In vitro turnover numbers do not reflect in vivo activities of yeast enzymes

Yu Chen* and Jens Nielsen**

*Department of Biology and Biotechnological Engineering, Chalmers University of Technology, SE412 96 Gothenburg, Sweden; **Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK2800 Kgs. Lyngby, Denmark; and †BioInnovation Institute, DK2200 Copenhagen N, Denmark

Edited by James J. Collins, Massachusetts Institute of Technology, Cambridge, MA, and approved July 9, 2021 (received for review May 5, 2021)

Turnover numbers ($k_{app}$ values) quantitatively represent the activity of enzymes, which are mostly measured in vitro. While a few studies have reported in vivo catalytic rates ($k_{cat}$ values) in bacteria, a large-scale estimation of $k_{app}$ in eukaryotes is lacking. Here, we estimated $k_{app}$ of the yeast Saccharomyces cerevisiae under diverse conditions. By comparing the maximum $k_{app}$ across conditions with in vitro $k_{cat}$, we found a weak correlation in log scale of $R^2 = 0.28$, which is lower than for Escherichia coli ($R^2 = 0.62$). The weak correlation is caused by the fact that many in vitro $k_{cat}$ values were measured for enzymes obtained through heterologous expression. Removal of these enzymes improved the correlation to $R^2 = 0.41$ but still not as good as for E. coli, suggesting considerable deviations between in vitro and in vivo enzyme activities in yeast. By parameterizing an enzyme-constrained metabolic model with our $k_{app}$ dataset we observed better performance than the default model with in vitro $k_{cat}$ in predicting proteomics data, demonstrating the strength of using the dataset generated here.

Saccharomyces cerevisiae | turnover number | $k_{cat}$ | proteomics | metabolism

Enzyme turnover numbers, also termed $k_{cat}$ values, are fundamental parameters that specify the maximum rates of enzymatic reactions and hence determine the rates of biological processes such as metabolism. Determining $k_{cat}$ is therefore essential for quantitatively understanding, modeling, and engineering cells. Traditionally, $k_{cat}$ values are measured in vitro, which might differ from the in vivo situation. In addition, the coverage of measured $k_{cat}$ is poor even for well-studied organisms (1). To address these issues, an approach for estimating in vivo enzyme catalytic rates, also termed $k_{app}$ values, is to use the equation

$$v = k_{app} \cdot E,$$  

where $v$ is the metabolic flux through the enzyme and $E$ the enzyme abundance (2). This approach was used to estimate $k_{app}$ for Escherichia coli using absolute proteomics and flux data from various sources (2, 3), and it was found that the maximum $k_{app}$ values across conditions, defined as $k_{max}$ values, correlate well with in vitro $k_{cat}$ values in log scale (2). Here, we generate a $k_{app}$ dataset for Saccharomyces cerevisiae under diverse conditions. We analyze our dataset and correlate in vivo with in vitro enzyme activities in yeast. Finally, we compare the predictive power of an enzyme-constrained metabolic model using in vivo and in vitro kinetic data.

Results and Discussion

To generate the yeast $k_{app}$ dataset we collected absolute proteomics data of S. cerevisiae under diverse conditions (4–7) (Dataset S1). The absolute protein abundance can be directly adopted as enzyme abundance in Eq. 1. Note that we did not consider enzyme complexes composed of multiple distinct subunits due to the difficulty in calculating the abundance of catalytic sites (2). To determine the metabolic flux in Eq. 1, we performed flux balance analysis (FBA), as done previously (2), using the latest genome-scale metabolic model (GEM) of S. cerevisiae Yeast8 (8) (SI Appendix).

Using the absolute proteomics and flux data, we calculated $k_{app}$ for 358 metabolic reactions under 26 conditions (Dataset S2). By correlating the estimated $k_{app}$ in log10 scale, we found that yeast $k_{app}$ varied between conditions with the lowest $R^2$ being around 0.4 (Dataset S3). This is different from the findings for E. coli, where log-transformed $k_{app}$ values correlate strongly across conditions with the lowest $R^2$ being above 0.9 (9).

By comparing $k_{max}$ (Dataset S4), i.e., maximum $k_{app}$ across all the studied conditions, with the corresponding in vitro $k_{cat}$ (Dataset S5) we obtained a fairly weak correlation in log scale with $R^2 = 0.28$ for S. cerevisiae (Fig. 4), which is much lower than that of E. coli ($R^2 = 0.62$) (2). A weak correlation was also reported for the plant Arabidopsis thaliana (10). By examining in vitro $k_{cat}$, we found that some were estimated using purified enzymes obtained through heterologous expression in E. coli, with the others being estimated from yeast extracts. We therefore divided the in vitro $k_{cat}$ dataset into two groups, i.e., heterologous and homologous expression. We found that there was no correlation for the heterologous expression group (Fig. 1B), suggesting that in vivo $k_{cat}$ values obtained through heterologous expression poorly represent in vivo catalytic rates of yeast enzymes. This might be due to the lack of natural posttranslational modifications (PTMs) in the expression organism, which could regulate enzyme activity. Indeed, we found that 27 out of the 29 reactions that reported PTMs on the enzymes (Dataset S6), indicating that these PTMs could functionally affect enzyme activity (11). In the homologous expression group we observed an improved correlation of $R^2 = 0.41$ in log scale (Fig. 1C) and thus identified the data obtained through heterologous expression as the main source of deviations.

To evaluate how uncertainties in the FBA-based flux may impact our dataset we first investigated the effect of flux variability on the estimated $k_{max}$ (SI Appendix). We found that less than 8% of $k_{max}$ values could differ, due to flux variability, by more than one order of magnitude (Dataset S4), and after removing these data we found the correlation between in vivo and in vitro values in log scale to be almost unchanged, i.e., $R^2$ (all data) = 0.27, $R^2$ (heterologous data) = 0.12, and $R^2$ (homologous data) = 0.39. Second, we provided another set of $k_{max}$ values (Dataset S7) estimated using unbiased flux random sampling (SI Appendix), which correlate strongly ($R^2 = 0.97$) with the FBA-based $k_{max}$ values in log scale. To evaluate the effect of a single high outlier value due to protein measurement we also correlated in log scale the second largest $k_{app}$ across all conditions with in vitro $k_{cat}$ but found similar $R^2$ values, i.e., $R^2$ (all data) = 0.25, $R^2$ (heterologous data) = 0.1, and $R^2$ (homologous data) = 0.39. Moreover, by correlating $k_{app}$ of...
each condition with in vitro $k_{\text{cat}}$ we found most correlations to be poor in log scale (Dataset S8). A recent study, using another proteomic dataset and a different GEM, also showed a poor correlation in log scale of $R^2 = 0.27$ between yeast $k_{\text{app}}$ and in vitro $k_{\text{cat}}$ at one condition (12).

We analyzed the yeast $k_{\text{app}}$ under various conditions based on the ratio of condition-specific $k_{\text{app}}$ over $k_{\text{max}}$ for each reaction. By plotting the average ratio for each condition versus the corresponding growth rate ($\mu$), we observed an increasing trend (Fig. 1D), which is in line with findings for E. coli (13). Furthermore, we compared the ratio of individual reactions between slow ($\mu \leq 0.2/h$) and fast ($\mu > 0.2/h$) growth and found that $k_{\text{app}}$ is significantly higher in faster-growing cells (Fig. 1E). We therefore conclude that $k_{\text{app}}$ of yeast enzymes increase with growth rate. This suggests that proteome is more efficiently used at faster growth (13) and also indicates that growth could be controlled by efficiency of specific enzymes independent of conditions.

As turnover numbers are essential parameters in enzyme-constrained GEMs (ecGEMs) (9), we tested the use of $k_{\text{app}}$ and in vitro $k_{\text{cat}}$ in a yeast ecGEM ecYeast8 (8). We parameterized the model with 1) an assumed same $k_{\text{cat}}$ for all enzymes, 2) default in vitro $k_{\text{cat}}$ in the original ecYeast8, 3) general $k_{\text{max}}$ and 4) $\mu$-dependent $k_{\text{max}}$. Note that $\mu$-dependent $k_{\text{max}}$ is defined as the maximum $k_{\text{app}}$ across the conditions under which growth rate is not greater than the given $\mu$. To compare model performance, we used the model to predict proteomics data for growth on various carbon sources (14, 15), and we compared with the data not used to estimate our $k_{\text{app}}$ dataset. We found that $k_{\text{max}}$ outperforms the assumed $k_{\text{cat}}$ and default in vitro $k_{\text{cat}}$ (Fig. 2), confirming our estimation of $k_{\text{max}}$ to be reliable. Notably, $\mu$-dependent $k_{\text{max}}$ can further improve the predictions (Fig. 2), meaning that it is more effective to use $\mu$-dependent $k_{\text{max}}$ values than condition-independent maximum values, which are adopted in most published ecGEMs (9). In addition, we found that default in vitro $k_{\text{cat}}$ in the original ecYeast8 outperforms the assumed same $k_{\text{cat}}$ for all enzymes (Fig. 2), meaning that it is still acceptable to use in vitro $k_{\text{cat}}$ when $k_{\text{app}}$ values are unavailable.

Overall, we present a $k_{\text{app}}$ dataset of S. cerevisiae under various conditions, which can be used by ecGEMs for simulating the corresponding conditions. As $k_{\text{app}}$ depends generally on growth rates rather than conditions (Fig. 1D), we believe that our $\mu$-dependent

![Fig. 2. Predictions of proteomics data on various carbon sources by ecYeast8 parameterized with an assumed same $k_{\text{cat}}$, default in vitro $k_{\text{cat}}$, general $k_{\text{max}}$, and $\mu$-dependent $k_{\text{max}}$. Model performance is evaluated by root-mean-square error (RMSE) between predicted and measured protein levels on a log10 scale. N is the number of proteins with predicted nonzero concentrations by four parameterization strategies.](https://doi.org/10.1073/pnas.2108391118)
$k_{\text{max}}$ can be used to parameterize ecGEMs for predicting other conditions that are not involved in our dataset.

**Materials and Methods**

The metabolic fluxes were simulated using Yeast8 constrained by measurements. The absolute proteomics data were processed, i.e., the units of protein abundances were converted to millimoles per gram cell dry weight (gCDW). Details of all the materials and methods are provided in SI Appendix.

**Data Availability.** The data and codes are available at https://github.com/SysBioChalmers/Yeast_kapp. All other study data are included in the article and/or supporting information.

**ACKNOWLEDGMENTS.** We acknowledge funding from the European Union’s Horizon 2020 research and innovation program under grant agreement 686070. We also acknowledge funding from the Novo Nordisk Foundation (grant NNF10CC1016517).