

RESEARCH

Open Access



Construction of an enzyme-constrained metabolic network model for *Myceliophthora thermophila* using machine learning-based k_{cat} data

Yutao Wang^{1,2,3,5†}, Zhitao Mao^{4,5†}, Jiacheng Dong^{1,2,3,5}, Peiji Zhang^{4,5}, Qiang Gao¹, Defei Liu^{2,3,5*},
Chaoguang Tian^{2,3,5*} and Hongwu Ma^{4,5*}

Abstract

Background Genome-scale metabolic models (GEMs) serve as effective tools for understanding cellular phenotypes and predicting engineering targets in the development of industrial strain. Enzyme-constrained genome-scale metabolic models (ecGEMs) have emerged as a valuable advancement, providing more accurate predictions and unveiling new engineering targets compared to models lacking enzyme constraints. In 2022, a stoichiometric GEM, iDL1450, was reconstructed for the industrially significant fungus *Myceliophthora thermophila*. To enhance the GEM's performance, an ecGEM was developed for *M. thermophila* in this study.

Results Initially, the model iDL1450 underwent refinement and updates, resulting in a new version named iYW1475. These updates included adjustments to biomass components, correction of gene-protein-reaction (GPR) rules, and a consensus on metabolites. Subsequently, the first ecGEM for *M. thermophila* was constructed using machine learning-based k_{cat} data predicted by TurNuP within the ECMpy framework. During the construction, three versions of ecGEMs were developed based on three distinct k_{cat} collection methods, namely AutoPACMEN, DLKcat and TurNuP. After comparison, the ecGEM constructed using TurNuP-predicted k_{cat} values performed better in several aspects and was selected as the definitive version of ecGEM for *M. thermophila* (ecMTM). Comparing ecMTM to iYW1475, the solution space was reduced and the growth simulation results more closely resembled realistic cellular phenotypes. Metabolic adjustment simulated by ecMTM revealed a trade-off between biomass yield and enzyme usage efficiency at varying glucose uptake rates. Notably, hierarchical utilization of five carbon sources derived from plant biomass hydrolysis was accurately captured and explained by ecMTM. Furthermore, based on enzyme cost considerations, ecMTM successfully predicted reported targets for metabolic engineering modification and introduced some new potential targets for chemicals produced in *M. thermophila*.

[†]Yutao Wang and Zhitao Mao have equally contributed to this work.

*Correspondence:

Defei Liu

liudf@tib.cas.cn

Chaoguang Tian

tian_cg@tib.cas.cn

Hongwu Ma

ma_hw@tib.cas.cn

Full list of author information is available at the end of the article



Conclusions In this study, the incorporation of enzyme constraint to iYW1475 not only improved prediction accuracy but also broadened the model's applicability. This research demonstrates the effectiveness of integrating of machine learning-based k_{cat} data in the construction of ecGEMs especially in situations where there is limited measured enzyme kinetic parameters for a specific organism.

Keywords Enzyme-constrained model, *Myceliophthora thermophila*, Metabolic engineering, Machine learning, Carbon source hierarchy utilization

Background

Myceliophthora thermophila, a thermophilic filamentous fungus, thrives at high temperatures (45–50 °C) which possesses a remarkable ability to secrete various glycoside hydrolases and auxiliary oxidation enzymes, making it an efficient plant biomass degrader [1, 2]. These unique characteristics make *M. thermophila* a highly promising candidate for biotechnological applications in biomass conversion and high-temperature fermentation [2]. Notably, *M. thermophila* has been engineered to be a cell factory for the production of enzyme [3] and various chemicals, including fumarate [4], succinic acid, malate [5], malonic acid [6], 1,2,4-butanetriol [7] and ethanol [8, 9]. Additionally, efforts have been made to engineer this fungus to be an outstanding consolidated bioprocessing (CBP) strain to produce chemicals from biomass sources [5, 10].

Genome-scale Metabolic Models (GEMs) are effective tools for elucidating cellular phenotypes and predicting potential engineering targets to guide development of industrial strains [11–13]. In 2022, a manually curated genome-scale metabolic model, iDL1450, was constructed for *M. thermophila* [14]. With the help of iDL1450, an optimized rational design for important bulk chemicals was simulated and metabolic differences under different temperature conditions were analyzed. However, GEMs only consider stoichiometric constraints, which may not accurately capture the intracellular conditions. To enhance the performance of GEMs, the integration of enzyme constraints into the models has been explored. Several methods have been developed to incorporate the enzyme constraints, considering enzyme concentration, enzyme catalytic efficiency, and enzyme molecular weight [15–21]. In 2007, the first mathematical framework, Flux Balance Analysis with Molecular Crowding (FBAwMC), was established by Beg et al. taking into account enzyme constraints based on macromolecular crowding [22]. FBAwMC imposes constraints on enzyme concentrations at a physical level by introducing crowding coefficients, thereby achieving an overall constraint on enzyme activity. Subsequently, several methods, such as MOMENT [23], GECKO [24], AutoPACMEN [25], and ECMpy [17], has been developed to integrate enzyme constraints into GEMs.

In 2017, based on FBAwMC, Sánchez et al. developed the GECKO toolbox, which extends GEMs by adding new rows to the S-matrix that represent the enzymes and new columns representing each enzyme's usage [24]. The enzyme-constrained GEM (ecGEM) ecYeast7 revealed improved predictive performance compared with the Yeast7 and identified enzyme limitation as a major driving force behind enzymatic protein reallocation. Bekiaris et al. Combined the MOMENT and GECKO methods to introduce AutoPACMEN, a method capable of automatically retrieving enzyme data from the BRENDA [26] and SABIO-RK [27] databases, marking a significant step in automating ecGEM construction [25]. Additionally, in 2021, Mao et al. presented ECMpy, an automated method for ecGEM construction, which simplified the workflow without modifying the S-matrix [17]. ECMpy was employed to construct an ecGEM for *Escherichia coli*, demonstrating improved prediction accuracy for various cellular phenotypes [17]. Subsequently, ECMpy was utilized to develop ecGEMs for *Bacillus subtilis* [20] and *Corynebacterium glutamicum* [19], providing more precise predictions and guiding the rational design of microbial cell factories.

In this study, we first updated and modified iDL1450 to a new version called iYW1475. Based on this GEM, an enzyme-constrained model for *M. thermophila* was constructed using the ECMpy workflow. During the construction, enzyme turnover numbers (k_{cat}) were gathered in three distinct methods (AutoPACMEN [25], DLKcat [28] and TurNuP [29]), resulting in three versions of ecGEMs (eciYW1475_AP, eciYW1475_DL and eciYW1475_TN). After comparison, eciYW1475_TN was selected as the final ecGEM version for *M. thermophila* (ecMTM). Simulation of substrate hierarchy utilization and prediction of metabolic engineering targets were performed using ecMTM.

Methods

Measurements of RNA and DNA content in *M. thermophila*

In order to determine RNA and DNA content in *M. thermophila*, the wild-type strain ATCC 42464 was grown on Vogel's minimal medium supplemented with 2% glucose (GMM) containing 1.5% agar at 35 °C for 7 days

to obtain mature conidia. Conidia was inoculated into 50 mL GMM to a final concentration of 1×10^6 conidia/mL in a 250-mL Erlenmeyer flask. Then, liquid cultures were incubated under 45 °C at 150 rpm in a rotary shaker for 20 h. RNA content was quantified using the method described in report [30, 31]. A 2 mL sample was collected and certificated at 10,000×g for 5 min. The resulting pellet was washed three times with 3 mL of cold 0.7 M HClO₄. Subsequently, the mycelium was resuspended in 3 mL of 0.3 M KOH and incubated at 37 °C for 60 min with occasional shaking. Following cooling to room temperature, the samples were neutralized by adding 1.0 mL 3 M HClO₄, followed by centrifugation. The supernatant was collected, and the pellet underwent two washes with 4 mL of cold 0.5 M HClO₄. The supernatants were combined, and the volume was adjusted to 15 mL with 0.5 M HClO₄. Finally, the samples were clarified by centrifugation, and the absorbance A_{260nm} was measured using a UV spectrometry (Nanodrop 2000c spectrophotometer). For the determination of biomass dry weight, mycelium was collected using vacuum filtration, washed three times with distilled water, and rapidly frozen in liquid nitrogen. Then the biomass samples were lyophilized under -40 °C until a constant weight was achieved.

To measure DNA content, approximately 0.01 g of lyophilized mycelium was ground into powder in liquid nitrogen in a mortar. Following this, 1 mL extraction buffer (200 mM Tris-HCl pH 8.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) [32] was added and incubated at 60 °C for 30 min. The sample was subjected to two extractions by adding an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v) with certification at 10,000×g for 10 min. The supernatant was mixed with 1/10 volume of 3 M sodium acetate (pH 5.3) and 2.5 volumes ethanol to precipitate DNA at -20 °C for 1 h. After precipitation, the DNA was resuspended in TNE (1 M NaCl, 10 mM EDTA, 0.1 M Tris-HCl, pH 7.4) and treated with RNase, followed by a second precipitation and wash to remove degraded RNA. Finally, the DNA content was determined using a Nanodrop spectrophotometer.

Model update and format modification

Prior to establishing ecGEM, the initial model iDL1450 underwent comprehensive updates and format modifications to align with ECMpy requirements. In detail, biomass components were adjusted based on experimental data (Additional file 1: Table S1). Precise corrections were applied to RNA, DNA, protein and amino acid content [33], according to measured data. Subsequently, the content of lipids and cell wall was adjusted according to literatures [34] and [30], respectively.

Some redundant metabolites were identified by their corresponding IDs and names and were manually consolidated into singular entities (Additional file 2: Table S2). Besides, gene-protein-reaction (GPR) relationships were checked and updated based on experiment data from previous reports [4, 7, 35] alongside KEGG annotation [36], mainly including glycolysis, tricarboxylic acid cycle (TCA) and oxidative phosphorylation pathways (Additional file 3: Table S3). After the revision, gene number was increased from 1450 to 1475 and the initial model was renamed iYW1475.

To meet requirement of ECMpy, several changes had been made. Primarily, metabolite names were mapped to those in BiGG database [37] using KEGG [36] identifier, CHEBI IDs [38] and metabolite names (Additional file 4: Table S4). Then the format of modified GEM was converted from XML to JavaScript Object Notation (JSON) format. Furthermore, protein IDs in UniProt database [39] were collected and integrated into JSON format model as a part of gene annotation.

Construction of ecGEM and calibration of enzyme kinetic parameters

The ecGEM was constructed based on iYW1475 using the ECMpy workflow (see Fig. 1). Initially, reversible reactions were transformed into pairs of irreversible reactions, and reactions catalyzed by multiple isoenzymes were segregated into distinct reactions. The enzyme mass fraction (f) was determined as 0.55, calculated using Eq. (1) based on unpublished proteomic data measured in our lab, where A_i and A_j represent the abundances (mole ratio) of the i -th protein (p_num represents proteins expressed in the model) and j -th protein (g_num represents proteins expressed in the whole proteome of *M. thermophila*), respectively; MW_i is molecular weight of an enzyme catalyzing reaction i . Additionally, enzyme subunit information was sourced from the 'Interaction information' section in the UniProt database (see Additional file 5: Table S5).

For enzyme kinetic parameter data, three methods were employed: AutoPACMEN was utilized to collect enzyme kinetic parameter data based on EC numbers, and machine learning-based tools (DLKcat [28] and TurNuP [29]) were used to predict enzyme kinetic parameter information. Finally, the enzyme-constrained model was developed by incorporating enzyme constraints (see Eq. 2) into GEM, where v_i represents flux of i -th reaction in the model; σ_i denotes the saturation coefficient for i -th enzyme, with an average value of 0.5 assigned to all enzymes; $ptot$ represents the total protein fraction (0.4653) in *M. thermophila* and f represents the mass fraction of enzymes; $k_{cat,i}$ represents the turnover number of the enzyme catalyzing reaction i .

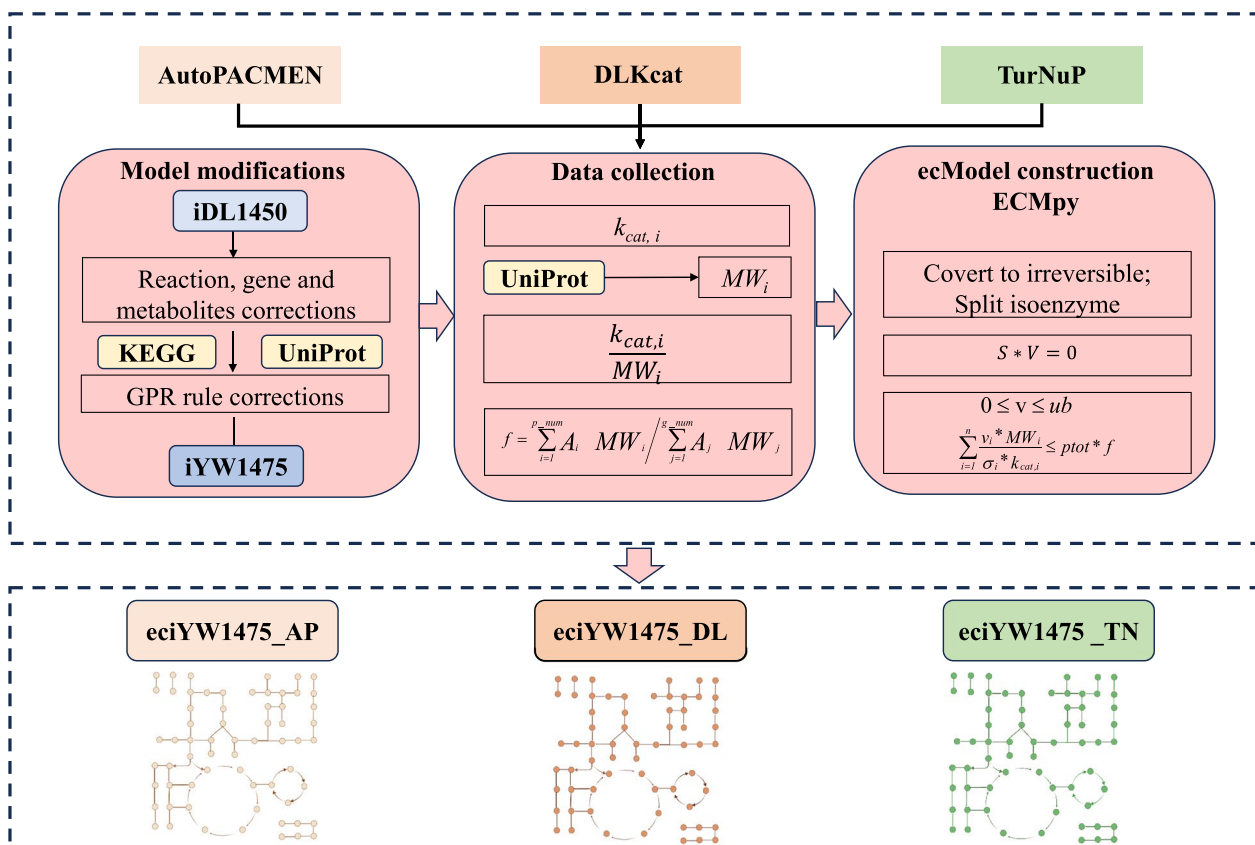


Fig. 1 Workflow for the construction of eciYW1475_AP, eciYW1475_DL and eciYW1475_TN

$$f = \frac{\sum_{i=1}^{p_num} A_i MW_i}{\sum_{j=1}^{g_num} A_j MW_j} \tag{1}$$

$$\sum_{i=1}^n \frac{v_i * MW_i}{\sigma_i * k_{cat,i}} \leq ptot * f \tag{2}$$

To enhance the alignment between model predictions and experimental data, additional fine-tuning of the original k_{cat} values was necessary for the enzyme-constrained model. The iterative calibration for k_{cat} was carried out utilizing two methods outlined in ECMpy [17]: the enzyme usage method and the ^{13}C flux consistency method, relying on measured ^{13}C flux data [33].

Phenotype phase plane (PhPP) analysis

Cells deploy distinct metabolic strategies contingent upon varying oxygen and glucose levels. To simulate this dynamic cellular responses, PhPP analysis was conducted to explore metabolic behavior [40]. During the analysis, uptake flux ranges for oxygen and glucose were set to 0–50 mmol/gDW/h and 0–10 mmol/gDW/h,

respectively. Parsimonious Flux Balance Analysis (pFBA) was used for PhPP analysis with the objective of maximum biomass [41].

Flux variability analysis (FVA)

Flux balance analysis (FBA) using linear programming with GEMs might generate diverse optimal solutions for the same objective value [42, 43]. To assess the changes in solution space after introducing enzyme constraint, flux variability ranges between iYW1475 and ecGEM were compared using FVA method [43] with some modifications [19]. Specifically, for reactions involving isozymes, the maximal flux variability range within the isozymes was used (Eq. (3)). For reversible reactions in ecGEMs, the corresponding flux variability ranges were solved using Eq. (4) [19], where $v_{i,isoj}^{max}$ and $v_{i,isoj}^{min}$ represent the maximum and minimum fluxes, respectively, of the j -th reaction among the reactions associated with isozyme i ; v_i^{max} and v_i^{min} represent the maximum and minimum fluxes of reaction i ; $v_{i,isoj}^{max}$ and $v_{i,REV}^{min}$ represent the maximum and minimum fluxes of the reversible reaction i .

$$FV_i = \max(v_{i,isoj}^{\max} - v_{i,isoj}^{\min}), j \in m \quad (3)$$

$$FV_i = (v_i^{\max} - v_i^{\min}) - (v_{i,REV}^{\max} - v_{i,REV}^{\min}) \quad (4)$$

Metabolic adjustment simulation

GEMs or ecGEMs were always employed to investigate and explicate metabolic adaptations [15–21]. In this study, metabolic adjustment was simulated using ecMTM solving by pFBA to maximize biomass production with substrate uptake rates ranging from 0 to 5 mmol/gDW/h. To further elucidate the metabolic adaptation to varying glucose uptake rates, detailed explorations were performed, encompassing alterations in biomass yield (Eq. 5), where $v_{biomass}$ represents the flux of biomass and $v_{glucose}$ denotes the glucose uptake rate, $MW_{glucose}$ is the molecular weight of glucose; enzyme efficiency for biomass synthesis ($\varepsilon_{biomass}$, Eq. 6), where $E_{\min, biomass}$ was determined using the minimum enzyme amount algorithm of ECMPy for biomass synthesis; energy synthesis enzyme cost (Eq. 7), where $E_{reaction\ for\ ATP\ production, i}$ represents the enzyme level of the i -th ATP production reaction, and $V_{net_generated_ATP, i}$ represents the flux of the i -th ATP production reaction; and oxidative phosphorylation ratio (proportion of ATP produced by oxidative phosphorylation to the total ATP production) [20].

$$\text{Biomass yield} = \frac{v_{biomass}}{v_{glucose} * MW_{glucose}} \quad (5)$$

$$\varepsilon_{biomass} = \frac{v_{biomass}}{E_{\min, biomass}} \quad (6)$$

$$\text{Energy synthesis enzyme cost} = \sum_{i=1}^n \frac{E_{reaction\ for\ ATP\ production, i}}{V_{net_generated_ATP, i}} \quad (7)$$

Substrate hierarchy utilization measured in vivo

During cellular growth on various carbon sources, a hierarchy exists in their utilization. Utilization pattern of five carbon sources in *M. thermophila* was measured in this study. For measurement consumption of carbon sources, inoculation of *M. thermophila* ATCC42464 conidia into 100 mL 1×Vogel's minimal medium supplemented with five carbon sources (glucose, xylose, galactose, arabinose, cellobiose, 0.5% (w/v) for each substrate) to a final concentration of 1×10^6 conidia/mL in a 250-mL Erlenmeyer flask. Then the liquid cultures were incubated at 45 °C and 150 rpm in a rotary shaker. Sample was centrifuged at 10,000×g for 5 min and the supernatant was filtered with 0.22- μ m cellulose acetate syringe filters before analysis. The concentration of cellobiose was determined

by HPLC (e2695; Waters, Manchester, United Kingdom) equipped with a Waters 2414 refractive index (RI) detector and an Aminex HPX-87H column (Bio-Rad) at 35 °C; 5 mM H₂SO₄ was used as the mobile phase with constant flow rate 0.5 mL/min. Other carbon sources concentrations were determined by an ICS6000 high-performance anion exchange chromatography system equipped with a Dionex CarboPac™ PA200 column (3×250 mm), a pulsed amperometric detector (HPAEC-PAD) featuring a gold working electrode and a silver/silver chloride reference electrode (Thermo Fisher Scientific, Waltham, MA, USA). The column temperature was 30 °C, the injection volume was 10 μ L, and the flow rate was 0.3 mL/min. The mobile phases were 200 mM NaOH (A) and 10 mM NaOH (B).

Substrate hierarchy utilization simulated in silico

In a recent study, Wang et al. developed a coarse-grained model to explain the sequential consumption of two carbon sources, providing a quantitative framework for microbial carbon source utilization [44]. Here, substrate hierarchy utilization of carbon sources for *M. thermophila* were explained using ecGEM in a quantitative framework. The enzyme efficiency in producing 12 biomass precursor metabolites G6P (Glucose 6-phosphate), F6P (Fructose 6-phosphate), G3P (Glyceraldehyde 3-phosphate), 3PG (Glycerate 3-phosphate), PEP (Phosphoenolpyruvate), PYR (Pyruvate), ACCOA (Acetyl-CoA), OAA (Oxaloacetate), AKG (2-Oxoglutarate), SUCC (Succinate), E4P (Erythrose 4-phosphate) and R5P (Ribose 5-phosphate) for five distinct carbon sources was calculated using ecMTM with substrate carbon atom uptake rate setting at 15 mmol/gDW/h. Then a comparative analysis was conducted to determine the order of substrate utilization. The enzyme efficiency for substrate synthesis precursors ($\varepsilon_{biomass\ precursor}$) is calculated using Eq. (8):

$$\varepsilon_{biomass\ precursor} = \frac{v_{biomass\ precursor}}{E_{\min, precursor}} \quad (8)$$

where $v_{biomass\ precursor}$ and $E_{\min, precursor}$ respectively represent the flux of biomass precursor and the minimum enzyme cost for biomass precursor synthesis.

Prediction of metabolic engineering targets

A distinctive feature of ecGEM lies in its capability to calculate the enzyme cost of reactions, facilitating the identification of key enzymes within pathways [19]. In this study, potential modification targets were identified using two methods: one is enzyme cost-based sorting method (Method 1) in which the first 15 proteins were classified as top-demanded proteins and selected as potential targets for metabolic engineering; The second is the enzyme

cost differences in different conditions (HGLP (high growth low product generation)/LGHP (low growth high product generation)) method (Method 2, reactions with a flux greater than 1 mmol/gDW/h are selected as potential targets for engineering modifications (HGLP involves the flux at biomass = 10%, while LGHP involves the flux at biomass = 100%) [19].

Statistical analysis

All experiments were carried out in three independent repeated assays.

Results

Model update and format modification

The biomass components in iDL1450 were adjusted based on measured data for protein, RNA, and DNA content. Subsequent modifications of additional components were based on literature data, all of which are detailed in the supplementary file (Additional file 1: Table S1). After that, 96 duplicate metabolites were identified and consolidated into singular entities with identical identifiers and names (Additional file 2: Table S2). Additionally, gene-protein-reaction (GPR) rules, particularly reactions with ‘and’ relationships, were modified using information from literature and KEGG data. A total of 29 reactions with ‘and’ relationships in their GPR associations were revised (Additional file 3: Table S3). Through iterative corrections, a refined model, named iYW1475, was obtained containing 1475 genes, 2591 reactions, and 1700 unique metabolites. Additionally, format adjustments were executed to comply with ECMpy criteria, and leveraging multiple identifiers mapping, 55.2% of metabolite IDs were substituted with BiGG IDs (Additional file 4: Table S4). After these updates and format modifications, iYW1475 was transitioned from XML format to JSON format (all workflows are showed in Fig. 1).

Construction of ecGEM for *M. thermophila*

Before integrating enzyme kinetic parameters, reactions in iYW1475 were transformed based on ECMpy workflow. After the transformation, the model was expanded to 6690 reactions (Table 1). Then, enzyme data were collected using AutoPACMEN method, accumulating a total of 4481 k_{cat} data ranging from 10^{-3} to 10^6 , leaving 1059 reactions without enzyme data. The missing values were filled with median value of all AutoPACMEN-collected k_{cat} data (Additional file 6: Table S6). Additionally, the molecular weights of the enzymes spanned from 8 to 1800 kDa (Fig. 2B). The constructed ecGEM from this workflow was denoted as eciYW1475_AP.

Table 1 Basic information of four models

Model	Reactions	Source of k_{cat}	Number of k_{cat}
iYW1475	2591	–	0
eciYW1475_AP	6690	BRENDA or SABIO-RK	4481
		Median imputation	1059
eciYW1475_DL	6690	DLKcat	3251
		BRENDA or SABIO-RK	52
		Median imputation	2237
eciYW1475_TN	6690	TurNuP	5519
		BRENDA or SABIO-RK	21

However, 19.1% of k_{cat} data cannot be collected through the AutoPACMEN method. Furthermore, only one k_{cat} value originated from the native species, while the remaining values were obtained from other species, indicating a low availability of data for enzyme kinetic parameters. Therefore, machine learning-based methods DLKcat and TurNuP were employed to predict k_{cat} data. DLKcat captured 67.7% k_{cat} data while TurNuP successfully predicted all k_{cat} values for reactions in iYW1475 (Additional file 6: Table S6). Consequently, a second ecGEM was constructed using DLKcat predicted data with missing values filled in using the median of DLKcat predicted k_{cat} resulting in eciYW1475_DL. Meanwhile, TurNuP-predicted data was chosen to construct a third ecGEM, named eciYW1475_TN. The distribution of k_{cat} values (Fig. 2A) indicated a more concentrated k_{cat} distribution in eciYW1475_TN and eciYW1475_DL, ranging from 10^{-2} to 10^3 compared to those in eciYW1475_AP.

Next, k_{cat} values were calibrated using enzyme usage method until the growth rate approximated the measured value (Additional file 7: Table S7). After that, the ^{13}C flux consistency method was employed to ensure flux consistency between ecGEM and measured fluxes (Additional file 7: Table S7). After the calibration of the k_{cat} values, growth and flux predictions were compared. Results showed that (Fig. 2C) the predicted biomass fluxes in eciYW1475_AP and eciYW1475_TN were both similar to experimental value. However, for eciYW1475_DL, the predicted biomass was much smaller than the experimental value due to missing values being filled with the median value of 7.8391 s^{-1} leading to an over-constrained model. Compared with eciYW1475_DL and eciYW1475_AP, eciYW1475_TN indicated a higher consistency with the ^{13}C data (Fig. 2C, Additional file 8: Table S8). Discrepancies were observed between the predicted fluxes and the experimental data for eciYW1475_DL and

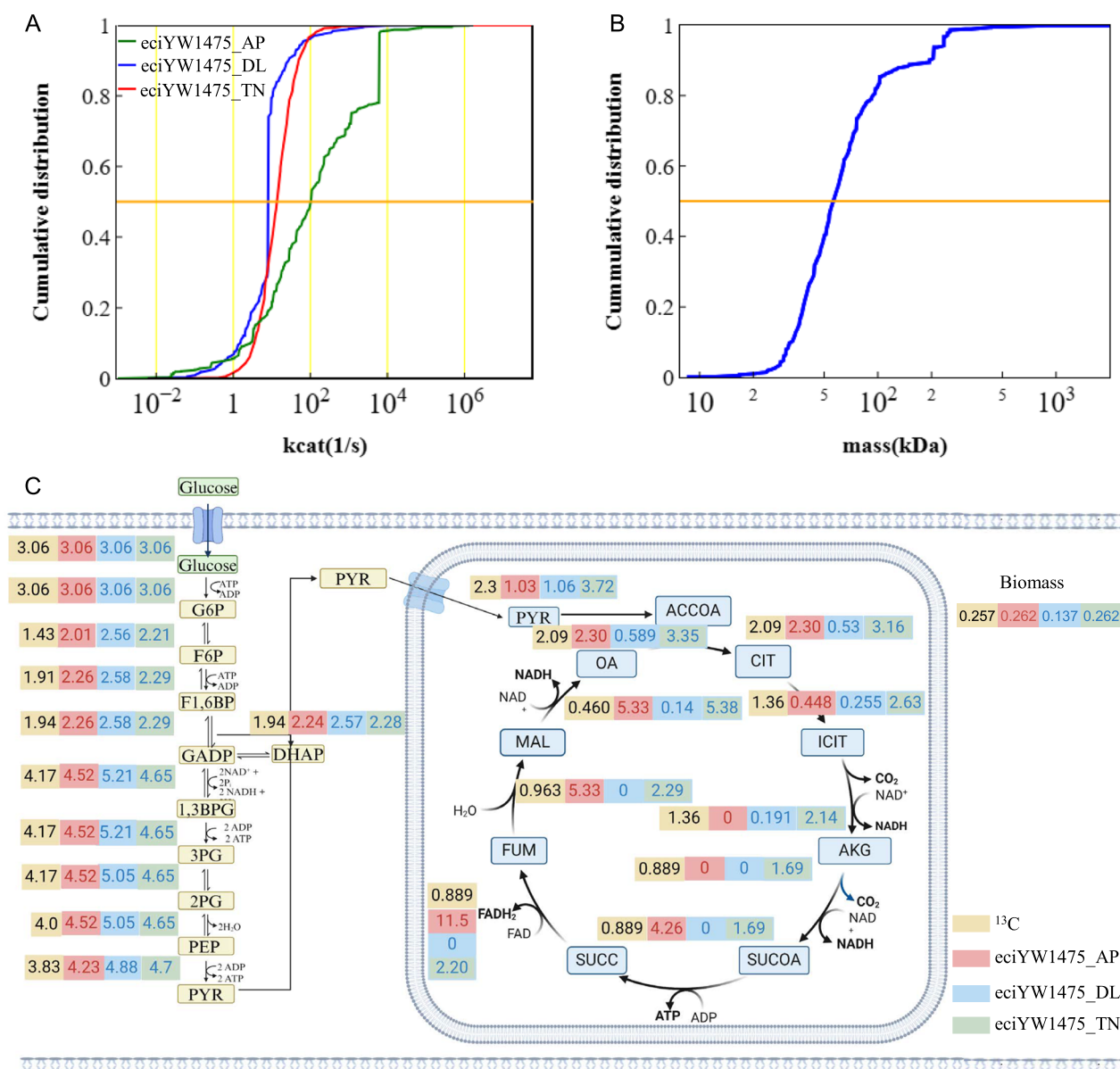


Fig. 2 Basic information of eciYW1475_AP, eciYW1475_DL and eciYW1475_TN. **A** Cumulative distribution of k_{cat} values for three ecGEMs. **B** Cumulative distribution of molecular weights. **C** Flux comparison of three ecGEMs

eciYW1475_AP (Fig. 2C, Additional file 8: Table S8). It was noted that some predicted fluxes for the TCA pathway were zero when the glucose uptake rate was 3.0676 mmol/gDW/h. Considering k_{cat} coverage and performance of these three ecGEMs, the revised eciYW1475_TN was then used as the final version of ecGEM for *M. thermophila*, named ecMTM. All subsequent simulations were carried out with ecMTM.

Introducing enzyme constraints reduced the solution space and improved prediction accuracy

When a GEM is given a specific objective, solving it through FBA often results in alternate optimal solutions, which constitutes a limitation of GEMs [43]. To address this challenge and narrow the range of flux variability, various strategies, including the incorporation of enzyme constraints, have been explored in previous studies [17, 19–21]. In this study, the cumulative flux variability

ranges between iYW1475 and ecMTM were compared using Flux Variability Analysis (FVA) with setting glucose uptake rate at 10 mmol/gDW/h. Results demonstrated that a substantial two-order reduction in the median flux variability range after introducing enzyme constraints. Furthermore, cumulative distribution indicated that there were 5% with totally variable flux (reactions that can carry any flux between -1000 and 1000 mmol/gDW/h) in iYW1475, while no such extreme variability range was observed in ecMTM (Fig. 3A, Additional file 9: Table S9).

In experimental condition, growth rate of *M. thermophila* is 0.2573 h^{-1} with maximum glucose uptake rate of $3.0676 \text{ mmol/gDW/h}$ when abundant oxygen and glucose are supplied [33]. The PhPP analysis of iYW1475 exhibited linear growth as a function of carbon source and oxygen uptake rates (Fig. 3B), which was inconsistent

with experimental observations. In contrast, ecMTM constrained the maximal growth rate with increasing carbon source availability (Fig. 3C), indicating a significant reduction in the solution space. All these results suggest that incorporating enzyme constraints into GEMs make the simulation results closer to realistic cellular phenotypes.

Simulating metabolic strategy adjustment

Enzyme-constrained models provide valuable insights into how cells adjust their metabolic pathways based on enzyme resources in response to increasing carbon source uptake. Previous studies have employed ecGEMs to simulate the metabolic adjustments in Yeast [24] and *E. coli* [17], elucidating overflow metabolism phenomena in these organisms. Simulation of metabolic strategy

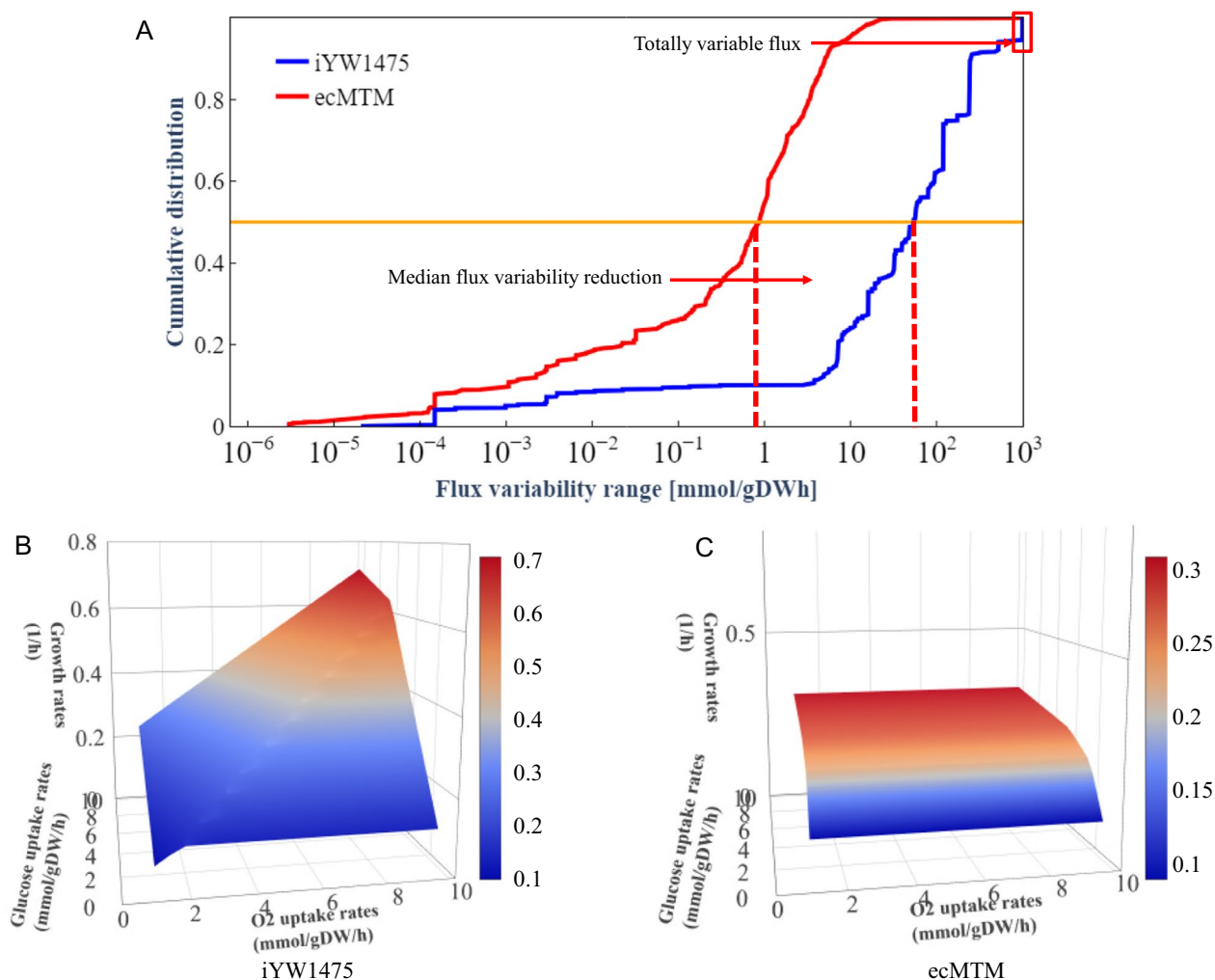


Fig. 3 Simulation of solution space and growth rate. **A** Cumulative distribution of flux variability at high growth rates for iYW1475 and ecMTM. Simulation of growth rates at different glucose and oxygen uptake rates using iYW1475 (**B**) and ecMTM (**C**)

adjustment were conducted with a glucose uptake rate ranging from 0 to 6 mmol/gDW/h.

The results illustrated that metabolic pathways were dynamically adjusted with an increase in glucose uptake rate. These adjustments can be divided into substrate-limited, metabolic adjustment stage, and metabolic overflow stage (Fig. 4A). In the first stage (glucose uptake rate less than 2.5 mmol/gDW/h), the growth rate exhibited a linear relationship with glucose uptake, aligning with the behavior observed in iYW1475 (Fig. 4A, B). In the second stage (glucose uptake rate 2.5 to 5 mmol/gDW/h), due to the enzyme resource constraint, growth rate was decreased compared with iYW1475 along with the gradual increase of substrate supply (Fig. 4B). Meanwhile, metabolic pathways were adjusted to these with higher enzyme efficiency and more carbon loss resulting in lower biomass yield (Fig. 4B, C). Moving on to the third stage, as substrate supply continued to increase, a shift towards a more enzymatically efficient ethanol production pathway was

observed. This shift resulted in an increase in ethanol production flux from 0 to 2.17 mmol/gDW/h. Overall, the results depicted a trade-off between enzyme efficiency and growth rate.

To investigate energy pathway regulation strategies during metabolic adjustments, energy synthesis enzyme costs and oxidative phosphorylation rates were calculated [20]. Cells showed a preference for energy-efficient respiratory oxidative phosphorylation at low growth rates, particularly in the substrate limited stage (Fig. 4D, Additional file 10: Table S10). However, as the glucose uptake rate increased during the metabolic adjustment stage, the ratio of oxidative phosphorylation with high enzyme consumption decreased, resulting in a decline of enzyme cost for energy synthesis (Fig. 4D, Additional file 10: Table S10). Due to enzyme limitation, more energy was produced through the glycolysis pathway. The regulation of the energy pathway indicated that enzyme limitation is a major driving

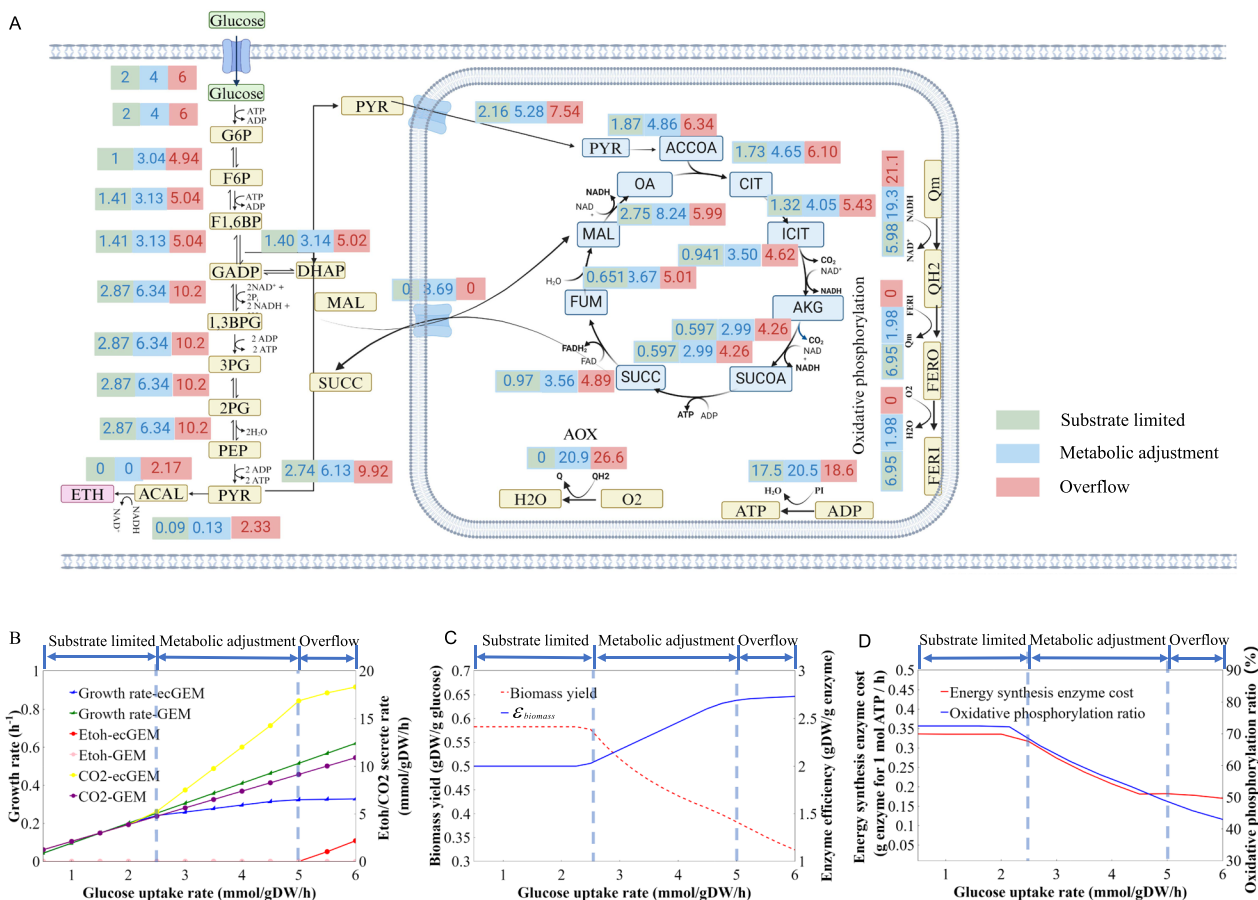


Fig. 4 Simulation of metabolic adjustments. **A** Metabolic flux map of the three stages. **B** Comparison of metabolic adjustments between iYW1475 and ecMTM. **C** Trade-off phenomenon simulated by ecMTM. **D** Energy synthesis enzyme cost and energy production ratio of oxidative phosphorylation simulated by ecMTM

force for the reallocation of enzymatic proteins for energy production.

Simulation of substrate hierarchy utilization

As a highly efficient cellulose and hemicellulose degrader, *M. thermophila* exhibits proficiency in utilizing various oligosaccharides and monosaccharides derived from plant biomass hydrolysis. The primary degradation products include cellobiose, glucose, xylose, arabinose and galactose [45]. Investigating the preferential utilization of five substrates in *M. thermophila*, essential for environmental adaptation, was performed by simulating hierarchical substrate utilization using ecMTM from the perspective of enzyme efficiency (substrate carbon atom uptake was scaled to 15 mmol/gDW/h). According to the assessment principles [44], if the enzyme efficiency for synthesizing biomass precursors from one substrate is higher than that from another substrate, the former is sequentially consumed before the latter; otherwise, they are deemed to be simultaneously consumed (co-utilization).

There are mainly two pathways for cleaving cellobiose in *M. thermophila*, namely, hydrolytic pathway and phosphorolytic pathway [35] (Fig. 5A). In the hydrolytic pathway, β -glucosidase catalyzes one cellobiose molecule into two molecules glucose, while in the phosphorolytic pathway, cellobiose phosphorylase converts one cellobiose molecule into one glucose molecule and one glucose-1-phosphate (G1P) molecule. Glucose generated in these pathways is further catalyzed into G6P with consuming ATP whereas G1P, produced in the phosphorolytic pathway, is catalyzed into G6P by phosphoglucomutase without ATP consumption. In comparison to the hydrolytic pathway, 11 biomass precursors, excluding PYR, are synthesized from one cellobiose molecule through the phosphorolytic pathway, which consumes one less molecule of ATP. This reduction in ATP usage saves enzyme resources by 12.5% (in the case of G6P), as ATP production involves large amounts of enzyme resource through oxidative phosphorylation (Additional file 11: Table S11). As a result, there is a higher enzyme efficiency in synthesizing 11 biomass precursors from cellobiose compared with glucose (Fig. 5B). However, the hydrolytic pathway is chosen for pyruvate synthesis from cellobiose since ATP production and consumption are balanced through this pathway. The cellobiose hydrolytic pathway requires an additional β -glucosidase, accounting for 4.5% of the total enzyme resources, leading to lower enzyme efficiency in PYR synthesis compared with glucose. Meanwhile, enzyme efficiencies for synthesizing all the 12 biomass precursors with cellobiose and glucose are higher than those with xylose, galactose and arabinose. Therefore,

glucose and cellobiose are first co-utilized showing a preferential utilization for the other three sugars.

Concerning galactose and xylose, the synthesis enzyme efficiency with the former is higher than that of the latter for 10 biomass precursors and is lower for 2, indicating a co-utilization relationship between the two sugars (Fig. 5B). Additionally, all enzyme efficiencies for the 12 precursors with xylose and galactose are higher than that of arabinose, resulting in arabinose being utilized last. Compared with xylose, there are two additional enzyme reactions for biomass precursor synthesis using arabinose which constitute 4.7% (in the case of G6P) of the total enzyme resources, resulting in lower enzyme efficiency (Additional file 11: Table S11). Overall, these simulation results are consistent with experimental measurements illustrated in Fig. 5C. The quantitative simulation elucidates the hierarchical utilization of different carbon sources derived from plant biomass in this fungus from the perspective of protein resource allocation.

Exploration of metabolic engineering targets based on enzyme cost

GEMs function as valuable predictive tools for identifying potential targets in metabolic engineering. In the case of *B. subtilis*, an ecGEM was applied to identify novel gene deletion targets, resulting in an increased yield of poly- γ -glutamic acid [18]. Based on enzyme costs, two methods (outlined in Method section) were employed to predict potential targets for three products: ethanol, malate and fumarate. The biosynthetic pathways for these three products were illustrated in Fig. 6A. In total, 12 out of 32 previously overexpressing targets were predicted by ecMTM using the first method as depicted in Fig. 6B. Modifying these 12 targets showed notable improvements in the production of target chemicals [4, 5, 8, 9].

Target prediction method 2 presents a modification strategy that includes both enhancing and weakening targets (Additional file 12: Table S12). Six enhancing targets for ethanol were successfully predicted and verified by overexpression target genes [8, 9]. Among these predicted targets, both *pdh* (pyruvate decarboxylase encoding gene) and *adh* (alcohol dehydrogenase encoding gene) were identified by these two methods, representing the final crucial steps in ethanol synthesis. Overexpressing PDC, which decarboxylates pyruvate to acetaldehyde, and ADH, which reduces acetaldehyde to ethanol, can directly enhance the ethanol synthesis capability. This strategy has been supported by previous studies [8, 9]. Additionally, two new enhancing targets, *eno* and *gapdh*, were predicted but have not been modified yet.

For malate, four enhancing targets were identified, including *pyc* (pyruvate carboxylase encoding gene), *mdh* (malate dehydrogenase encoding gene), and *hxt*

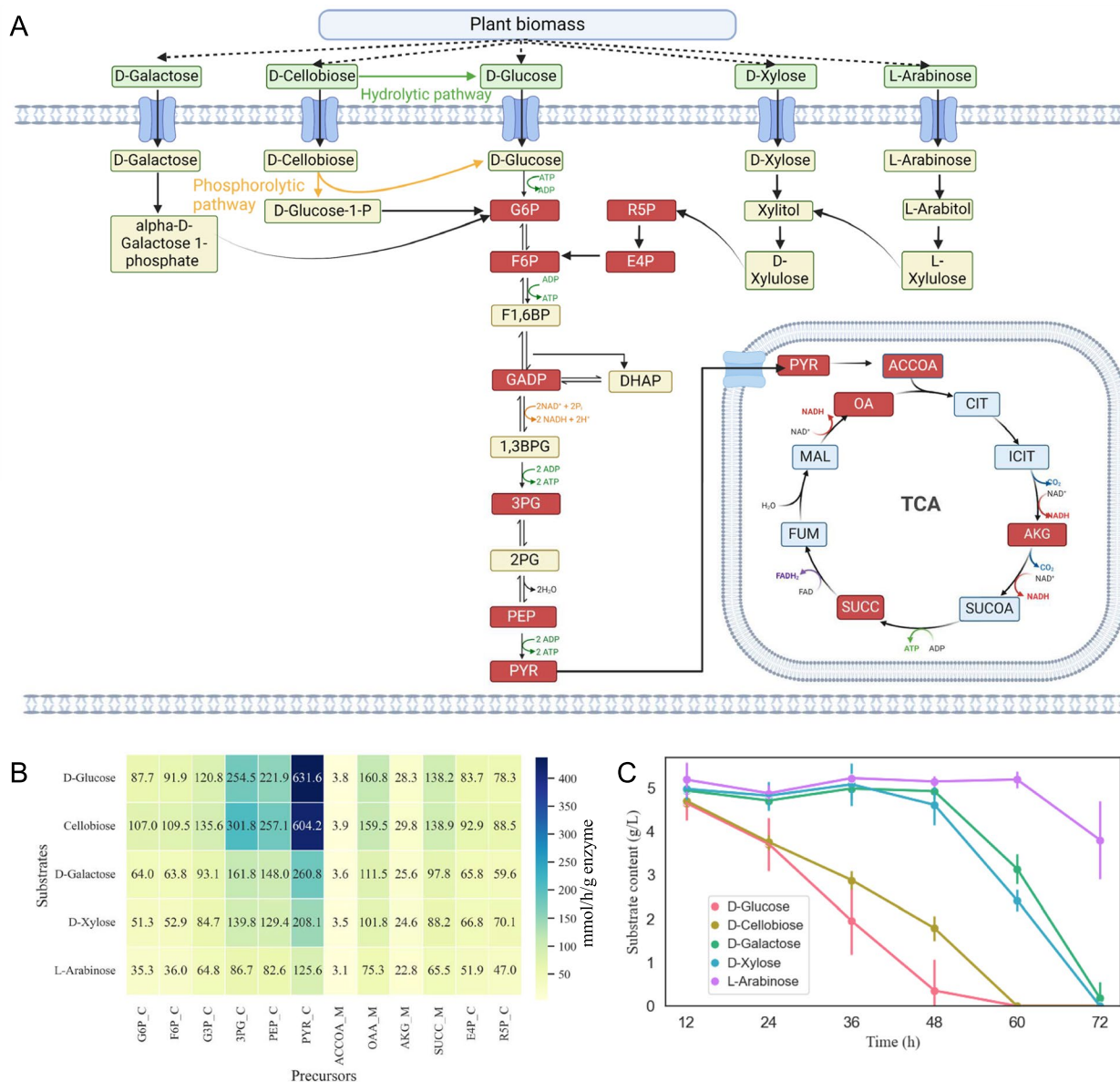


Fig. 5 Substrate hierarchy utilization simulation. **A** Substrate utilization pathways for five carbon resources. Metabolites with red background represent 12 biomass precursors. **B** Simulation of substrate hierarchy utilization. Numbers represent enzyme efficiency. **C** Measured data of substrate utilization

(hexose transporter encoding gene), predicted by both methods. In microbes, PYC catalyzes the conversion of pyruvic acid to oxaloacetic acid with the fixation of CO₂; the oxaloacetic acid is then converted into malate by MDH. PYC and MDH are two key steps in produce of malate through the reductive tricarboxylic acid (rTCA) pathway in cytoplasm [5]. Due to limited experiments on fumarate, only two predicted targets by these two methods were verified, leaving the remaining

targets serve as potential objectives for fumarate modification.

Notably, among the predicted targets, *hxt* and *eno* were shared by these three chemicals. HXT is responsible for transporting glucose, a key factor in increasing glucose uptake. This has been confirmed by studies showing that overexpressing HXT can elevate ethanol and malate yield [5, 8, 9]. As a vital step of glycolysis, ENO is regarded as a key enzyme due to its low *k_{cat}*. Therefore, *eno* is a potential target to increase precursor metabolite flux

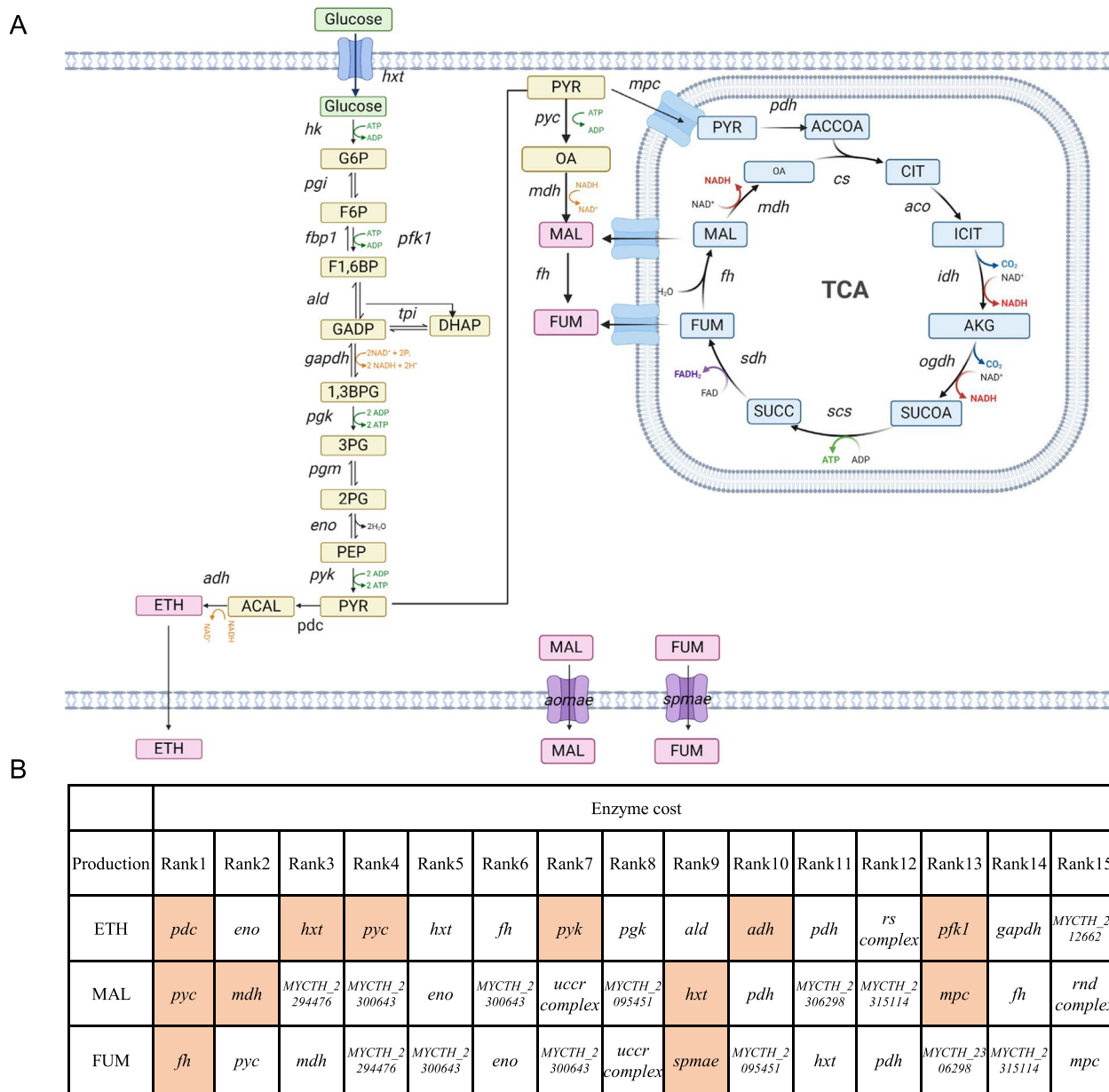


Fig. 6 Metabolic engineering targets prediction for ethanol, malate and fumarate in *M. thermophila*. **A** Depicts the biosynthetic pathways for ethanol, malate and fumarate in *M. thermophila*. **B** Metabolic engineering targets for three products predicted using ecMTM. The targets reported in the literature are highlighted in orange. The reactions *uccr* (Ubiquinol-cytochrome c reductase complex), *rs* (RNA synthesis complex), and *rnd* (Respiratory-chain NADH dehydrogenase complex) are involved multiple genes

from glycolysis for enhancing product synthesis, which is worth trying in future.

In the results predicted by method 2, putative oxaloacetate transporter was identified as a weakening target which contributes to the enhancement of malate and fumarate production. Knocking out this target in *Aspergillus carbonarius* has been shown to boost malate

production [46], making it a prospective modification target for improving malate synthesis. In addition, most of predicted weakening targets for these three products were focused on the respiratory chain, suggesting that modifying these reactions may allocate more enzyme resources for product synthesis.

Discussion

In this research, an enzyme-constrained model ecMTM was constructed based on iDL1450 under the ECMpy framework. Enzyme turnover number k_{cat} is a fundamental parameter in the quantitative study of enzyme activity and is crucial for understanding cellular metabolism, physiology, and resource allocation [15–21]. Several databases such as BRENDA, UniProt, and SABIO-RK offer enzyme parameter queries, while automated pipelines like AutoPACMEN facilitate the automated acquisition of k_{cat} data for ecGEMs. Besides, k_{cat} values can be predicted through methods based on machine learning algorithms, such as DLKcat and TurNuP. During construction of ecMTM, three versions of ecGEMs were developed using k_{cat} values obtained through AutoPACMEN, DLKcat and TurNuP. After comparison, ecGEM constructed using TurNuP-predicted data performed better in several aspects and was selected as the final version of ecGEM for *M. thermophila*. Nowadays, the machine learning based methods has gradually used in ecGEMs with improved performance [28, 47].

The constructed ecMTM demonstrated improved accuracy in simulating strain behavior compared to iYW1475. Incorporating enzyme constraints into iYW1475 notably reduced the solution space, yielding predictions closer to experimental observations. However, there are still some disagreements that could not explain by ecMTM. In particular, metabolic strategy adjustment simulations within the model revealed an ethanol overflow, as was observed in yeast ecGEM [24]. Although ethanol overflow is important for biofuel production from plant biomass, it is not observed in *M. thermophila* wild type strain. These discrepancies suggest that factors beyond enzyme constraints influence cell phenotype. Nonetheless, metabolic adjustment simulation provided insight into potential strategies to increase ethanol flux.

Due to enzyme resource limitations and quantified enzyme efficiencies, ecMTM could predict hierarchy substrate utilization in a quantitative perspective. *M. thermophila* typically found in low-carbon-content soil environments requires lignocellulolytic enzymes for lignocellulose degradation [48]. In this niche, the utilization of carbon sources with low cost of enzyme is important for the survival of *M. thermophila*. Hierarchy substrate utilization was determined in liquid shake flask and simulated with ecMTM. The simulation unveiled the enzyme-cost-driven hierarchy in substrate utilization by *M. thermophila*. In previously research, Ramkrishna et al. [49] developed a cybernetic model to simulate mixed-substrate growth dynamics of *E. coli*. This model utilizes dynamic equations to describe the rates of individual reactions in metabolic pathways, taking into

account substrate concentration, enzyme concentration and enzyme activity. It requires a large amount of experimental data to fit model parameters. On the other hand, Enzyme efficiency method determines the hierarchical utilization of carbon sources by calculating the enzyme efficiency for biomass precursors without relying on complex dynamic equations or experimental data. Nevertheless, it ignores the dynamic regulation processes in metabolic pathways. Therefore, the cybernetic model is better suitable for studying the dynamic regulation processes of metabolic pathways in biological systems, while the enzyme efficiency method is more suitable for predicting substrate hierarchy utilization.

Based on enzyme costs, metabolic engineering targets were predicted using ecMTM. Notably, the first prediction method revealed several shared targets across the three products, including *hxt* and *eno*. Similarly, the second prediction method predominantly highlighted targets within the glycolysis pathway to enhance these products. These predictions underscore the pivotal role of glucose transport and the rapid glycolysis flux in driving the production of these specific products.

After integration of enzyme constraints, the predictive capabilities of GEM were significantly enhanced. As we all know, biological systems are extremely complex with various constraints and hierarchical regulations. Enzyme constraints, though influential, might not comprehensively capture the intricacies of these systems. Furthermore, *M. thermophila* exhibits special phenotypes, such as thermophily and excellent ability of cellulose degradation. To elucidate these characteristics, it may be necessary to introduce additional constraints, such as temperature constraints, or development of comprehensive metabolic network models like ME models [50] and etcGEM models [51].

Conclusion

In this study, an enzyme-constrained genome-scale metabolic model ecMTM was constructed for *M. thermophila* within ECMpy framework using machine learning-based k_{cat} data predicted by TurNuP. Simulation results demonstrated that enzyme constraints could reduce the model's solution space, leading to more realistic pathway predictions. The ecMTM successfully predicted and elucidated the hierarchical utilization of five carbon sources derived from lignocellulose. Moreover, key enzymes were identified for three chemicals, providing rational metabolic engineering modifications to increase the yield of valuable compounds.

Abbreviations

GEMs	Genome-scale metabolic models
ecGEMs	Enzyme-constrained genome-scale metabolic models
GPR	Gene-protein-reaction

CBP	Consolidated bioprocessing
GMM	Vogel's minimal medium supplemented with 2% glucose
TCA	Tricarboxylic acid cycle
rTCA	Reductive tricarboxylic acid cycle
PhPP	Phenotype phase plane
pFBA	Flux balance analysis
FBA	Flux balance analysis
FVA	Flux variability analysis
G6P	Glucose 6-phosphate
F6P	Fructose 6-phosphate
G3P	Glyceraldehyde 3-phosphate
3PG	Glycerate 3-phosphate
PEP	Phosphoenolpyruvate
PYR	Pyruvate
ACCOA	Acetyl-CoA
OAA	Oxaloacetate
AKG	2-Oxoglutarate
SUCC	Succinate
E4P	Erythrose 4-phosphate
R5P	Ribose 5-phosphate
G1P	Glucose-1-phosphate
DW	Dry cell weight

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12934-024-02415-z>.

Additional file 1: Table S1. Biomass composition.
 Additional file 2: Table S2. Metabolite consensus.
 Additional file 3: Table S3. Refined GPR associations.
 Additional file 4: Table S4. Metabolite ID mapping to BiGG ID.
 Additional file 5: Table S5. Subunit number information from UniProt.
 Additional file 6: Table S6. k_{cat} data obtained from three methods.
 Additional file 7: Table S7. Calibration of k_{cat} for three ecGEMs by two methods.
 Additional file 8: Table S8. Flux comparison of three ecGEMs.
 Additional file 9 Table S9. FVA of iYW1475 and ecMTM.
 Additional file 10: Table S10. ATP generation in three stages.
 Additional file 11 Table S11. Reaction fluxes and enzyme amounts for precursors synthesis of 12 biomass precursors from five carbon sources.
 Additional file 12 Table S12. Targets prediction by enzyme allocation method.

Acknowledgements

Not applicable.

Author contributions

Y. W. and Z. M. carried out the model reconstruction and validation. P. Z. contributed to substrate hierarchy utilization simulation. Y. W., Z. M. and D. L. wrote the manuscript. D. L., C. T. and H. M. participated in the design of the study. J. D. participated in data collection and analysis. Q. G. contributed to the analysis and discussion. All authors read and approved the final manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (2021YFC2100700), the Key Project of the Ministry of Science and Technology of China (2023YFC3403602 and 2018YFA0900500), the National Natural Science Foundation of China (32071424, 32270100 and 32300529), the Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project (TSBICIPKJGG-015 and TSBICIP-PTJJ-007-12), the Innovation fund of Haihe Laboratory of Synthetic Biology (22HHSWSS00014), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDC0110301), the project of Key Laboratory of Tobacco Processing of Zhengzhou Tobacco Research Institute of CNTC (202022AWCX02), the Key research and

development project of China National Tobacco Corporation (110202202004), Guangxi Science and Technology Major Program (Guike-AA22117013).

Data availability

The models and code used in this study can be found at: <https://github.com/wangtao-cell/ecMTM>.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no conflict of interest.

Author details

¹Key Laboratory of Industrial Fermentation Microbiology of the Ministry of Education, Tianjin Key Laboratory of Industrial Microbiology, College of Biotechnology, Tianjin University of Science and Technology, Tianjin 300457, China. ²Haihe Laboratory of Synthetic Biology, Tianjin 300308, China. ³Key Laboratory of Engineering Biology for Low-Carbon Manufacturing, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China. ⁴Biodesign Center, Key Laboratory of Engineering Biology for Low-carbon Manufacturing, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China. ⁵National Technology Innovation Center of Synthetic Biology, Tianjin 300308, China.

Received: 4 February 2024 Accepted: 4 May 2024

Published online: 15 May 2024

References

1. Maheshwari R, Bharadwaj G, Bhat MK. Thermophilic fungi: their physiology and enzymes. *Microbiol Mol Biol Rev.* 2000;64:461–88. <https://doi.org/10.1128/mmr.64.3.461-488.2000>.
2. Singh B. *Myceliophthora thermophila* syn Sporotrichum thermophile: a thermophilic mould of biotechnological potential. *Crit Rev Biotechnol.* 2016. <https://doi.org/10.3109/07388551.2014.923985>.
3. Zhu Z, Zhang M, Liu D, Liu D, Sun T, Yang Y, Dong J, Zhai H, Sun W, Liu Q, Tian C. Development of the thermophilic fungus *Myceliophthora thermophila* into glucoamylase hyperproduction system via the metabolic engineering using improved AsCas12a variants. *Microb Cell Fact.* 2023;22:150. <https://doi.org/10.1186/s12934-023-02149-4>.
4. Gu S, Li J, Chen B, Sun T, Liu Q, Xiao D, Tian C. Metabolic engineering of the thermophilic filamentous fungus *Myceliophthora thermophila* to produce fumaric acid. *Biotechnol Biofuels.* 2018;11:323. <https://doi.org/10.1186/s13068-018-1319-1>.
5. Li J, Lin L, Sun T, Xu J, Ji J, Liu Q, Tian C. Direct production of commodity chemicals from lignocellulose using *Myceliophthora thermophila*. *Metab Eng.* 2020;61:416–26. <https://doi.org/10.1016/j.ymben.2019.05.007>.
6. Gu S, Zhao Z, Yao Y, Li J, Tian C. Designing and constructing a novel artificial pathway for malonic acid production biologically. *Front Bioeng Biotechnol.* 2021;9: 820507. <https://doi.org/10.3389/fbioe.2021.820507>.
7. Liu D, Zhang Y, Li J, Sun W, Yao Y, Tian C. The Weimberg pathway: an alternative for *Myceliophthora thermophila* to utilize D-xylose. *Biotechnol Biofuels Bioprod.* 2023;16:13. <https://doi.org/10.1186/s13068-023-02266-7>.
8. Li J, Zhang Y, Li J, Sun T, Tian C. Metabolic engineering of the cellulolytic thermophilic fungus *Myceliophthora thermophila* to produce ethanol from cellobiose. *Biotechnol Biofuels.* 2020;13:23. <https://doi.org/10.1186/s13068-020-1661-y>.
9. Zhang Y, Sun T, Wu T, Li J, Hu D, Liu D, Li J, Tian C. Consolidated bioprocessing for bioethanol production by metabolically engineered cellulolytic fungus *Myceliophthora thermophila*. *Metab Eng.* 2023;78:192–9. <https://doi.org/10.1016/j.ymben.2023.06.009>.
10. Li J, Chen B, Gu S, Zhao Z, Liu Q, Sun T, Zhang Y, Wu T, Liu D, Sun W, Tian C. Coordination of consolidated bioprocessing technology and carbon dioxide fixation to produce malic acid directly from plant biomass in *Myceliophthora thermophila*. *Biotechnol Biofuels.* 2021;14:186. <https://doi.org/10.1186/s13068-021-02042-5>.

11. Fondi M, Lio P. Genome-scale metabolic network reconstruction. *Methods Mol Biol.* 2015;1231:233–56. https://doi.org/10.1007/978-1-4939-1720-4_15.
12. Gu C, Kim GB, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. *Genome Biol.* 2019;20:121. <https://doi.org/10.1186/s13059-019-1730-3>.
13. Succurro A, Ebenhoh O. Review and perspective on mathematical modeling of microbial ecosystems. *Biochem Soc Trans.* 2018;46:403–12. <https://doi.org/10.1042/BST20170265>.
14. Liu D, Xu Z, Li J, Liu Q, Yuan Q, Guo Y, Ma H, Tian C. Reconstruction and analysis of genome-scale metabolic model for thermophilic fungus *Myceliophthora thermophila*. *Biotechnol Bioeng.* 2022;119:1926–37. <https://doi.org/10.1002/bit.28080>.
15. Arend M, Zimmer D, Xu R, Sommer F, Mühlhaus T, Nikoloski Z. Proteomics and constraint-based modelling reveal enzyme kinetic properties of *Chlamydomonas reinhardtii* on a genome scale. *Nat Commun.* 2023;14:4781. <https://doi.org/10.1038/s41467-023-40498-1>.
16. Caivano A, van Winden W, Dragone G, Mussatto SI. Enzyme-constrained metabolic model and in silico metabolic engineering of *Clostridium ljungdahlii* for the development of sustainable production processes. *Comput Struct Biotechnol J.* 2023;21:4634–46. <https://doi.org/10.1016/j.csbj.2023.09.015>.
17. Mao Z, Zhao X, Yang X, Zhang P, Du J, Yuan Q, Ma H. ECMpy, a Simplified workflow for constructing enzymatic constrained metabolic network model. *Biomolecules.* 2022;12:65. <https://doi.org/10.3390/biom12010065>.
18. Massaiu I, Pasotti L, Sonnenschein N, Rama E, Cavaletti M, Magni P, Calvio C, Herrgard MJ. Integration of enzymatic data in *Bacillus subtilis* genome-scale metabolic model improves phenotype predictions and enables in silico design of poly-gamma-glutamic acid production strains. *Microb Cell Fact.* 2019;18:3. <https://doi.org/10.1186/s12934-018-1052-2>.
19. Niu J, Mao Z, Mao Y, Wu K, Shi Z, Yuan Q, Cai J, Ma H. Construction and analysis of an enzyme-constrained metabolic model of *Corynebacterium glutamicum*. *Biomolecules.* 2022;12:1499. <https://doi.org/10.3390/biom12101499>.
20. Wu K, Mao Z, Mao Y, Niu J, Cai J, Yuan Q, Yun L, Liao X, Wang Z, Ma H. ecBSU1: a genome-scale enzyme-constrained model of *Bacillus subtilis* based on the ECMpy workflow. *Microorganisms.* 2023;11:178. <https://doi.org/10.3390/microorganisms11010178>.
21. Zhou J, Zhuang Y, Xia J. Integration of enzyme constraints in a genome-scale metabolic model of *Aspergillus niger* improves phenotype predictions. *Microb Cell Fact.* 2021;20:125. <https://doi.org/10.1186/s12934-021-01614-2>.
22. Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabási AL, Oltvai ZN. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc Natl Acad Sci USA.* 2007;104:12663–8. <https://doi.org/10.1073/pnas.0609845104>.
23. Adadi R, Volkmer B, Milo R, Heinemann M, Shlomi T. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. *PLoS Comput Biol.* 2012;8: e1002575. <https://doi.org/10.1371/journal.pcbi.1002575>.
24. Sánchez BJ, Zhang C, Nilsson A, Lahtvee PJ, Kerkhoven EJ, Nielsen J. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol.* 2017. <https://doi.org/10.15252/msb.20167411>.
25. Bekiaris PS, Klamt S. Automatic construction of metabolic models with enzyme constraints. *BMC Bioinform.* 2020;21:19. <https://doi.org/10.1186/s12859-019-3329-9>.
26. Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* 2019;47:D542–9. <https://doi.org/10.1093/nar/gky1048>.
27. Wittig U, Rey M, Weidemann A, Kania R, Müller W. SABIO-RK: an updated resource for manually curated biochemical reaction kinetics. *Nucleic Acids Res.* 2018;46:D656–60. <https://doi.org/10.1093/nar/gkx1065>.
28. Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MKM, Kerkhoven EJ, Nielsen J. Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nat Catal.* 2022;5:662–72. <https://doi.org/10.1038/s41929-022-00798-z>.
29. Kroll A, Rousset Y, Hu X-P, Liebrand NA, Lercher MJ. Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning. *Nat Commun.* 2023;14:4139. <https://doi.org/10.1038/s41467-023-39840-4>.
30. Castillo S, Barth D, Arvas M, Pakula TM, Pitkanen E, Blomberg P, Seppanen-Laakso T, Nygren H, Sivasiddharthan D, Penttila M, Oja M. Whole-genome metabolic model of *Trichoderma reesei* built by comparative reconstruction. *Biotechnol Biofuels.* 2016;9:252. <https://doi.org/10.1186/s13068-016-0665-0>.
31. Carnicer M, Baumann K, Toplit I, Sanchez-Ferrando F, Mattanovich D, Ferrer P, Albiol J. Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. *Microb Cell Fact.* 2009;8:65. <https://doi.org/10.1186/1475-2859-8-65>.
32. Raeder U, Broda P. Rapid preparation of DNA from filamentous fungi. *Lett Appl Microbiol.* 1985;1:17–20. <https://doi.org/10.1111/j.1472-765X.1985.tb01479.x>.
33. Jiang J, Liu D, Tian C, Xia J. Elucidation of the metabolic mechanism for malate production in *Myceliophthora thermophila* via ¹³C metabolic flux analysis. *Res Sq.* 2022. <https://doi.org/10.21203/rs.3.rs-2123109>.
34. Iantsevich EA, Danilova OA, Groza NV, Kotlova ER, Tereshina VM. Heat shock response of thermophilic fungi: membrane lipids and soluble carbohydrates under elevated temperatures. *Microbiology.* 2016;162:989–99. <https://doi.org/10.1099/mic.0.000279>.
35. Li J, Gu S, Zhao Z, Chen B, Liu Q, Sun T, Sun W, Tian C. Dissecting cellobiose metabolic pathway and its application in biorefinery through consolidated bioprocessing in *Myceliophthora thermophila*. *Fungal Biol Biotechnol.* 2019;6:21. <https://doi.org/10.1186/s40694-019-0083-8>.
36. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 2021;49:D545–51. <https://doi.org/10.1093/nar/gkaa970>.
37. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE. BIGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* 2016;44:D515–522. <https://doi.org/10.1093/nar/gkv1049>.
38. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* 2016;44:D1214–1219. <https://doi.org/10.1093/nar/gkv1031>.
39. Consortium U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49:D480–9. <https://doi.org/10.1093/nar/gkaa1100>.
40. Edwards JS, Ramakrishna R, Palsson BO. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol Bioeng.* 2002;77:27–36. <https://doi.org/10.1002/bit.10047>.
41. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRAPy: Constraints-based reconstruction and analysis for python. *BMC Syst Biol.* 2013;7:74. <https://doi.org/10.1186/1752-0509-7-74>.
42. Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. *Curr Opin Biotechnol.* 2003;14:491–6. <https://doi.org/10.1016/j.copbio.2003.08.001>.
43. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng.* 2003;5:264–76. <https://doi.org/10.1016/j.ymben.2003.09.002>.
44. Wang X, Xia K, Yang X, Tang C. Growth strategy of microbes on mixed carbon sources. *Nat Commun.* 2019;10:1279. <https://doi.org/10.1038/s41467-019-09261-3>.
45. Béguin P, Aubert JP. The biological degradation of cellulose. *FEMS Microbiol Rev.* 1994;13:25–58. <https://doi.org/10.1111/j.1574-6976.1994.tb00033.x>.
46. Yang L, Linde T, Hossain AH, Lübeck M, Punt PJ, Lübeck PS. Disruption of a putative mitochondrial oxaloacetate shuttle protein in *Aspergillus carbonarius* results in secretion of malic acid at the expense of citric acid production. *BMC Biotechnol.* 2019;19:72. <https://doi.org/10.1186/s12896-019-0572-0>.
47. Chen Y, Nielsen J. Mathematical modeling of proteome constraints within metabolism. *Curr Opin Syst Biol.* 2021;25:50–6. <https://doi.org/10.1016/j.coisb.2021.03.003>.
48. Bhat KM, Maheshwari R. *Sporotrichum thermophile* growth, cellulose degradation, and cellulase activity. *Appl Environ Microbiol.* 1987;53:2175–82. <https://doi.org/10.1128/aem.53.9.2175-2182.1987>.
49. Ramakrishna R, Ramkrishna D, Konopka AE. Cybernetic modeling of growth in mixed, substitutable substrate environments: preferential and

simultaneous utilization. *Biotechnol Bioeng.* 1996;52:141–51. [https://doi.org/10.1002/\(SICI\)1097-0290\(19961005\)52:1%3c141::AID-BIT14%3e3.0.CO;2-R](https://doi.org/10.1002/(SICI)1097-0290(19961005)52:1%3c141::AID-BIT14%3e3.0.CO;2-R).

50. Lloyd CJ, Ebrahim A, Yang L, King ZA, Catoiu E, O'Brien EJ, Liu JK, Palsson BO. COBRAme: A computational framework for genome-scale models of metabolism and gene expression. *PLoS Comput Biol.* 2018;14: e1006302. <https://doi.org/10.1371/journal.pcbi.1006302>.
51. Li G, Hu Y, Jan Z, Luo H, Wang H, Zelezniak A, Ji B, Nielsen J. Bayesian genome scale modelling identifies thermal determinants of yeast metabolism. *Nat Commun.* 2021;12:190. <https://doi.org/10.1038/s41467-020-20338-2>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.