

EnzymeTuning Improves Enzyme-constrained Metabolic Modeling and Proteome Abundance Prediction Through Deep Learning

Corresponding Author: Professor Guan Wang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This manuscript presents a framework for enzyme-level kinetic parameter optimization using a conditional GAN (cGAN). While the study addresses an important problem in computational systems biology—namely, improving kinetic parameter estimation and model predictive power—it suffers from several major issues in methodological transparency, result interpretation, and biological validation. The paper's claims about significant model improvements and biological insights are not fully supported by the presented evidence. Substantial revision is necessary before the work can be considered for publication.

Major Comments

1. Insufficient Transparency in GAN Training Details

The manuscript lacks sufficient information regarding the cGAN training configuration, which raises concerns about reproducibility. Specifically:

- The choice of generator/discriminator activation functions, batch size, learning rate, and learning rate decay strategy are not reported. These parameters critically influence GAN stability and convergence.
- The criteria for selecting the “high-performance kcat set” are vague. The text only mentions datasets that “accurately predict growth and protein abundance,” without quantitative thresholds (e.g., RMSE < X, PCC > Y). This prevents independent reproduction of the training dataset.
- The claim that the GAN “converged after 50 epochs” lacks evidence. Please provide loss curves showing the generator and discriminator losses over iterations to substantiate this statement.

2. Incorrect Description of Figure S1C

The text states that the model successfully reproduced the Crabtree effect (ethanol production under high glucose), but Figure S1C shows no ethanol generation. This contradiction must be corrected.

3. Overstated Claims Regarding Pearson Correlation Coefficients (PCCs)

The PCC of predicted versus measured protein abundance is the central metric supporting the study's conclusions, yet the reported correlations are weak.

- Across Figures 3(g), 4(c), and 5(c), the PCC values are consistently around 0.3–0.4, which indicates only weak correlation.
- Some figures appear to exaggerate small differences by truncating the y-axis (e.g., Figure 5c, third panel, where PCCs of 0.16 vs. 0.23 are visually shown as distinct).
- Statements such as “PCC improved significantly” (page 8, line 166) are misleading; the mean PCC increases only from ~0.37 to 0.38, which is negligible despite a small p-value. All descriptions involving PCC improvement should be carefully revised to reflect the actual effect sizes.

4. Lack of Biological Basis for Sensitivity Analysis Ranges

The rationale for choosing kcat perturbation ranges (10–30%, 50–70%, etc.) is unclear. Please justify the biological relevance of these intervals. Moreover, Figure 2(c) lacks the curve corresponding to the 70% kcat perturbation, which should be included for completeness.

5. Ambiguous Definition of “Extreme Enzymes”

The definition of “extreme enzymes” is inconsistent.

- Page 6 (line 120) defines them as predicted/measured abundance ratios >100 or <0.01.
- Page 8 (line 158) redefines them as >50 or <0.02.
- Figure S2(f–h) shows that performance improves as the threshold increases, suggesting that >150 or >200 might yield

better performance. Please clarify and justify the final definition used throughout the paper.

6. Language and Typographical Errors

There are several typographical issues that require careful proofreading, e.g.,

- “Figure 2h. Comparison of the number of extreme enzymes using using different fitness function.” → remove redundant “using.”

7. Errors and Ambiguities in Figure 3

- The captions of Figures 3a and 3b are mislabeled. Figure 3b corresponds to PCC analysis.
- Additional explanation is needed for Figure 3a—its meaning and data representation are unclear.
- Contrary to the text claim (“PCC steadily increased”), PCC does not increase beyond epoch 5 in Figure 3b.

8. Unjustified Selection of Top 100 vsyn Values

The selection of “top 100 vsyn values” as the optimal dataset is not adequately supported.

- Figure S6 suggests a local turning point between 50–100, but not necessarily that 100 is optimal.
- Moreover, kdeg values are aggregated from different studies without addressing methodological heterogeneity (e.g., differences in degradation measurement techniques), which could bias vsyn estimation.

9. GAN Convergence Issue (Figure 4b)

The discriminator loss at epoch 500 still shows strong fluctuations and does not stabilize near a fixed value, indicating that the GAN has not fully converged. Therefore, conclusions drawn from this model state require caution.

10. Insufficient Figure Legends and Clarity in Figure 5

- Figure 5h lacks an informative title and explanation of what is being displayed.
- The meaning of colors, clustering method, and the variable represented by “Protein” on the y-axis should be clearly stated.

11. Lack of Clarity in Figure 6e

- The y-axis label and the fitting method used to generate the curve are not described. Additional details are needed to interpret this figure.

12. Lack of Experimental Validation

The manuscript proposes several enzyme usage mechanisms dependent on C/N ratio but provides no experimental validation. At least one such prediction should be experimentally tested to substantiate the model’s biological relevance.

13. Limited Demonstration of Practical Utility

The current study stops at “model performance improvement” without showing translational potential for metabolic engineering.

- A case study is recommended: use EnzymeTuning-optimized ecGEMs to predict the production yield of a high-value metabolite, comparing pre- and post-optimization results.
- Cross-species validation (e.g., *E. coli* ecGEMs) would further demonstrate generalizability and strengthen the manuscript’s impact.

14. The introduction section of the manuscript provides an incomplete review of AI-based kinetic parameter prediction methods. Several relevant tools have already been published, such as UniKP, CatPred, etc.

15. Unavailable Code Repository

The provided repository (<https://github.com/hongzhonglu/EnzymeTuning>) is inaccessible, making it impossible to evaluate or reproduce the computational workflow. Full access to the code and datasets must be ensured.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

The provided repository (<https://github.com/hongzhonglu/EnzymeTuning>) is inaccessible.

Reviewer #3

(Remarks to the Author)

The authors present EnzymeTuning, a conditional GAN-based framework for global optimisation of kcat parameters in enzyme-constrained genome-scale metabolic models (ecGEMs). The approach integrates phenotype constraints, proteomics matching, and protein synthesis/degradation considerations (vsyn) to improve predictive accuracy across yeast species. The results show improved Pearson correlation coefficients for proteome abundance prediction, reduced numbers of extreme enzyme predictions, improved cross-species generalizability, and new insights into metabolic adjustments under varying C/N ratios.

The manuscript is timely and addresses a critical limitation - inaccurate, sparse, and non-physiological kinetic parameter sets. Collecting in vivo and in vitro enzyme turnover, but also enzyme degradation rate data, is time-consuming and often unfeasible for many interested microbes. Developing computational approaches that provide more accurate and scalable parameter estimation is essential for the field. The presented framework is technically sound and potentially impactful.

However, a few issues must be addressed to improve rigour, clarity, and reproducibility.

Major comments:

1. The introduction provides good context on current deep-learning approaches (DLKcat, TurNup) but the novelty claim of EnzymeTuning relative to Bayesian methods and earlier DL approaches is somewhat unclear, how the presented model

differs in terms of runtime, robustness, or biological interpretability of the results?

2. [L487-490] It is not clear from which source the protein degradation rates are adapted, as a review article is referenced. The manuscript claims to incorporate literature-derived kdeg values to estimate vsyn; however, it does not specify the number of proteins with experimentally measured kdeg, how the missing values were handled, or what the distribution of assumed vs measured degradation rates looked like.

3. Although RMSE and PCC are used extensively (e.g., Figure 3, S5), the manuscript does not clearly specify the number of enzymes considered "predictable," nor whether PCC was computed on log-transformed or raw abundance values.

4. It is not clear whether any validation experiments were performed on a dataset which was not included in the initial training. Such validation would be important to validate the predictability of the proposed model.

Minor comments:

5. Figures have often very small and unreadable axis names (particularly Figures 3, 5 and 6).

6. "fundings" → "findings" (Figure 4 discussion)

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have made extensive revision on the manuscript to address the reviewer's comments and the quality of the manuscript is improved. kcat prediction from enzymeTurning might be better than previous methods but still far from "high-performance". As stated in the answers to reviewer 1: "a kcat parameter set is classified as high-performance if it satisfies $RMSE < 2$ for growth phenotype prediction and an overall Pearson correlation coefficient (PCC) > 0.4 for protein abundance prediction." A PCC value of 0.4 is still very low, especially considering that it is computed using log-transformed values of both predicted and measured protein abundances. I think RMSE was calculated based on eq.5 (line 605). How many exchange rates were used in the calculation (value of n)? Why value of 2 was considered as acceptable? The authors may need to give more details on the calculation, for example providing the experimental exchange rates and predicted rates using different sets of kcat values.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

I can recommend the manuscript for publication.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I do not have any further comments and can recommend the manuscript for publication.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This manuscript presents a framework for enzyme-level kinetic parameter optimization using a conditional GAN (cGAN). While the study addresses an important problem in computational systems biology—namely, improving kinetic parameter estimation and model predictive power—it suffers from several major issues in methodological transparency, result interpretation, and biological validation. The paper's claims about significant model improvements and biological insights are not fully supported by the presented evidence. Substantial revision is necessary before the work can be considered for publication.

Major Comments

1. Insufficient Transparency in GAN Training Details

The manuscript lacks sufficient information regarding the cGAN training configuration, which raises concerns about reproducibility. Specifically:

- The choice of generator/discriminator activation functions, batch size, learning rate, and learning rate decay strategy are not reported. These parameters critically influence GAN stability and convergence.
- The criteria for selecting the “high-performance k_{cat} set” are vague. The text only mentions datasets that “accurately predict growth and protein abundance,” without quantitative thresholds (e.g., $RMSE < X$, $PCC > Y$). This prevents independent reproduction of the training dataset.
- The claim that the GAN “converged after 50 epochs” lacks evidence. Please provide loss curves showing the generator and discriminator losses over iterations to substantiate this statement.

Response: Thank you for this important comment. We fully agree that these details are critical and have revised the manuscript accordingly. Our responses are provided point by point below.

(1) cGAN training configuration

We have expanded the Methods section to explicitly report the key hyperparameters used for cGAN training (lines 633-643):

“The cGAN consisted of two fully connected neural networks: a generator and a discriminator, which conditioned on the labels during the training. The discriminator consisted of three dense layers of increasing complexity: the first layer contained 32 units, 64 units, and 128 units, each followed by a Dropout layer (rate = 0.5). The generator network comprised three dense layers with progressive depth and regularization: 128 units, 256 units, and 512 units, each followed by Batch Normalization and Dropout (rate = 0.5). All models were implemented using Keras with a TensorFlow-GPU backend. Both the generator and discriminator were trained using binary cross-entropy as the loss function and optimized with the Adam optimizer using a fixed learning rate of 2×10^{-4} and a batch size of 50.”

(2) Quantitative definition of “high-performance k_{cat} sets”

We agree that the original description of “high-performance k_{cat} sets” lacked quantitative rigor. In the revised manuscript, we now explicitly define the selection criteria. For optimization scenarios that consider both phenotype and proteome constraints, a k_{cat} parameter set is classified as high-performance if it satisfies $RMSE < 2$ for growth phenotype prediction and an overall Pearson correlation coefficient (PCC) > 0.4 for protein abundance prediction. We have expanded the criteria in the Results section (lines 118-122):

“In **Step 1**, high-performance parameter sets were identified based on their ability to predict growth phenotypes (the root mean squared error, $RMSE < 2$) and protein abundances (Pearson correlation coefficient, $PCC > 0.4$), forming a labeled training dataset for a conditional generative adversarial network (cGAN).”

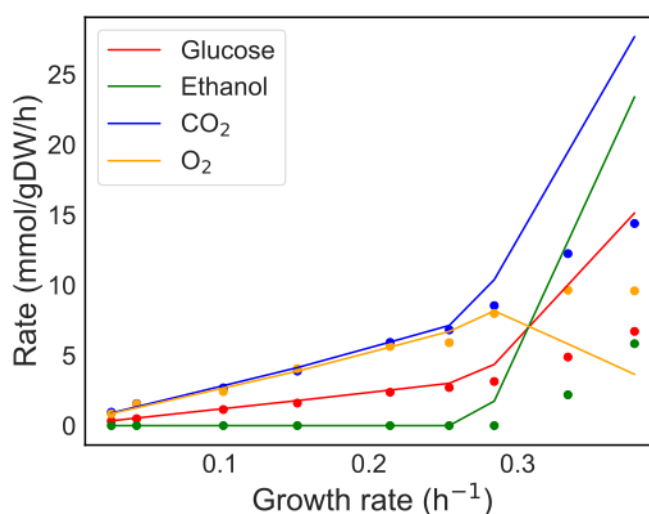
(3) Clarification of “convergence after 50 epochs”

We appreciate you for pointing out the potential ambiguity of this statement. In the original manuscript, “50 epochs” did not refer to the number of epochs in a single neural network training process. Instead, it denoted the number of outer-loop iterations of the EnzymeTuning workflow, in which new k_{cat} sets are generated by the cGAN, evaluated using the ecGEM, and then fed back into subsequent training rounds. To avoid confusion, we have adopted the term “50 training generations”.

2. Incorrect Description of Figure S1C

The text states that the model successfully reproduced the Crabtree effect (ethanol production under high glucose), but Figure S1C shows no ethanol generation. This contradiction must be corrected.

Response: We thank you for pointing out this inconsistency. We confirm that this issue was caused by an incorrect version of **Fig. S1C** being inadvertently included in the Supplementary Information, rather than a problem with the model predictions themselves. In our analysis, the Bayesian-parameterized ecGEM correctly predicted ethanol production, thereby reproducing the Crabtree effect. In the revised manuscript, **Fig. S1C** has been replaced with the correct version. The corresponding text and figure legend have also been revised to ensure full consistency between the description and the displayed results.



Supplemental Figure 1 Baseline performance of ecGEM prior to k_{cat} optimization. **c** Predicted exchange rates for the Crabtree effect based on the Bayesian-posterior-parameterized k_{cat} values (lines) compared with experimental data (dots).

3. Overstated Claims Regarding Pearson Correlation Coefficients (PCCs)

The PCC of predicted versus measured protein abundance is the central metric supporting the study's conclusions, yet the reported correlations are weak.

- Across Figures 3(g), 4(c), and 5(c), the PCC values are consistently around 0.3–0.4, which indicates only weak correlation.
- Some figures appear to exaggerate small differences by truncating the y-axis (e.g., Figure 5c, third panel, where PCCs of 0.16 vs. 0.23 are visually shown as distinct).
- Statements such as “PCC improved significantly” (page 8, line 166) are misleading; the mean PCC increases only from ~0.37 to 0.38, which is negligible despite a small p-value. All descriptions involving PCC improvement should be carefully revised to reflect the actual effect sizes.

Response: Thank you so much for this important and constructive comment. We agree that, when considered in isolation, the Pearson correlation coefficients (PCCs) between predicted and measured protein abundances remain modest (typically in the range of 0.3–0.4). Given the inherent limitations of ecGEMs, achieving a PCC value of ~0.4 or higher can be established as a standard indicating relatively good predictive performance for proteome-level abundance.

Accordingly, we have revised the manuscript to quantify these improvements precisely. The descriptions such as “significantly improved” have been replaced with specific numerical increases (e.g., PCC increased from 0.33 to 0.37) and their corresponding statistical significance (P value < 0.05), like the description in lines 360-363:

“Correspondingly, prediction accuracy improved in all three yeasts, with PCC values increasing from ~0.33 to ~0.37 in *K. marxianus* (P value < 0.05), from ~0.21 to ~0.29 in *K. lactis* (P value < 0.05), reflecting enhanced proteome-level performance (**Fig. 5c**).”

Importantly, the conclusions of this study are not based on PCC alone. Instead, the performance of EnzymeTuning was evaluated using multiple complementary criteria, including

- (i) the reduction in the number of extreme enzymes with highly unrealistic predicted-to-measured protein abundance ratios,
- (ii) the increase in the number of enzymes for which protein abundances could be reliably predicted, and
- (iii) improved agreement with global protein abundance distributions, reflecting enhanced overall quantitative consistency rather than rank correlation alone.

Taken together, these metrics indicate a meaningful improvement in model behavior relative to the baseline ecGEM, even when PCC gains are moderate.

We further note that the absolute values of protein abundance prediction are inherently constrained by the underlying ecGEMs and available experimental data. For several organisms, such as *K. lactis*, the baseline predictive accuracy of the original ecGEM is low, which limits the maximum achievable improvement in absolute PCC values regardless of

the optimization strategy. We have now explicitly acknowledged these limitations in the revised results (lines 366-370):

“Nevertheless, the improvements observed in other yeast species were less pronounced than those in *S. cerevisiae*. These results suggested that the efficacy of k_{cat} optimization was critically modulated by the quality of the GEMs and the availability of experimental constraints for each species, which remained relatively limited for less-studied organisms.”

4. Lack of Biological Basis for Sensitivity Analysis Ranges

The rationale for choosing k_{cat} perturbation ranges (10–30%, 50–70%, etc.) is unclear. Please justify the biological relevance of these intervals. Moreover, Figure 2(c) lacks the curve corresponding to the 70% k_{cat} perturbation, which should be included for completeness.

Response: Thank you for this constructive comment regarding the choice of k_{cat} perturbation ranges and the completeness of Figure 2c.

We would like to clarify that the purpose of this analysis was to perform a sensitivity analysis to evaluate how perturbations in k_{cat} values affect growth phenotypes, rather than to assume that these perturbation ranges directly represent the true biological variability of k_{cat} .

The original perturbation intervals (e.g., 10–30%, 50–70%) were chosen to systematically probe model sensitivity across moderate and large relative changes in k_{cat} , following grouping strategies commonly adopted in previous enzyme-constrained modeling studies¹. These ranges were intended to capture progressive levels of deviation from the baseline k_{cat} values and to assess the nonlinear response of growth predictions.

We further acknowledge that experimental k_{cat} measurements can vary by orders of magnitude across studies and conditions. To address this point and strengthen the biological relevance of the analysis, we have extended the sensitivity analysis in the revised manuscript to include two additional perturbation ranges representing extreme under- and over-estimation scenarios (1–3% and 990–1100% of the original k_{cat} values). Notably, the qualitative trends observed under these extended ranges are consistent with those obtained using the original perturbation groups, supporting the robustness of our conclusions.

Finally, we apologize for the confusion regarding the “missing” curve for the 70% perturbation group. This was a labeling error in the figure legend; the data was present in the original plot but was inadvertently mislabeled. We have corrected the legend in the revised Figure 2c to clearly identify the curve corresponding to the 70% perturbation group.

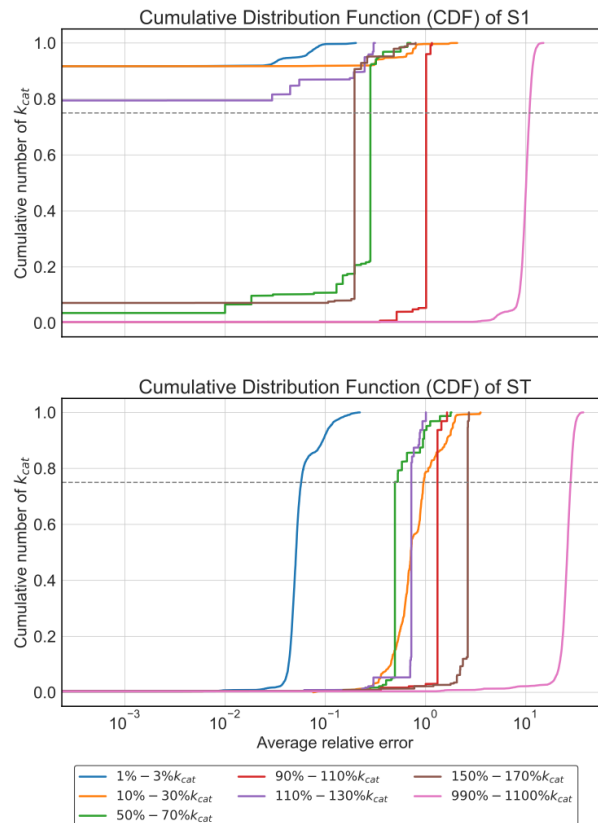


Fig. 2 Sensitivity analysis of the k_{cat} optimization process and the comparison of the influence of the fitness function. **c** Cumulative Distribution Function (CDF) curves for the sensitivity analysis results.

5. Ambiguous Definition of “Extreme Enzymes”

The definition of “extreme enzymes” is inconsistent.

- Page 6 (line 120) defines them as predicted/measured abundance ratios >100 or <0.01 .
- Page 8 (line 158) redefines them as >50 or <0.02 .
- Figure S2(f–h) shows that performance improves as the threshold increases, suggesting that >150 or >200 might yield better performance. Please clarify and justify the final definition used throughout the paper.

Response: We thank the reviewer for carefully pointing out the inconsistency in the definition of “extreme enzymes.” In the revised manuscript, we have clarified that the definition of extreme enzymes is fixed throughout the study as enzymes whose predicted-to-measured protein abundance ratios > 100 or < 0.01 . The alternative thresholds (e.g., >50 or <0.02) were introduced to evaluate how different scales of k_{cat} correction affect model performance. These analyses were not intended to redefine extreme enzymes but rather to assess the robustness of the observed trends across different threshold choices. We have now explicitly stated this distinction in the Results (lines 139-143; 167-170).

“Relative to the proteomics dataset, the updated ecGEM exhibited a high proportion of extreme enzyme abundance predictions- defined as predicted/measured abundance ratios greater than 100 or less than 0.01- which accounted for approximately one-fourth to one-third of all predictable enzymes.”

“Here, we focused on reducing the extreme enzyme values (**Eq. 1**) by progressively optimizing enzyme subsets defined at five levels of fold changes between predicted and measured abundances: >100 or <0.01; >50 or <0.02; >10 or <0.1; >5 or <0.2; and >2 or <0.5.”

6. Language and Typographical Errors

There are several typographical issues that require careful proofreading, e.g.,

- “Figure 2h. Comparison of the number of extreme enzymes using using different fitness function.” → remove redundant “using.”

Response: Thank you for pointing out this. we have already fixed it.

7. Errors and Ambiguities in Figure 3

- The captions of Figures 3a and 3b are mislabeled. Figure 3b corresponds to PCC analysis.
- Additional explanation is needed for Figure 3a—its meaning and data representation are unclear.
- Contrary to the text claim (“PCC steadily increased”), PCC does not increase beyond epoch 5 in Figure 3b.

Response: Thank you so much for carefully checking **Fig. 3** and for pointing out these important issues. We have addressed each point in the revised manuscript.

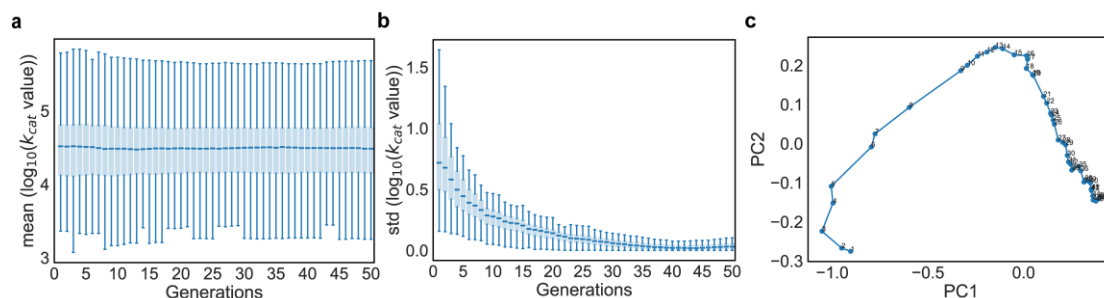
First, we acknowledge that the captions of **Fig. 3a** and **Fig. 3b** were incorrectly labeled in the original version. This labeling error has been corrected so that **Fig. 3a** and **Fig. 3b** now accurately correspond to their respective analyses (lines 902-905).

“**a** Changes in the criteria values calculated via **Eq. 1** during GAN training. **b** Changes in Pearson correlation coefficients for protein abundance predictions in the training process.”

Second, we have added **Eq. 1** to the revised manuscript. This equation explicitly defines the mathematical formula used to calculate the data plotted in the **Fig. 3a**, ensuring its meaning and representation are now clearly presented. Specifically, **Fig. 3a** reports a fold-change–based deviation metric between predicted and measured protein abundances, defined as the square root of the mean squared difference between log-transformed predicted and measured values across all considered enzymes.

$$fold\ change = \sqrt{\frac{1}{n} \sum_{i=0}^n \frac{\log predicted\ abundance_i}{\log measured\ abundance_i}} \quad (1)$$

Third, we agree with the reviewer that the Pearson correlation coefficient (PCC) does not increase substantially beyond approximately generation 5. We have therefore revised the text to avoid overstatement and to accurately reflect the observed trend. While PCC reaches a plateau at later generations, we observed that the generated k_{cat} distributions become increasingly stable during continued training. To support this point, we have added additional analyses in **Fig. S4**, which show that the mean and variance of the generated k_{cat} values converge and that the distributions exhibit increased robustness at later generations (lines 230-237).



Supplemental Fig. 4 Convergence and stability of EnzymeTuning-optimized k_{cat} parameters across training generations. **a** Distribution of the mean log-transformed k_{cat} values across all k_{cat} variables at each training generation. **b** Distribution of the dispersion (standard deviations, std) of k_{cat} parameters across training generations. **c** Principal component analysis (PCA) of iteration-wise mean k_{cat} vectors.

“While the PCC remained largely stable in later generations, analysis of the generated k_{cat} distributions showed that the mean values gradually converge and the associated standard deviations decreased as training proceeded (**Fig. S4a, b**). Moreover, the PCA trajectory demonstrated that iteration-wise k_{cat} mean vectors collapsed into a compact region in parameter space during later generations (**Fig. S4c**). These results indicated that EnzymeTuning generated robust and stable k_{cat} distributions, and thus effectively improved the accuracy of protein abundance predictions.”

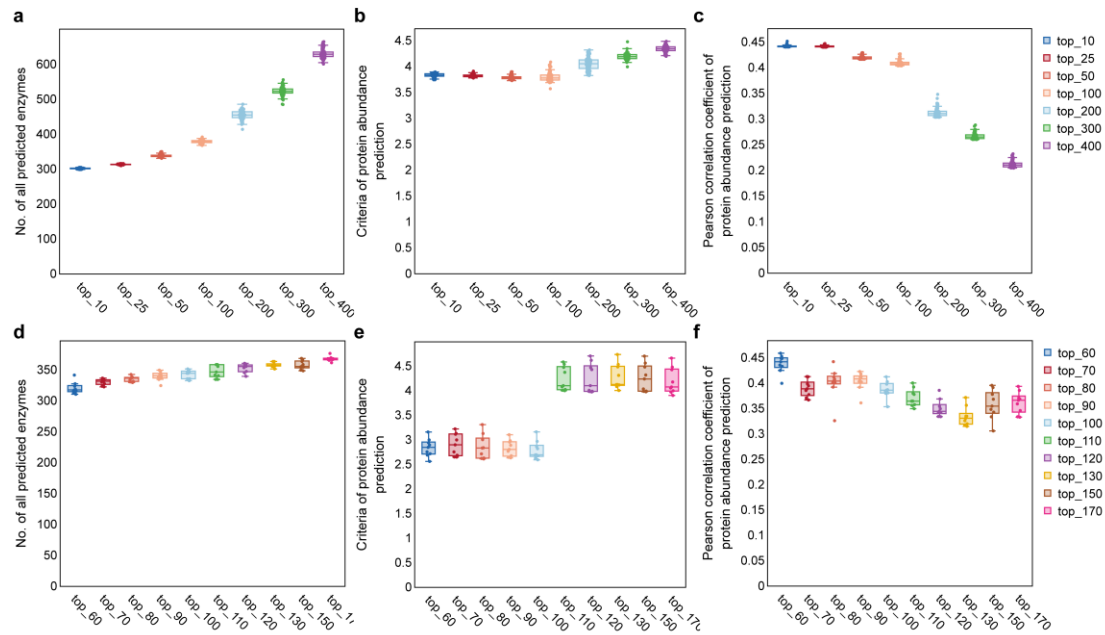
8. Unjustified Selection of Top 100 vsyn Values

The selection of “top 100 vsyn values” as the optimal dataset is not adequately supported.

- Figure S6 suggests a local turning point between 50–100, but not necessarily that 100 is optimal.
- Moreover, kdeg values are aggregated from different studies without addressing methodological heterogeneity (e.g., differences in degradation measurement techniques), which could bias vsyn estimation.

Response: Thank you for raising this important point. We have addressed both concerns in the revised manuscript.

We fully agree that **Fig. S6** alone indicates a local transition region between approximately 50 and 100 v_{syn} values, rather than uniquely identifying 100 as the optimal cutoff. To systematically assess this region, we extended the analysis by explicitly evaluating intermediate subset sizes (top 60, 70, 80, 90, 110, 120, 130, 150, and 170 v_{syn} values). The results of this expanded analysis are now presented in **Fig. S6**.



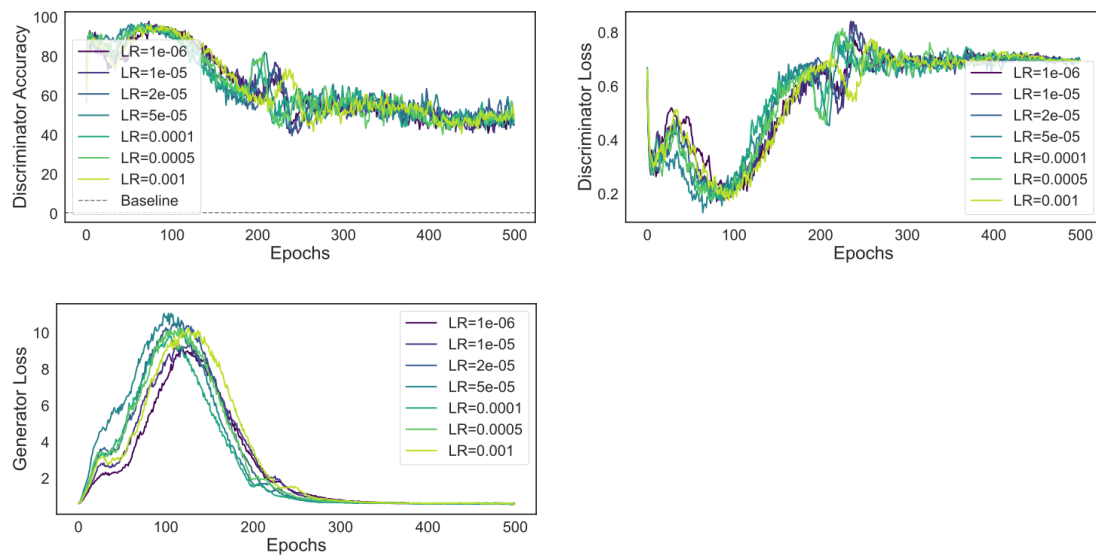
Supplemental Fig. 6 Comparison of EnzymeTuning in predicting protein synthesis rates at different scales. **a** Comparison of the total number of predictable enzymes cross a wide range of protein synthesis rate scaling sizes with the GAN method. **b** Comparison of the criteria values calculated via Eq. 1 across a wide range of protein synthesis rate scaling sizes with the GAN method. **c** Comparison of Pearson correlation coefficients for protein abundance predictions cross a wide range of protein synthesis rate scaling sizes with the GAN method. **d** Comparison of the total number of predictable enzymes cross an intermediate range of protein synthesis rate scaling sizes with the GAN method. **e** Comparison of the criteria values calculated via Eq. 1 across an intermediate range of protein synthesis rate scaling sizes with the GAN method. **f** Comparison of Pearson correlation coefficients for protein abundance predictions cross an intermediate range of protein synthesis rate scaling sizes with the GAN method. Second, regarding the aggregation of k_{deg} values from different studies ^{2,3}, we note that the k_{deg} data used in this work were derived from two primary experimental studies that employed the same pulse-chase-based proteome turnover measurement strategy. Although the two studies differ slightly in how k_{deg} was calculated—specifically, whether growth-rate-dependent dilution effects were explicitly subtracted—the underlying experimental methodology is consistent. We have clarified this methodological difference in the revised Methods section (lines 662-669).

“The initial v_{syn} values were derived from literature-reported protein degradation constants (k_{deg}), which were collected from two independent large-scale proteome turnover studies employing distinct computational strategies ^{2,3}. These two studies obtained k_{deg} constants by fitting the temporal profiles of labeling incorporation (relative isotope abundance, RIA) to an exponential equation. However, one study defined the k_{deg} constants on protein turnover rates inferred from dynamic labeling experiments, whereas the other estimated the k_{deg} constants by explicitly subtracting growth-rate-dependent dilution effects.”

9. GAN Convergence Issue (Figure 4b)

The discriminator loss at epoch 500 still shows strong fluctuations and does not stabilize near a fixed value, indicating that the GAN has not fully converged. Therefore, conclusions drawn from this model state require caution.

Response: Thank you so much for highlighting the convergence issue in the original **Fig. 4b**. We agree that the observed fluctuations in discriminator loss at later epochs indicated suboptimal stability, necessitating a re-evaluation of the training hyperparameters. To address this, we performed a sensitivity analysis on the learning rate to optimize training stability. We compared the convergence trajectories across different learning rates and found that by reducing the learning rate to 1×10^{-4} , we achieved significantly smoother convergence. We have included a comparison of the convergence curves under different learning rates in the following figure to demonstrate this improvement.



Consequently, we have updated **Fig. 4b** in the revised manuscript to reflect the results from the optimized model 1×10^{-4} . We confirm that all subsequent analyses and conclusions in the manuscript have been updated based on this fully converged model to ensure the reliability of our findings.

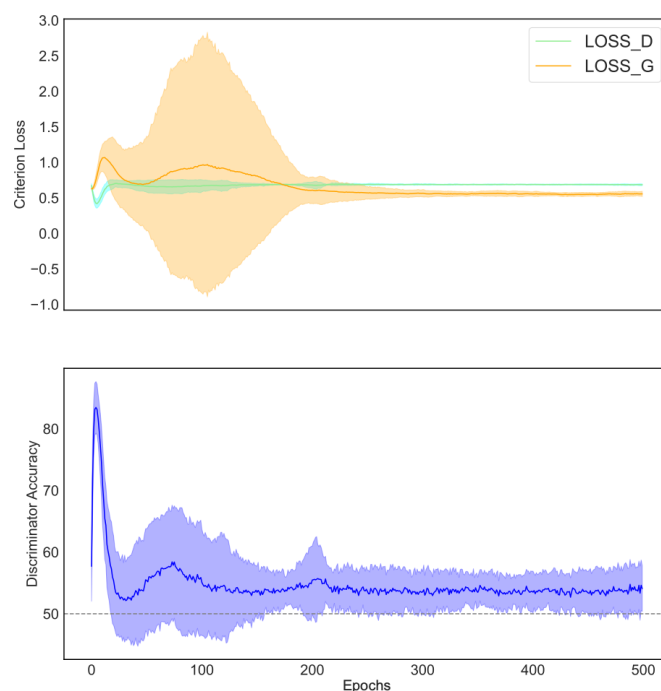


Fig. 4 Performance of the GAN approach for protein synthesis rate prediction. **b** Discriminator loss (green), generator loss (orange), and discriminator accuracy (blue) over training epochs.

10. Insufficient Figure Legends and Clarity in Figure 5

- Figure 5h lacks an informative title and explanation of what is being displayed.
- The meaning of colors, clustering method, and the variable represented by “Protein” on the y-axis should be clearly stated.

Response: Thank you for pointing out the lack of clarity in **Fig. 5h**. We agree that the previous labels and figure legend were insufficient to fully interpret the heatmap. To address this, we have made the following specific revisions to the figure and its legend: We have updated the y-axis label from “protein” to “ $\log_{10}(k_{cat}$ value) of conserved enzymes” to clearly indicate that each row represents an individual enzyme shared across the four yeast species. We have rewritten the legend for **Fig. 5h** to provide a detailed explanation. The revised legend now specifies that the heatmap shows hierarchically clustered k_{cat} profiles, explains the meaning of the color gradient (blue to red), and describes the clustering method used (lines 940-946).

(h)

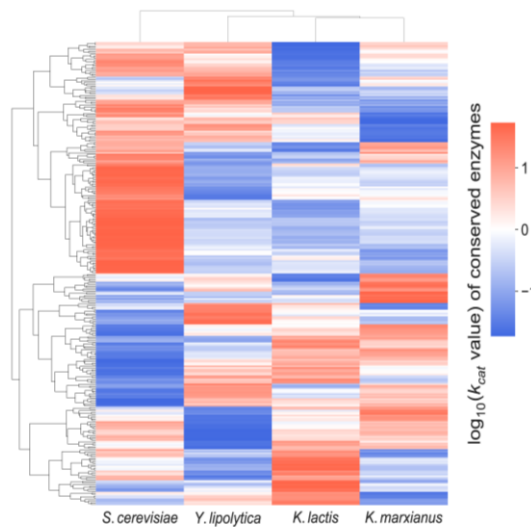


Fig. 5 Performance of the GAN approach in other yeasts and comparative analysis. **h** Hierarchical clustering analysis of k_{cat} values for the 742 conserved enzymes shared across the four yeast species. The heatmap displayed the $\log_{10}$ -transformed k_{cat} values (s^{-1}), where rows represented individual enzymes and columns represented species. The color scale ranges from blue (low) to red (high). The hierarchical clustering of species (top) and enzymes (left) based on kinetic similarity, highlighting the divergence between Crabtree-positive (*S. cerevisiae*) and Crabtree-negative yeasts.

11. Lack of Clarity in Figure 6e

- The y-axis label and the fitting method used to generate the curve are not described. Additional details are needed to interpret this figure.

Response: Thank you for requesting clarification on **Fig. 6e**. We acknowledge that the description of the y-axis and the curve fitting method was missing in the original manuscript.

1. The figure displays ridge plots representing the distribution of optimized k_{cat} values. The y-axis represents individual enzymes, while the height of each curve represents the probability density of the k_{cat} values for that specific enzyme.

2. The distributions were generated using Kernel Density Estimation (KDE) based on the top 100 high-performing k_{cat} parameter sets obtained from the EnzymeTuning optimization process. This visualization illustrates the variability and potential shifts in catalytic efficiency under different conditions (C-lim vs. N-lim). We have revised the legend of **Fig. 6e** to explicitly state the plotting method and the definition of the axes. Due to the rearrangement of the manuscript, this figure has been moved to Supplementary Figure 8a (lines 1046-1052).

a Ridge plots showing the distributions of log-transformed k_{cat} values (h^{-1}) for representative enzymes enriched in lipid, nucleobase, and amino acid metabolic processes. The distributions were derived from the top 100 GAN-optimized parameter sets. Curves were generated using Kernel Density Estimation (KDE) to visualize the probability density of k_{cat} values under carbon-limited (green) and nitrogen-limited (orange) chemostat conditions. The y-axis indicated specific enzymes, and the x-axis represented the $\log(k_{cat})$

magnitude."

12. Lack of Experimental Validation

The manuscript proposes several enzyme usage mechanisms dependent on C/N ratio but provides no experimental validation. At least one such prediction should be experimentally tested to substantiate the model's biological relevance.

Response: Thank you for this critical comment. We fully agree that validating the biological relevance of the proposed enzyme usage mechanisms is essential. While the primary scope of this study is the development of the EnzymeTuning computational framework, which has been extensively validated via phenotype and proteome abundance predictions across multiple species (**Figs. 2-5**), we acknowledge the need to substantiate the specific mechanistic insights derived from the C/N ratio analysis.

To address this without conducting new wet-lab experiments, we have implemented a two-fold validation strategy in the revised manuscript, leveraging both independent high-quality experimental datasets from the literature and a new robustness analysis under anaerobic conditions.

1. Validation via Consistency with Independent Experimental Studies

In the revised Results section, we have explicitly mapped our model-predicted adaptive strategies to established experimental findings in *S. cerevisiae*. We demonstrate that our predictions are not merely theoretical but are corroborated by quantitative proteomic studies (lines 445-462):

"To decipher the specific regulatory mechanism, we analyzed the density distributions of relative enzyme usage for key functional modules (**Fig. S8a**). Crucially, the predicted patterns showed high consistency with established physiological responses reported in the literature, substantiating the biological relevance of our model. The model predicted a significant reduction in the relative usage of enzymes involved in energetically costly lipid pathways (e.g., glycerophospholipid and sterol biosynthesis) under N-lim chemostat conditions. This prediction aligned with the "proteome economy" principle observed in quantitative proteomic studies, which reported that yeast cells actively downregulated non-critical biosynthetic machinery to preserve proteome resources under nitrogen scarcity ^{4,5}. (Yu R, et al. *Nature Communications*, 2020; Xia J, et al. *Nature Communications*, 2022)

Conversely, enzymes facilitating nitrogen recycling and redox balance, such as ornithine acetyltransferase (ARG8) and aldehyde dehydrogenase (ALD2), displayed increased utilization under N-lim chemostat conditions. This upregulation mirrored experimental findings, where nitrogen limitation triggers the induction of scavenging pathways (e.g., the Arg/Orn cycle) to mobilize intracellular nitrogen reservoirs ^{6,7}. (Boer VM, et al. *Molecular Biology of the Cell*, 2010; Kolkman A, et al. *Molecular Systems Biology*, 2006)

Additionally, the increased usage of glycolysis and glycerol metabolism enzymes (e.g., GPD2) under N-lim chemostat conditions suggested a shift towards maintaining redox homeostasis and precursor supply, a strategy consistent with previously described metabolic remodeling mechanisms under nutrient stress ^{8,9}. (Kumar R, Lahtvee P-J. *bioRxiv*, 2020; Li Y, Lai YH, Lu T. *ACS Synthetic Biology*, 2025)"

2. Extended Validation: Robustness in Anaerobic Conditions (New Fig. S8)

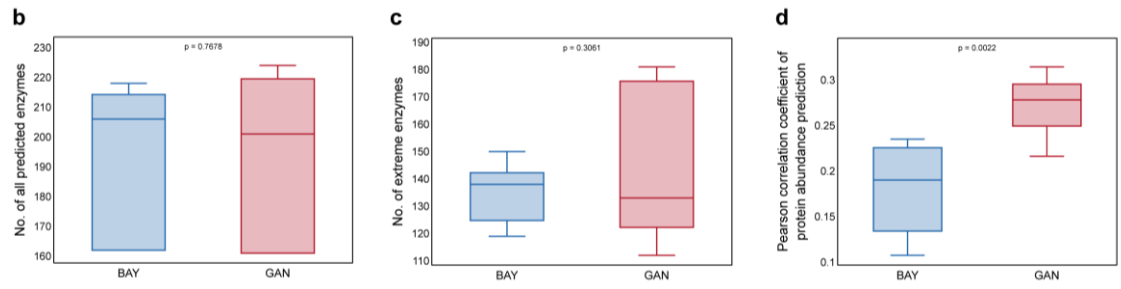
To further substantiate the model's biological relevance beyond the specific C/N case, we extended our application to anaerobic carbon-limited fermentation across varying growth rates. We selected anaerobic conditions as a "stress test". If the model captures true biological principles rather than just fitting noise, it should produce physiologically consistent predictions under these distinct conditions (lines 463-492).

"To demonstrate that the robustness of EnzymeTuning extends beyond aerobic environments, we applied the framework to simulate metabolic phenotypes and protein abundances under anaerobic carbon-limited (C-lim) chemostat conditions across five specific growth rates (0.05, 0.10, 0.15, 0.20, and 0.25 h⁻¹). As shown in **Fig. S8b-d**, no significant differences were observed between the Bayesian-based and EnzymeTuning-optimized ecGEMs in terms of the total number of predictable enzymes or the number of extreme enzymes across growth rates (**Fig. S8b, c**). In contrast, EnzymeTuning achieved a clear improvement in protein abundance prediction accuracy. The PCC between predicted and measured protein abundances increased from 0.19 to 0.28 across all growth rates (P value <0.05; **Fig. S8d**). These results indicated that the enzyme-constrained models incorporating K_{cat} values optimized by EnzymeTuning also achieved improved predictive performance in protein abundance prediction under anaerobic conditions.

Furthermore, we analyzed metabolic flux distributions to assess the physiological consistency of the model predictions. As illustrated in **Fig. 6d-f**, the Bayesian-based ecGEM exhibited unrealistic "flux shifting" artifacts in key enzymes, including alcohol dehydrogenase (ADH1), hexokinase (HXK2), and pyruvate decarboxylase (PDC1). Under anaerobic carbon-limited chemostat conditions, the flux of HXK2 increased linearly with the specific growth rate to match the glucose uptake rate, while the fluxes of PDC1 and ADH1 concomitantly rose to regenerate cytosolic NAD⁺ and generate ATP ^{10, 11}. (Nielsen J, *et al. Microbiology*, 1997; Diderich JA, *et al. Journal of Biological Chemistry*, 1999)

The Bayesian model violated this principle, showing abrupt oscillations or zero-flux artifacts at intermediate growth rates (e.g., 0.15 h⁻¹), whereas EnzymeTuning generated smooth, continuous trajectories. The subsystem-level analysis also revealed that EnzymeTuning correctly captured the global flux redistribution, which was consistent with the quantitative metabolic flux analysis (MFA) studies ¹². (Jouhten P, *et al. BMC Systems Biology*, 2008)

Specifically, EnzymeTuning correctly captured the global flux redistribution characteristic of anaerobic metabolism. The carbon flux was predominantly redirected towards glycolysis/gluconeogenesis to support growth, while the tricarboxylic acid (TCA) cycle operated as a minimized, branched pathway for biosynthesis rather than energy generation ¹². In contrast, the Bayesian-based ecGEM incorrectly predicted high respiratory activity (oxidative phosphorylation) and unstable TCA fluxes."



Supplemental Fig. 8 The k_{cat} distributions under various C/N conditions and performance of EnzymeTuning in improving protein abundance prediction anaerobic conditions. **b** Comparison of the total number of predictable enzymes using the GAN method versus the Bayesian approach. **c** Comparison of the number of extreme enzymes using the GAN method and the Bayesian approach. **d** Comparison of Pearson correlation coefficients of protein abundance predictions between the GAN method and the Bayesian approach.

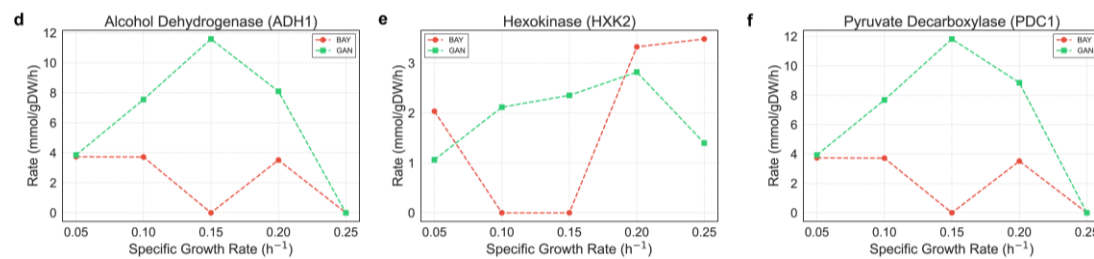


Fig. 6 The k_{cat} distribution and functional enrichment analysis of differential proteins under various C/N conditions and the metabolic flux of key enzymes under anaerobic conditions. **d-f** Metabolic flux of key enzymes (ADH1 (**d**), HXK2 (**e**), and PDC1 (**f**)) across specific growth rates under anaerobic conditions. Green dashed line: GAN-based ecGEM; red dashed line: Bayesian-based ecGEM.

3. Future Experimental Roadmap

We fully acknowledge the reviewer's point that direct experimental validation remains the gold standard. To address this constructively, we have added a dedicated paragraph in the Discussion section explicitly outlining a roadmap for future validation (lines 545-552).

“While these consistencies substantiate the biological relevance of ecGEM, future experimental validation remains essential to fully corroborate the inferred catalytic plasticity. We propose that specific enzymatic activity assays should be conducted on key responsive enzymes (e.g., GPD2, ARG8) with controlled C/N gradients. Such experimental validation would allow for the direct observation of the predicted continuous shifts in enzyme usage, thus establishing a definitive link between computational inferences and *in vivo* physiological adaptations.”

13. Limited Demonstration of Practical Utility

The current study stops at “model performance improvement” without showing translational potential for metabolic engineering.

- A case study is recommended: use EnzymeTuning-optimized ecGEMs to predict the production yield of a high-value metabolite, comparing pre- and post-optimization results.

- Cross-species validation (e.g., *E. coli* ecGEMs) would further demonstrate generalizability and strengthen the manuscript's impact.

Response: Thank you very much for this constructive suggestion regarding the practical utility of our framework. We agree that demonstrating the translational potential of EnzymeTuning for metabolic engineering is crucial.

1. To address your recommendation, we highlight that EnzymeTuning has already been successfully applied in the “StrainOptimizer” framework¹³. As referenced in the revised manuscript, the k_{cat} parameters optimized by our tool enabled StrainOptimizer to accurately predict yield and targets for multiple products. We further performed a specific case study on threonine target identification to demonstrate the reliability of our predictions. We compared our EnzymeTuning-predicted targets with results from ecFactory, a well-established computational strain design tool. We found that EnzymeTuning successfully identified key high-priority targets consistent with ecFactory results, such as HOM2, HOM3, and THR1 (lines 494-515).

“The potential applications of EnzymeTuning for metabolic engineering has been verified by one recent study from our group¹³. In that study, EnzymeTuning-optimized parameters were crucial for enabling the ecGEMs to accurately predict production targets for various high-value chemicals, validating the reliability of our toolbox in broader application. To further evaluate the utility of EnzymeTuning in identifying gene targets, we conducted one additional case study on threonine production. We applied the FSEOF algorithm using both Bayesian-based and EnzymeTuning-optimized ecGEMs¹⁴. Comparison of the two models revealed that EnzymeTuning identified a broader landscape of potential targets (28 vs. 20 proteins in Bayesian-based ecGEM), showing high consistency with the state-of-the-art tool ecFactory¹⁵. Specifically, EnzymeTuning correctly prioritized high-value candidates within the direct biosynthetic pathway, including HOM3 (aspartokinase), HOM2 (aspartate-semialdehyde dehydrogenase), and THR1 (homoserine kinase). Notably, HOM3 encodes the key rate-limiting enzyme governing the entry of carbon flux into the aspartate family pathway, and its overexpression is a well-established strategy for enhancing threonine production¹⁶. Additionally, EnzymeTuning uniquely highlighted critical upstream targets that were missed by the Bayesian approach, such as PYC1 (pyruvate carboxylase). This finding aligns with the fact that enhancing precursor supply via PYC1 is crucial for high-yield threonine production¹⁷. This capability to simultaneously identify direct biosynthetic and distal precursor-supply targets demonstrated that EnzymeTuning offered a comprehensive and mechanistically sound roadmap for rational strain design.”

2. Regarding the reviewer's suggestion on cross-species validation, we would like to highlight that we have extensively validated EnzymeTuning on the prokaryotic model organism *Escherichia coli* (lines 410-424).

“To further assess the general applicability of EnzymeTuning beyond yeast species, we extended the framework to a prokaryotic system and applied it to *Escherichia coli*. Specifically, we constructed an ecGEM for *E. coli* based on the genome-scale metabolic model iML1515 and optimized the k_{cat} parameters under 17 distinct nutrient conditions. These conditions spanned a diverse set of carbon sources, enabling a systematic evaluation of model behavior in a bacterial system.

As shown in **Fig. 5j-l**, EnzymeTuning consistently outperformed the Bayesian approach. As shown in **Fig. 5j**, the two methods exhibited comparable performance, with negligible differences observed in the total number of predictable enzymes. The number of extreme enzymes decreased from a median of approximately 41 to 20 (**Fig. 5k**), suggesting a more rational k_{cat} assignment. Consistently, the accuracy of protein abundance prediction was improved with the PCC increasing from 0.37 to 0.47 (**Fig. 5l**). These results indicated that EnzymeTuning enabled more effective calibration of k_{cat} parameters, thus improving the proteome-level prediction performance in bacterial ecGEMs”

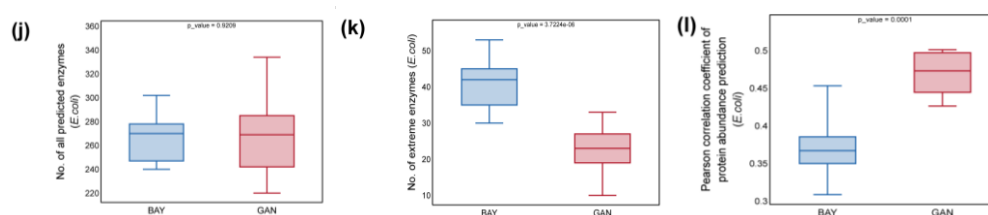


Fig. 5 Performance of the GAN approach in other yeasts and comparative analysis of k_{cat} distributions across Crabtree-positive and Crabtree-negative yeasts. **j** Comparison of the total number of predictable enzymes using the GAN method versus the Bayesian approach in *E. coli*. **k** Comparison of the number of extreme enzymes using the GAN method and the Bayesian approach in *E. coli*. **l** Comparison of Pearson correlation coefficients of protein abundance predictions between the GAN method and the Bayesian approach in *E. coli*.

14. The introduction section of the manuscript provides an incomplete review of AI-based kinetic parameter prediction methods. Several relevant tools have already been published, such as UniKP, CatPred, etc.

Response: Thank you so much for this valuable suggestion. We agree that the original Introduction did not provide a sufficiently comprehensive overview of AI-based approaches for kinetic parameter prediction. In the revised manuscript, we have expanded the Introduction to include and discuss several recently published representative tools, such as UniKP, CatPred, and DEKP, briefly summarizing their methodological characteristics and application scopes in enzyme kinetic parameter prediction. Furthermore, we have also extended the discussion to include “model-based refinement strategies” (e.g., PRESTO, Bayesian inference, and CAMEL). By distinguishing between these “bottom-up” sequence methods and “top-down” methods, we provide a comprehensive context that clearly delineates the methodological positioning of EnzymeTuning (lines 52-80; 91-99).

“To overcome these limitations, deep learning (DL)-based approaches have emerged as promising tools for large-scale kinetic parameter inference^{18, 19}. Sequence-based tools, such as DLKcat and TurNup, leverage graph neural networks (GNNs) or Transformer architectures to predict k_{cat} values from protein sequences and substrate structures^{20, 21}.

More recently, advanced frameworks like UniKP, CatPred and DEKP have further integrated pre-trained protein language models and 3D structural information to enhance prediction robustness and coverage^{18, 22, 23}. While these “bottom-up” methods provided valuable initial estimates for uncharacterized enzymes, they inherently generated point

estimates based on physicochemical properties, often neglecting the system-level physiological constraints (e.g., metabolic flux balance and proteome capacity) that govern *in vivo* enzyme behavior. Consequently, direct integration of these predicted enzyme kinetic parameter values into ecGEMs often led to suboptimal phenotype and protein abundance predictions.

To address the discrepancy between *in vitro* potentials and *in vivo* states, model-based refinement strategies (top-down methods) have been developed to calibrate k_{cat} values, thereby narrowing the gap between model predictions and experimentally measured values. Early approaches, such as the heuristic tuning in GECKO, manually adjusted the k_{cat} of growth-limiting enzymes to match experimental growth rates, but often failed to capture global proteome allocation patterns²⁴. Statistical inference methods, such as Bayesian approaches and PRESTO, offered more rigorous corrections by integrating quantitative proteomics data to infer posterior parameter distributions^{20, 25}. Additionally, hybrid frameworks like CAMEL coupled ecGEMs with machine learning (ML) regressors to improve abundance predictions²⁶. However, these approaches were constrained by inherent methodological limitations. Bayesian frameworks were often computationally intensive and sensitive to prior distributions, whereas optimization-based methods (e.g., PRESTO) typically yielded deterministic point estimates that neglected biological variability. Furthermore, hybrid ML predictors often operated as 'black boxes,' failing to provide the explicit, mechanistic parameters necessary for rational strain design."

"In this study, we developed a GAN-based framework named EnzymeTuning for global k_{cat} optimization. Distinct from sequence-based methods (e.g., DLKcat, UniKP) that focused on single enzymes, EnzymeTuning functioned as a "model refiner" that enforced global consistency between metabolic fluxes and proteome resources. Unlike sampling-based methods that were often limited by strong prior dependencies, EnzymeTuning enabled a global and allocated inference of kinetic parameters. This formulation allowed the model to learn biologically feasible distributions constrained by global proteomic resources, providing a robust alternative that reflected context-specific metabolic regulations."

15. Unavailable Code Repository

The provided repository (<https://github.com/hongzhonglu/EnzymeTuning>) is inaccessible, making it impossible to evaluate or reproduce the computational workflow. Full access to the code and datasets must be ensured.

Response: Thank you for bringing this issue to our attention. We confirm that the inaccessibility was caused by a temporary repository permission setting, which has now been corrected. The GitHub repository is currently fully public and accessible.

Reviewer #2 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #2 (Remarks on code availability):

The provided repository (<https://github.com/hongzhonglu/EnzymeTuning>) is inaccessible.

Response: We confirm that the inaccessibility was caused by a temporary repository permission setting, which has now been corrected. The GitHub repository is currently fully public and accessible.

Reviewer #3 (Remarks to the Author):

Major comments:

1. The introduction provides good context on current deep-learning approaches (DLKcat, TurNup) but the novelty claim of EnzymeTuning relative to Bayesian methods and earlier DL approaches is somewhat unclear, how the presented model differs in terms of runtime, robustness, or biological interpretability of the results?

Response: Thank you so much for this insightful comment. In the revised Introduction, we have restructured the literature review to categorize existing methods into “sequence-based methods” (bottom-up) and “model-based methods” (top-down). We have explicitly described how EnzymeTuning addressed the limitations of these approaches in terms of computational efficiency, robustness, and biological interpretability. As this point overlaps with Review 1’s Comment 14, we have provided a comprehensive response and detailed validation results in our reply to that comment. Please refer to the response to Review 1, Comment 14 for further details.

2. [L487-490] It is not clear from which source the protein degradation rates are adapted, as a review article is referenced. The manuscript claims to incorporate literature-derived k_{deg} values to estimate v_{syn} ; however, it does not specify the number of proteins with experimentally measured k_{deg} , how the missing values were handled, or what the distribution of assumed vs measured degradation rates looked like.

Response: Thank you for raising this important point regarding the source and handling of protein degradation rates. We would like to clarify that the k_{deg} values were derived from two independent experimental studies that employed the same proteome turnover measurement strategy based on pulse-chase with labeled amino acids ^{2, 3}. These two studies differ slightly in the definition and calculation of k_{deg} , which we now explicitly describe and discuss in the Methods section (lines 662-669).

“The initial v_{syn} values were derived from literature-reported protein degradation constants (k_{deg}), which were collected from two independent large-scale proteome turnover studies employing distinct computational strategies ^{2, 3}. These two studies obtained k_{deg} constants by fitting the temporal profiles of labeling incorporation (relative isotope abundance, RIA) to an exponential equation. However, one study defined the k_{deg} constants on protein turnover rates inferred from dynamic labeling experiments, whereas the other estimated the k_{deg} constants by explicitly subtracting growth-rate-dependent dilution effects.”

3. Although RMSE and PCC are used extensively (e.g., Figure 3, S5), the manuscript does not clearly specify the number of enzymes considered “predictable,” nor whether PCC was computed on log-transformed or raw abundance values.

Response: Thank you for pointing this out. We have now clarified both points in the revised manuscript.

First, “predictable enzymes” are defined as enzymes for which both model-predicted protein (i.e., non-zero flux) and experimentally measured protein abundance are available under the same condition. Enzymes with zero predicted usage or missing experimental measurements were excluded from the correlation and error analyses.

Second, Pearson correlation coefficients (PCC) were computed using log-transformed values of both predicted and measured protein abundances. Specifically, PCC was calculated based on log10-transformed data to reduce the influence of scale differences and heteroscedasticity commonly observed in quantitative proteomics datasets (lines 594-598).

“In this study, predictable enzymes were defined as the subset of enzymes for which both non-zero predicted protein abundances (derived from non-zero protein exchange fluxes) and experimentally measured protein abundances were available under identical conditions. All PCCs were calculated using log-transformed abundance values.”

4. It is not clear whether any validation experiments were performed on a dataset which was not included in the initial training. Such validation would be important to validate the predictability of the proposed model.

Response: Thank you very much for this insightful comment regarding the necessity of validation on independent datasets. We agree that demonstrating the model's predictability on data not included in the initial training is crucial for establishing its robustness and generalizability. To address this, we have performed validation on two independent datasets that were strictly excluded from the model's training phase: (1) independent quantitative proteomic studies from literature under different C/N ratios and (2) a distinct physiological condition (anaerobic condition). As this point overlaps with Review 1's Comment 12, we have provided a comprehensive response and detailed validation results in our reply to that comment. Please refer to the response to Review 1, Comment 12 for further details.

Minor comments:

5. Figures have often very small and unreadable axis names (particularly Figures 3, 5 and 6).

Response: Thank you for pointing out this. we have already fixed it.

6. “fundings” → “findings” (Figure 4 discussion)

Response: Thank you for pointing out this. we have already fixed it.

References

1. Fan X, Cao L & Yan X. Sensitivity analysis and adaptive mutation strategy differential evolution algorithm for optimizing enzymes' turnover numbers in metabolic models. *Biotechnol Bioeng* **120**, 2301-2313 (2023).
2. Christiano R, Nagaraj N, Frohlich F & Walther TC. Global proteome turnover analyses of the Yeasts *S. cerevisiae* and *S. pombe*. *Cell Rep* **9**, 1959-1965 (2014).
3. Martin-Perez M & Villen J. Determinants and Regulation of Protein Turnover in Yeast. *Cell Syst* **5**, 283-294 (2017).
4. Yu R, *et al.* Nitrogen limitation reveals large reserves in metabolic and translational capacities of yeast. *Nat Commun* **11**, 1881 (2020).
5. Xia J, Sánchez BJ, Chen Y, Campbell K, Kasvandik S & Nielsen J. Proteome allocations change linearly with the specific growth rate of *Saccharomyces cerevisiae* under glucose limitation. *Nat Commun* **13**, 2819 (2022).
6. Boer VM, Crutchfield CA, Bradley PH, Botstein D & Rabinowitz JD. Growth-limiting intracellular metabolites in yeast growing under diverse nutrient limitations. *Mol Biol Cell* **21**, 198-211 (2010).
7. Kolkman A, *et al.* Proteome analysis of yeast response to various nutrient limitations. *Mol Syst Biol* **2**, 2006 (2006).
8. Kumar R & Lahtvee P-J. Proteome overabundance enables respiration but limitation onsets carbon overflow. *Preprint at* <https://doi.org/10.1101/2020.02.20.957662> (2020).
9. Li Y, Lai YH & Lu T. Coarse-grained modeling elucidates differential metabolism of *Saccharomyces cerevisiae* under varied nutrient limitations. *ACS Synth Biol* **14**, 1523-1532 (2025).
10. Nissen TL, Schulze U, Nielsen J & Villadsen J. Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces Cerevisiae*. *Microbiology* **143**, 203-218 (1997).
11. Diderich JA, *et al.* Glucose uptake kinetics and transcription of HXTGenes in chemostat cultures of *Saccharomyces cerevisiae*. *J Biol Chem* **274**, 15350-15359 (1999).
12. Jouhten P, *et al.* Oxygen dependence of metabolic fluxes and energy generation of *Saccharomyces cerevisiae* CEN.PK113-1A. *BMC Systems Biology* **2**, 60 (2008).
13. Wang H, *et al.* StrainOptimizer empowers rational cell factory design through multi-scale metabolic models with expression and proteome constraints. *Preprint at* <https://doi.org/10.1101/2025.11.03.685948> (2025).
14. Choi HS, Lee SY, Kim TY & Woo HM. *In silico* identification of gene amplification targets for improvement of lycopene production. *Appl Environ Microbiol* **76**, 3097-3105 (2010).
15. Ishchuk OP, *et al.* Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **119**, e2108245119 (2022).
16. Farfán Ma-J, Aparicio L & Calderón IL. Threonine overproduction in yeast strains carrying the HOM3-R2 mutant allele under the control of different inducible promoters. *Appl Environ Microbiol* **65**, 110-116 (1999).
17. Kramer BP, Fischer M & Fussenegger M. Semi-synthetic mammalian gene regulatory networks. *Metab Eng* **7**, 241-250 (2005).
18. Boorla VS & Maranas CD. CatPred: a comprehensive framework for deep learning in

- vitro enzyme kinetic parameters. *Nat Commun* **16**, 2072 (2025).
19. Wang T, *et al.* DeepEnzyme: a robust deep learning model for improved enzyme turnover number prediction by utilizing features of protein 3D-structures. *Briefings in Bioinformatics* **25**, bbae409 (2024).
 20. Li F, *et al.* Deep learning-based k_{cat} prediction enables improved enzyme-constrained model reconstruction. *Nat Catal* **5**, 662-672 (2022).
 21. Kroll A, Rousset Y, Hu XP, Liebrand NA & Lercher MJ. Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning. *Nat Commun* **14**, 4139 (2023).
 22. Yu H, Deng H, He J, Keasling JD & Luo X. UniKP: a unified framework for the prediction of enzyme kinetic parameters. *Nat Commun* **14**, 8211 (2023).
 23. Wang Y, Cheng L, Zhang Y & Cao Y, Alghazzawi D. DEKP: a deep learning model for enzyme kinetic parameter prediction based on pretrained models and graph neural networks. *Brief Bioinform* **26**, bbaf187 (2025).
 24. Chen Y, *et al.* Reconstruction, simulation and analysis of enzyme-constrained metabolic models using GECKO Toolbox 3.0. *Nat Protoc* **19**, 629-667 (2024).
 25. Wendering P, Arend M, Razaghi-Moghadam Z & Nikoloski Z. Data integration across conditions improves turnover number estimates and metabolic predictions. *Nat Commun* **14**, 1485 (2023).
 26. Moura Ferreira MA, Wendering P, Arend M, Batista da Silveira W & Nikoloski Z. Accurate prediction of *in vivo* protein abundances by coupling constraint-based modelling and machine learning. *Metab Eng* **80**, 184-192 (2023).

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have made extensive revision on the manuscript to address the reviewer's comments and the quality of the manuscript is improved. k_{cat} prediction from EnzymeTuning might be better than previous methods but still far from "high-performance". As stated in the answers to reviewer 1: "a k_{cat} parameter set is classified as high-performance if it satisfies $RMSE < 2$ for growth phenotype prediction and an overall Pearson correlation coefficient (PCC) > 0.4 for protein abundance prediction." A PCC value of 0.4 is still very low, especially considering that it is computed using log-transformed values of both predicted and measured protein abundances. I think RMSE was calculated based on eq.5 (line 605). How many exchange rates were used in the calculation (value of n)? Why value of 2 was considered as acceptable? The authors may need to give more details on the calculation, for example providing the experimental exchange rates and predicted rates using different sets of k_{cat} values.

Response: We sincerely thank the reviewer for this important and constructive comment. We agree that additional clarification regarding the definition of "high-performance", together with a more explicit description of the RMSE calculation and additional supporting data, would greatly improve the clarity of our method. Accordingly, we have revised the manuscript and added new methodological explanations and supplementary analyses, as detailed below.

1. Clarification of the definition of "high-performance" k_{cat} parameter sets

We appreciate the reviewer's concern regarding our previous description that a "high-performance" k_{cat} parameter set satisfies $RMSE < 2$ for phenotype prediction and $PCC > 0.4$ for protein abundance prediction. In the revised manuscript, we have clarified that these two criteria were used only as empirical reference values for the initial Bayesian-derived k_{cat} parameter sets, rather than as the final definition of high-performance sets used in EnzymeTuning. The labels used for model training were assigned based on the discriminator score described in **Eq. 2-4**, which jointly considers phenotype prediction and protein abundance prediction performance.

$$\min Normalized(1 - PCC) + Normalized(fold\ change) \quad (2)$$

$$\min Normalized(1 - PCC) + Normalized(extreme_number) \quad (3)$$

$$\min Normalized(1 - PCC) + Normalized(extreme_number) + Normalized(RMSE) \quad (4)$$

Specifically, parameter sets in the better-performing half of the initial parameter sets were labeled as "high-performance", whereas those in the poorer-performing half were labeled as "low-performance". In addition, the criteria of $RMSE < 2$ and $PCC > 0.4$ was originally determined with reference to the performance observed in the previously reported Bayesian-based method, which served as the starting point of our optimization framework. We have revised the corresponding text in the Methods section (lines 636-643) and Results section (lines 118-122) to make this point explicit and to avoid the misleading impression.

Lines 636-643:

“The initial parameter sets were selected using the criteria of $RMSE < 2$ and $PCC > 0.4$, which served as empirical reference of Bayesian-derived parameter sets. Subsequently, these selected parameter sets were assigned a discriminator score that integrated growth phenotype prediction and protein abundance prediction performance, as described in **Eq. 2-4**, where lower scores correspond to superior predictive accuracy. Parameter sets with the lowest scores were classified as the high-performance group, whereas those with higher scores were designated as the low-performance group.”

Lines 118-122:

“In **Step 1**, the initial parameter sets were first screened using empirical reference of Bayesian method for predicting growth phenotypes (the root mean squared error, $RMSE < 2$) and protein abundances (Pearson correlation coefficient, $PCC > 0.4$). Subsequently, high-performance sets were identified based on a comprehensive discriminator score, forming a labeled training dataset for a conditional generative adversarial network (cGAN)¹.”

2. Interpretation of protein abundance prediction performance

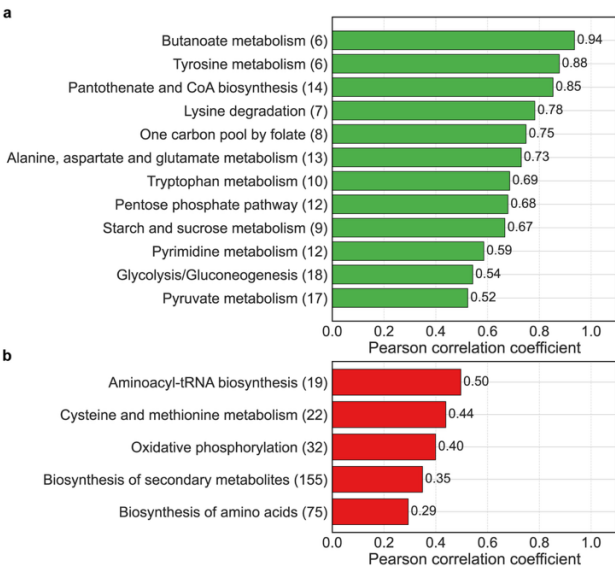
We also agree that a PCC value around 0.4 is relatively low and does not represent strong predictive performance in absolute terms. However, to the best of our knowledge, predicting intracellular protein abundances using genome-scale enzyme-constrained metabolic models remains highly challenging because it is influenced by multiple sources of uncertainty, including incomplete enzyme annotation, variability in quantitative proteomics measurements, condition-dependent regulation, and structural simplifications in metabolic models^{2, 3}. In ecGEMs, correlations between predicted enzyme usage or protein abundance and experimental proteomics data are often modest, and PCC values around 0.4-0.5 are generally considered to represent relatively strong performance in this context. We also compared EnzymeTuning with three representative state-of-the-art methods. In terms of PCC, the practical performance of existing ecGEM-based frameworks for protein abundance prediction generally remains at a modest level, typically around 0.3-0.5. In this context, EnzymeTuning outperformed the Bayesian-, CAMEL-, and PRESTO-based parameterizations in overall protein abundance prediction accuracy, as summarized in **Table 1**.

Table 1 Comparison of EnzymeTuning with representative k_{cat} inference approaches in *S. cerevisiae* under carbon-limited chemostat cultivation.

Methods		Bayesian ⁴	PRESTO ⁵	CAMEL ²	EnzymeTuning
Phenotype Accuracy	RMSE	0.732	1.170	1.495	0.789
	PCC	0.301	0.371	0.356	0.426
Proteome Accuracy	Predictable Number	262	260	259	265
	Extreme Enzyme	53	76	59	46
Efficiency	Runtime	~24 h	~7 h	~6 h	~24 h

To further demonstrate the advantage of our method, we also added a new supplementary analysis showing KEGG pathway-level prediction accuracy for protein abundances. This analysis demonstrates that, in addition to improving global PCC, the GAN-based ecGEM achieved relatively high prediction accuracy for several important metabolic pathways, including glycolysis/gluconeogenesis, pyruvate metabolism, and the pentose phosphate pathway. These additions provide a more comprehensive evaluation of the predictive performance. (lines 257-263)

“In addition, pathway-level analysis using EnzymeTuning further demonstrated that high prediction accuracy was achieved for central carbon metabolic pathways, including glycolysis/gluconeogenesis (PCC = 0.54, P value < 0.05), pyruvate metabolism (PCC = 0.52, P value < 0.05), and the pentose phosphate pathway (PCC = 0.68, P value < 0.05) (Fig. S7). These results suggested that EnzymeTuning not only enhanced the overall protein abundance prediction but also effectively captured the protein allocation patterns in the central carbon metabolic pathways.”



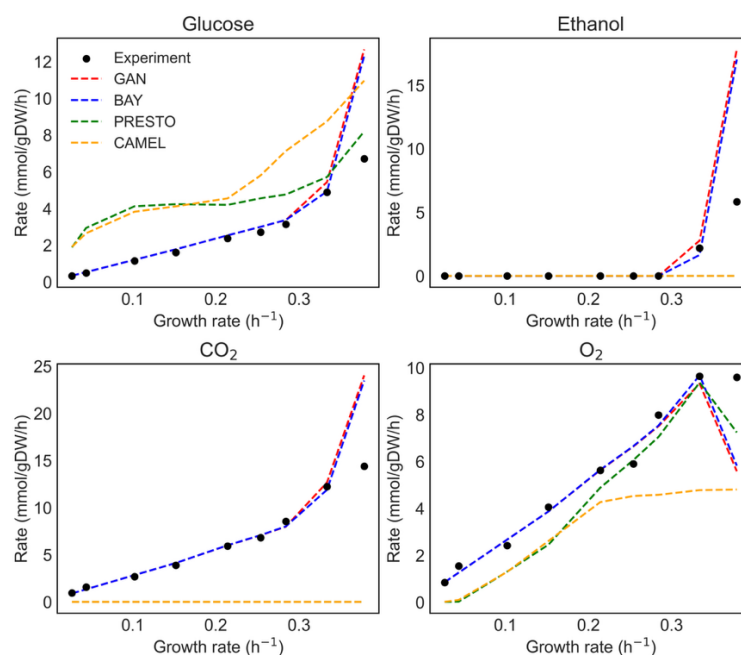
Supplemental Fig. 7 KEGG pathway-level evaluation of protein abundance prediction accuracy using the GAN method. **a** and **b** KEGG pathways associated with proteins that were predicted with relatively high accuracy (**a**) and relatively low accuracy (**b**) using the GAN method. Bars indicate the Pearson correlation coefficient between predicted and experimentally measured protein abundances for proteins mapped to each pathway. Numbers in parentheses indicate the number of proteins in each pathway.

3. Clarification of the RMSE calculation

We thank the reviewer for pointing out the lack of clarity in **Eq. 5**. In the revised Methods section, we now explicitly define n as the number of experimentally measured phenotypic variables included in the RMSE calculation for the experimental dataset. These variables include the specific growth rate and the measured exchange rates, such as ethanol, acetate, pyruvate, succinate, glycerol, carbon dioxide, and oxygen. Only variables with corresponding experimental measurements were included in the RMSE calculation; therefore, the value of n is dataset-dependent rather than fixed. In the training dataset used in this study, the RMSE calculation primarily considered the specific growth rate and the exchange rates of glucose, ethanol, CO₂, and O₂. (lines 613-616)

“Here, $y_{i,pre}$ and $y_{i,exp}$ represent the predicted and experimentally measured values, respectively. n is the number of experimentally measured phenotypic variables such as the specific growth rate and exchange fluxes for ethanol, glycerol, carbon dioxide and oxygen.” In addition, a comparison between the experimentally measured exchange rates and those predicted by the four different methods has been provided in **Fig. S6**. Notably, compared to the other approaches, EnzymeTuning yields the most accurate predictions, demonstrating its enhanced capability in reproducing experimental trends, such as glucose uptake rate and oxygen consumption rate, under carbon-limited uptake conditions. (lines 251-254)

“Beyond the overall RMSE comparison, the Bayesian and GAN-based methods were superior in simulating the trends of specific exchange rate, such as glucose and oxygen uptake rates under carbon-limited chemostat conditions.”



Supplemental Fig. 6 Performance comparison of different k_{cat} parameterization strategies in predicting exchange rates under carbon-limited chemostat conditions. Experimentally measured and predicted exchange rates of glucose, ethanol, CO₂, and O₂ are plotted against growth rate for ecGEMs parameterized using EnzymeTuning (GAN) (red), Bayesian optimization (BAY) (blue), PRESTO (green), and CAMEL (orange). Black dots represent experimental measurements, and dashed lines represent model predictions. Overall, we believe that these revisions substantially improve the clarity, transparency, and reproducibility of the study, and we sincerely thank the reviewer for these valuable suggestions.

Response: We also sincerely thank Reviewer 2 and Reviewer 3 for your insightful comments during the previous round, which have significantly strengthened the quality of our manuscript. Again, we thank you for your time and your favorable assessment of our revisions.

Reviewer #2 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3 (Remarks to the Author):

I can recommend the manuscript for publication.

References

1. Choudhury, S., et al. Reconstructing Kinetic Models for Dynamical Studies of Metabolism using Generative Adversarial Networks. *Nat Mach Intell* **4**, 710-719 (2022).
2. Moura Ferreira, M. A., Wendering, P., Arend, M., Batista da Silveira, W., Nikoloski, Z. Accurate prediction of in vivo protein abundances by coupling constraint-based modelling and machine learning. *Metab Eng* **80**, 184-192 (2023).
3. Sjoberg, G., Rekena, A., Fornstad, M., Lahtvee, P. J., van Maris, A. J. A. Evaluation of enzyme-constrained genome-scale model through metabolic engineering of anaerobic co-production of 2,3-butanediol and glycerol by *Saccharomyces cerevisiae*. *Metab Eng* **82**, 49-59 (2024).
4. Li, F., et al. Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nature Catalysis* **5**, 662-672 (2022).
5. Wendering, P., Arend, M., Razaghi-Moghadam, Z., Nikoloski, Z. Data integration across conditions improves turnover number estimates and metabolic predictions. *Nat Commun* **14**, 1485 (2023).