Genome-scale modeling of yeast metabolism: retrospectives and perspectives

FEMS Yeast Research, 2022, 22, 1–9
DOI: 10.1093/femsyr/foac003
Advance access publication date: 29 January 2022
Minireview

Genome-scale modeling of yeast metabolism: retrospectives and perspectives

Yu Chen [1], Feiran Li [1] and Jens Nielsen [1,2,*]

[1]Department of Biology and Biological Engineering, Chalmers University of Technology, SE412 96 Gothenburg, Sweden
[2]BioInnovation Institute, DK2200 Copenhagen N, Denmark
*Corresponding author: Kemivägen 10, SE412 96 Gothenburg, Sweden. Tel: +46-(0)31-772-3804; E-mail: nielsenj@chalmers.se
One sentence summary: Historical development, recent applications, current status and future directions of yeast genome-scale metabolic models.
Editor: John Morrissey

Abstract

Yeasts have been widely used for production of bread, beer and wine, as well as for production of bioethanol, but they have also been designed as cell factories to produce various chemicals, advanced biofuels and recombinant proteins. To systematically understand and rationally engineer yeast metabolism, genome-scale metabolic models (GEMs) have been reconstructed for the model yeast *Saccharomyces cerevisiae* and nonconventional yeasts. Here, we review the historical development of yeast GEMs together with their recent applications, including metabolic flux prediction, cell factory design, culture condition optimization and multi-yeast comparative analysis. Furthermore, we present an emerging effort, namely the integration of proteome constraints into yeast GEMs, resulting in models with improved performance. At last, we discuss challenges and perspectives on the development of yeast GEMs and the integration of proteome constraints.

Keywords: constraint-based modeling, genome-scale metabolic model, metabolic engineering, proteome constraints

Introduction

The yeast *Saccharomyces cerevisiae* is one of the earliest domesticated organisms, as it has been used in food fermentation such as brewing and baking since ancient times. In the modern era, this yeast is used as a cell factory, by engineering its metabolism, to produce diverse compounds, e.g. sustainable biofuels, high-value chemicals and recombinant proteins (Wang, Huang and Nielsen 2017; Liu *et al.* 2021; Patra *et al.* 2021). Gaining extensive knowledge of the metabolism of this yeast is therefore of high importance to improve the performance of yeast cell factories (Nielsen and Keasling 2016). Furthermore, understanding the metabolism of *S. cerevisiae*, which serves as a eukaryal model organism, is instrumental for studying fundamental pathways in other eukaryotes, especially pathways associated with human diseases (Ferreira, Limeta and Nielsen 2019; Yu and Nielsen 2020).

To investigate metabolism, genome-scale metabolic models (GEMs) have been widely used during the past decades (Kim, Kim and Lee 2017). A GEM contains a whole set of an organism's metabolic reactions with gene–protein–reaction associations (GPRs), which is a valuable knowledgebase of the organism's metabolism. In addition, the GEM can be converted into a computable format based on reaction stoichiometry, and therefore allows to predict genome-scale metabolic fluxes, i.e. rates of metabolic reactions, using constraint-based modeling methods (Lewis, Nagarajan and Palsson 2012), typically flux balance analysis (FBA) (Orth, Thiele and Palsson 2010).

Given the important role in understanding metabolism, GEMs have been built for various yeast species, including *S. cerevisiae* and various nonconventional yeasts (Lopes and Rocha 2017). Here, we review the historical development of yeast GEMs and very recent applications. Subsequently, we review emerging modeling frameworks that improve the predictive power and expand the scope of basic yeast GEMs. Finally, we shed light on challenges and future perspectives of modeling yeast metabolism.

History of yeast GEMs

The first yeast GEM named iFF708 was published in 2003, which was reconstructed for the most studied yeast *S. cerevisiae* based on genomic, biochemical and physiological information (Förster *et al.* 2003). Originating from this first effort, multiple updates have been released for the *S. cerevisiae* GEM, which have been reviewed (Österlund, Nookaew and Nielsen 2012; Sánchez and Nielsen 2015; Lopes and Rocha 2017; Chen, Li and Nielsen 2019; Lu, Kerkhoven and Nielsen 2021; Domenzain *et al.* 2021a). Here, we highlight the most impressive feature of each update (Fig. 1). There were three GEMs of *S. cerevisiae* directly derived from iFF708: iND750 introduced five additional compartments (Duarte, Herrgård and Palsson 2004), iLL672 fixed unlinked reactions by either removing or connecting them based on new biological knowledge (Kuepfer, Sauer and Blank 2005), and iIN800 included a detailed description of lipid metabolism, tRNA synthesis and transport processes (Nookaew *et al.* 2008). Subsequently, iMM904 modified existing reactions and GPRs of iND750 and expanded the content (Mo, Palsson and Herrgård 2009), and iAZ900 identified 120 corrections to iMM904 based on essentiality and synthetic lethality data (Zomorrodi and Maranas 2010). In order to combine all the knowledge into one model, Yeast1 was reconstructed as the first genome-scale metabolic reconstruction of *S. cerevisiae* using consensus annotation and standard terminology (Herrgård

Received: November 29, 2021. Revised: January 6, 2022. Accepted: January 27, 2022

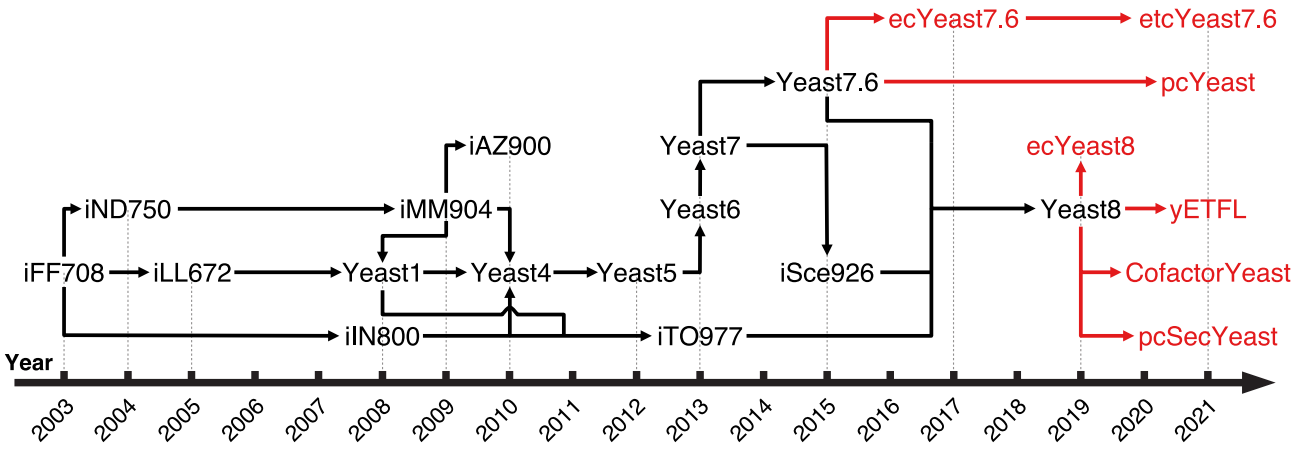

Figure 1. Development of GEMs of *S. cerevisiae*. The classical GEMs are in black, while the GEMs integrated with proteome constraints are in red. Arrow connects a GEM with its predecessor(s).

et al. 2008). The model could, however, not be used for constraint-based simulations due to gaps in the network. Yeast1 was used as a template for further addition of new metabolic information resulting in several updates, but Yeast4 was the first consensus GEM of *S. cerevisiae* that could be used for simulations as it contained enhanced network connectivity and a thorough representation of lipid metabolism (Dobson *et al.* 2010). Yeast5 (Heavner *et al.* 2012), Yeast6 (Heavner *et al.* 2013) and Yeast7 (Aung, Henry and Walker 2013) were successively updated consensus GEMs of *S. cerevisiae* with further expansions and refinements. Meanwhile, based on the consensus GEMs of *S. cerevisiae*, other models were developed, i.e. iTO977 (Österlund *et al.* 2013) and iSce926 (Chowdhury, Chowdhury and Maranas 2015), in which the former merged Yeast1 with iIN800 and contained the highest gene number at that time, while the latter updated Yeast7 by 50 literature-supported modifications.

Based on Yeast7.6, iTO977 and iSce926, the latest consensus GEM of *S. cerevisiae* was developed, namely Yeast8 (Lu *et al.* 2019). In addition to representing the most comprehensive reconstruction of yeast metabolism, the reconstruction of Yeast8 also addressed issues that emerged in the development of previous yeast GEMs, including adding missing transporters (Österlund, Nookaew and Nielsen 2012), updating the biomass equation (Lopes and Rocha 2017) and making the reconstruction processes transparent (Lopes and Rocha 2017). Yeast8 added transport reactions to fill gaps for metabolite transport between compartments, and also assigned new re-annotated transporters to reactions with missing information, which improved the gene coverage. Yeast8 also formulated a new biomass equation by introducing cofactors and metal ions, which can activate more reactions such as cofactor synthesis in growth simulations and thereby the model has improved prediction of essential genes. Lastly, to make the changes transparent Yeast8 is using GitHub (https://github.com/SysBioChalmers/yeast-GEM) to record updates including scripts, corrections, datasets and all released versions, which promotes open and parallel collaboration for the yeast modeling community.

With the multiple rounds of updates, the GEMs of *S. cerevisiae* have been of high quality and coverage, and therefore served as a starting point for the reconstruction of GEMs for non-conventional yeasts such as *Kluyveromyces marxianus* (Marcišauskas, Ji and Nielsen 2019), *Pichia pastoris* (Caspeta *et al.* 2012; Tomàs-Gamisans, Ferrer and Albiol 2018), *Rhodosporidium toruloides* (Dinh *et al.* 2019; Tiukova *et al.* 2019; Kim *et al.* 2021) and *Yarrowia lipolytica* (Loira *et al.* 2012; Kavščcek *et al.* 2015; Kerkhoven *et al.* 2016; Czajka, Oyetunde and Tang 2021). The development of the GEMs of non-conventional yeasts has been reviewed in detail (Lopes and Rocha 2017; Domenzain *et al.* 2021a) and will not be discussed further here.

Applications of yeast GEMs

While a few review papers (Österlund, Nookaew and Nielsen 2012; Sánchez and Nielsen 2015; Lopes and Rocha 2017; Chen, Li and Nielsen 2019; Lu, Kerkhoven and Nielsen 2021; Domenzain *et al.* 2021a) already discussed the applications of yeast GEMs, we focus on the recent applications published within 2 years, which can be classified into four areas: metabolic flux prediction, cell factory design, culture condition optimization and multi-yeast comparative analysis (Fig. 2).

Yeast GEMs are widely used for estimating metabolic fluxes. By constraining GEMs with experimentally measured data such as growth rate and exchange rates of exometabolites, genome-scale metabolic flux distribution can be calculated using constraint-based modeling methods (Fig. 2A), enabling quantifying metabolic changes in response to environmental and genetic perturbations (Liu *et al.* 2019; Olin-Sandoval *et al.* 2019; Lopes *et al.* 2020; Pinheiro *et al.* 2020; Qi *et al.* 2020; Tomàs-Gamisans *et al.* 2020; da Veiga Moreira *et al.* 2021; Henriques *et al.* 2021). Furthermore, calculated metabolic fluxes can serve as a new layer of omics, i.e. fluxomics that can be integrated with other omics data to gain new insight that a single omics type cannot provide. For example, a recent study combined GEM-based fluxomics with absolute proteomics data to estimate *in vivo* enzyme catalytic rates for *S. cerevisiae*, suggesting large deviations between *in vitro* and *in vivo* enzyme activities (Chen and Nielsen 2021a). Additionally, the predicted metabolic fluxes were demonstrated to be key features in machine learning (ML) models for predicting *S. cerevisiae* growth (Culley *et al.* 2020) and *Y. lipolytica* chemical titers (Czajka, Oyetunde and Tang 2021). However, despite the wide use of predicted metabolic fluxes, it should be noted that there are uncertainties in the predicted fluxomics as an infinite number of feasible flux distributions can be obtained by the model while only one of them is usually selected based on biased objective functions (Lewis, Nagarajan and Palsson 2012). Therefore, we recommend also performing unbiased constraint-based methods such as sampling algorithms (Bordel, Agren and Nielsen 2010; Haraldsdóttir *et al.* 2017) to ensure reliable flux predictions.

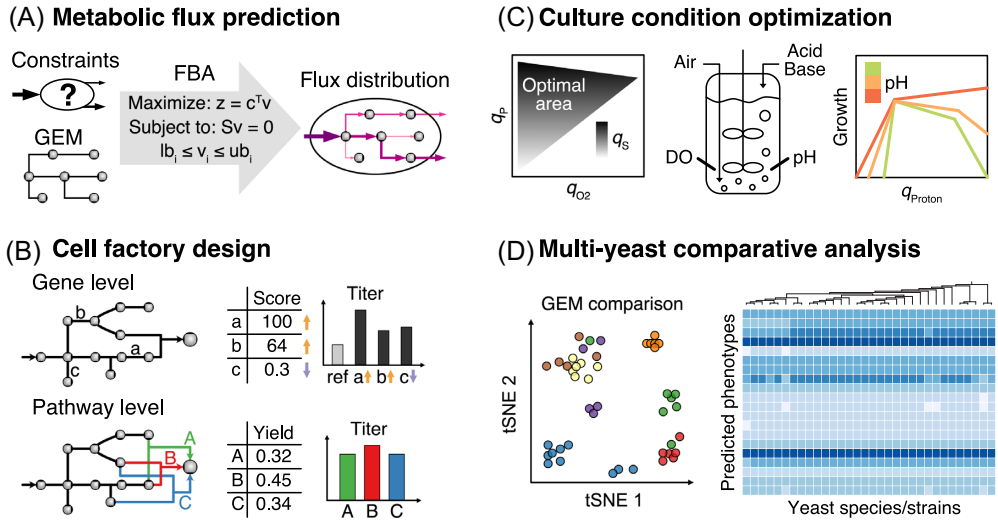

Figure 2. Recent applications of yeast GEMs. **(A)** Metabolic flux prediction. By constraining a GEM with experimentally measured data such as substrate uptake rates and product secretion rates, metabolic flux distribution can be simulated using constraint-based methods such as FBA, in which an objective function is optimized. **(B)** Cell factory design. GEMs can be used to predict gene targets for improving a product of interest and compare various pathways that assimilate the same substrate or synthesize the same product. **(C)** Culture condition optimization. Environmental parameters such as DO and pH can be integrated into GEMs to investigate their effects on metabolism. **(D)** Multi-yeast comparative analysis. Multiple yeast species and strains can be analyzed in a large scale based on GEM structures and simulations.

Another large portion of applications of yeast GEMs have focused on guiding the design of cell factories. Making cells into factories that overproduce native products or secret new products requires redirecting metabolic fluxes toward the desired products optimally, which usually takes a long time due to inefficient trial and error (Nielsen and Keasling 2016). GEMs enable analyzing engineering strategies *in silico*, and therefore have the potential to significantly reduce the time. In the past 2 years, GEMs have been used to aid in developing yeast cell factories by predicting gene targets and identifying optimal pathways (Fig. 2B). GEMs, with the aid of computational algorithms (Long, Ong and Reed 2015), were mostly used to predict targets on gene level, which could be candidates of upregulation, downregulation or knockout. While most algorithms predicted knockout targets that improved product production such as squalene (Paramasivan, Kumar HN and Mutturi 2019), L-phenylacetylcarbinol (Iranmanesh, Asadollahi and Biria 2020) and dicarboxylic acid (Pereira *et al.* 2021) in *S. cerevisiae* and lipids in *Y. lipolytica* (Kim *et al.* 2019), a few also predicted testable upregulation candidates. For example, a GEM-based approach can predict upregulation and downregulation candidates based on computed scores of genes, which guided the overproduction of 3-hydroxypropionic acid (Ferreira *et al.* 2019) and tryptophan (Zhang *et al.* 2020) by *S. cerevisiae*. Also, some already validated upregulation targets could be captured by yeast GEMs implemented with other algorithms for overproduction of lipids in *Y. lipolytica* (Kim *et al.* 2019) and *R. toruloides* (Dinh *et al.* 2019). On the other hand, yeast GEMs allow for comparison of various pathways such as substrate assimilation (Vartiainen *et al.* 2019) and product biosynthesis (Qin *et al.* 2021). For example, a recent study used Yeast7.6 to calculate maximum theoretical yields and pathway lengths of seven spermidine biosynthetic pathways and accordingly selected the most optimal one for further engineering (Qin *et al.* 2021).

In addition to cell factories, the cultivation conditions can also be investigated by yeast GEMs. GEMs enable simulating the effects of cultivation parameters such as dissolved oxygen (DO) and pH on yeast metabolism (Fig. 2C), which could guide process improvement and bioreactor control. For example, a GEM of *P. pastoris* revealed a relationship between DO level and growth-associated ATP requirement, providing practical insights into optimal conditions for protein production (Torres *et al.* 2019). Another study modified a GEM of *S. cerevisiae* by accounting for intracellular pH, which enabled understanding of the impact of pH on growth and introduced manipulations for ethanol overproduction (Ghaffarinasab and Motamedian 2021). In addition, GEMs can be directly adopted in bioreactor control system, i.e. online simulations by GEMs could be used to adjust cultivation parameters, enabling the condition updated based on the cellular need. This was demonstrated by a few studies in which GEM simulations were used to control oxygen and feed flow in micro-aerated ethanol fermentation by *S. cerevisiae* (Mesquita *et al.* 2019, 2021).

An emerging application of yeast GEMs is comparative analysis among yeast strains and species. Attributed to the increasing number of sequenced genomes in recent years, multiple GEMs of yeast strains and species can be generated in a large scale, allowing for conservation and diversity analysis (Fig. 2D). For example, strain-specific GEMs for 1011 *S. cerevisiae* strains were reconstructed, and while structural comparison of GEMs signified that *S. cerevisiae* metabolism is well conserved, predictions of GEMs captured great changes among various strains in some phenotypes such as biomass yield on glucose that could be explained by their ecological origins (Lu *et al.* 2019). In addition, species-specific GEMs for 33 yeasts and fungi were also reconstructed, which predicted different amino acid yields from glucose and therefore reflected evolution of metabolic networks (Correia and Mahadevan 2020). Moreover, the large-scale reconstruction of species-specific GEMs for 343 fungal species enabled comprehensive analyses of evolutionary diversification of substrate utilization (Lu *et al.* 2021).

Integration of proteome constraints into yeast GEMs

Besides the aforementioned applications, some yeast GEMs, as well-curated models, are also a proper platform to explore cutting-edge modeling frameworks that introduce more con-

straints and biological processes. While previous efforts extended yeast GEMs to include transcriptional regulatory constraints (Herrgård *et al.* 2006), *N*-glycosylation pathways (Irani *et al.* 2016), iron metabolism (Dikicioglu and Oliver 2019) and other cellular processes (Ye *et al.* 2020), here we focus on the integration of proteome constraints and related biological processes into yeast GEMs (Figs 1 and 3).

Proteome constraints define finite proteome resources within cells due to limited cell size and space, which should be optimally allocated among metabolic pathways and biological processes in response to various perturbations (Basan 2018). The proteome constraints were integrated with GEMs of many organisms (Chen and Nielsen 2021b), and the first effort for yeast was enzyme-constrained Yeast7.6, named ecYeast7.6 (Sánchez *et al.* 2017), while the latest version ecYeast8 was published in connection with the release of Yeast8 (Lu *et al.* 2019). ecYeast7.6 was generated using the GECKO toolbox (Sánchez *et al.* 2017), which defines an upper bound on each metabolic flux by its enzymatic capacity calculated by enzyme abundance and turnover number (k_{cat}). The enzyme abundance can be constrained by absolute proteomics data when available, otherwise an upper bound on the total abundance of all enzymes could be adopted (Fig. 3). Compared with Yeast7.6, ecYeast7.6 largely reduced simulated flux variability and performed better in predicting physiological behavior such as the Crabtree effect and growth on various carbon sources (Sánchez *et al.* 2017). Such enzyme-constrained models were also generated for some other yeast species including *Y. lipolytica* and *K. marxianus* with the release of the upgraded GECKO toolbox (Domenzain *et al.* 2021b).

Based on ecYeast7.6, a model was recently developed to account for temperature effects on *S. cerevisiae* metabolism, named etcYeast7.6, i.e. enzyme and temperature constrained Yeast7.6 (Li *et al.* 2021c). In addition to temperature-dependent maintenance energy, the temperature constraints were also imposed by making enzyme abundance and k_{cat} temperature-dependent based on enzyme thermal parameters such as melting and optimal temperature (Fig. 3). etcYeast7.6 captured metabolic shifts at high temperatures, which could therefore be explained by temperature-induced proteome constraints. Moreover, etcYeast7.6 identified squalene epoxidase to be the most rate limiting at superoptimal temperatures, which was experimentally validated.

While the enzyme-constrained GEMs account for proteome constraints in a coarse-grained manner, fine-grained approaches that explicitly formulate protein expression processes enable prediction and interpretation of cell behavior in more detail (Yang *et al.* 2018; Chen and Nielsen 2021b). pcYeast, i.e. proteome-constrained GEM of *S. cerevisiae*, was recently developed in such a fine-grained manner (Fig. 3), which extended Yeast7.6 with protein synthesis and degradation reactions for all proteins involved in metabolism, protein expression and degradation processes (Elsemman *et al.* 2021). In this model, metabolic fluxes are constrained by the synthesis rates of the corresponding enzymes, and there are also constraints on protein pools of compartments such as cytosol, plasma membrane and mitochondria. Besides predicting the Crabtree effect, pcYeast also suggested that the mitochondrial constraint could explain the onset of ethanol formation.

Recently another fine-grained model for *S. cerevisiae* was published, named yETFL (Oftadeh *et al.* 2021), which was constructed using Yeast8 based on the ETFL formulation (Salvy and Hatzimanikatis 2020). In addition to metabolism and protein expression, yETFL takes into account thermodynamic constraints that can be incorporated using metabolomics data, enabling to force thermodynamically consistent directionality of metabolic reactions (Fig. 3). yETFL showed better performance than Yeast8 in predicting maximum growth rate and the Crabtree effect.

With the formulation of protein expression, the fine-grained modeling frameworks can serve as a scaffold to mathematically describe other protein-related components and processes. For example, the model CofactorYeast was developed to integrate enzyme cofactors such as metal ions into Yeast8 by formulating protein translation and cofactor binding reactions (Chen *et al.* 2021) (Fig. 3). CofactorYeast can predict condition-specific abundances of metal ions binding on enzymes, which cannot be achieved by the base GEM Yeast8 with condition-independent composition of metal ions in the biomass equation. In addition, CofactorYeast captured dependence of yeast growth and metabolism on metal ions. Particularly, CofactorYeast predicted iron-dependent performance of the cell factory that harbors heterologous iron-containing enzymes to synthesize *p*-coumaric acid, which was experimentally validated.

The fine-grained framework of proteome-constrained genome-scale models was recently expanded to also cover the protein secretory pathway of *S. cerevisiae* in a model abbreviated pcSecYeast (Li *et al.* 2021a). In addition to protein expression and degradation, pcSecYeast also formulated within Yeast8 protein processing and secretion processes, e.g. translocation, posttranslational modification, folding, misfolding and so on (Fig. 3). pcSecYeast expanded applications in the prediction and interpretation of protein secretion related phenotypes such as the switch of glucose transporters in response to changing extracellular glucose concentrations and the growth reduction by protein misfolding. Moreover, pcSecYeast can predict overexpression targets for improving production of recombinant proteins, and the effectiveness of the targets for α-amylase overproduction were experimentally validated.

Challenges and future perspectives

Given that GEMs of *S. cerevisiae* have developed much further than those of the nonconventional yeasts, the presence of *S. cerevisiae*'s GEMs could indicate the future of the nonconventional yeasts' GEMs. For example, future directions of the nonconventional yeasts' GEMs could include model expansion and refinement with increasing knowledge of their metabolism and species-specific information. Therefore, we focus here on the challenges and future perspectives of the *S. cerevisiae* GEMs as well as GEMs integrated with proteome constraints.

What is the next step in the improvement of the *S. cerevisiae* GEM? It has been demonstrated that none of the published yeast GEMs are the best for all applications, and that there might be tradeoffs between model predictive accuracy and model coverage (Heavner and Price 2015). Thus, improving the model coverage might not be a necessary step in near future. Instead, endeavors would be to improve model quality based on increasing knowledge on yeast metabolism. First, it will be necessary to continuously fill gaps in the model. Even in the latest version 8.5.0 of Yeast8 there are 464 out of 2742 metabolites participating in only one reaction or can only be consumed or produced, and those so-called dead-end metabolites indicate missing reactions in the network, which could be linked by adding new reactions with strong evidence. Besides the addition of reactions, genes and proteins should be assigned to existing reactions once new evidence becomes available, among which transport reactions still need much attention as they are difficult to properly annotate (Zuñiga *et al.* 2018; Zuñiga, Tibocha-Bonilla and Betenbaugh 2021). In contrast to the addition efforts, removal of redundant entries and low

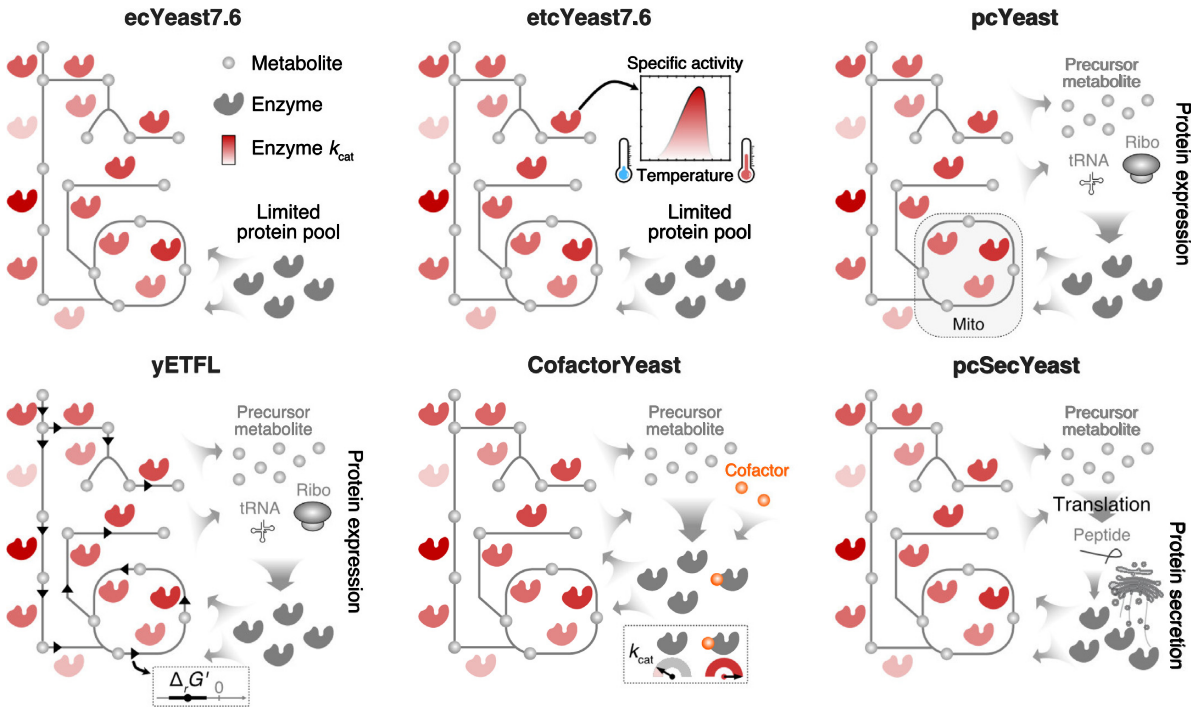

Figure 3. Yeast GEMs integrated with proteome constraints. ecYeast7.6 constrains metabolic fluxes based on enzyme turnover numbers (k_{cat}s) and abundance. etcYeast7.6 constrains metabolic fluxes based on enzyme thermal parameters. pcYeast formulates protein expression within Yeast7.6 and imposes constraints on protein pools of compartments. yETFL formulates protein expression within Yeast8 and enables incorporation of thermodynamic constraints. CofactorYeast formulates protein translation and cofactor binding within Yeast8. pcSecYeast formulates protein expression, processing and secretion within Yeast8.

confidence content was recently highlighted, which has already been observed during the update of yeast GEMs and is believed to be more efficiently performed in yeast GEM refinement with the proposed content removal framework (Seif and Palsson 2021). Another improvement could be on the biomass equation, which has been almost condition-independent and scarcely changed in yeast GEM sequels although Yeast8 has formulated metal ions and vitamins of the biomass composition. Given that variations in biomass equation can indeed affect model predictive accuracy (Dikicioglu, Kırdar and Oliver 2015), we foresee a wider use of condition-specific biomass equations in yeast GEMs in future studies.

In addition to the manual efforts based on increasing knowledge, computer-aided methods especially machine learning can also improve the model quality (Kim, Kim and Lee 2021). Machine learning has been used in various GEMs to fill gaps (Rana *et al.* 2020), improve enzyme annotation by predicting Enzyme Commission (EC) numbers (Ryu, Kim and Lee 2019) and specific subcellular localization (Almagro Armenteros *et al.* 2017; Savojardo *et al.* 2018; Jiang *et al.* 2021), which could hopefully aid in improving yeast GEMs. Note that all the computer-predicted additions should be recorded and require expert level verification.

There are also some model issues independent of the knowledge on yeast metabolism. The MEMOTE test (Lieven *et al.* 2020) shows that while scores generally increase with upgraded yeast GEMs, model annotation and consistency of Yeast8 can still be improved (Domenzain *et al.* 2021a). The model annotation represents the fraction of genes, metabolites and reactions annotated with standardized annotations. Improving annotations of the yeast GEM can facilitate not only its use such as omics data analysis, but also the role as a template for reconstructing other eukaryal GEMs. The model consistency score indicates stoichiometric consistency and balance of mass and charge. Since the publication of Yeast8, >1000 unbalanced reactions have been corrected (https://github.com/SysBioChalmers/yeast-GEM/pull/222), but this step is still not yet completed.

While it is an emerging effort to integrate proteome constraints into GEMs, a few challenges have already been proposed, among which the uncertainty in enzyme turnover numbers has received much attention (Chen and Nielsen 2021b). The enzyme turnover numbers play a key role in mathematically relating metabolic reaction rates to enzyme abundances (in coarse-grained models) or enzyme synthesis rates (in fine-grained models). However, in the modified Yeast8, in which all reversible reactions are split into forward and reverse reactions, only 343 out of 3445 enzymatic reactions have experimentally measured turnover numbers (Chen and Nielsen 2021a). For the reactions without measured values, the turnover numbers could be assigned from other organisms based on various criteria (Chen *et al.* 2021; Domenzain *et al.* 2021b), which may lead to uncertainties. Therefore, we see the need for a large-scale estimation of yeast-specific turnover numbers, i.e. yeast kcatome (Nilsson, Nielsen and Palsson 2017). In contrast to traditional approaches that characterize one enzyme at a time, two recent studies estimated enzyme turnover numbers for yeast in a high-throughput manner. One of these studies proposed a deep learning model that can predict enzyme turnover numbers for diverse yeast species based on protein sequences and substrate structures (Li *et al.* 2021b), and the other estimated *in vivo* enzyme catalytic rates for *S. cerevisiae* using absolute proteomics and flux data (Chen and Nielsen 2021a). The use of these two types of high-throughput data both improved the predictive power of enzyme-constrained GEMs (ecGEMs). Interestingly, the ecGEM parameterized with *in vivo* catalytic rates even outperformed a model with *in vitro* turnover numbers (Chen and Nielsen 2021a),

suggesting that *in vivo* kinetic data would be preferable in future studies.

Despite these achievements, the *in vivo* estimation for enzyme complexes composed of multiple distinct subunits remain to be solved as it is usually difficult to calculate the abundances of catalytic sites (Davidi *et al.* 2016) and sometimes the measured subunit stoichiometry based on proteomics data might not be as expected (Taggart *et al.* 2020; Yu *et al.* 2020). Also, the deep learning model does not handle enzyme complexes well as it only predicts turnover numbers for individual subunits (Li *et al.* 2021b). Therefore, sensitivity analysis of turnover numbers of enzyme complexes is required for model-driven research.

Another future step would be to integrate subcellular constraints into yeast GEMs, given that the integration of the mitochondrial and secretory constraints can improve model predictions and explain observed phenotypes (Elsemman *et al.* 2021; Li *et al.* 2021a). Theses constraints however rely heavily on experimental measurements. For example, the feasibility of imposing mitochondrial constraints could be supported by mitochondrial proteome quantification (Di Bartolomeo *et al.* 2020; Elsemman *et al.* 2021). Therefore, quantification of subcellular proteomes (Wiederhold *et al.* 2010; Valli *et al.* 2020) is expected to advance the development of yeast models.

In addition, the GEMs integrated with proteome constraints are expected to be a proper platform to incorporate regulation mechanisms such as gene expression, posttranslational modification and allosteric regulation (Chubukov *et al.* 2014), which can be readily implemented on enzyme abundances and activities in the model structure. Despite such a promising platform, there are however obstacles for modeling yeast regulation, among which the major one is the lack of a complete understanding of regulation mechanisms. We believe that techniques such as omics (Oliveira *et al.* 2012; Hackett *et al.* 2016) and ChIP-exo (Liu, Bergenholm and Nielsen 2016; Holland *et al.* 2019) could aid in identification of novel regulatory functions, and therefore give new impetus for incorporating regulatory information in yeast GEMs.

The efforts to integrate proteome constraints into yeast GEMs are currently in an initial stage and therefore focus on the improved performance over classical GEMs rather than the widespread use in the yeast modeling community. Accordingly, these advanced models have differences in nomenclatures and formalisms even though they account for overlapping biological components and processes. Given that standardization of models can facilitate effective communication within the community of constraint-based metabolic modeling (Carey *et al.* 2020), we expect that a consensus model would emerge by integrating existing yeast proteome-constrained models, which can hopefully formulate protein expression processes and implement proteome constraints in a standardized fashion.

Funding

This work was supported by the European Union's Horizon 2020 - Research and Innovation Framework Programme under grant agreement (no. 686070) and the Knut and Alice Wallenberg Foundation.

Conflict of interest. None declared.

Reference

Almagro Armenteros JJ, Sønderby CK, Sønderby SK *et al.* DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* 2017;**33**:3387–95.

Aung HW, Henry SA, Walker LP. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Ind Biotechnol* 2013;**9**:215–28.

Basan M. Resource allocation and metabolism: the search for governing principles. *Curr Opin Microbiol* 2018;**45**:77–83.

Bordel S, Agren R, Nielsen J. Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput Biol* 2010;**6**:16.

Carey MA, Dräger A, Beber ME *et al.* Community standards to facilitate development and address challenges in metabolic modeling. *Mol Syst Biol* 2020;**16**:e9235.

Caspeta L, Shoaie S, Agren R *et al.* Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and *in silico* evaluation of their potentials. *BMC Syst Biol* 2012;**6**:1–14.

Chen Y, Li F, Mao J *et al.* Yeast optimizes metal utilization based on metabolic network and enzyme kinetics. *Proc Natl Acad Sci USA* 2021;**118**:e2020154118.

Chen Y, Li G, Nielsen J. Genome-scale metabolic modeling from yeast to human cell models of complex diseases: latest advances and challenges. *Methods Mol Biol* 2019;**2049**:329–45.

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

Chen Y, Nielsen J. Mathematical modelling of proteome constraints within metabolism. *Curr Opin Syst Biol* 2021b;**25**:50–6.

Chowdhury R, Chowdhury A, Maranas CD. Using gene essentiality and synthetic lethality information to correct yeast and CHO cell genome-scale models. *Metabolites* 2015;**5**:536–70.

Chubukov V, Gerosa L, Kochanowski K *et al.* Coordination of microbial metabolism. *Nat Rev Microbiol* 2014;**12**:327–40.

Correia K, Mahadevan R. Pan-genome-scale network reconstruction: harnessing phylogenomics increases the quantity and quality of metabolic models. *Biotechnol J* 2020;**15**:1900519.

Culley C, Vijayakumar S, Zampieri G *et al.* A mechanism-aware and multiomic machine-learning pipeline characterizes yeast cell growth. *Proc Natl Acad Sci* 2020:202002959.

Czajka JJ, Oyetunde T, Tang YJ. Integrated knowledge mining, genome-scale modeling, and machine learning for predicting *Yarrowia lipolytica* bioproduction. *Metab Eng* 2021;**67**:227–36.

da Veiga Moreira J, Jolicoeur M, Schwartz L *et al.* Fine-tuning mitochondrial activity in Yarrowia lipolytica for citrate overproduction. *Sci Rep* 2021;**11**:1–11.

Davidi D, Noor E, Liebermeister W *et al.* Global characterization of *in vivo* enzyme catalytic rates and their correspondence to *in vitro* kcat measurements. *Proc Natl Acad Sci* 2016;**113**:3401–6.

Di Bartolomeo F, Malina C, Campbell K *et al.* Absolute yeast mitochondrial proteome quantification reveals trade-off between biosynthesis and energy generation during diauxic shift. *Proc Natl Acad Sci* 2020;**117**:7524–35.

Dikicioglu D, Kırdar B, Oliver SG. Biomass composition: the "elephant in the room" of metabolic modelling. *Metabolomics* 2015;**11**:1690–701.

Dikicioglu D, Oliver SG. Extension of the yeast metabolic model to include iron metabolism and its use to estimate global levels of iron-recruiting enzyme abundance from cofactor requirements. *Biotechnol Bioeng* 2019;**116**:610–21.

Dinh HV, Suthers PF, Chan SHJ *et al.* A comprehensive genome-scale model for *Rhodosporidium toruloides* IFO0880 accounting for functional genomics and phenotypic data. *Metab Eng Commun* 2019;**9**:e00101.

Dobson PD, Smallbone K, Jameson D *et al.* Further developments towards a genome-scale metabolic model of yeast. *BMC Syst Biol* 2010;**4**:1–7.

Domenzain I, Li F, Kerkhoven EJ *et al.* Evaluating accessibility, usability and interoperability of genome-scale metabolic models for diverse yeasts species. *FEMS Yeast Res* 2021a;**21**:2.

Domenzain I, Sánchez B, Anton M *et al.* Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *bioRxiv* 2021b. DOI: 2021.03.05.433259.

Duarte NC, Herrgård MJ, Palsson BØ. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* 2004;**14**:1298–309.

Elsemman IE, Prado AR, Grigaitis P *et al.* Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. *bioRxiv* 2021. DOI: 2021.06.11.448029.

Ferreira R, Limeta A, Nielsen J. Tackling cancer with yeast-based technologies. *Trends Biotechnol* 2019;**37**:592–603.

Ferreira R, Skrekas C, Hedin A *et al.* Model-assisted fine-tuning of central carbon metabolism in yeast through dCas9-based regulation. *ACS Synth Biol* 2019;**8**:2457–63.

Förster J, Famili I, Fu P *et al.* Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Res* 2003;**13**:244–53.

Ghaffarinasab S, Motamedian E. Improving ethanol production by studying the effect of pH using a modified metabolic model and a systemic approach. *Biotechnol Bioeng* 2021;**118**:2934–46.

Hackett SR, Zanotelli VRT, Xu W *et al.* Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* 2016;**354**:aaf2786.

Haraldsdóttir HS, Cousins B, Thiele I *et al.* CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics* 2017;**33**:1741–3.

Heavner BD, Price ND. Comparative analysis of yeast metabolic network models highlights progress, opportunities for metabolic reconstruction. *PLoS Comput Biol* 2015;**11**:e1004530.

Heavner BD, Smallbone K, Barker B *et al.* Yeast 5: an expanded reconstruction of the Saccharomyces cerevisiae metabolic network. *BMC Syst Biol* 2012;**6**:1–13.

Heavner BD, Smallbone K, Price ND *et al.* Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database* 2013;**2013**:bat059.

Henriques D, Minebois R, Mendoza SN *et al.* A multiphase multiobjective dynamic genome-scale model shows different redox balancing among yeast species of the Saccharomyces genus in fermentation. *mSystems* 2021;**6**:e0026021.

Herrgård MJ, Lee B-S, Portnoy V *et al.* Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. *Genome Res* 2006;**16**:627–35.

Herrgård MJ, Swainston N, Dobson P *et al.* A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* 2008;**26**:1155–60.

Holland P, Bergenholm D, Börlin CS *et al.* Predictive models of eukaryotic transcriptional regulation reveals changes in transcription factor roles and promoter usage between metabolic conditions. *Nucleic Acids Res* 2019;**47**:4986–5000.

Irani ZA, Kerkhoven EJ, Shojaosadati SA *et al.* Genome-scale metabolic model of pichia pastoris with native and humanized glycosylation of recombinant proteins. *Biotechnol Bioeng* 2016;**113**:961–9.

Iranmanesh E, Asadollahi MA, Biria D. Improving l-phenylacetylcarbinol production in saccharomyces cerevisiae by *in silico* aided metabolic engineering. *J Biotechnol* 2020;**308**:27–34.

Jiang Y, Wang D, Yao Y *et al.* MULocDeep: a deep-learning framework for protein subcellular and suborganellar localization prediction with residue-level interpretation. *Comput Struct Biotechnol J* 2021;**19**:4825–39.

Kavšcek M, Bhutada G, Madl T *et al.* Optimization of lipid production with a genome-scale model of Yarrowia lipolytica. *BMC Syst Biol* 2015;**9**:1–13.

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

Kim J, Coradetti ST, Kim Y-M *et al.* Multi-omics driven metabolic network reconstruction and analysis of lignocellulosic carbon utilization in rhodosporidium toruloides. *Front Bioeng Biotechnol* 2021;1484.

Kim M, Park BG, Kim E-J *et al.* In silico identification of metabolic engineering strategies for improved lipid production in Yarrowia lipolytica by genome-scale metabolic modeling. *Biotechnol Biofuels* 2019;**12**:1–14.

Kim WJ, Kim HU, Lee SY. Current state and applications of microbial genome-scale metabolic models. *Curr Opin Syst Biol* 2017;**2**:10–8.

Kim Y, Kim GB, Lee SY. Machine learning applications in genome-scale metabolic modeling. *Curr Opin Syst Biol* 2021;**25**:42–9.

Kuepfer L, Sauer U, Blank LM. Metabolic functions of duplicate genes in Saccharomyces cerevisiae. *Genome Res* 2005;**15**:1421–30.

Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* 2012;**10**:291–305.

Li F, Chen Y, Qi Q *et al.* Genome scale modeling of the protein secretory pathway reveals novel targets for improved recombinant protein production in yeast. *bioRxiv* 2021a. DOI: 2021.10.16.464630.

Li F, Yuan L, Lu H *et al.* Deep learning based kcat prediction enables improved enzyme constrained model reconstruction. *bioRxiv* 2021b. DOI: 2021.08.06.455417.

Li G, Hu Y, Zrimec J *et al.* Bayesian genome scale modelling identifies thermal determinants of yeast metabolism. *Nat Commun* 2021c;**12**:190.

Lieven C, Beber ME, Olivier BG *et al.* MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol* 2020;**38**:272–6.

Liu G, Bergenholm D, Nielsen J. Genome-wide mapping of binding sites reveals multiple biological functions of the transcription factor Cst6p in *Saccharomyces cerevisiae*. *mBio* 2016;**7**:e00559–16.

Liu Q, Yu T, Li X *et al.* Rewiring carbon metabolism in yeast for high level production of aromatic chemicals. *Nat Commun* 2019;**10**:4976.

Liu Z, Moradi H, Shi S *et al.* Yeasts as microbial cell factories for sustainable production of biofuels. *Renew Sustain Energy Rev* 2021;**143**:110907.

Loira N, Dulermo T, Nicaud JM *et al.* A genome-scale metabolic model of the lipid-accumulating yeast Yarrowia lipolytica. *BMC Syst Biol* 2012;**6**:1–9.

Long MR, Ong WK, Reed JL. Computational methods in metabolic engineering for strain design. *Curr Opin Biotechnol* 2015;**34**:135–41.

Lopes H, Rocha I. Genome-scale modeling of yeast: chronology, applications and critical perspectives. *FEMS Yeast Res* 2017;**17**:50.

Lopes HJS, Bonturi N, Kerkhoven EJ *et al.* C/N ratio and carbon source-dependent lipid production profiling in Rhodotorula toruloides. *Appl Microbiol Biotechnol* 2020;**104**:2639–49.

Lu H, Kerkhoven EJ, Nielsen J. Multiscale models quantifying yeast physiology: towards a whole-cell model. *Trends Biotechnol* 2021;**S0167–7799**:00152–9.

Lu H, Li F, Sánchez BJ *et al.* A consensus *S. cerevisiae* metabolic model yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat Commun* 2019;**10**:1–13.

Lu H, Li F, Yuan L *et al.* Yeast metabolic innovations emerged via expanded metabolic network and gene positive selection. *Mol Syst Biol* 2021;**17**:e10427.

Marcišauskas S, Ji B, Nielsen J. Reconstruction and analysis of a Kluyveromyces marxianus genome-scale metabolic model. *BMC Bioinf* 2019;**20**:1–9.

Mesquita TJB, Campani G, Giordano RC *et al.* Machine learning applied for metabolic flux-based control of micro-aerated fermentations in bioreactors. *Biotechnol Bioeng* 2021;**118**:2076–91.

Mesquita TJB, Sargo CR, Fuzer JR *et al.* Metabolic fluxes-oriented control of bioreactors: a novel approach to tune micro-aeration and substrate feeding in fermentations. *Microb Cell Fact* 2019;**18**:1–17.

Mo ML, Palsson BØ, Herrgård MJ. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* 2009;**3**:37.

Nielsen J, Keasling JD. Engineering cellular metabolism. *Cell* 2016;**164**:1185–97.

Nilsson A, Nielsen J, Palsson BO. Metabolic models of protein allocation call for the kinetome. *Cell Syst* 2017;**5**:538–41.

Nookaew I, Jewett MC, Meechai A *et al.* The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* 2008;**2**:1–15.

Oftadeh O, Salvy P, Masid M *et al.* A genome-scale metabolic model of Saccharomyces cerevisiae that integrates expression constraints and reaction thermodynamics. *Nat Commun* 2021;**12**:1–10.

Olin-Sandoval V, Yu JSL, Miller-Fleming L *et al.* Lysine harvesting is an antioxidant strategy and triggers underground polyamine metabolism. *Nature* 2019:1.

Oliveira AP, Ludwig C, Picotti P *et al.* Regulation of yeast central metabolism by enzyme phosphorylation. *Mol Syst Biol* 2012;**8**:623.

Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol* 2010;**28**:245–8.

Österlund T, Nookaew I, Bordel S *et al.* Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst Biol* 2013;**7**:36.

Österlund T, Nookaew I, Nielsen J. Fifteen years of large scale metabolic modeling of yeast: developments and impacts. *Biotechnol Adv* 2012;**30**:979–88.

Paramasivan K, Kumar HN P, Mutturi S. Systems-based Saccharomyces cerevisiae strain design for improved squalene synthesis. *Biochem Eng J* 2019;**148**:37–45.

Patra P, Das M, Kundu P *et al.* Recent advances in systems and synthetic biology approaches for developing novel cell-factories in non-conventional yeasts. *Biotechnol Adv* 2021;**47**:107695.

Pereira F, Lopes H, Maia P *et al.* Model-guided development of an evolutionarily stable yeast chassis. *Mol Syst Biol* 2021;**17**:e10253.

Pinheiro MJ, Bonturi N, Belouah I *et al.* Xylose metabolism and the effect of oxidative stress on lipid and carotenoid production in Rhodotorula toruloides: insights for future biorefinery. *Front Bioeng Biotechnol* 2020;**8**:1008.

Qi Q, Li F, Yu R *et al.* Different routes of protein folding contribute to improved protein production in Saccharomyces cerevisiae. *mBio* 2020;**11**:1–12.

Qin J, Krivoruchko A, Ji B *et al.* Engineering yeast metabolism for the discovery and production of polyamines and polyamine analogues. *Nat Catal* 2021;**4**:498–509.

Rana P, Berry C, Ghosh P *et al.* Recent advances on constraint-based models by integrating machine learning. *Curr Opin Biotechnol* 2020;**64**:85–91.

Ryu JY, Kim HU, Lee SY. Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers. *Proc Natl Acad Sci USA* 2019:201821905.

Salvy P, Hatzimanikatis V. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nat Commun* 2020;**11**:30.

Sánchez BJ, Nielsen J. Genome scale models of yeast: towards standardized evaluation and consistent omic integration. *Integr Biol* 2015;**7**:846–58.

Sánchez BJ, Zhang C, Nilsson A *et al.* Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol* 2017;**13**:935.

Savojardo C, Martelli PL, Fariselli P *et al.* BUSCA: an integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res* 2018;**46**:W459–66.

Seif Y, Palsson BØ. Path to improving the life cycle and quality of genome-scale models of metabolism. *Cell Syst* 2021;**12**:842–59.

Taggart JC, Zauber H, Selbach M *et al.* Keeping the proportions of protein complex components in check. *Cell Syst* 2020;**10**:125–32.

Tiukova IA, Prigent S, Nielsen J *et al.* Genome-scale model of Rhodotorula toruloides metabolism. *Biotechnol Bioeng* 2019;**116**:3396–408.

Tomàs-Gamisans M, Andrade CCP, Maresca F *et al.* 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.

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. *Microb Biotechnol* 2018;**11**:224–37.

Torres P, Saa PA, Albiol J *et al.* Contextualized genome-scale model unveils high-order metabolic effects of the specific growth rate and oxygenation level in recombinant Pichia pastoris. *Metab Eng Commun* 2019;**9**:e00103.

Valli M, Grillitsch K, Grünwald-Gruber C *et al.* A subcellular proteome atlas of the yeast *Komagataella phaffii*. *FEMS Yeast Res* 2020;**20**:foaa001.

Vartiainen E, Blomberg P, Ilmén M *et al.* Evaluation of synthetic formaldehyde and methanol assimilation pathways in Yarrowia lipolytica. *Fungal Biol Biotechnol* 2019;**6**:1–13.

Wang G, Huang M, Nielsen J. Exploring the potential of Saccharomyces cerevisiae for biopharmaceutical protein production. *Curr Opin Biotechnol* 2017;**48**:77–84.

Wiederhold E, Veenhoff LM, Poolman B *et al.* Proteomics of saccharomyces cerevisiae organelles. *Mol Cell Proteomics* 2010;**9**:431–45.

Yang L, Yurkovich JT, King ZA *et al.* Modeling the multi-scale mechanisms of macromolecular resource allocation. *Curr Opin Microbiol* 2018;**45**:8–15.

Ye C, Xu N, Gao C *et al.* Comprehensive understanding of *Saccharomyces cerevisiae* phenotypes with whole-cell model WM_S288C. *Biotechnol Bioeng* 2020;**117**:1562–74.

Yu R, Campbell K, Pereira R *et al.* Nitrogen limitation reveals large reserves in metabolic and translational capacities of yeast. *Nat Commun* 2020;**11**:1881.

Yu R, Nielsen J. Yeast systems biology in understanding principles of physiology underlying complex human diseases. *Curr Opin Biotechnol* 2020;**63**:63–9.

Zhang J, Petersen SD, Radivojevic T *et al.* Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism. *Nat Commun* 2020; **11**:4880.

Zomorrodi AR, Maranas CD. Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. *BMC Syst Biol* 2010;**4**:1–15.

Zuñiga C, Levering J, Antoniewicz MR *et al.* Predicting dynamic metabolic demands in the photosynthetic eukaryote *Chlorella vulgaris*. *Plant Physiol* 2018;**176**:450–62.

Zuñiga C, Tibocha-Bonilla JD, Betenbaugh MJ. Kinetic, metabolic, and statistical analytics: addressing metabolic transport limitations among organelles and microbial communities. *Curr Opin Biotechnol* 2021;**71**:91–7.