Advances in genome-scale metabolic models of industrially important fungi

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Advances in genome-scale metabolic models of industrially important fungi

Yichao Han[1,2], Albert Tafur Rangel[3,4], Kyle R Pomraning[1,2], Eduard J Kerkhoven[3,4,5] and Joonhoon Kim[1,2,6]

Many fungal species have been used industrially for production of biofuels and bioproducts. Developing strains with better performance in biomanufacturing contexts requires a systematic understanding of cellular metabolism. Genome-scale metabolic models (GEMs) offer a comprehensive view of interconnected pathways and a mathematical framework for downstream analysis. Recently, GEMs have been developed or updated for several industrially important fungi. Some of them incorporate enzyme constraints, enabling improved predictions of cell states and proteome allocation. Here, we provide an overview of these newly developed GEMs and computational methods that facilitate construction of enzyme-constrained GEMs and utilize flux predictions from GEMs. Furthermore, we highlight the pivotal roles of these GEMs in iterative design–build–test–learn cycles, ultimately advancing the field of fungal biomanufacturing.

Addresses
[1] Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, USA
[2] Agile BioFoundry, Department of Energy, Emeryville, CA, USA
[3] Department of Life Sciences, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden
[4] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
[5] SciLifeLab, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden
[6] Joint BioEnergy Institute, Department of Energy, Emeryville, CA, USA

Corresponding author: Kim, Joonhoon (joonhoon.kim@pnnl.gov)

Current Opinion in Biotechnology 2023, **84**:103005

This review comes from a themed issue on **Energy Biotechnology**

Edited by **Thomas Eng**, **Michelle Omalley** and **Sudeep Agarwala**

For complete overview of the section, please refer to the article collection, "Energy Biotechnology (2024)"

Available online 3 October 2023

https://doi.org/10.1016/j.copbio.2023.103005

Introduction

Fungi possess exceptional metabolic versatility, robustness, and secretory capacity, which offer the potential to sustainably produce biofuels and bioproducts [1]. Their unmatched biosynthetic capacity has derived a wide range of products, which include not only bread, beer, and wine in daily foods and beverages [2], but also organic acids, proteins, and secondary metabolites with broad-reaching applications [3]. Fungal biomanufacturing can promote the transition from our petroleum-based economy to a bio-based circular economy [3].

Understanding metabolism in fungi can enable their potential biosynthetic capacity and enhance bioproduction via rational design. Genome-scale metabolic models (GEMs) provide a holistic view of all interconnected pathways in an organism. Moreover, GEMs enable prediction of phenotypic steady states via constraint-based modeling methods. Different types of omics data can be integrated into GEMs to improve prediction via more constraints or estimate metabolic kinetics.

In this review, we summarize advances in industrially relevant fungal GEMs and computational methods developed associated with GEMs. We then highlight their recent applications in iterative design–build–test–learn (DBTL) cycles to improve bioproduction with a focus on yeasts and filamentous fungi. Finally, we discuss challenges and future perspectives on computational methods that have not yet been tested in fungi and combined use of GEMs and machine learning (ML) approaches.

Recent genome-scale metabolic model development in industrially relevant fungi

Many fungal GEMs have been developed and updated in recent years. Here, we focus on industrially relevant fungi and classify them based on the group of molecules they typically overproduce (Table 1). More information about these GEMs can be found in Supplementary Table S1.

Oleaginous yeasts accumulate lipids or triacylglycerides to more than 20% of their cell dry weight and are promising microbial hosts for oleochemical production. GEMs are available for some oleaginous yeasts, including *Yarrowia lipolytica*, *Rhodotorula toruloides*, *Papiliotrema laurentii*, and *Cutaneotrichosporon oleaginosus*. Since 2016, several GEMs have been reconstructed and updated for *Y. lipolytica* for different purposes. Among these, iYali4, the fourth published GEM of *Y. lipolytica*,

Table 1

Recent developments and updates in genome-scale models of industrially important fungi.

Organism	Strains	GEMs and references	MEMOTE scores[a]	Notes
S. cerevisiae	S288C	Yeast8 [60] ecYeastGEM [42]	68%	Model organism
Y. lipolytica	W29	iYali4 [4] eciYali [42]	47%	Regulation of amino acid metabolism
	CLIB122	iYL_2.0 [6]	N/A	Strain design for triacylglycerol production
	CLIB122	iYLI647 [7]	25%	Strain design for dicarboxylic acid production
	W29	iYli21 [5]	17%	Improved prediction of growth capabilities
R. toruloides	NP11	Rhto-GEM [8] ecRhtoGEM [61]	66%	Detailed representation of lipid metabolism
	IFO0880	iRhto1108 [9]	87%	Strain design for triacylglycerol production
	IFO0880	RT_IFO0880 [10]	93%	Lignocellulosic carbon source utilization
P. laurentii	UFV-1	papla-GEM [11]	48%	Strain design for triacylglycerol production
C. oleaginosus	ATCC 20509	iNP636 [12]	19%	Simulation of optimal lipid production condition
A. niger	CBS 513.88	iHL1210 [14]	N/A	Validation with ^{13}C metabolic flux data
	ATCC 1015 and CBS 513.88	iJB1325 [15] eciJB1325 [62]	29%	Validation against extensive experimental data
	ATCC 1015	iDU1756 [16]	22%	Strain design for organic acid production
I. orientalis	SD108	iIsor850 [17]	84%	Strain design for organic acid production
Alternaria sp.	MG1	iYL1539 [21]	N/A	Media and strain design for resveratrol production
P. rubens	Wisconsin 54-1255	Penicillium-GEMs [19]	22%	Reconstruction of 24 Penicillium models
	Wisconsin 54-1255	iPrub22 [63]	74%	Prediction of specialized metabolite production
P. pastoris	GS115 and DSMZ 70382	iMT1026 [22]	27%	Consensus model of three previous models
	GS115	iRY1243 [24]	N/A	Improved prediction of growth capabilities
	GS115 and DSMZ 70382	iMT1026v3 [23]	19%	Improved prediction of growth on methanol or glycerol
O. polymorpha	NCYC 495	iUL909 [25]	42%	Strain design for organic acid production

[a] MEMOTE benchmarks metabolic models and reports an overall score from standardized tests for annotation, formal correctness, biomass reaction, and stoichiometry.

was derived from the consensus GEM of *Saccharomyces cerevisiae* to analyze the regulation of amino acid metabolism during lipid accumulation [4]. Using iYali4 as a scaffold, the GEM iYL21 specific for strain W29 was developed with experimental data curation, which well-predicted growth and gene essentiality [5•]. For another *Y. lipolytica* strain CLIB122, two GEMs, iYL_2.0 [6] and iYLI647 [7], were concurrently developed using different previous models as templates, and were respectively utilized for metabolic engineering of triacylglycerol and dicarboxylic acid. Three GEMs have been independently developed for *R. toruloides*: Rhto-GEM, iRhto1108, and RT_IFO0880. Among these three, Rhto-GEM had the most comprehensive representation of lipid metabolism for lipidomics data integration and predicted growth on glucose, xylose, and glycerol [8]. iRhto1108 had improved gene essentiality predictions and introduced context-specific biomass composition for carbon or nitrogen-limited conditions [9]. RT_IFO0880 described lignocellulosic carbon utilization pathways identified from multi-omics datasets and improved growth predictions on different nutrients using growth phenotyping and fitness data [10]. More recently, two GEMs (Papla-GEM [11] and iNP636 [12]) have been respectively developed for other oleaginous yeasts, *P. laurentii* and *C. oleaginous*, which can grow on industrially relevant feedstocks. These models largely enhanced our understanding of lipid production in oleaginous yeasts.

Acid-tolerant filamentous fungi (*Aspergillus niger*) and yeast (*Issatchenkia orientalis* also known as *Pichia kudriavzevii*) are ideal hosts for industrial organic acid production. To date, four GEMs have been developed for *A. niger* [13–16]. The first *A. niger* GEM iMA871 [13] has served as a scaffold to construct other high-quality GEMs. Updates to this model were produced by independent groups taking advantage of increased biochemical data and availability of *Aspergillus* genomes. The dual-purpose model iJB1325 [15] includes independent models for organic acid (ATCC 1015) and protein (CBS 513.88) production strains and updates to core and secondary metabolite pathways. Concurrently, iHL1210 [14] was produced with updated annotations from the model protein production strain (CBS 513.88) and was used as a scaffold to produce iDU1756 [16] based on the citric acid production strain (ATCC 1015). While iJB1325 may be more useful for general-purpose metabolic engineering due to its greater number of represented metabolites and reactions, iDU1756 may be more appropriate for accurate modeling of industrial organic acid production as flux predictions from its precursor iHL1210 were verified by ^{13}C metabolic flux analysis. In contrast to *A. niger* GEMs with extensive

studies, the first GEM for *I. orientalis* (*iIsor*850) was produced only recently based on the Yeast 7.6 scaffold with refinements from the SD108 genome [17]. Biomass composition and ATP maintenance requirements were determined experimentally and substrate utilization and gene essentiality assays were used to validate model predictions, making this a high-quality model for yeast-based organic acid production.

Developing GEMs for fungi with the potential to produce complex natural products is challenging, because it requires specific emphasis on incorporation of reactions and metabolites in secondary metabolism that may not be well-predicted from genomic data. Recent development has focused on updates to the *Penicillium rubens* model used for β-lactam antibiotics and *Alternaria* sp. MG1 for production of resveratrol. *P. rubens* (Wisconsin 54-1255) has a long history of development and modeling to support industrial production of antibiotics. Modern GEMs for *P. rubens* are based on iAL1006, which was produced as a demonstration of the RAVEN toolbox [18]. Updates to this model have been generated by incorporating additional data on secondary metabolite and natural product pathways from MetaCyc to produce models for *P. rubens* and 23 additional *Penicillium* species (Penicillium-GEMs) using automated pipelines [19]. iPrub22 builds upon this automatic reconstruction by dramatically increasing the size of the model to 5919 reactions but currently lacks experimental validation [20•]. While the newer GEMs better represent secondary metabolite pathways, iAL1006 remains the gold standard for prediction accuracy as it was optimized to perform precise simulations based on experimental data. The first GEM for an endophytic *Alternaria* species (iYL1539) was produced recently based on reactions from *Aspergillus terreus*, *P. rubens*, and *S. cerevisiae* models and additional natural product pathways based on KEGG and literature sources [21]. This model will provide a high-quality starting point but may require refinement and experimentally derived constraints to target specific natural product pathways.

Pichia pastoris is used for industrial production of recombinant and total protein products. Modern GEMs representing its metabolism are based on iMT1026 [22], which was produced as a consensus model by merging aspects of metabolism from three previous *P. pastoris* GEMs (iPP668, PpaMBEL1254, and iLC915). iMT1026 has been updated to v3.0 [23], which incorporates additional reactions, experimentally determined bounds, and biomass compositions for growth on methanol and glycerol. iRY1243 was independently produced as an update to iMT1026 v1.0 that incorporates additional reactions, including complex sugar utilization pathways [24]. iMT1026 v3.0 is recommended as a high-quality *P. pastoris* model that has gone through iterative refinements and incorporates accurate representations of growth on industrial feedstocks. iUL909 has also been recently developed for another protein production host *Ogataea polymorpha* (also known as *Pichia angusta*) and tested against phenotype microarray using 190 substrates as carbon sources [25].

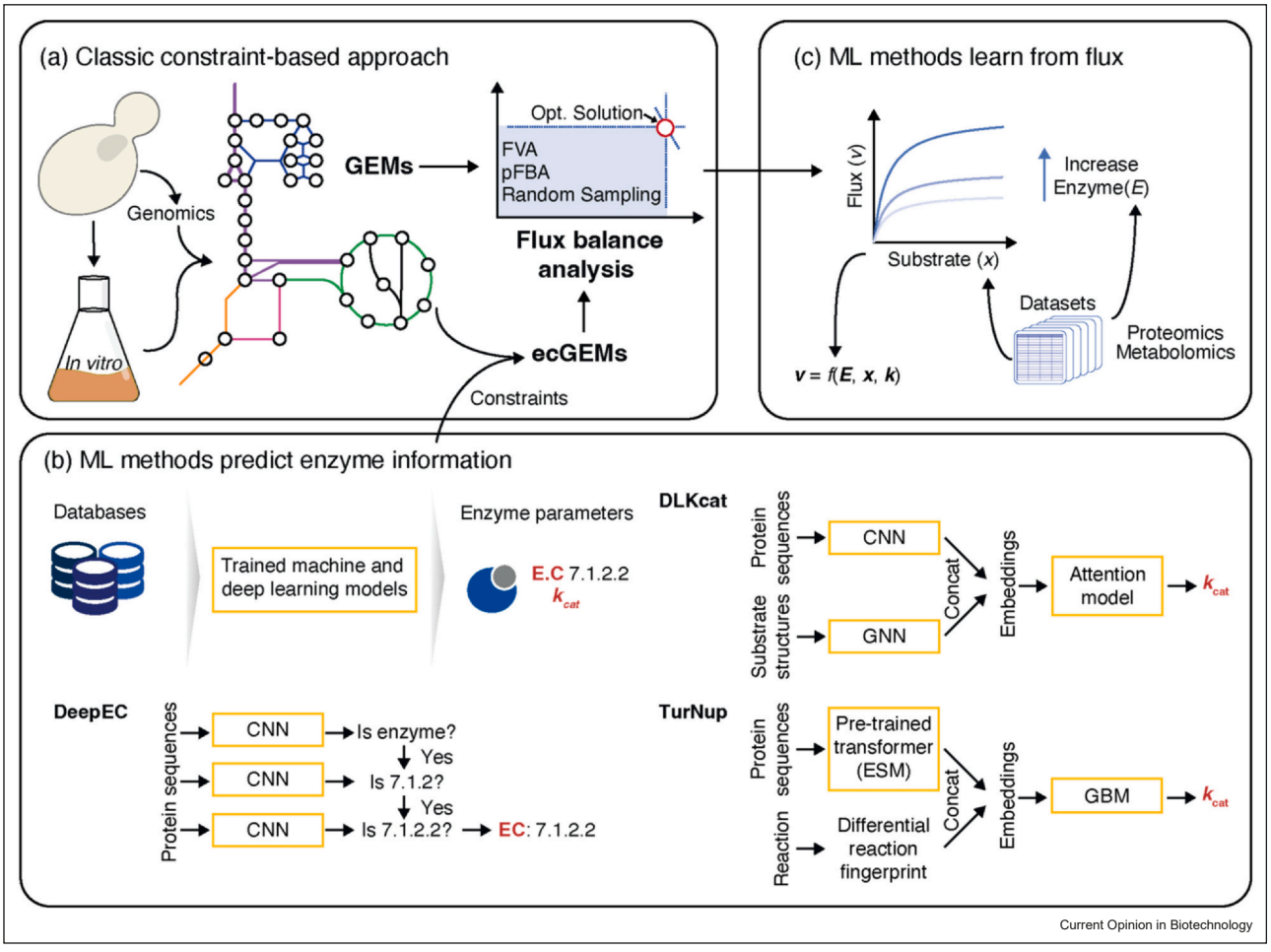
These GEMs were assessed using MEMOTE [26], a tool developed for standardized testing of GEMs. MEMOTE performs various tests and quantifies the results to calculate an overall score. In Table 1, the MEMOTE scores were included for the GEMs available in Systems Biology Markup Language format. It is important to note that while some GEMs had lower scores than others, this does not necessarily indicate lower quality or unsuitability for use, because the MEMOTE score is the weighted sum of many different metrics. A low score might mean that certain modeled entities, such as metabolites, are not linked to various databases, even if the metabolic network itself accurately reflects reality. In general, focusing on only one validation method cannot provide a comprehensive assessment of the overall model performance. It would be good practice to validate the model under diverse conditions. For instance, assessing growth with diverse substrates can activate and test different parts of metabolism. Models that pass multiple validation tests may still have large solution spaces. Integrating new constraint layers will further enhance the predictive power of such GEMs.

Advances in computational methods

To exploit GEMs, a variety of computational methods have been developed. Flux balance analysis (FBA), a commonly employed method to predict metabolic flux distributions, serves as the basis for many other methods. To find flux distribution solutions, FBA defines an objective function to be optimized subject to certain constraints. Owing to numerous possible metabolic states in the solution space given by the network topology, some variations on FBA have been proposed to identify alternative flux distributions [27–30] (Figure 1a). Among them, the cost-weighted FBA considered the potential contribution of alternative pathways in the network, which increases the biological relevance of the flux predictions [29]. To consider the cost of expressing metabolic enzymes, enzyme-constrained GEMs (ec-GEMs) determine the metabolic fluxes incorporating enzyme kinetics, and constraining the total amount of proteins. These GEMs strongly rely on enzyme turnover numbers (k_{cat}), which mostly come from *in vitro* measurements and are chosen according to the Enzyme Commission (EC) number, substrate and organism.

Since there exist limited experimental records of EC number and k_{cat}, statistical and ML models have been employed to provide inferences on enzyme information (Figure 1b). For example, DeepEC used three

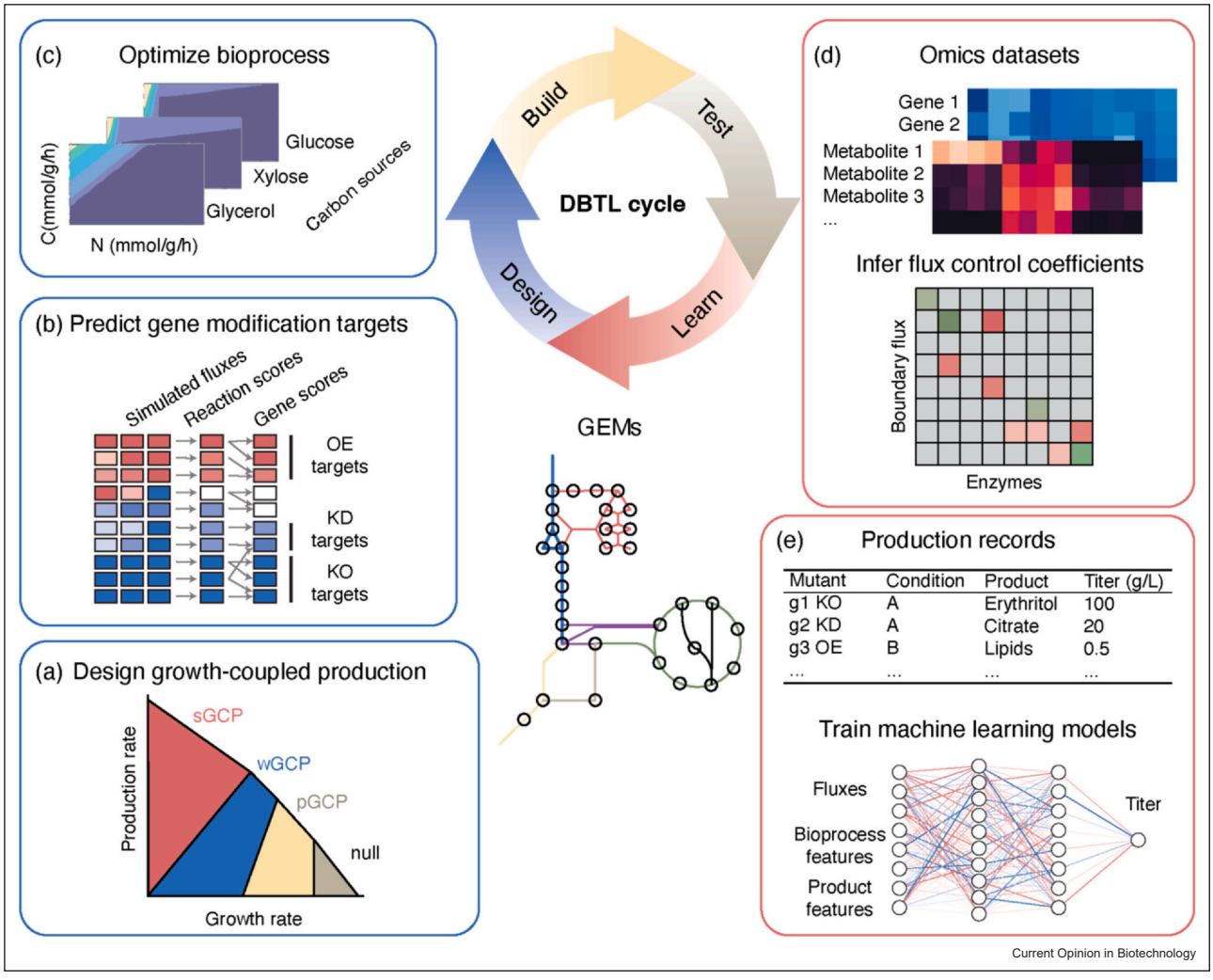
Figure 1



Overview of advances in genome-scale modeling methods. **(a)** Classic approach to build and analyze GEMs or ecGEMs. **(b)** ML approaches predict enzyme information and add constraints to ecGEMs. **(c)** ML approaches learn from GEM predictions in hybrid modeling framework.

convolutional neural networks (CNNs) to predict EC numbers with high precision, true positives/(true positives + false positives) = 0.920, based on the protein sequence as input [31]. Recently, Bayesian multilevel models were developed to estimate k_{cat} values and their uncertainty using EC numbers, identifiers, and protein families as inputs [32]. Additionally, k_{cat} can also be predicted using ML methods. An ML model trained on less than 200 observations with manually curated features (metabolic fluxes, enzyme structural and biochemical properties, and assay conditions), showed R^2 of 0.76 and 0.31 for predictions of apparent catalytic rates (k_{app}) and *in vitro* k_{cat}, respectively. A deep learning model called DLKcat [33•] was trained on enzyme datasets in BRENDA [34] and SABIO-RK [35] databases to predict turnover numbers. The model, which combined a graph neural network (GNN) and a CNN, used substrate SMILES and enzyme protein sequences as input, significantly increasing the number of available training entries. With good predictive performance (R^2 of 0.50 on the test dataset), DLKcat has generated approximately 25.7 million turnover numbers [36] compared with the 86 919 available in BRENDA [34]. The recently updated GECKO 3.0 incorporated DLKcat as a key component [37], which facilitates the reconstruction of ecGEMs for any organism. More recently, a machine and deep learning hybrid approach called TurNup was developed to predict *in vitro* k_{cat}, which outperformed DLKcat, especially for enzymes without close homologs in the training dataset and enzymes catalyzing unseen reactions [38••]. This model first utilized reaction fingerprints and a pretrained transformer model for protein sequence embedding to generate enzyme-reaction representation. The representation was further used as inputs to train a gradient boosting model (GBM) for k_{cat} prediction. However, it should be noted that many

Figure 2



Representative applications of fungal GEMs in biomanufacturing. Applications in the 'design' and 'learn' stages are respectively surrounded by a blue line and a red line. **(a)** Design grow-coupled production. sGCP, strongly GCP (producing the target in all flux states with growth); wGCP, weakly GCP (producing the target in all growth-maximal flux states); pGCP, potentially GCP (producing the target in at least one growth-maximal flux state). **(b)** Predict genetic modification targets. KO, knockout; KD, knockdown. **(c)** Evaluate carbon sources and C/N ratios to optimize bioprocess. **(d)** Infer FCCs from omics datasets. **(e)** Train ML models from production records.
Figures in (a), (b), and (c) adapted from Refs. [51,54••,12].

turnover numbers were obtained *in vitro* and may not accurately reflect the *in vivo* activities [39,40]. Large-scale *in vivo* k_{cat} can be estimated based on k_{app} using absolute proteomics and fluxomics data from either FBA or ^{13}C metabolic flux analysis [39–41].

When the growth or production rate predicted by an ecGEM does not align with experimentally observed values, a k_{cat} correction procedure may be necessary. Some computational methods have been developed for automatic curation by substituting or relaxing the k_{cat} value for the enzyme with the highest impact on growth. In the first case, k_{cat} value will be substituted by the largest value in BRENDA for this enzyme across all organisms [42••] or with the mean value of the k_{cat} collected based on nearest species [43]. In the second case, the k_{cat} value will be increased by n-fold the initial value [37]. This procedure is repeated until the experimental growth rate can be predicted. As an alternative method, PRESTO utilized an optimization framework by minimizing a weighted sum of two objectives: the relative error to measured growth rates across multiple conditions and the sum of positive k_{cat} corrections [44••].

Flux predictions from constraint-based methods can be subsequently used as input for statistical and ML models in hybrid modeling frameworks (Figure 1c).

Metabolic flux information is an important part in metabolic control analysis (MCA), where flux control coefficients (FCC) describe how flux changes in response to changes in enzyme concentrations or activities. From multi-omics datasets and fluxes determined by FBA, Bayesian MCA (BMCA) was developed to predict FCC for a GEM using linear-logarithmic (linlog) kinetics to approximate mechanistic enzyme kinetic equations [42]. Additionally, supervised ML models can train on flux predictions from FBA as inputs and production performance in literature as outputs [44]. This type of method can estimate production performance and guide metabolic engineering.

Applications in metabolic engineering

GEMs serve multiple purposes in iterative DBTL cycles. In the initial round of the DBTL cycle, GEMs can guide the design process. First, they can determine the optimal biosynthetic routes. For instance, based on pathway lengths and maximum product yields calculated by yeast GEM, SPD-3 and SPD-4 were selected as initial production targets for the development of a yeast-based polyamine production platform [45]. Second, GEMs can predict genetic engineering targets based solely on the stoichiometry of metabolic networks by employing classical methods such as OptKnock [46], Minimization Of Metabolic Adjustment (MOMA) [47], and Flux Scanning based on Enforced Objective Flux (FSEOF) [48]. Many strain design methods were recently implemented in Python packages [49,50]. OptKnock was utilized on the *I. orientalis* GEM iIsor850 to identify three degrees (i.e. strongly, weakly, and potentially) of growth-coupled production (GCP) designs for 22 organic acid products from both glucose and xylose as carbon substrates [51] (Figure 2a). MOMA was performed on *P. pastoris* GEM iMT1026 v3.0 to evaluate the impact of cytosolic and mitochondrial reduced nicotinamide adenine dinucleotide (NADH) kinase overexpression (OE) on production of an antibody fragment [52]. FSEOF was employed on GEMs iUL909 and papla-GEM to identify reactions that enhance desired bioproduction in *O. polymorpha* [25] and *P. laurentii* [11], respectively. Recently, a FSEOF variant, later called ecFactory [53••], was developed on top of ecGEMs to identify a set of minimum genetic modifications, which prevents the arbitrary selection of the number of gene targets. Using this approach with ecYeast8, 84 genes (62 for OE, 14 for repression, and eight for deletion) were identified to enhance heme production [54••] (Figure 2b). The author further optimized the combination of gene modification through model-guided design, resulting in 70-fold improvement in heme production. The authors also used ecFactory to predict engineering targets for production of 102 different chemicals in yeast and found shared gene targets for each group of chemicals [53••]. Third, GEMs can optimize production conditions during the fermentation process (Figure 2c). Such attempts in oleaginous yeasts *C. oleaginosus* and *R. toruloides* have been made to investigate the effects of different carbon sources and carbon-to-nitrogen (C/N) ratios on lipid production using GEMs iNP636 [12] and rhto-GEM [55], respectively.

In the learning stage of the DBTL cycle, GEMs serve as a data integration platform for understanding the metabolism and guiding subsequent rounds of design. GEMs can incorporate kinetic parameters from batch fermentation to analyze metabolism changes. In a recent example, flux distribution changes in *Y. lipolytica* were analyzed by dynamic FBA using GEM iYali4 [56]. This analysis identified metabolic pathways that contribute to citrate production and suggested a strategy to achieve a twofold increase in citrate titer. Additionally, genome-scale multi-omics datasets can be integrated with GEMs to recommend genetic engineering targets. An evolutionary algorithm that incorporated transcriptomic data was developed to identify targets for increasing citric acid productivity from lignocellulosic hydrolysate in *A. niger* using GEM iDU1327 [57]. BMCA was applied to a few itaconic acid-producing *Y. lipolytica* strains with the central metabolism part of GEM iYLI649 [58•]. Based on FCCs inferred from multi-omics data, they identified enzymes in rate-limiting reactions that can improve itaconic acid production. Last, with more production data available, metabolic fluxes calculated from GEMs are important features in ML. For instance, ML models were trained on flux distributions generated from FBA with *Y. lipolytica* GEM iYLI647 and other bioprocess parameters to predict product titers [59••].

Conclusion and future perspectives

An increasing number of GEMs have been reconstructed and updated in industrially relevant fungi for biomanufacturing. These GEMs provide a potent computational platform enabling model-driven design. To enhance simulation performance, ecGEMs impose additional constraints on finite proteome, enzyme concentration, and catalytic activity. Despite these advances, ecGEMs are available for only a few fungi (*A. niger*, *R. toruloides*, *S. cerevisiae*, and *Y. lipolytica* in Table 1), likely due to the lack of enzyme information for most fungi. By integrating ML methods that address missing enzyme information, tools have been developed to facilitate the construction of ecGEMs. GEMs often benchmark against growth phenotyping data across different nutrients and gene fitness/essentiality data, which in turn allows subsequent curation and improves model quality with more experimental data.

Classical approaches such as OptKnock [46], MOMA [47], and FSEOF [48] still dominate the target predictions for metabolic engineering, although a few variants

have been developed. Among more recent innovations, BMCA is a promising approach to identify gene targets. It considers enzyme kinetics by incorporating omics data, but does not require absolute quantification. All these methods are compatible with ecGEMs, which can yield more accurate predictions. A combination of multiple methods can be considered to identify better targets, because no single method can address all problems in metabolic engineering.

Data Availability

No data were used for the research described in the article.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding: This work was supported by the U.S. Department of Energy (DOE) Agile BioFoundry, Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, under Contract no. DE-NL0030038; the DOE Joint BioEnergy Institute, Office of Science, Office of Biological and Environmental Research, under contract no. DE-AC02-05CH11231; and the Novo Nordisk Foundation under Grant no. NNF20CC0035580. Pacific Northwest National Laboratory is a multi-program national laboratory operated by Battelle for the DOE under Contract no. DE-AC05-76RLO 1830.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.copbio.2023.103005.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest.

1. Meyer V, Andersen MR, Brakhage AA, Braus GH, Caddick MX, Cairns TC, de Vries RP, Haarmann T, Hansen K, Hertz-Fowler C, *et al.*: **Current challenges of research on filamentous fungi in relation to human welfare and a sustainable bio-economy: a white paper**. *Fungal Biol Biotechnol* 2016, **3**:6.

2. Ghorai S, Banik SP, Verma D, Chowdhury S, Mukherjee S, Khowala S: **Fungal biotechnology in food and feed processing**. *Food Res Int* 2009, **42**:577-587.

3. Meyer V, Basenko EY, Benz JP, Braus GH, Caddick MX, Csukai M, de Vries RP, Endy D, Frisvad JC, Gunde-Cimerman N, *et al.*: **Growing a circular economy with fungal biotechnology: a white paper**. *Fungal Biol Biotechnol* 2020, **7**:5.

4. Kerkhoven EJ, Pomraning KR, Baker SE, Nielsen J: **Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica***. *NPJ Syst Biol Appl* 2016, **2**:16005.

5. • Guo Y, Su L, Liu Q, Zhu Y, Dai Z, Wang Q: **Dissecting carbon metabolism of *Yarrowia lipolytica* type strain W29 using genome-scale metabolic modelling**. *Comput Struct Biotechnol J* 2022, **20**:2503-2511.
An updated GEM for *Yarrowia lipolytica* strain W29 was developed by improving growth predictions using Biolog phenotype microarrays. The GEM was used to integrate transcriptomics data for studying carbon metabolism.

6. Wei S, Jian X, Chen J, Zhang C, Hua Q: **Reconstruction of genome-scale metabolic model of *Yarrowia lipolytica* and its application in overproduction of triacylglycerol**. *Bioresour Bioprocess* 2017, **4**:1-9.

7. Mishra P, Lee N-R, Lakshmanan M, Kim M, Kim B-G, Lee D-Y: **Genome-scale model-driven strain design for dicarboxylic acid production in *Yarrowia lipolytica***. *BMC Syst Biol* 2018, **12**:9-20.

8. Tiukova IA, Prigent S, Nielsen J, Sandgren M, Kerkhoven EJ: **Genome-scale model of *Rhodotorula toruloides* metabolism**. *Biotechnol Bioeng* 2019, **116**:3396-3408.

9. Dinh HV, Suthers PF, Chan SHJ, Shen Y, Xiao T, Deewan A, Jagtap SS, Zhao H, Rao CV, Rabinowitz JD, *et al.*: **A comprehensive genome-scale model for *Rhodosporidium toruloides* IFO0880 accounting for functional genomics and phenotypic data**. *Metab Eng Commun* 2019, **9**:e00101.

10. Kim J, Coradetti ST, Kim Y-M, Gao Y, Yaegashi J, Zucker JD, Munoz N, Zink EM, Burnum-Johnson KE, Baker SE, *et al.*: **Multi-omics driven metabolic network reconstruction and analysis of lignocellulosic carbon utilization in *Rhodosporidium toruloides***. *Front Bioeng Biotechnol* 2020, **8**:612832.

11. • Ventorim RZ, de M. Ferreira MA, de Almeida ELM, Kerkhoven EJ, da Silveira WB: **Genome-scale metabolic model of oleaginous yeast *Papiliotrema laurentii***. *Biochem Eng J* 2022, **180**:108353.
The first GEM for oleaginous yeast *Papiliotrema laurentii* was developed with a detailed representation of lipid metabolism. All major biomass components were measured, and the GEM was evaluated with growth parameters from batch and chemostat cultivation.

12. Pham N, Reijnders M, Suarez-Diez M, Nijsse B, Springer J, Eggink G, Schaap PJ: **Genome-scale metabolic modeling underscores the potential of *Cutaneotrichosporon oleaginosus* ATCC 20509 as a cell factory for biofuel production**. *Biotechnol Biofuels* 2021, **14**:2.

13. Andersen MR, Nielsen ML, Nielsen J: **Metabolic model integration of the bibliome, genome, metabolome and reactome of *Aspergillus niger***. *Mol Syst Biol* 2008, **4**:178.

14. Lu H, Cao W, Ouyang L, Xia J, Huang M, Chu J, Zhuang Y, Zhang S, Noorman H: **Comprehensive reconstruction and in silico analysis of *Aspergillus niger* genome-scale metabolic network model that accounts for 1210 ORFs**. *Biotechnol Bioeng* 2017, **114**:685-695.

15. Brandl J, Aguilar-Pontes MV, Schäpe P, Noerregaard A, Arvas M, Ram AFJ, Meyer V, Tsang A, de Vries RP, Andersen MR: **A community-driven reconstruction of the *Aspergillus niger* metabolic network**. *Fungal Biol Biotechnol* 2018, **5**:1-12.

16. Upton DJ, McQueen-Mason SJ, Wood AJ: **In silico evolution of *Aspergillus niger* organic acid production suggests strategies for switching acid output**. *Biotechnol Biofuels* 2020, **13**:1-21.

17. Suthers PF, Dinh HV, Fatma Z, Shen Y, Chan SHJ, Rabinowitz JD, Zhao H, Maranas CD: **Genome-scale metabolic reconstruction of the non-model yeast *Issatchenkia orientalis* SD108 and its application to organic acids production**. *Metab Eng Commun* 2020, **11**:e00148.

18. Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J: **The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum***. *PLoS Comput Biol* 2013, **9**:e1002980.

19. Prigent S, Nielsen JC, Frisvad JC, Nielsen J: **Reconstruction of 24 Penicillium genome-scale metabolic models shows diversity based on their secondary metabolism**. *Biotechnol Bioeng* 2018, **115**:2604-2612.

20. •• Nègre D, Larhlimi A, Bertrand S: **Reconciliation and evolution of *Penicillium rubens* genome-scale metabolic networks-What about specialised metabolism?** *PLoS One* 2023, **18**:e0289757.

A new GEM was reconstructed for *Penicillium rubens* with significantly improved metabolic coverage. The GEM was used to investigate the production of secondary metabolites and natural products.

21. Lu Y, Ye C, Che J, Xu X, Shao D, Jiang C, Liu Y, Shi J: **Genomic sequencing, genome-scale metabolic network reconstruction, and in silico flux analysis of the grape endophytic fungus Alternaria sp. MG1**. *Micro Cell Fact* 2019, **18**:1-16.

22. Tomàs-Gamisans M, Ferrer P, Albiol J: **Integration and validation of the genome-scale metabolic models of *Pichia pastoris*: a comprehensive update of protein glycosylation pathways, lipid and energy metabolism**. *PLoS One* 2016, **11**:e0148031.

23. Tomàs-Gamisans M, Ferrer P, Albiol J: **Fine-tuning the *P. pastoris* iMT1026 genome-scale metabolic model for improved prediction of growth on methanol or glycerol as sole carbon sources**. *Micro Biotechnol* 2018, **11**:224-237.

24. Ye R, Huang M, Lu H, Qian J, Lin W, Chu J, Zhuang Y, Zhang S: **Comprehensive reconstruction and evaluation of *Pichia pastoris* genome-scale metabolic model that accounts for 1243 ORFs**. *Bioresour Bioprocess* 2017, **4**:1-12.

25. • Liebal UW, Fabry BA, Ravikrishnan A, Schedel CV, Schmitz S, Blank LM, Ebert BE: **Genome-scale model reconstruction of the methylotrophic yeast *Ogataea polymorpha***. *BMC Biotechnol* 2021, **21**:23.

The first GEM for methylotrophic yeast *Ogataea polymorpha* was developed. The GEM was tested with Biolog phenotype microarrays and growth rates on different carbon sources.

26. Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, Bartell JA, Blank LM, Chauhan S, Correia K, *et al.*: **MEMOTE for standardized genome-scale metabolic model testing**. *Nat Biotechnol* 2020, **38**:272-276.

27. Price ND, Schellenberger J, Palsson BO: **Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies**. *Biophys J* 2004, **87**:2172-2186.

28. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, *et al.*: **Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models**. *Mol Syst Biol* 2010, **6**:390.

29. Cheung CYM, Ratcliffe RG, Sweetlove LJ: **A method of accounting for enzyme costs in flux balance analysis reveals alternative pathways and metabolite stores in an illuminated Arabidopsis leaf**. *Plant Physiol* 2015, **169**:1671-1682.

30. Kenefake D, Armingol E, Lewis NE, Pistikopoulos EN: **An improved algorithm for flux variability analysis**. *BMC Bioinform* 2022, **23**:550.

31. Ryu JY, Kim HU, Lee SY: **Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers**. *Proc Natl Acad Sci USA* 2019, **116**:13996-14001.

32. Gollub MG, Backes T, Kaltenbach H-M, Stelling J: ENKIE: A package for predicting enzyme kinetic parameter values and their uncertainties. *bioRxiv*; 2023. ⟨doi:10.1101/2023.03.08.531697⟩.

33. • 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**:662-672.

A deep learning model combining a CNN for protein sequences and a GNN for substrate structures was trained to predict k_{cat} values.

34. Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, Neumann-Schaal M, Jahn D, Schomburg D: **BRENDA, the ELIXIR core data resource in 2021: new developments and updates**. *Nucleic Acids Res* 2021, **49**:D498-D508.

35. 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-D660.

36. Li F, Chen Y, Anton M, Nielsen J: **GotEnzymes: an extensive database of enzyme parameter predictions**. *Nucleic Acids Res* 2022, **51**:D583-D586.

37. Kerkhoven E, Sánchez B, Domenzain I, johan-gson, Rangel AT, Anton M, Redestig H, Robinson J, Beber ME, Sulheim S, et al.: **SysBioChalmers/GECKO: GECKO 3.1.1. Zenodo**; 2023.

38. •• 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.

A machine and deep learning hybrid model was trained to predict k_{cat} values. Protein sequences and reactions were first embedded using existing methods and used to train a GBM for k_{cat} prediction.

39. Davidi D, Noor E, Liebermeister W, Bar-Even A, Flamholz A, Tummler K, Barenholz U, Goldenfeld M, Shlomi T, Milo R: **Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements**. *Proc Natl Acad Sci USA* 2016, **113**:3401-3406.

40. Chen Y, Nielsen J: **In vitro turnover numbers do not reflect in vivo activities of yeast enzymes**. *Proc Natl Acad Sci USA* 2021, **118,**:e2108391118.

41. Heckmann D, Campeau A, Lloyd CJ, Phaneuf PV, Hefner Y, Carrillo-Terrazas M, Feist AM, Gonzalez DJ, Palsson BO: **Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers**. *Proc Natl Acad Sci USA* 2020, **117**:23182-23190.

42. •• Domenzain I, Sánchez B, Anton M, Kerkhoven EJ, Millán-Oropeza A, Henry C, Siewers V, Morrissey JP, Sonnenschein N, Nielsen J: **Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0**. *Nat Commun* 2022, **13**:1-13.

GECKO 2.0 was developed to facilitate the construction of ecGEMs. A k_{cat} value correction procedure was introduced.

43. Bekiaris PS, Klamt S: **Automatic construction of metabolic models with enzyme constraints**. *BMC Bioinforma* 2020, **21**:19.

44. • Wendering P, Arend M, Razaghi-Moghadam Z, Nikoloski Z: **Data integration across conditions improves turnover number estimates and metabolic predictions**. *Nat Commun* 2023, **14**:1485.

A method called PRESTO was developed to correct k_{cat} values for genome-scale models using an optimization framework.

45. Qin J, Krivoruchko A, Ji B, Chen Y, Kristensen M, Özdemir E, Keasling JD, Jensen MK, Nielsen J: **Engineering yeast metabolism for the discovery and production of polyamines and polyamine analogues**. *Nat Catal* 2021, **4**:498-509.

46. Burgard AP, Pharkya P, Maranas CD: **Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization**. *Biotechnol Bioeng* 2003, **84**:647-657.

47. Segrè D, Vitkup D, Church GM: **Analysis of optimality in natural and perturbed metabolic networks**. *Proc Natl Acad Sci USA* 2002, **99**:15112-15117.

48. Choi HS, Lee SY, Kim TY, Woo HM: **In silico identification of gene amplification targets for improvement of lycopene production**. *Appl Environ Microbiol* 2010, **76**:3097-3105.

49. Schneider P, Bekiaris PS, von Kamp A, Klamt S: **StrainDesign: a comprehensive Python package for computational design of metabolic networks**. *Bioinformatics* 2022, **38**:4981-4983.

50. Pereira V, Cruz F, Rocha M: **MEWpy: a computational strain optimization workbench in Python**. *Bioinformatics* 2021, **37**:2494-2496.

51. Suthers PF, Maranas CD: **Examining organic acid production potential and growth-coupled strategies in Issatchenkia orientalis using constraint-based modeling**. *Biotechnol Prog* 2022, **38**:e3276.

52. Tomàs-Gamisans M, Andrade CCP, Maresca F, Monforte S, Ferrer P, Albiol J: **Redox engineering by ectopic overexpression of NADH kinase in recombinant Pichia pastoris (Komagataella phaffii): impact on cell physiology and recombinant production of secreted proteins**. *Appl Environ Microbiol* 2020, **86**:e02038-19.

53. •• Domenzain I, Lu Y, Shi J, Lu H, Nielsen J: Computational biology predicts metabolic engineering targets for increased production of 102 valuable chemicals in yeast. *bioRxiv*; 2023. ⟨doi:10.1101/2023.01.31.526512⟩..

A FSEOF-based computational method was developed on top of the enzyme-constrained model for strain design. The method was applied to the prediction of engineering targets for 102 different chemicals using *S. cerevisiae*.

54. •• Ishchuk OP, Domenzain I, Sánchez BJ, Muñiz-Paredes F, Martínez JL, Nielsen J, Petranovic D: **Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae***. *Proc Natl Acad Sci USA* 2022, **119**:e2108245119.

The authors improved heme production in *S. cerevisiae* by 70-fold using a GEM-guided design.

55. Lopes HJS, Bonturi N, Kerkhoven EJ, Miranda EA, Lahtvee P-J: **C/N ratio and carbon source-dependent lipid production profiling in *Rhodotorula toruloides***. *Appl Microbiol Biotechnol* 2020, **104**:2639-2649.

56. da Veiga Moreira J, Jolicoeur M, Schwartz L, Peres S: **Fine-tuning mitochondrial activity in *Yarrowia lipolytica* for citrate overproduction**. *Sci Rep* 2021, **11**:878.

57. Upton DJ, Kaushal M, Whitehead C, Faas L, Gomez LD, McQueen-Mason SJ, Srivastava S, Wood AJ: **Integration of *Aspergillus niger* transcriptomic profile with metabolic model identifies potential targets to optimise citric acid production from lignocellulosic hydrolysate**. *Biotechnol Biofuels Bioprod* 2022, **15**:4.

58. • McNaughton AD, Bredeweg EL, Manzer J, Zucker J, Munoz Munoz N, Burnet MC, Nakayasu ES, Pomraning KR, Merkley ED, Dai Z, *et al.*: **Bayesian inference for integrating *Yarrowia lipolytica* multiomics datasets with metabolic modeling**. *ACS Synth Biol* 2021, **10**:2968-2981.

Bayesian metabolic flux control analysis was performed in *Yarrowia lipolytica* to predict gene targets for improved itaconic acid production.

59. •• Czajka JJ, Oyetunde T, Tang YJ: **Integrated knowledge mining, genome-scale modeling, and machine learning for predicting *Yarrowia lipolytica* bioproduction**. *Metab Eng* 2021, **67**:227-236.

The authors trained ML models on production performance and associated product and bioprocess information in the literature. Flux distribution predicted by FBA was used as a feature as well.

60. Lu H, Li F, Sánchez BJ, Zhu Z, Li G, Domenzain I, Marcišauskas S, Anton PM, Lappa D, Lieven C, *et al.*: **A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism**. *Nat Commun* 2019, **10**:3586.

61. Reķēna A, Pinheiro MJ, Bonturi N, Belouah I, Tammekivi E, Herodes K, Kerkhoven EJ, Lahtvee P-J: **Genome-scale metabolic modeling reveals metabolic trade-offs associated with lipid production in *Rhodotorula toruloides***. *PLoS Comput Biol* 2023, **19**:e1011009.

62. Zhou J, Zhuang Y, Xia J: **Integration of enzyme constraints in a genome-scale metabolic model of *Aspergillus niger* improves phenotype predictions**. *Micro Cell Fact* 2021, **20**:1-16.

63. Nègre D, Larhlimi A, Bertrand S: **Reconciliation and evolution of *Penicillium rubens* genome-scale metabolic networks — what about specialised metabolism?** *Plos one* 2023,e0289757.