of optimal growth without affecting protein abundance (E^{eng}) compared to the wild type. The small parameter α quantifies the insignificant deviation from the enzyme abundances in the wild-type model.

<https://doi.org/10.1371/journal.pcbi.1012576.g001>

(see Fig 1A, Methods). In the second step, with the obtained enzyme abundances in the wild-type model, we constrain the abundance ranges in the model of the engineered strain, to ensure no significant change in comparison to the wild type strain (Methods). By introducing a binary variable for every turnover number, indicating its modification or lack thereof, and a tuneable parameter, specifying the threshold at which changes to the turnover numbers are considered significant, we keep track of whether or not a turnover number is considered changed. Another tuneable parameter is in turn used to define the admissible range for the modified turnover number and guarantees that its value remains within this range (Fig 1B, Methods). The second step of OKO then minimizes the number of significantly changed turnover numbers while ensuring that a desired level of chemical production is achieved at a factor of optimal growth without affecting protein abundance compared to the wild type (Fig 1C, Methods).

Application of OKO to two cell factories

Modification of enzyme turnover numbers enhances the capacity of wild type cell factories and marks a departure from typical engineering strategies that rely on manipulation of enzyme abundance in prioritizing product formation at the expense of biomass growth. As a result, we expect that by using OKO, the optimal growth rate in the engineered model may remain almost unchanged in comparison to that of the wild-type, thus breaking the trade-off between product production and growth (or biomass formation) [42]. The reason for this claim lies in the fact that the modification of the turnover numbers, in the absence of optimization of a product of interest, may increase growth in comparison to the wildtype; hence, diverting part of this growth increase towards the production of the chemical of interest can lead to almost unaltered growth while increasing the target production rate compared to the wildtype.

Here, we applied the proposed computational approach, OKO, to the ecYeastGEM_v8.3.4 [31] of the yeast *Saccharomyces cerevisiae* and the eciML1515 [32] of the bacterium *Escherichia coli* as cell factories. We aimed to determine modification strategies that rely on turnover numbers for the overproduction of multiple native compounds in these hosts. The ecYeastGEM

v8.3.4 model contains 1148 enzymes, of which 968 are linked to k_{cat} values, resulting in a total of 4261 available k_{cat} values that can be modified. The eciML1515 model contains 1505 enzymes with 3392 available k_{cat} values linked to 1259 enzymes in the model.

Domenzain *et al.* [43] proposed a computational method to predict optimal combinations of gene engineering targets and applied it to the ecYeastGEM_v8.3.4 model to increase the production of 102 different chemicals. Their method involves scoring genes within the model to find the magnitude and directionality of genetic modifications, and then discards genes that encode enzymes that are unfavorable for the production of the target metabolite. Finally, the approach identifies the minimal set of modifications for the transition of cells from biomass formation to metabolic production. Of the chemicals studied in [43], 49 compounds, including amino acids, are native metabolites in *Saccharomyces cerevisiae*. Mapping of these native compounds to metabolites in eciML1515 using KEGG IDs, CHEBI IDs and chemical formulae resulted in 41 metabolites to which the OKO method was applied (see [S1 Table](#)).

We first calculated the maximum production of each native compound at optimal growth rates, i.e. 0.38 per hour for ecYeastGEM_v8.3.4 and 0.57 per hour for eciML1515. Enzyme abundances were then calculated by using the first step of OKO (see [Methods](#)). To demonstrate that the resulting enzyme abundances can be used as a reference point, we conducted a robustness analysis relying on tailored variability analysis that specifies admissible ranges for enzyme abundance (see [Methods](#)). The results show that, on average, the logarithmic ratio of the maximum and minimum values for each enzyme abundance at the maximum production of the different chemicals of interest is 0.015 for the ecYeastGEM_v8.3.4 model and varies from 6×10^{-4} to 0.011 for the eciML1515 model (see [S1](#) and [S2](#) Tables, respectively). This demonstrates that the estimated maximum and minimum abundance values for every enzyme are very close to each other, allowing the usage of the estimated enzyme abundances in the remaining computational steps of OKO. Note that, like in other variability analyses in constraint-based modelling, these estimated ranges are direct consequence of the applied objective and the imposed constraints in the first step of OKO.

We then used OKO to search for strategies with the minimum number of changes in k_{cat} values that at least double the compound production while maintaining almost the same growth rates as in the wild type. The minimum number of modified k_{cat} values for overproduction of the chemicals of interest ranges from 7 to 338 for the ecYeastGEM_v8.3.4 model and from 4 to 39 for the eciML1515 model (see [S1](#) and [S2](#) Tables, respectively). The two compounds with the least number of modifications required are choline and palmitoleate in the ecYeastGEM_v8.3.4 model, with seven modified turnover numbers, and proline and serine in the eciML1515 model, with four and five modified turnover numbers, respectively (see [S3 Table](#) for the involved enzymes).

Since our goal is to double the production of a chemical of interest in comparison to the wild type, while ensuring certain growth, one may expect that only allowing for an increase in turnover numbers would suffice to address these objectives. Interestingly, however, the increase in production of the considered compounds does not necessarily entail only increases in the turnover numbers. We found that in the ecYeastGEM v8.3.4 model on average 35.5% of the enzymes require an increase in their k_{cat} values, while in the eciML1515 model, this number is 38.5% to double the production of the selected compounds. Further, since in metabolic engineering it is easier to knock out an enzyme than to reduce its turnover number via enzyme engineering, we also identified the cases where the reduction in turnover number could be replaced by knocking out the enzyme. To the end, we inspected if an enzyme requiring a reduction in k_{cat} value was associated with a flux of zero ($< 10^{-6} \text{ mmol gDW}^{-1} \text{ h}^{-1}$) in the engineered model. On average, 53.9% of the reductions in the ecYeastGEM_v8.3.4 model and 12.9% in the eciML1515 model fell into this category. This finding indicated that decrease in

turnover number, without blocking of reactions, are required to simultaneously achieve different objectives, including growth and production of the chemical of interest. Indeed, the growth rates of the engineered ecYeastGEM_v8.3.4 and eciML1515 models are on average 88% of the optimal growth rates of the wild-type models, indicating that the increase in the production of compounds following the proposed strategies does come at some, but not drastic cost of growth.

Promiscuous enzymes have been deemed reservoirs of functions that can be used in metabolic engineering, especially if particular catalytic activities are modified in a desired direction [2]. Hence, we also investigated whether enzymes targeted for modification of their turnover number were more likely to be promiscuous. At fixed enzyme abundances, our results imply that the turnover numbers of non-promiscuous enzymes in the model are mostly limiting the over production of the investigated compounds. We found, on average, 26.5% and 24.3% of the enzymes in the proposed strategies for the ecYeastGEM_v8.3.4 and eciML1515 models, respectively, were exhibited promiscuous functions. We further examined the target reactions in the modification strategies to determine whether the estimated turnover numbers suggested by the OKO approach were associated with any promiscuous enzymes for these reactions. In 47% of the proposed modification strategies for the ecYeastGEM_v8.3.4 model and in 62% of the strategies for the eciML1515 model, more than 50% of the estimated turnover numbers were associated with promiscuous enzymes (S1 and S2 Tables).

OKO identifies strategies for overproduction of amino acids

Razaghi-Moghadam et al. [11] investigated the participation of enzymes in different reactions and the inclusion of complex gene-protein-reaction rules, and showed that the inherent complexity of biological systems is extensively reflected in metabolic models. Therefore, strategies that build on existing enzymes and pathways often lead to conflicts in the manipulation of enzyme abundances. For example, using their proposed approach, it was found that there is no feasible strategy at the native enzyme level to increase production of almost all amino acids in *E. coli* and *S. cerevisiae*. This finding motivated us to further investigate the performance of OKO to increase the production of amino acids in these cell factories. To this end, we first reviewed the primary results of OKO, which indicated the minimum number of modified turnover numbers for amino acid overproduction. This number ranges from 9 to 20 for the ecYeastGEM_v8.3.4 model and from 4 to 25 for the eciML1515 model (Fig 2A). We found that, on average, 35% and 37.4% of the proposed enzymes in the ecYeastGEM_v8.3.4 and eciML1515 models, respectively, required an increase in their k_{cat} values.

Furthermore, we observed that the proposed amino acid modification strategies involve a total of 83 enzymes in the ecYeastGEM v8.3.4 model and 80 enzymes in the eciML1515 model. Of these, 15 enzymes in the ecYeastGEM v8.3.4 model and 17 enzymes in the eciML1515 model are required for the overproduction of at least three amino acids. These enzyme names are illustrated in Fig 2B and 2C, where the number of amino acids for which each enzyme modification is required and the log-fold changes of the enzymes are visually depicted. Moreover, and most importantly, the relative changes in the k_{cat} values are small, reaching the defined limits (either $(\beta_2)^{-1}$ or β_2) only 0.6% and 7.6% of the time for the ecYeastGEM v8.3.4 and eciML1515 models, respectively (see S4 and S5 Tables).

In addition, as shown in Fig 2B and 2C, for each enzyme included in the proposed amino acid modification strategies, the log-fold changes are predominantly distributed to one side of zero, indicating that the proposed modifications are biased towards either increasing or decreasing the k_{cat} values to increase the production of each amino acid. Furthermore, the box-plots in Fig 2B and 2C show a small spread, indicating the uniformity of the estimated k_{cat}

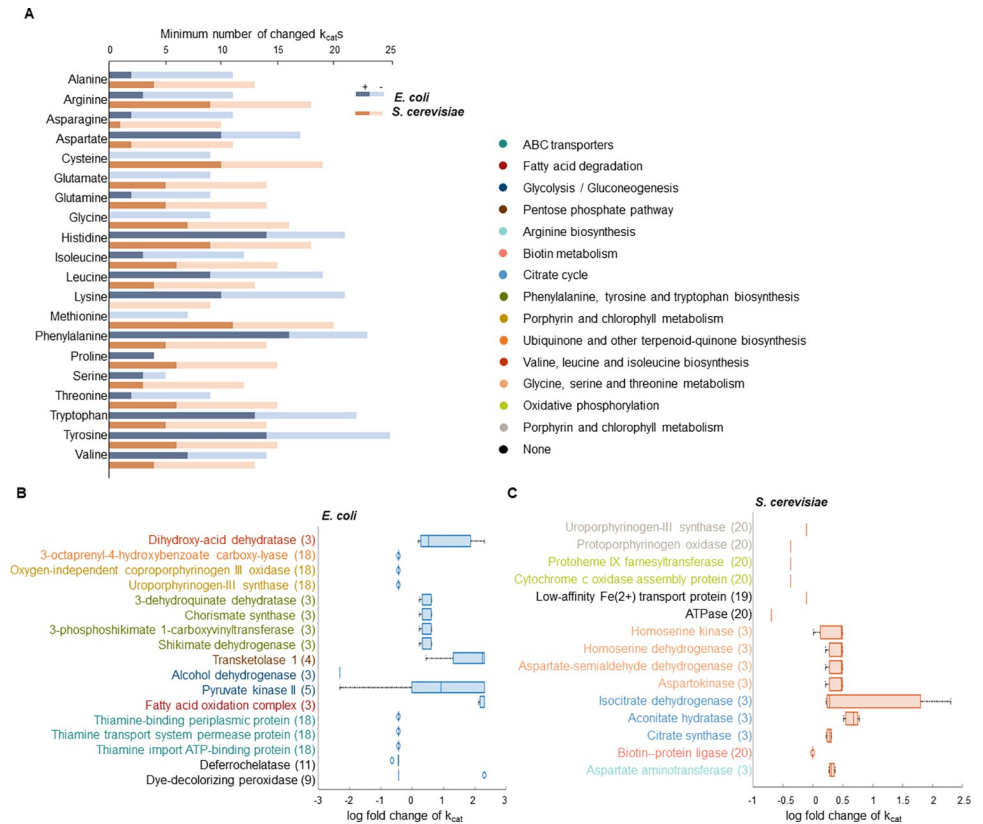


Fig 2. Performance of OKO for amino acids overproduction. (A) The minimum number of modified k_{cat} s for amino acid overproduction is shown for both *E. coli* (blue) and *S. cerevisiae* (orange) ecGEMs. Darker (lighter) colors correspond to the number of increased (decreased) k_{cat} s. In (B) and (C) illustrated are enzymes with modified turnover numbers as part of proposed strategies for overproduction of at least three amino acids in *E. coli* and *S. cerevisiae* models, respectively. The box plots in each case show the log-fold changes for the enzymes listed in the proposed modification strategies over different amino acids. The numbers in the parentheses indicate the number of modifications for the respective enzymes in the different engineering strategies. The different colors of the enzymes correspond to the biological pathways in which they are involved, described in the legend.

<https://doi.org/10.1371/journal.pcbi.1012576.g002>

values for each enzyme. Therefore, *in vivo* enzyme engineering following the proposed strategy for overproduction of amino acids is expected to benefit the increase of production of other amino acids that also relies on that enzyme.

As depicted in Fig 2B and S4 Table, the eciML1515 model demonstrates two exceptions to the consistency of change directions. The k_{cat} value of dye-decolorizing peroxidase YfeX enzyme (Uniprot ID P76536) in the ferrochelatase reaction is initially 25999.42 (s^{-1}) in the eciML1515 model. For the overproduction of alanine, aspartate, cysteine, glutamine, glycine, threonine, tryptophan, and tyrosine, this value must be reduced by a factor of 0.7. Conversely, overproduction of proline requires a 10-fold increase in the k_{cat} value. In addition, pyruvate kinase II (Uniprot ID P21599) with an initial k_{cat} value of 3204.01 (s^{-1}) in various pyruvate kinase reactions of the eciML1515 model needs to be increased for arginine and isoleucine overproduction and decreased for tyrosine overproduction. In the proposed strategy for arginine overproduction, the k_{cat} value of pyruvate kinase II needs to be increased 10-fold in two reactions and 2.5-fold in another reaction.

In the ecYeastGEM v8.3.4 model, as shown in Fig 2C and detailed in S5 Table, the estimated changes in k_{cat} values for two enzymes show greater variability. The isocitrate dehydrogenase [NADP] cytoplasmic (IDH) enzyme (Uniprot ID P41939) has a k_{cat} value of 58.30 (s^{-1}) in the

isocitrate dehydrogenase (NADP) reaction. This value needs to be increased by a factor of 1.3 for the overproduction of glutamine and proline, and by a factor of 10 for the overproduction of glutamate. In addition, the homoserine kinase (HK) (HSK) enzyme (Uniprot ID P17423), which has a k_{cat} value of 40.80 (s^{-1}) in the homoserine kinase reaction, needs to be increased 1.6-fold for glycine and threonine overproduction, whereas only a 1.01-fold increase is suggested for isoleucine overproduction. These examples demonstrate that while changes in some turnover numbers can generally accommodate usage of the engineered enzymes in the implementation of engineering strategies for different compounds, there are few, notable exceptions.

As discussed above, the decrease in turnover number is required to simultaneously achieve different objectives, including growth and production of the chemical of interest. We further examined the enzymes for which a reduction in the turnover number was proposed to increase the production of at least one amino acid (see S6 and S7 Tables). In the case of ecYeast-GEM_v8.3.4, there were eight different enzymes of this type involved in *porphyrin and chlorophyll metabolism, oxidative phosphorylation and biotin metabolism*. For eciML1515, 13 enzymes were targeted for reduction in turnover number, mainly involved in *porphyrin and chlorophyll metabolism, ubiquinone and other terpenoid-quinone biosynthesis and ABC transporters*.

This study presents the first *in silico* optimization method targeting enzyme activity rather than flux or enzyme abundance. Consequently, direct comparisons with existing approaches may not be entirely appropriate. Nonetheless, to check the extent to which the manipulation of the catalytic activity of enzymes improves the proposed metabolic engineering strategies, we compared the strategies proposed by OKO with those of OptReg [8], an alternative approach that only focuses on *in silico* manipulation of fluxes (see S8 Table). The results showed that, on average, in ecYeastGEM_v8.3.4 and eciML1515, the number of required manipulations proposed by OptReg was 12 and 17.22 times higher than those proposed by OKO, respectively.

Setting the ground for experimental validation

Metabolic engineering approaches that aim to use enzymes with the optimized k_{cat} values can leverage either existing nonnative enzymes or engineered enzymes with modified substrate specificity. Methods that rely on existing enzymes have the flexibility to reuse the enzyme repository in nature and facilitate the creation of synthetic pathways for engineering purposes. This motivated us to investigate whether the identified k_{cat} values in our engineering strategies are close to the turnover numbers of these enzymes in other organisms.

Enzyme databases such as BRENDA [17] and SABIO-RK [18] contain extensive repositories of k_{cat} values. However, there is still a lack of high-throughput techniques for measuring k_{cat} values, resulting in a notable disparity in the number of k_{cat} values available compared to the vast number of organisms and metabolic enzymes in existence [29]. To fill this gap, deep learning and machine learning approaches have been proposed to predict the k_{cat} values for different enzymes in different organisms, and have shown high performance [19–22]. To investigate whether the k_{cat} values estimated by OKO could be found for the respective enzymes in other organisms, we used the predicted k_{cat} values from the DLKcat [19] and TurNup [20] approaches. DLKcat includes the prediction of k_{cat} values for 651 metabolic enzymes in 343 organisms (see S9 Table). We employed another deep learning method, TurNup [20], to predict turnover values for the same 343 organisms for which turnover numbers were predicted using the DLKcat. Altogether, TurNup includes the prediction of k_{cat} values for 630 metabolic enzymes in 343 organisms (see S10 Table).

Having mapped the two sets of k_{cat} values from ecGEMs and the DLKcat results (see Methods), we examined each enzyme with an estimated k_{cat} value from OKO to check if it was

Table 1. Deep-learning-driven identification of suitable turnover numbers from other organisms. Shown is the number of enzyme-reaction pairs in the proposed strategies for which turnover numbers can be predicted by DLKcat and TurNup. The table also includes the average distance of the k_{cat} values predicted by DLKcat and TurNup and the modifications resulting from OKO and OKO⁺, respectively.

	OKO		OKO ⁺ (using DLKcat)		OKO ⁺ (using TurNup)	
	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>E. coli</i>	<i>S. cerevisiae</i>
Found enzyme-reaction pairs	57	137	354	817	365	1559
Missing enzyme-reaction pairs	179	111	0	0	0	0
Average distance (s^{-1})	32188.07	503.15	2.35	4.92	61.31	346.62

<https://doi.org/10.1371/journal.pcbi.1012576.t001>

included in the list of enzymes with a predicted k_{cat} value from the DLKcat approach. If an enzyme was included, we then identified the closest k_{cat} value to the estimated one across all organisms. In the proposed amino acid modification strategies, we identified 248 and 236 estimated k_{cat} values for the ecYeastGEM_v8.3.4 and eciML1515 models, corresponding to 83 and 80 different enzymes, respectively. As shown in Table 1, only 55.2% and 24.2% of these enzymes were included in the list of enzymes with predicted k_{cat} values by the DLKcat approach. Additionally, the average distance of the closest k_{cat} values to the estimated values across all organisms was 503.15 (s^{-1}) and 32188.07 (s^{-1}) for the ecYeastGEM_v8.3.4 and eciML1515 models, respectively. These findings indicated that the enzymes with k_{cat} values estimated by the OKO approach were rarely found in other organisms and could not be directly used for *in vivo* engineering. Specifically, these results show that the OKO approach is effective in identifying optimal kinetic parameters that may not occur naturally, demonstrating its strength in predicting novel enzyme functionalities. However, it also highlights the challenge of translating these theoretical predictions into practical applications, requiring the design or engineering of new enzymes to achieve the desired catalytic efficiency *in vivo*. This highlights the need to impose constraints on the estimated turnover numbers, which we address in the following section by detailing the methodology used for these constraints and assessing their impact on the results.

Refinement of OKO enhances its applicability

As it is more convenient to construct synthetic pathways using existing enzymes, we also proposed a refinement of OKO, termed OKO⁺. In OKO⁺, for each enzyme-reaction pair with an available k_{cat} value in an ecGEM, we first extract the minimum and maximum of the predicted k_{cat} values in all other organisms using the mapping established between ecGEM and DLKcat results. In total, there are 4261 and 3392 enzyme-reaction pairs with available k_{cat} values in the ecYeastGEM_v8.3.4 and eciML1515 models, respectively, of which 1676 and 469 respected values are found in the predicted k_{cat} values from the DLKcat approach (see S11 Table). We introduce two additional sets of constraints in OKO⁺ to ensure that only enzymes with found predicted turnover numbers can be targeted: (1) for the enzyme-reaction pairs with available turnover predictions from DLKcat or other resources, the predicted k_{cat} should remain either unchanged or within the range extracted from the values obtained from DLKcat, and (2) the enzyme-reaction pairs with available k_{cat} values in the ecGEM, for which values from DLKcat (from other organisms) are not available, are not allowed to be modified.

The introduction of these two additional sets of constraints resulted in infeasibilities to identify engineering strategies that yield the desired levels of product formation and growth. Therefore, in OKO⁺, we also allow for variation in enzyme abundance, modelled by introduction of a binary variable for each enzyme abundance indicating if it is significantly altered (with respect to a tunable parameter). This approach then minimizes the weighted average of the number of modified turnover numbers and the number of modified abundances, with preference for the former (see Methods).

By deploying OKO⁺ with the same models we found that the minimum number of modified k_{cat} values for overproduction of the chemicals of interest ranges from 0 to 85 for the ecYeastGEM_v8.3.4 model and from 0 to 44 for the eciML1515 model (see S1 and S2 Tables), while also requiring changes in enzyme abundance. The ecYeastGEM_v8.3.4 model requires different number of changes in enzyme abundance, ranging from 2 to 240 modifications, while this number varies from 3 to 286 for the eciML1515 model. On average, 48.8% and 77% of the enzymes in the proposed strategies for the ecYeastGEM_v8.3.4 and eciML1515 models were promiscuous enzymes, respectively.

In addition, for each k_{cat} value estimated by OKO⁺ in amino acid modification strategies, we identified the closest k_{cat} value across all organisms. Clearly, due to the constraints imposed in OKO⁺, all enzymes with newly estimated k_{cat} values were included in the list of enzymes with predicted k_{cat} values from the DLKcat approach. The average distance of the estimated values to their closest counterparts across all organisms was 2.35 for the ecYeastGEM_v8.3.4 model and 4.92 for the eciML1515 model (see Table 1). The small average distance value indicates that OKO⁺ successfully identified a suitable enzyme for another organism with closely matching activity.

Inspection of the proposed strategies to increase the production of amino acids showed that the existing nonnative enzymes with k_{cat} values closest to the predicted by OKO⁺ originate from different organisms (see S1, S2 and S12 Tables). In total, 103 different organisms contribute to the proposed modification strategies for the ecYeastGEM_v8.3.4 model and 53 for the eciML1515 model (see S12 Table). Each organism contributes a different number of enzymes, ranging from 1 to 50 for the proposed modifications in the ecYeastGEM_v8.3.4 model and from 1 to 29 for the eciML1515 model.

To further analyze the applicability of the strategies resulting from OKO⁺, we constructed two distinct ecGEMs by substituting the k_{cat} values in the original ecGEMs with the values estimated by OKO⁺, as well as the closest values from other organisms. We then assessed the performance of the strategies with the so-modified turnover numbers in producing the compounds of interest. In each model constructed and for each compound of interest, the biomass level was set to match that of the proposed engineering strategy for the overproducing that chemical. In addition, the number of modifications to the enzyme abundance was constrained by the number of modifications in the corresponding proposed engineering strategy. Clearly, this evaluation is only applicable in cases where at least one modification to the enzyme activity was proposed by OKO⁺ and this is for 19 amino acids in the ecYeastGEM_v8.3.4 model and 12 in the eciML1515 model. With these constraints, we calculated the maximum production of each chemical of interest in the two constructed ecGEMs and determined their ratio $\left(\frac{v_{max}^{DLKcat}}{v_{max}^{OKO^+}}\right)$ (see S1 and S2 Tables). In the majority of cases the two constructed ecGEMs yielded the same levels of product formation. Small differences in product formation levels were only observed for arginine in the eciML1515 model, with the ratio being 0.97.

Effect of deep learning model for turnover numbers on predicted engineering strategies

To evaluate the impact of deep and machine learning methods on the engineering strategies proposed by OKO⁺, we employed another deep learning method, TurNup [20], to predict turnover values for the same 343 organisms used with DLKcat. We used the predicted values to constrain the optimization approach. TurNup predicted 2912 and 943 turnover numbers for enzyme-reaction pairs with available k_{cat} values in the ecYeastGEM_v8.3.4 and eciML1515 models, respectively (see S13 Table).

According to the constraints in OKO+, the higher number of predicted turnover numbers from TurNup compared to DLKcat should theoretically provide more opportunities for modification, thereby reducing the number of changes needed. However, in the case of ecYeastGEM_v8.3.4, out of 1470 shared enzyme-reaction pairs from both approaches, 854 pairs have ranges of change from the TurNup approach that fall within the ranges from the DLKcat approach. This can result in fewer opportunities for modification and an increase in the number of proposed changes (see S1 and S2 Tables). Similarly, for eciML1515, there are 469 shared enzyme-reaction pairs between the two approaches, of which 282 pairs have ranges of change from the TurNup approach that are encompassed by the ranges from the DLKcat approach. This indicates that it is not reasonable to compare the results of OKO+ using different prediction models, and highlights the need for more refined prediction models to effectively improve optimization strategies.

The average difference of the estimated values to their closest counterparts across all organisms was 346.62 (s^{-1}) for the ecYeastGEM_v8.3.4 model and 61.31 (s^{-1}) for the eciML1515 model (see Table 1). The average distance values indicate that OKO+ identified closer enzymes from another organism with matching activity using the list of predicted turnover numbers from DLKcat compared to that from TurNup. In total, using the predicted turnover numbers by TurNup, 203 different organisms contribute to the proposed modification strategies for the ecYeastGEM_v8.3.4 model and 33 for the eciML1515 model (see S14 Table). Each organism contributes a different number of enzymes, ranging from 1 to 86 for the proposed modifications in the ecYeastGEM_v8.3.4 model and from 1 to 27 for the eciML1515 model.

Feasibility of strategies designed by employing OKO+

OKO+ was employed to design engineering strategies for the overproduction of 49 compounds in the ecYeastGEM_v8.3.4 model and 41 compounds in the eciML1515 model. The total number of modifications, including enzyme abundance and turnover numbers, varied from 2 to 293 in the ecYeastGEM_v8.3.4 model and from 3 to 316 in the eciML1515 model (see S1 and S2 Tables). Further analysis focused on the engineering strategies of OKO+ that are more feasible to implement, specifically those with a total number of modifications less than or equal to seven. Six such compounds were identified in the ecYeastGEM_v8.3.4 model and five in the eciML1515 model (see Table 2 and Fig 3).

It is noteworthy that all modification strategies proposed by OKO+ to overproduce these eleven compounds involved an increase in enzyme abundance. Furthermore, the proposed

Table 2. Proposed metabolic engineering strategies with feasible implementation. Shown are the compounds for which the proposed engineering strategies involve a total of seven or fewer modifications. The number of modifications to the turnover number and to the enzyme abundance are included.

		modifications of turnover number	Modifications of enzyme abundance
<i>E. coli</i>	Aspartate	0	5
	Citrulline	0	7
	Glycine	0	3
	Homoserine	2	3
	(S)- malate	0	6
	Acetate	0	6
<i>S. cerevisiae</i>	Isoamylol	1	4
	Isoleucine	1	6
	Lysine	0	2
	Spermine	1	5
	(S)-malate	1	2

<https://doi.org/10.1371/journal.pcbi.1012576.t002>

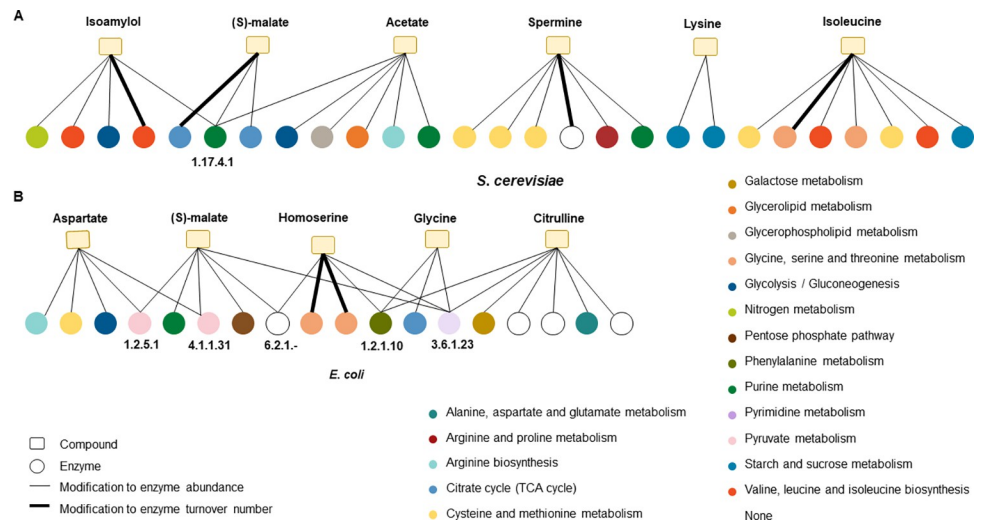


Fig 3. Proposed metabolic engineering strategies presented as bipartite graphs. Bipartite graphs are constructed to represent associations between enzymes to be modified for engineering purposes and compounds for which the OKO⁺ engineering strategies involve at most seven modifications. Six such compounds were identified in the *S. cerevisiae* model (A) and five in the *E. coli* model (B). The EC numbers are provided only for those enzymes to which modifications contribute to the overproduction of more than one compound. Thicker edges are associated with modifications to the enzyme turnover numbers. The different colors of the enzymes correspond to the biological pathways in which they are involved.

<https://doi.org/10.1371/journal.pcbi.1012576.g003>

engineering strategies do not only involve modification of enzyme abundance, but also adjustment of enzyme turnover numbers. For instance, the overproduction of isoamylol requires a 1.4-fold increase in the k_{cat} value of the 3-isopropylmalate dehydrogenase enzyme (Uniprot ID P04173), with the closest comparable value found in *Candida rhagii*. Similarly, for the overproduction of isoleucine, a 1.2-fold increase in the k_{cat} value of homoserine dehydrogenase (Uniprot ID P31116) is required, and the closest identified value comes from *Blastobotrys raffinofermentans*.

In the proposed strategy for spermine overproduction, a 1.5×10^{-4} -fold reduction in the k_{cat} value of fumarate reductase 2 (Uniprot ID P21375) was proposed, with the closest corresponding value identified in *Ogataea trehaloabstinens*. Finally, for (S)-malate overproduction, the proposed strategy involved a 1.6-fold increase in the k_{cat} value of pyruvate carboxylase 1 (Uniprot ID P11154), with the closest comparable value coming from *Lachancea waltii*.

Additionally, for increasing homoserine production in the ecIML1515 model, OKO⁺ proposed an engineering strategy involving a 1.3-fold increase in the k_{cat} value of aspartate-semialdehyde dehydrogenase (Uniprot ID P0A9Q9) and a 2.9-fold increase in the k_{cat} value of the bifunctional aspartokinase/homoserine dehydrogenase 1 (Uniprot ID P00561). The closest comparable k_{cat} values to these proposed modifications are identified in *Saturnispora zaruensis* and *B. anomalus*, respectively. These results jointly indicate that feasible engineering strategies can be obtained when combining overexpression of enzymes with enzymes engineered towards an increase in turnover numbers.

Conclusions

Availability of enzyme turnover numbers, estimated by coupling omics data with models or relying on classical experimental techniques or predicted by machine and deep learning approaches, holds the promise for their usage in the design and deployment of precise metabolic engineering strategies. However, despite this promise, metabolic engineering has not

capitalized from the availability of repositories of and deep learning models for turnover numbers. We address this gap by proposing, implementing, and testing a constraint-based approach, termed OKO, for the design of metabolic engineering strategies that identifies which turnover numbers need to be modified alongside the extent of the change. In experimental setting, the identified modifications can be achieved by using either the two well-established strategies for protein engineering, i.e. rational protein design and directed evolution [44], or relying on proteins from other organisms for which turnover numbers can be predicted by deep learning models [19–22]. In the first case, one can make use of advances in CRISPR-Cas9 system for precise single-base mutations (and combinations thereof) in combination with the pocketome databases [45] to identify active site residues that lead to the proposed increased or decreased catalytic rates of a protein for a given reaction. In the second case, one can knock-out the promiscuous enzyme in the host and make use of heterologous non-promiscuous enzymes for the reactions of interest, if such can be identified with computational search (as done in our study, relying on models for prediction of catalytic rates). Despite the existence of these options, further advances in structural biology and synthetic biology will be needed to overcome challenges in implementing the proposed strategies.

Our findings indicated that the use of OKO with enzyme-constrained metabolic models of *E. coli* and *S. cerevisiae* resulted in metabolic engineering strategies that overcome the trade-off between growth and increased production of a chemical of interest. Importantly, we found that the overproduction of chemicals of interest does not only entail an increase in selected enzymes. Motivated by the need to consider the available estimates and predictions of turnover numbers for selected enzymes, we also presented a refinement of OKO that combines the modification of turnover numbers (from a well-defined range) with alterations of enzyme abundances. The resulting approach resulted in feasible strategies for overproduction of amino acids in the two hosts, which was not possible to obtain with other constrained-based approaches focused only on the alteration of gene expression (as a proxy for enzyme abundance). However, our findings indicated that the metabolic engineering strategies strongly depend on the predictions of the deep learning models for turnover numbers. It is important to note that the effectiveness of the proposed strategies also depends on the accuracy of the initial turnover numbers used in the model. Altogether, our study paves the way for the deployment of precise metabolic engineering strategies that rely on the natural diversity of turnover numbers for given enzymes or that can use techniques for enzyme engineering.

Methods

Formulation of OKO

To determine the abundance of enzymes at the maximum product yield ($v_{pro,wt}^{max}$) in the wild-type model, while ensuring optimal growth rate ($v_{bio,wt}^{opt}$), we solve the following linear programming problem (LP1):

$$\min_{v,E} \sum_{i=1}^p [E_i] \quad (1)$$

s.t

$$\mathbf{N} \cdot \mathbf{v} = 0 \quad (2)$$

$$v_j^{min} \leq v_j \leq v_j^{max}, 1 \leq j \leq r \quad (3)$$

$$v_{pro} \geq 0.98 \cdot v_{pro,wt}^{max} \tag{4}$$

$$v_{bio} \geq 0.98 \cdot v_{bio,wt}^{opt} \tag{5}$$

$$v_j = k_{cat}^{ij} \cdot [E_i], 1 \leq j \leq r, 1 \leq i \leq p \tag{6}$$

$$\sum_{i=1}^p MW_i \cdot [E_i] \leq \sigma \cdot f \cdot P_{total} \tag{7}$$

where $[E_i]$ is the abundance of enzyme i , p denotes the number of enzymes, \mathbf{N} is the stoichiometric matrix, \mathbf{v} is the flux distribution vector, v_j is the metabolic flux through reaction j , r is the number of reactions, v_{pro} and v_{bio} are the fluxes through the synthesis of the product of interest and the biomass reaction, respectively, k_{cat}^{ij} is the catalytic rate of enzyme i catalysing reaction j , MW_i is the molecular weight of enzyme i , σ is the average *in vivo* enzyme saturation, f is the mass fraction of all measured enzymes included in the model, and P_{total} is the total protein content.

We use the resulting obtained enzyme abundances ($[E_i^{wt}]$) to constrain the abundance range in the engineered model, using the following constraint:

$$(1 - \alpha) \cdot ([E_i^{wt}]) \leq [E_i^{eng}] \leq (1 + \alpha) \cdot ([E_i^{wt}]), 1 \leq i \leq p \tag{8}$$

where α is a small parameter ($\approx 10^{-1}$) that quantifies the insignificant deviation from the enzyme abundances in the wild-type model. Therefore, for $[E_i^{eng}]$, we still allow a deviation from $[E_i^{wt}]$, but not a large one, as controlled by the parameter α .

To overcome the shortcomings of abundance-based strategies [7–11], OKO opt to develop strategies that manipulate the turnover numbers. Accordingly, in the engineered model, the constraint in Eq (6) is replaced by the following constraint:

$$v_j = (k_{cat}^{ij} + \delta^{ij}) \cdot [E_i^{eng}], 1 \leq j \leq r, 1 \leq i \leq p \tag{9}$$

where δ^{ij} shows the deviation from the turnover number of the wild-type enzyme, and it can be positive or negative. Eq (9) contains a bi-linear term which includes δ^{ij} and $[E_i^{eng}]$. The Petersen linearization scheme, previously used in metabolic modeling [46], is applicable when one variable is binary, which is not the case here. Therefore, here to linearize this constraint a new variable $\delta_e^{ij} = \delta^{ij} \cdot [E_i^{eng}]$ and the following constraints are added to the optimization problem:

$$v_j = k_{cat}^{ij} \cdot [E_i^{eng}] + \delta_e^{ij}, 1 \leq j \leq r, 1 \leq i \leq p \tag{9}$$

$$-M \cdot [E_i^{eng}] \leq \delta_e^{ij} \leq M \cdot [E_i^{eng}], 1 \leq j \leq r, 1 \leq i \leq p \tag{10}$$

where M is a large number (10^6) chosen to be greater than all upper bound [46], and Eq (10) guaranties that $[E_i^{eng}] = 0$ implies $\delta_e^{ij} = 0$. As a result, Eq (9) is replaced by Eq (9) to remove the bi-linear constraint. The optimized turnover numbers can then be calculated by $k_{cat}^{ij*} = k_{cat}^{ij} + \delta_e^{ij} / [E_i^{eng}]$. For every k_{cat}^{ij} value, a binary variable y^{ij} is introduced; it takes the value of 0, if the catalytic rate is significantly changed, and 1, otherwise. For each k_{cat}^{ij} , we also define two intervals (see Fig 1B) within which the optimized value may fall. These intervals are adjusted to force the optimized value to be significantly, but not excessively away from k_{cat}^{ij} : if $(1 - \beta_1) \cdot k_{cat}^{ij} \leq k_{cat}^{ij*} \leq (1 + \beta_1) \cdot k_{cat}^{ij}$, then we consider the value of k_{cat}^{ij} to not have

significantly changed and y^{ij} equals 1. Further, we prevent large relative changes by ensuring that $(\beta_2)^{-1} \cdot k_{cat}^{ij} \leq k_{cat}^{ij*} \leq \beta_2 \cdot k_{cat}^{ij}$. The two tuneable parameters β_1 and β_2 define the allowed intervals for the optimized catalytic rate (k_{cat}^{ij*}), i.e. $\beta_1 (\approx 10^{-8})$ specifies the initial threshold for what is considered a change and $\beta_2 (\approx 10)$ sets the upper limit for acceptable change. Based on the following constraints, the binary variable y^{ij} enables us to switch between intervals and determine whether the k_{cat}^{ij} can be changed or not:

$$-\delta_e^{ij} - \beta_1 \cdot k_{cat}^{ij} \cdot [E_i^{eng}] \leq M \cdot (1 - y^{ij}), 1 \leq j \leq r, 1 \leq i \leq p \tag{11}$$

$$\delta_e^{ij} - \beta_1 \cdot k_{cat}^{ij} \cdot [E_i^{eng}] \leq M \cdot (1 - y^{ij}), 1 \leq j \leq r, 1 \leq i \leq p \tag{12}$$

$$k_{cat}^{ij} \cdot ((\beta_2)^{-1} - 1) \cdot [E_i^{eng}] \leq \delta_e^{ij} \leq k_{cat}^{ij} \cdot (\beta_2 - 1) \cdot [E_i^{eng}], 1 \leq j \leq r, 1 \leq i \leq p \tag{13}$$

Here, Eq (13) guaranties that the optimized turnover number (k_{cat}^{ij*}) is in the allowed range. OKO finds a solution with the minimum number of modified turnover numbers (which is equivalent to the maximum number of unchanged turnover numbers ($\sum_{1 \leq j \leq r, 1 \leq i \leq p} y^{ij}$), while guaranteeing that: (i) the flux towards the objective is at least a given factor, $f_{pro} (\geq 2)$ of the maximum product yield and (ii) a certain factor, $\alpha_{bio} (\approx 1)$, of optimal growth is achieved. These are enforced by the following constraints:

$$v_{pro} \geq f_{pro} \cdot v_{pro,wt}^{max} \tag{14}$$

$$v_{bio} \geq \alpha_{bio} \cdot v_{bio,wt}^{opt} \tag{15}$$

The combination of the linear objective function of $\max_{v,E,\delta_e,y} \sum_{1 \leq j \leq r, 1 \leq i \leq p} y^{ij}$ with the linear constraints in Eqs (2, 3, 7, 8, 9', 10–15) leads to our proposed Mixed-Integer Linear Programming (MILP) formulation of OKO (see Fig 1C).

Formulation of OKO+

In OKO+, we also allow for variation in enzyme abundance. To this end, for each $[E_i^{eng}]$ value ($1 \leq i \leq p$), we introduce an interval of $((1 - \beta_e) \cdot [E_i^{wt}], (1 + \beta_e) \cdot [E_i^{wt}])$ which, similarly to OKO, defines the range of unchanged values for the abundance. In addition, a binary variable y_e^i is introduced which takes the value of 1, if the enzyme abundance falls within the interval, and 0, otherwise. Here, $\beta_e (\approx 10^{-1})$ is a tuneable parameter that quantifies the deviation from the enzyme abundances in the wild-type model. Based on the following constraints, the binary variable y_e^i enables us to determine whether the $[E_i^{eng}]$ changes or not:

$$(1 - \beta_e) \cdot [E_i^{wt}] - [E_i^{eng}] \leq M \cdot (1 - y_e^i), 1 \leq i \leq p \tag{16}$$

$$-(1 + \beta_e) \cdot [E_i^{wt}] + [E_i^{eng}] \leq M \cdot (1 - y_e^i), 1 \leq i \leq p \tag{17}$$

where M is a large number (10^6). In OKO+, the constraint in Eq (8) is replaced by the constraints in the Eqs (16) and (17) and the objective is changed to $\max_{v,E,\delta_e,y} \sum_{1 \leq j \leq r, 1 \leq i \leq p} y^{ij} + w \sum_{1 \leq i \leq p} y_e^i$, where w is the weight reflecting the priority given to the number of changes for enzyme abundance or activity. Given that the primary aim of OKO is to identify modification strategies at the level of catalytic rates, the weight w was set to 10 to prioritize minimizing the number of modifications on the enzyme abundance.

Robustness analysis of enzyme abundance

To assess the precision of the estimated enzyme abundances, for each enzyme abundance we determine the range over which the minimum enzyme usage, from LP1, remains valid. This is done by calculating the minimum and the maximum values the abundance of each enzyme can take at minimum enzyme usage (sum_e^*) obtained from solving LP1, and can be done by solving the following linear programming problems (LP2):

$$\min_{v,E} / \max [E_i] \quad 1 \leq i \leq p \quad (18)$$

s.t

$$\sum_{i=1}^p [E_i] = sum_e^* \quad (19)$$

Eqs (2), (3), (4), (5), (6), (7)

Given that the minimum value is usually zero and the maximum is a very small number, we use the log-sum-exp trick [47] to calculate the logarithmic ratio between the maximum and minimum values to deal with the potential numerical instability. As an additional validation step, we compared the variability of enzyme use under both the equality and inequality conditions in Eq 6 and found that the results remained consistent, confirming that our initial assumption did not affect the conclusions.

OKO and OKO⁺ implementation

OKO and OKO⁺ both have MILP formulations and are implemented in MATLAB R2022b, available at <https://github.com/MonaRazaghi/OKO/>. These approaches take as input an ecGEM together with one or a list of compound indices in ecGEM. The output is a list of enzymes with their respective modifications, either on enzyme turn over number or abundance. Both OKO and OKO⁺ use the Gurobi Optimizer v9.5.2 [48] as their solver to efficiently solve the corresponding MILP problems.

OKO includes several tuneable parameters, α , β_1 and β_2 . To increase the likelihood of finding a solution for the MILP formulation of OKO, the tuneable parameters α and β_1 were set to 10^{-1} and 10^{-8} , respectively. In addition, the parameter β_2 was set to 10, to avoid large relative changes in k_{cat} values. In the case of OKO⁺, an additional tunable parameter β_e is included. This parameter was set to 10^{-1} to have reasonable intervals for enzyme abundance.

Mapping of k_{cat} values from ecGEMs and DLKcat

The k_{cat} values in ecGEMs are linked to enzyme-reaction pairs, whereas the predicted k_{cat} values from the DLKcat approach are linked to enzyme-metabolite pairs. To establish a mapping between the two sets of k_{cat} values, we first associated the enzymes with their respective EC numbers using the Uniprot database [49]. Enzyme-reaction pairs were then converted to enzyme-metabolite pairs by including all substrates involved in the reaction. Metabolite names from the ecGEMs were mapped to the metabolites present in the DLKcat results using MetaNetX IDs.

Supporting information

S1 Table. The result of applying the OKO method on ecML1515 model for overproducing 41 native metabolites of *E. coli*. Columns A-D present the name, KEGG ID, CHEBI ID and formulae of each metabolite, respectively. The maximum production level of each metabolite at optimal growth rate is shown in column E. Column F presents the production level of the

metabolite in the engineered model, where the biomass level is presented in column G. The numbers of Kcat that are required to increase and decrease are respectively given in columns H and I, and the sum of fold-changes for Kcats is given in column M. Choline is one metabolite mapped in *eciML1515* for which the maximum production is zero in both the wild-type and the engineered model. So, this metabolite is excluded from further analysis.

(XLSX)

S2 Table. The result of applying the OKO method on *ecYeastGEM_v8.3.4* model for overproducing 49 native metabolites of *S. cerevisiae*. Columns A-D present the name, KEGG ID, CHEBI ID and formulae of each metabolite, respectively. The maximum production level of each metabolite at optimal growth rate is shown in column E. Column F presents the production level of the metabolite in the engineered model, where the biomass level is presented in column G. The numbers of Kcat that are required to increase and decrease are respectively given in columns H and I, and the sum of fold-changes for Kcats is given in column M.

(XLSX)

S3 Table. Proposed engineering strategies for overproduction of compounds with the least number of modifications required.

(XLSX)

S4 Table. The log fold change of predicted kcat values for enzymes required for the overproduction of at least three amino acids in *eciML1515* model.

(XLSX)

S5 Table. The log fold change of predicted kcat values for enzymes required for the overproduction of at least three amino acids in *ecYeastGEM_v8.3.4* model.

(XLSX)

S6 Table. Enzymes required reduction in their turnover number for the overproduction of at least one amino acid in *eciML1515* model.

(XLSX)

S7 Table. Enzymes required reduction in their turnover number for the overproduction of at least one amino acid in *ecYeastGEM_v8.3.4* model.

(XLSX)

S8 Table. Comparison of OKO and OptReg.

(XLSX)

S9 Table. The number of genes and proteins with predicted kcat values in each species, using DLKcat.

(XLSX)

S10 Table. The number of genes and proteins with predicted kcat values in each species, using TurNup.

(XLSX)

S11 Table. Enzyme-reaction pair with available k_{cat} values in the DLKcat results with their corresponding range.

(XLSX)

S12 Table. Number of enzymes taken from each organism for amino acid modifications, using DLKcat.

(XLSX)

S13 Table. Enzyme-reaction pair with available k_{cat} values in the TurNup results with their corresponding range.

(XLSX)

S14 Table. Number of enzymes taken from each organism for amino acid modifications, using TurNup.

(XLSX)

Author Contributions

Conceptualization: Zahra Razaghi-Moghadam, Zoran Nikoloski.

Funding acquisition: Zahra Razaghi-Moghadam, Zoran Nikoloski.

Investigation: Zahra Razaghi-Moghadam.

Methodology: Zahra Razaghi-Moghadam.

Project administration: Zahra Razaghi-Moghadam, Zoran Nikoloski.

Software: Fayaz Soleymani Babadi.

Supervision: Zahra Razaghi-Moghadam, Zoran Nikoloski.

Validation: Zahra Razaghi-Moghadam, Zoran Nikoloski.

Writing – original draft: Zahra Razaghi-Moghadam, Zoran Nikoloski.

Writing – review & editing: Zahra Razaghi-Moghadam, Fayaz Soleymani Babadi, Zoran Nikoloski.

References

1. Cho JS., Kim GB., Eun H., Moon CW., Lee SY. Designing microbial cell factories for the production of chemicals. *JACS Au*. 2022, 4; 2(8), 1781–1799. <https://doi.org/10.1021/jacsau.2c00344> PMID: 36032533
2. Erb TJ., Jones PR., Bar-Even A. Synthetic metabolism: metabolic engineering meets enzyme design, *Current Opinion in Chemical Biology*, 2017, 37, 56–62. <https://doi.org/10.1016/j.cbpa.2016.12.023> PMID: 28152442
3. Monk JM., Lloyd C., Brunk E., Mih N., Sastry, A., King, Z. et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol*. 2017, 11; 35(10), 904–908. <https://doi.org/10.1038/nbt.3956> PMID: 29020004
4. Lu H., Li F., Sánchez BJ., Zhu, Z., Li, G., Domenzain, I. et al. A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun*. 2019, 10 (1), 3586. <https://doi.org/10.1038/s41467-019-11581-3> PMID: 31395883
5. Oh YK., Palsson BO., Park SM., Schilling CH., Mahadevan R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J Biol Chem*. 2007, 28; 282(39), 28791–28799. <https://doi.org/10.1074/jbc.M703759200> PMID: 17573341
6. Zhang Y., Cai J., Shang X., Wang B., Liu S., Chai X. et al. A new genome-scale metabolic model of *Corynebacterium glutamicum* and its application. *Biotechnol Biofuels*, 2017, 10, 169. <https://doi.org/10.1186/s13068-017-0856-3> PMID: 28680478
7. Burgard AP., Pharkya, P., Maranas, CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*. 2003, 20; 84(6):647–57. <https://doi.org/10.1002/bit.10803> PMID: 14595777
8. Pharkya P., Maranas CD. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab Eng*. 2006, 8(1):1–13. <https://doi.org/10.1016/j.ymben.2005.08.003> PMID: 16199194
9. Ranganathan S., Suthers PF., Maranas CD. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol*. 2010, 15; 6(4): e1000744. <https://doi.org/10.1371/journal.pcbi.1000744> PMID: 20419153

10. von Kamp A., Klamt S. Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nat Commun*. 2017, 22; 8:15956. <https://doi.org/10.1038/ncomms15956> PMID: 28639622
11. Razaghi-Moghadam Z., Nikoloski Z. GeneReg: a constraint-based approach for design of feasible metabolic engineering strategies at the gene level, *Bioinformatics*, 2021, 37(12), 1717–1723, <https://doi.org/10.1093/bioinformatics/btaa996> PMID: 33245091
12. Küken A., Nikoloski Z. Computational Approaches to Design and Test Plant Synthetic Metabolic Pathways, *Plant Physiology*, 2109, 179(3), Pages 894–906, <https://doi.org/10.1104/pp.18.01273> PMID: 30647083
13. Lewis NE., Nagarajan H., Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol*. 2012, 27; 10(4):291–305. <https://doi.org/10.1038/nrmicro2737> PMID: 22367118
14. Goldsmith M., Tawfik DS. Enzyme engineering: reaching the maximal catalytic efficiency peak. *Curr Opin Struct Biol*. 2017, 47:140–150. <https://doi.org/10.1016/j.sbi.2017.09.002> PMID: 29035814
15. Sánchez BJ, Zhang, C., Nilsson, A., Lahtvee, P., Kerkhoven, EJ., Nielsen, J. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol*. 2017, 13, 935.
16. Bekiaris PS., Klamt S. Automatic construction of metabolic models with enzyme constraints. *BMC Bioinformatics*. 2020, 14; 21(1):19. <https://doi.org/10.1186/s12859-019-3329-9> PMID: 31937255
17. Jeske L., Placzek S., Schomburg I., Chang A. & Schomburg D. BRENDA in 2019: a European ELEXIR core data resource. *Nucleic Acids Res*. 2019, 47, D542–D549.
18. Wittig U., Rey M., Weidemann A., Kania R., Müller W. SABIO-RK: an updated resource for manually curated biochemical reaction kinetics. *Nucleic Acids Research*, 2018, 46, D656–D660, <https://doi.org/10.1093/nar/gkx1065> PMID: 29092055
19. Li F., Yuan L., Lu H., Li G., Chen Y., Engqvist MKM. et al. Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nat Catal* 2022, 5, 662–672 <https://doi.org/10.1038/s41929-022-00798-z>.
20. Kroll A., Rousset Y., Hu XP., Liebrand NA., Lercher MJ. Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning. *Nat Commun* 2023, 14, 4139, <https://doi.org/10.1038/s41467-023-39840-4> PMID: 37438349
21. Yu H., Deng H., He J., Keasling J., Luo X. Highly accurate enzyme turnover number prediction and enzyme engineering with PreKcat. *Research Square*. 2023, <https://doi.org/10.21203/rs.3.rs-2749688/v1>
22. Kroll A., Lercher MJ. Machine learning models for the prediction of enzyme properties should be tested on proteins not used for model training bioRxiv 2023.02.06.526991; <https://doi.org/10.1101/2023.02.06.52699>
23. Davidi D., Noor E., Liebermeister W., Bar-Even A., Flamholz A., Tumbler K., et al. Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro kcat measurements. *Proc. Natl. Acad. Sci*. 2016, 113, 3401–3406. <https://doi.org/10.1073/pnas.1514240113> PMID: 26951675
24. Heckmann D., Campeau A., Lloyd CJ., Phaneuf PV., Hefner Y., Carrillo-Terrazas M, et al. Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers. *Proceedings of the National Academy of Sciences*, 2020, 117(37), 23182–23190. <https://doi.org/10.1073/pnas.2001562117> PMID: 32873645
25. Xu R., Razaghi-Moghadam Z., Nikoloski Z. Maximization of non-idle enzymes improves the coverage of the estimated maximal in vivo enzyme catalytic rates in *Escherichia coli*. *Bioinformatics*, 2021, 37 (21), 3848–3855 <https://doi.org/10.1093/bioinformatics/btab575> PMID: 34358300
26. Küken A., Gennermann K. and Nikoloski Z. Characterization of maximal enzyme catalytic rates in central metabolism of *Arabidopsis thaliana*. *The Plant Journal*, 2020, 103(6), 2168–2177. <https://doi.org/10.1111/tbj.14890> PMID: 32656814
27. Hackett SR., Zanutelli VRT, Xu W., Goya J., Park JO., Perlman DH, et al. Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science*, 2016, 354, 631. <https://doi.org/10.1126/science.aaf2786> PMID: 27789812
28. Arend M., Zimmer D., Xu R., Sommer F., Mühlhaus T., Nikoloski Z. Proteomics and constraint-based modelling reveal enzyme kinetic properties of *Chlamydomonas reinhardtii* on a genome scale. *Nat Commun* 2023, 14, 4781 <https://doi.org/10.1038/s41467-023-40498-1> PMID: 37553325
29. Ferreira MAM., Silveira WBD., Nikoloski Z. Protein constraints in genome-scale metabolic models: Data integration, parameter estimation, and prediction of metabolic phenotypes. *Biotechnol Bioeng*. 2024, <https://doi.org/10.1002/bit.28650> PMID: 38178617

30. Wendering P., Arend M., Razaghi-Moghadam Z., Nikoloski Z. Data integration across conditions improves turnover number estimates and metabolic predictions. *Nat Commun* 2023, 14, 1485 <https://doi.org/10.1038/s41467-023-37151-2> PMID: 36932067
31. Domenzain I., Sánchez B., Anton M., Kerkhoven E.J., Millán-Oropeza A, Henry C., et al. Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *Nat. Commun.* 2022, 13, 3766. <https://doi.org/10.1038/s41467-022-31421-1> PMID: 35773252
32. Ye C., Luo Q., Guo L., Gao C., Xu N., Zhang L., et al. Improving lysine production through construction of an Escherichia coli enzyme-constrained model. *Biotechnol Bioeng.* 2020, 117(11):3533–3544. <https://doi.org/10.1002/bit.27485> PMID: 32648933
33. Yeo HC., Hong J., Lakshmanan M., Lee DY. Enzyme capacity-based genome scale modelling of CHO cells. *Metab Eng.* 2020, 60:138–147. <https://doi.org/10.1016/j.ymben.2020.04.005> PMID: 32330653
34. Robinson JL., Kocabaş P., Wang H., Cholley PE, Cook D., Nilsson A., et al. An atlas of human metabolism. *Sci. Signal.* 2020, 13 <https://doi.org/10.1126/scisignal.aaz1482> PMID: 32209698
35. Massaiu I., Pasotti L., Sonnenschein N., Rama E., Cavaletti M., Magni P., et al. Integration of enzymatic data in Bacillus subtilis genome-scale metabolic model improves phenotype predictions and enables in silico design of poly-γ-glutamic acid production strains. *Microb Cell Fact.* 2019, 18, 3 <https://doi.org/10.1186/s12934-018-1052-2>
36. Niu J., Mao Z., Mao Y., Wu K., Shi Z., Yuan Q., et al. Construction and Analysis of an Enzyme-Constrained Metabolic Model of Corynebacterium glutamicum. *Biomolecules*, 2022, 12, 1499 <https://doi.org/10.3390/biom12101499> PMID: 36291707
37. Wendering P., Nikoloski Z. Genome-Scale Modeling Specifies the Metabolic Capabilities of Rhizophagus irregularis. *mSystems.* 2022, 22; 7(1):e0121621. <https://doi.org/10.1128/msystems.01216-21> PMID: 35076269
38. Jumper J., Evans R., Pritzel A., Green T., Figurnov M., Ronneberger O., et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2> PMID: 34265844
39. Griesbeck O. CRISPR/Cas9-based directed evolution in mammalian cells, *Current Opinion in Structural Biology*, 2021, 69, 35–40. <https://doi.org/10.1016/j.sbi.2021.02.005> PMID: 33784525
40. Orth J., Thiele I. & Palsson B. What is flux balance analysis? *Nat Biotechnol* 2010, 28, 245–248 <https://doi.org/10.1038/nbt.1614> PMID: 20212490
41. Segre D., Vitkup D., Church GM. Analysis of optimality in natural and perturbed metabolic networks, *PNAS* 2002, 99(23) 15112–15117 <https://doi.org/10.1073/pnas.232349399> PMID: 12415116
42. Hashemi S., Laitinen R., Nikoloski Z. Models and molecular mechanisms for trade-offs in the context of metabolism. *Molecular Ecology*, 2022, 1–11 <https://doi.org/10.1111/mec.16879>
43. Domenzain I., Lu Y., Shi J., Lu H., Nielsen J. Computational biology predicts metabolic engineering targets for increased production of 102 valuable chemicals in yeast. 2023, <https://doi.org/10.1101/2023.01.31.526512>
44. Zhao H., Lee S.Y., Nielsen J., Stephanopoulos G. Protein engineering, Tools and applications, *Wiley-VCH*, 2021, Vol. 10.
45. Bhagavat R., Sankar S., Srinivasan N., Chandra N. An Augmented Pocketome: Detection and Analysis of Small-Molecule Binding Pockets in Proteins of Known 3D Structure, *Structure* 2018, 26(3):499-512. e2. <https://doi.org/10.1016/j.str.2018.02.001> PMID: 29514079.
46. Salvy P., Hatzimanikatis V. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nat Commun* 2020, 11, 30 <https://doi.org/10.1038/s41467-019-13818-7> PMID: 31937763
47. McElreath R. Statistical rethinking (1st ed.). Safari an O'Reilly Media Company Chapman and Hall/ CRC 2018, <https://www.safaribooksonline.com/library/view/9781482253481/?ar>.
48. Gurobi Optimization L. Gurobi Optimizer Reference Manual 2021, <https://www.gurobi.com>.
49. The UniProt Consortium, UniProt: the Universal Protein Knowledgebase in 2023, *Nucleic Acids Research*, 2023, 51, D1 D523–D531. <https://doi.org/10.1093/nar/gkac1052> PMID: 36408920