



## High-throughput and reliable acquisition of *in vivo* turnover number fuels precise metabolic engineering

Zhenghong Li<sup>a,b,1</sup>, Chengyu Zhang<sup>b,c,1</sup>, Zhengduo Wang<sup>b,c,1</sup>, Chuan Li<sup>b</sup>, Zhiheng Yang<sup>b</sup>, Zilong Li<sup>c</sup>, Lixin Zhang<sup>b</sup>, Weishan Wang<sup>c,d,\*</sup>

<sup>a</sup> Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 518055, Shenzhen, China

<sup>b</sup> State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China

<sup>c</sup> State Key Laboratory of Bioreactor Engineering, And School of Biotechnology, East China University of Science and Technology (ECUST), Shanghai, 200237, China

<sup>d</sup> University of Chinese Academy of Sciences, Beijing 100049, China

### ARTICLE INFO

#### Keywords

Turnover number  
Genome scale models  
High throughput  
Metabolic reconstitution  
Machine learning

### ABSTRACT

As synthetic biology enters the era of quantitative biology, mathematical information such as kinetic parameters of enzymes can offer us an accurate knowledge of metabolism and growth of cells, and further guidance on precision metabolic engineering.  $k_{cat}$ , termed the turnover number, is a basic parameter of enzymes that describes the maximum number of substrates converted to products each active site per unit time. It reflects enzyme activity and is essential for quantitative understanding of biosystems. Usually, the  $k_{cat}$  values are measured *in vitro*, thus may not be able to reflect the enzyme activity *in vivo*. In this case, Davidi et al. defined a surrogate  $k_{max}^{vivo}$  ( $k_{app}$ ) for  $k_{cat}$  and developed a high throughput method to acquire  $k_{max}^{vivo}$  from omics data. Heckmann et al. and Chen et al. proved that the surrogate parameter can be a good embodiment of the physiological state of enzymes and exhibit superior performance for enzyme-constrained metabolic model to the default one. These breakthroughs will fuel the development of system and synthetic biology.

Traditional synthetic biology and metabolic engineering focus more on the trial-and error-based pathway engineering, which consumes lots of time and effort. With the development of big data and machine learning, synthetic biology has evolved to the era of quantitative biology, and it is possible to conduct precise metabolic engineering guided by rational design like multi-scale data processing and learning from microbial intelligence. For this purpose, mathematical parameters are crucial for a better knowledge of cell growth and metabolism [1,2]. Among them,  $k_{cat}$  is considered as one of the most important parameters that illustrate the basic metabolic activities.

$k_{cat}$  values, also called the turnover numbers, are parameters describe the maximum number of substrates converted to products per unit time per active site. It is a quantitative measurement of enzyme activity. The  $k_{cat}$  values are defined by equation (1),

$$k_{cat} = \frac{v_{max}}{E_0} \quad (1)$$

where  $v_{max}$  is the maximum reaction rate and  $E_0$  is the concentration of

the enzyme. In system and computational biology, the  $k_{cat}$  values are adopted for enzyme-constrained genome scale models, which offers a better understanding of metabolic activities of living cells [2]. Also,  $k_{cat}$  values can provide quantitative guidance for pathway engineering and metabolic reconstitution (Fig. 1). To this end, it is important to develop strategies for  $k_{cat}$  acquisition.

$k_{cat}$  values are normally obtained from protein expression assays by a low throughput way (Fig. 1), which is labor-intensive and time-consuming [3]. Indeed, the *in vitro*  $k_{cat}$  measurements differ among literatures and only 9% of  $k_{cat}$  values of the well-studied *Escherichia coli* are available [4]. Furthermore, the *in vitro*  $k_{cat}$  may not be a true reflection for *in vivo* scenarios since the conditions are significantly changed [5]. Actually, the  $k_{cat}$  value varies according to pH, buffer, temperature, and immobilization. For example, the *in vitro*  $k_{cat}$  of catalase varies from 11 s<sup>-1</sup> to 151 s<sup>-1</sup> due to changes in pH and buffer in *E. coli* [6]. Therefore, it is an urgent necessity for a universal method to estimate  $k_{cat}$  data that well matching the *in vivo* scenarios.

To address the aforementioned issues, Davidi *et al.* defined  $k_{app}$ , as a

\* Corresponding author. State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China.

E-mail address: [wangws@im.ac.cn](mailto:wangws@im.ac.cn) (W. Wang).

<sup>1</sup> These authors contribute equally.

surrogate for in vitro  $k_{cat}$  [7]:

$$k_{app,ij} = \frac{v_{ij}}{E_{ij}} = \eta k_{cat,ij} \tag{2}$$

$$k_{app,maxi} = k_{cat,i} \tag{3}$$

where  $v_{ij}$  is the reaction rate of  $i$ -th enzymatic reaction under  $j$ -th cultivation conditions.  $E_{ij}$  is the enzyme concentration.  $\eta$  is a condition-dependent function, ranging between 0 and 1, which describes the decrease in the catalytic rate. The  $k_{app,maxi}$  values represent the maximum turnover number under optimal state among varied cultivation conditions. It can be determined when the growth condition pools are large enough. Davidi et al. tested 31 different growth conditions and proteomic data  $E_{ij}$  were extracted (Fig. 1). For the reaction rate, flux balance analysis (FBA) was adopted for the calculation of  $v_{ij}$ . The  $k_{app,maxi}$  values were then determined. Comparison of  $k_{app,maxi}$  values with in vitro  $k_{cat}$  values for *E. coli* yielded a correlation factor of  $R^2 = 0.62$ , indicating a good agreement for in vivo and in vitro situations [7]. Chen et al. used the same method to obtain the  $k_{app}$  data of yeast and the correlation analysis results in a  $R^2 = 0.26$ , which suggested a high in vivo and in vitro discrepancy. The in vitro  $k_{cat}$  values of yeast were further analyzed by Chen et al. and the authors found that the weak correlation is caused by the heterologous expression of enzymes. Exclusion of these heterologous expressed enzymes led to a better  $R^2 = 0.41$  but still not as good as for *E. coli* [8].

However, since the metabolic flux  $v_{ij}$  is calculated by FBA in the method, such estimation can be limited by poor FBA accuracy or incomplete genome-scale model, and the robustness of the  $k_{app}$  in

response to various perturbations remains unclear [9]. Therefore, Heckmann et al. proposed a  $^{13}C$  metabolic flux analysis (MFA) based method for  $k_{app}$  predication [9] (Fig. 1). To obtain the  $v_{ij}$  data, *E. coli* cells with various knockouts of central metabolism were cultivated in minimum medium with glucose. The resulting cultures were then subjected to MFA to generate  $v_{ij}$  of 21 strains from adaptive laboratory evolution (ALE) [10,11]. The MFA based  $k_{app,maxi}$  showed good consistency with  $k_{app,maxi}$  by FBA with  $R^2 = 0.9$ , indicating that a good complementation for the systems lacking FBA data. Heckmann et al. [9] and Chen et al. [8] then applied  $k_{app}$  to the enzyme-constrained metabolic model and found that model with  $k_{app}$  showed better performance in compared with the original one. It is clear that flux data and proteome data from the well characterized species under the certain cultivation conditions are available and reliable. Although getting omics data under different cultivation conditions is still a tedious process for non-model organisms, the work flow and protocols, as described in this comment, are available to some extent.

The above summarized methods can effectively and unfailingly measure the  $k_{app}$  in a high-throughput way currently. However, the coverage is still limited. To address this issue, Chew et al. developed a 3D convolutional neural network to predict enzymatic catalysis rates based on experimental reaction data and corresponding molecular dynamics simulation data, which can be used to predict flux data [12]. It is worth mentioning that Heckman et al. [9] made use of machine learning to extrapolate the  $k_{app}$  to genome scale, which combined with the 3D structure and biochemical characteristics of enzymes in the database to further estimate the  $k_{app}$  for enzymes suffering from coverage issue of proteomic techniques [13]. This method offered another perspective for

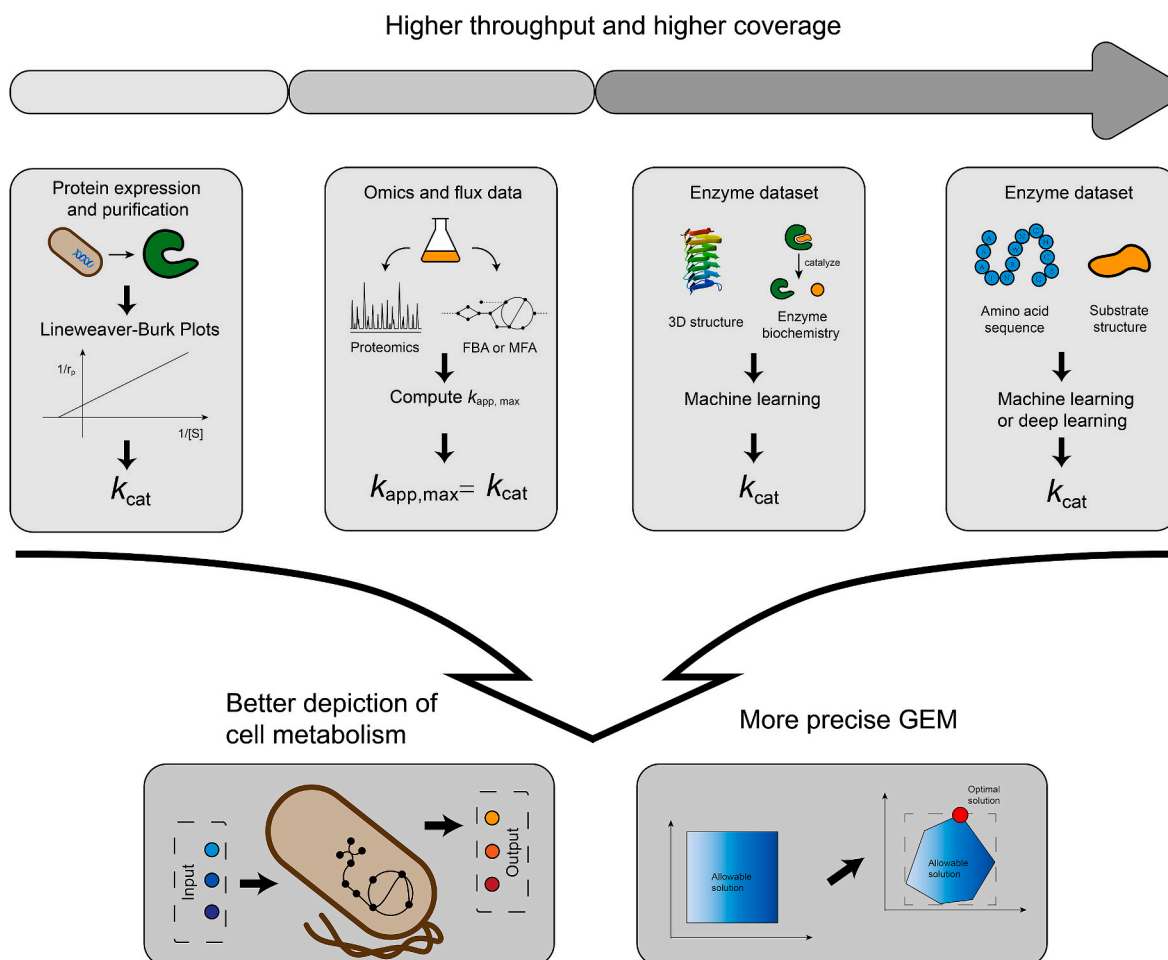


Fig. 1. Stages of the development of  $k_{cat}$  extraction methods for higher throughput and better coverage.

$k_{app}$  calculation despite the availability of the structural information and biochemical property is limited for some enzymes of interest. Thus, Feiran Li *et al* have developed a deep learning based  $k_{app}$  data prediction, which exploited the most extensive and common information, like amino acid sequences and substrate structures to detect the  $k_{cat}$  values for better coverage of the enzymes in the system [14].

In summary, the development of  $k_{cat}$  data can be divided into three stages (Fig. 1). For the original  $k_{cat}$  measurement, in vitro assays were utilized, and the reaction data was fit to Michaelis-Menten equation [15–17]. Then methods integrating proteomics data together with FBA or MFA were introduced based on equation (1), (2) and (3) [7–9]. Currently, a more coverage method regarding the machine learning is being developed for generic  $k_{app}$  prediction. This shows that  $k_{app}$  data extraction methods are transitioning from a low-throughput, low-coverage approach to a high-throughput, high-coverage, high-accuracy approach.

In the outlook, the accurate and complete dataset of  $k_{cat}/k_{app}$  provides us another mathematical measurement in addition to transcriptome, proteome and metabolome and reveals the quantitative picture of genome scale metabolic network. On the one hand,  $k_{app}$  based enzyme-constrained metabolic model can presumably be improved for better depiction of intrinsic growth and metabolic activities of cells. On the other hand, kinetic features can effectively provide theoretical guidance for metabolic engineering modification and further realize the transformation from random trial-and-error to rational design. The development of  $k_{cat}$  values will support a revolutionary progress for synthetic biology and metabolic engineering.

#### Declaration of competing interest

The authors indicate that they have no conflict of interest.

#### Acknowledgements

This work was supported by the National Natural Science Foundation of China (31922002, 32101174) to W.W and Z.L, the National Key R&D program of China (2020YFA0907800), the 111 Project (B18022), the Youth Innovation Promotion Association CAS (Y202027) to W.W and the China Postdoctoral Science Foundation (2021M703379) to Z.L.

#### References

- [1] Bekiaris PS, Klamt S. Automatic construction of metabolic models with enzyme constraints. *Bmc Bioinformatics* 2020;21. <https://doi.org/10.1186/s12859-019-3329-9>.
- [2] Sanchez BJ, Salazar MJ, Weng LL, et al. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol* 2017;13(935). <https://doi.org/10.15252/msb.20167411>.
- [3] NilssonNielsen AJ, Palsson BO. Metabolic Models of Protein Allocation Call for the Kinetome. *Cell Syst* 2017;5:538–41. <https://doi.org/10.1016/j.cels.2017.11.013>.
- [4] Davidi D, Milo R. Lessons on enzyme kinetics from quantitative proteomics. *Curr Opin Biotechnol* 2017;46:81–9. <https://doi.org/10.1016/j.copbio.2017.02.007>.
- [5] van Eunen K, Bouwman J, Daran-Lapujade P, et al. Measuring enzyme activities under standardized in vivo-like conditions for systems biology. *FEBS J* 2010;277:749–60. <https://doi.org/10.1111/j.1742-4658.2009.07524.x>.
- [6] Moore RL, Powell LJ, Goodwin DC. The kinetic properties producing the perfunctory pH profiles of catalase-peroxidases. *Biochim Biophys Acta* 2008;1784(6):900–7. <https://doi.org/10.1016/j.bbapap.2008.03.008>.
- [7] Davidi D, Noor E, Liebermeister W, et al. Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro  $k_{cat}$  measurements. *Proc Natl Acad Sci U S A* 2016;113:3401–6. <https://doi.org/10.1073/pnas.1514240113>.
- [8] Chen Y, Nielsen J. In vitro turnover numbers do not reflect in vivo activities of yeast enzymes. *Proc Natl Acad Sci U S A* 2021;118. <https://doi.org/10.1073/pnas.2108391118>.
- [9] Heckmann D, CampeuLloyd A, CJ, et al. Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers. *Proc Natl Acad Sci U S A* 2020;117:23182–90. <https://doi.org/10.1073/pnas.2001562117>.
- [10] Sandberg TE, SalazarWeng MJ, LL, et al. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab Eng* 2019;56:1–16. <https://doi.org/10.1016/j.ymben.2019.08.004>.
- [11] McCloskey D, Xu S, Sandberg TE, et al. Adaptive laboratory evolution resolves energy depletion to maintain high aromatic metabolite phenotypes in *Escherichia coli* strains lacking the Phosphotransferase System. *Metab Eng* 2018;48:233–42. <https://doi.org/10.1016/j.ymben.2018.06.005>.
- [12] Chew AK, Jiang S, Zhang W, et al. Fast predictions of liquid-phase acid-catalyzed reaction rates using molecular dynamics simulations and convolutional neural networks. *Chem Sci* 2020;11(46):12464–76. <https://doi.org/10.1039/d0sc03261a>.
- [13] Heckmann D, Lloyd CJ, Mih N, et al. Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nat Commun* 2018;9(1):5252. <https://doi.org/10.1038/s41467-018-07652-6>.
- [14] Li F, Yuan L, Hongzhong L, et al. <https://doi.org/10.1101/2021.08.06.455417>.
- [15] Wang ML, Moynie L, Harrison PJ, et al. Using the pimeloyl-CoA synthetase adenylation fold to synthesize fatty acid thioesters. *Nature Chemical Biology* 2017;13:660. <https://doi.org/10.1038/nchembio.2361>.
- [16] Yu YL, Zhang HL, Guo FG, et al. A Unique Hexokinase in *Cryptosporidium parvum*, an Apicomplexan Pathogen Lacking the Krebs Cycle and Oxidative Phosphorylation. *Protist* 2014;165:701–14. <https://doi.org/10.1016/j.protis.2014.08.002>.
- [17] Yin Y, Kirsch JF. Identification of functional paralog shift mutations: conversion of *Escherichia coli* malate dehydrogenase to a lactate dehydrogenase. *Proc Natl Acad Sci USA* 2007;104:17353–7. <https://doi.org/10.1073/pnas.0708265104>.