Review

Multiscale models quantifying yeast physiology: towards a whole-cell model

Hongzhong Lu,[1] Eduard J. Kerkhoven,[1] and Jens Nielsen [1,2,*]

The yeast *Saccharomyces cerevisiae* is widely used as a cell factory and as an important eukaryal model organism for studying cellular physiology related to human health and disease. Yeast was also the first eukaryal organism for which a genome-scale metabolic model (GEM) was developed. In recent years there has been interest in expanding the modeling framework for yeast by incorporating enzymatic parameters and other heterogeneous cellular networks to obtain a more comprehensive description of cellular physiology. We review the latest developments in multiscale models of yeast, and illustrate how a new generation of multiscale models could significantly enhance the predictive performance and expand the applications of classical GEMs in cell factory design and basic studies of yeast physiology.

Emergence of multiscale models for yeast

Being widely used as a **microbial cell factory** (see Glossary), metabolism in the yeast *Saccharomyces cerevisiae* has been extensively studied and engineered with the purpose of improving its properties. To this end, various types of computational models have been leveraged to quantitatively characterize yeast physiology and to guide **metabolic engineering**. Among these, **genome-scale metabolic models (GEMs)** have been most widely used (Box 1). A benefit of this **constraint**-based modeling concept is that it readily allows continuous model expansion when new experimental evidence becomes available, and consequently GEMs for *S. cerevisiae* have been frequently updated with more than 14 published versions between 2003 and 2019 [1,2], thereby typically yielding improved model performance.

Even though GEMs are instrumental to investigate what the yeast metabolic network can achieve, a drawback of such classical GEMs is that they only consider the stoichiometry of the metabolic network. In reality, the fluxes through a metabolic network are constrained by many more aspects that are by design neglected in classical GEMs, such as metabolic regulation caused by gene expression and post-translational modifications, as well as information about enzymatic properties defined, among others, by the protein 3D structures. Incorporating information about these cellular processes and protein structures would allow integrative analysis of **multilayer omics** data, thereby enabling the models to provide mechanical insight into the basic principles of the regulation and evolution of complex cellular metabolism. This has been recognized, and has resulted in the gradual development of **multiscale models** [3] through the addition of enzyme kinetics, 3D structures, and **heterogeneous networks** into classical GEMs, thereby laying the groundwork for holistic and accurate simulations of whole-cell behaviors.

[7], and metabolic engineering [10]. They also represent significant breakthroughs in exploring the complex relations between cellular genotype and phenotype. Similar to bacterial multiscale models, an ME-model [11] and a whole-cell model [12] have recently been constructed and evaluated for *S. cerevisiae*. Although these models are significant advances that place eukaryal multiscale models on a par with their bacterial equivalents, the models still lack integration of regulatory information from heterogeneous networks, not only for yeast but also for various other model organisms. In addition, significant challenges remain for reconstructing comprehensive multiscale models for non-model species owing to the lack of data.

We review here progress in yeast multiscale metabolic modeling and show how multiscale models are constructed by gradually encompassing additional constraints. We demonstrate how this approach greatly enhances model predictions, thereby accelerating model-based biological discoveries and pioneering work in systematic metabolic engineering. Because these bases are founded on classical GEMs, we first briefly outline the latest developments in GEMs and heterogeneous network reconstructions for yeast. In particular, strategies to expand yeast GEMs with enzyme parameters and heterogeneous networks to yield multiscale models are evaluated and discussed. Next, applications of GEMs and multiscale models for yeast in multi-omics integrative analysis and *in silico* cell factory design are highlighted. Finally, strategies and directions to develop future generations of multiscale models are set forth.

Continuous growth of GEMs for model and non-model yeast species

When the whole-genome sequence of *S. cerevisiae* became available in 1996 [13] it was possible to reconstruct the first GEM for yeast, published in 2003 as iFF708 [14]. Since then a series of updated *S. cerevisiae* GEMs have been released [1,2], where each new version has represented a gradual improvement of previous models. Among these models, the consensus yeast GEMs (Yeast1 to Yeast8 [15–20]) are a series of community-curated models, and progress made before Yeast7.6 has been intensively reviewed previously [2,21]. In this section we therefore only evaluate recent developments in *S. cerevisiae* consensus GEMs since Yeast7.6, which was published in 2013 [18]. Beyond Yeast7.6, the coverage and quality in description of lipid, flavor, **cofactor**, and substrate metabolism has been improved in recent years. The large number of unique metabolite species in lipid metabolism make this part of the metabolic network not straightforward to represent, in particular in a format that is readily constrained by measured lipidomics data. Based on detailed curation of lipid metabolism [22], novel approaches have been developed for modeling lipid metabolism in yeast, for example, the SLIMEr [23] formulation of lipid reactions and an alternative object-oriented stochastic strategy [24]. The pathways relevant for flavor formation were recently curated and extended in *S. cerevisiae* GEM iWS902, which provided mechanical insights underlying aroma formation during industrial applications [25]. To cover cofactor metabolism, a network covering yeast iron metabolism was recently integrated into Yeast7.6 [26]. In Yeast8, the reported cofactor concentrations were further used to update biomass composition, and consequently related sub-pathways that did not carry metabolic fluxes [19] in previous GEMs were now activated. Yeast8 was further expanded to simulate a wider range of substrate utilization based on *in vivo* substrate usage data from Biolog experiments. Furthermore, 13 additional aroma compounds and their associated reactions were added to extend the application of Yeast8 to industrial wine production [27]. Notably, the development of the consensus yeast GEM has been reproducibly tracked since Yeast8 through Git- and GitHub-hosted versioning systems, enabling community-driven model improvements and accessibility to the wide research community (https://github.com/SysBioChalmers/yeast-GEM). So far this has resulted in engagement of 11 researchers and the release of 23 updated versions of the model, including the current version 8.4.2.

Glossary

Cofactor: a non-protein compound that is necessary to allow or improve the catalytic efficiency of an enzyme in specific biochemical reactions.
Constraints: *in vivo* metabolic fluxes cannot take on any value but are constrained to minimum and maximum values. For example, an irreversible reaction cannot have a negative flux value, while a cell is not able to take up nutrients at an infinitely high rate. The constraints for each reaction are dictated by genetics, environment, network topology, and physicochemical laws, which can be regarded as different types of constraints. If substrate-uptake rates are experimentally measured, they can be used to set the lower and upper bounds of the corresponding transport reactions in a metabolic model.
Genome-scale metabolic models (GEMs): when a whole-genome annotation is available, all metabolic enzymes present in a specific organism can be identified and combined to reconstruct a GEM that encompasses the metabolic network and all gene–protein–reaction associations. The GEM is an organism-specific knowledgebase, but can also be used to predict cellular phenotypes under various constraints, for example, exchange reaction rates measured from fermentation experiments.
Heterogeneous networks: in addition to the cellular metabolic networks that can be described in genome-scale metabolic models, various other molecular networks are present in the cell, including transcriptional regulatory networks and signal transduction networks. Although these networks affect cellular metabolism in distinct ways, they furthermore interact with each other to determine the final phenotypic output.
Metabolic engineering: a strategy where multiple rounds of gene manipulation are employed guided by omic analysis, flux simulation, and/or *in silico* strain design, with the objective of optimizing a microbial cell factory to overcome bottlenecks in the production of a desired product.
Metabolism and expression (ME)-models: in contrast to GEMs, ME-models combine a genome-scale description of metabolism with stoichiometric representations of gene transcription and protein translation. In comparison to GEMs, ME-models are

Although the GEM of *S. cerevisiae* as a model yeast has been most extensively curated, GEMs have also been constructed and applied for many other yeast species, including but not limited to *Lachancea kluyveri* [28], *Exophiala dermatitidis* [29], *Issatchenkia orientalis* [30], and *Cutaneotrichosporon oleaginosus* [31]. Progress in reconstructing classical GEMs for these non-model yeast species is thoroughly reviewed elsewhere [1,2,32]. Although the remainder of this review will mostly focus on *S. cerevisiae*, the methods and approaches discussed can also be applied to GEMs of non-model yeasts, thereby providing a solid basis for the development of the next generation of multiscale models.

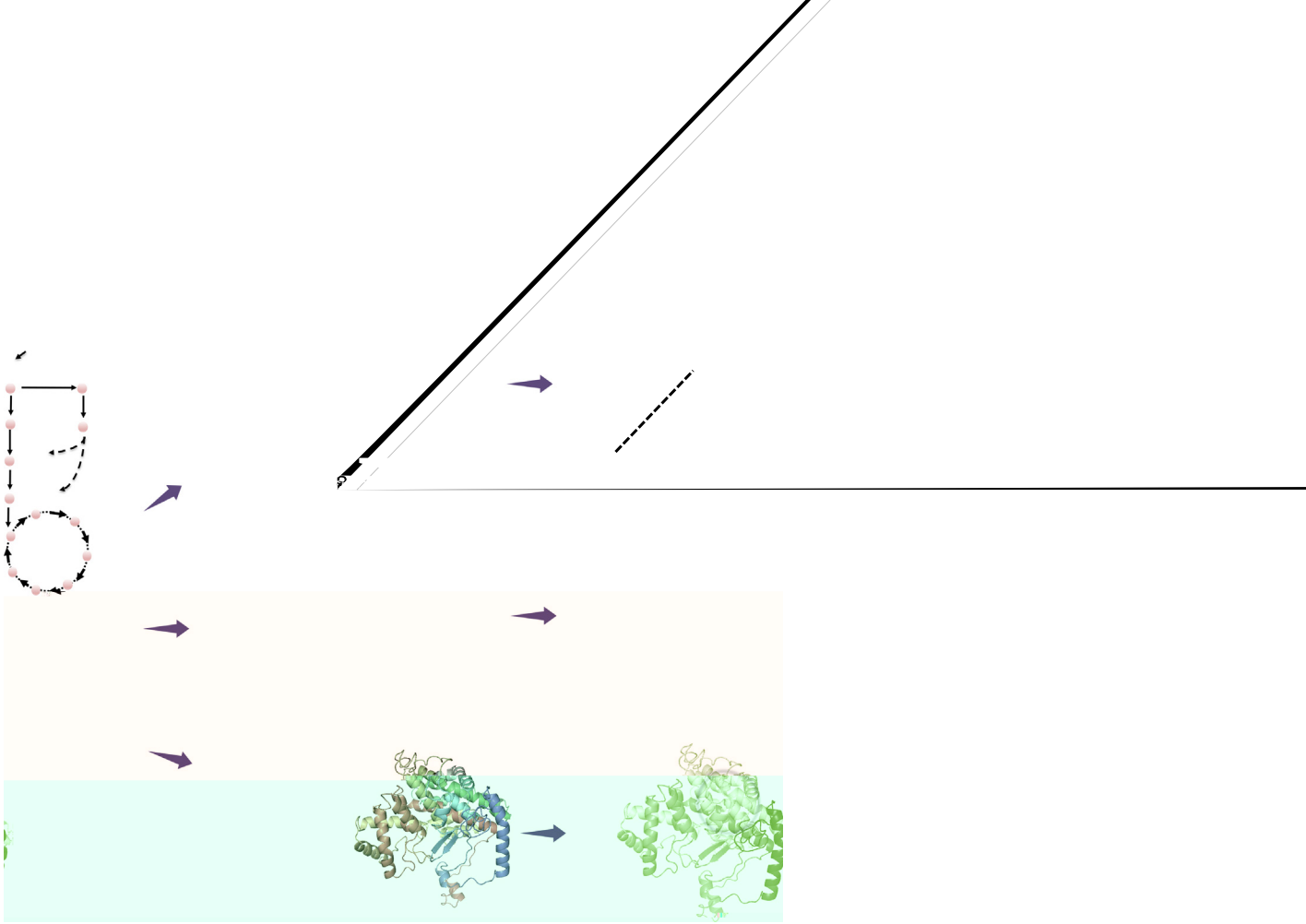
Enhanced yeast GEMs with constraints from kinetics and proteome

Classical GEMs mainly rely on flux balance constraints, but the distribution of metabolic flux through different branches of the metabolic network is additionally determined by enzyme kinetics and enzyme abundances. Furthermore, enzyme activities are dictated by their 3D protein structures which by themselves are linked to their primary structure (i.e., protein sequence). As distinct phenotypes can be governed by variations in enzyme activities, the mapping of protein sequence variations to altered fluxes can provide new insight into these connections. The integration of reaction kinetics, protein abundances, and 3D protein structures with GEMs will therefore enhance its prediction capabilities (Figure 1).

Adding reaction kinetics

Kinetic models have long been developed to describe yeast metabolism. In their early stages such models primarily encompassed specific sub-pathways, for example, trehalose metabolism [33], glycolysis [34], and sphingolipid biosynthesis [35]. Although these models can predict cellular dynamics under environmental or genetic perturbations, efforts have been made to expand their limited scope towards describing larger metabolic networks. As part of this, a kinetic model was established that considered the core metabolic pathways including the glycolysis pathway, the pentose phosphate pathway, an2(],)5.2(6.9(i)15.576.3(th15.576.(h)-16.8.5(t(un)-12i(e)21(6.(h)8.5(n)-16)17.6(.8.5(t)5.2(6.(un)-12i(e)15.576.1.(un)-

aspect that cellular metabolism has a limited catalytic capacity that is dictated by constraints on the proteome, and they thereby display strong potential to improve the prediction capabilities of GEMs [41,43



rates, and overflow metabolism [11]. Altogether, these proteome-constrained models are powerful extensions beyond classical GEMs and can simulate and characterize yeast metabolism, which was not possible when only considering the metabolic network stoichiometry.

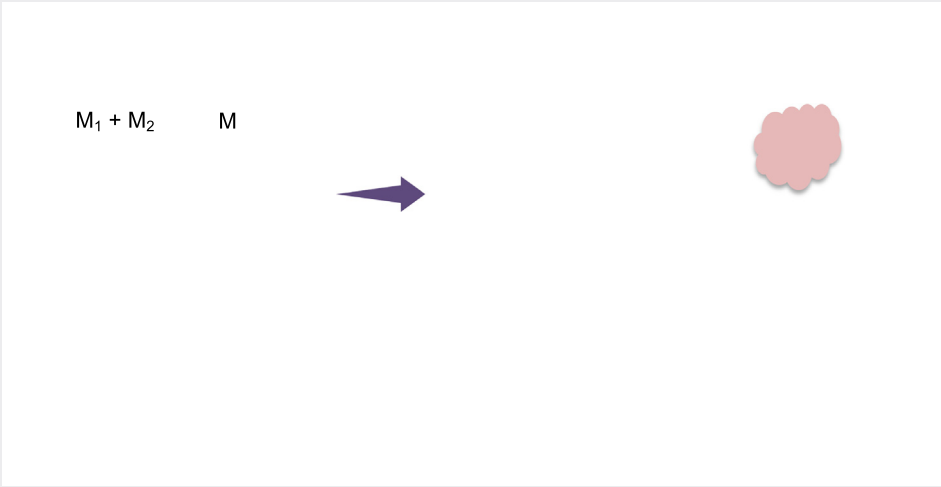
Incorporation of protein 3D structure information

Although enzyme kinetics and abundances influence the fluxes through the reactions that constitute a metabolic model, consideration of protein 3D structures would enable exploration of how sequence differences might affect interactions between proteins and metabolites as well as protein activity and stability. Protein structure information can be connected with GEMs through the use of the gene–protein–reaction relationships (GPRs) that are an essential part of GEMs [49], and thereby provide additional constraints for model predictions (Figure 1).

So far, the lack of full integration of yeast protein 3D structures with GEMs is largely due to the insufficient quality and coverage of 3D structures for all metabolic enzymes in GEMs. The first attempt to perform large-scale protein 3D structure modeling of *S. cerevisiae* was in 1998 with the reconstruction of all-atom 3D models for 1071 (17%) of the yeast proteins [50]. Because

Box 2. Enzyme-constrained GEMs (ecGEMs)

What separates classical GEMs (Box 1) from ecGEMs is that both enzyme kinetics and abundances are considered in the latter. Because no nonlinear relationships are introduced when adding enzyme constraints, these ecGEMs can be simulated and analyzed by using the same constraint-based algorithms that are applied to GEMs (Box 1). In contrast to classical GEMs, in ecGEMs fluxes through each enzyme-catalyzed reaction are constrained by their turnover numbers, namely enzyme-specific k_{cat} values multiplied by the enzyme abundances (Figure I). Meanwhile, the total quantity of enzymes that can catalyze all metabolic reactions is constrained by a protein-pool pseudo-metabolite, whose usage is restricted by an upper bound that is in accordance with experimental total protein measurements. When no quantitative proteomics data are available, only the total protein-pool usage is constrained, while in the model the amount of protein can be freely distributed across all enzymes (Figure I). If quantitative proteomics data are available for a condition of interest, then individual enzyme usages can be constrained to their corresponding measured abundances. GECKO, a MATLAB and Python toolbox, has been developed to construct and simulate such ecGEMs [45]. The latest iteration of this toolbox, version 2 [117], has placed particular focus on the reconstruction of ecGEMs for non-model organisms.



protein 3D structures are more conserved than amino acid sequences, this showed that protein function prediction can be aided by clues from the folds, active sites, and binding sites extracted from the protein 3D structures. More recently, the Rosetta *de novo* structure-prediction method was used to predict the structure of 3338 protein domains that were parsed from the whole yeast proteome, among which 581 domains could be assigned to novel Structural Classification of Proteins (SCOP) superfamilies [51].

Experimentally determined yeast protein structures are readily available from the Protein Data Bank (PDB) database [52], while homology-derived protein structures can be queried from various sources including SWISS-MODEL [53] and Modbase [54]. As a typical example, a total of 3846 experimentally determined protein structures are available from the PDB database for 1543 *S. cerevisiae* S288c proteins. Some of these structures are at low resolution or include mutations compared to the reference sequence, such that quality analysis and homology modeling are necessary to ensure that advanced models are based on high-quality structures [55]. Meanwhile, high-quality experimentally determined yeast protein structures continue to accumulate, and homology modeling approaches have significantly advanced. It has now

become feasible to collect protein 3D structures at the proteome scale for *S. cerevisiae,* and this forms the basis for building GEMs constrained (or accompanied) by additional parameters from protein 3D structures in the near future.

Connecting yeast GEMs with heterogeneous cellular networks

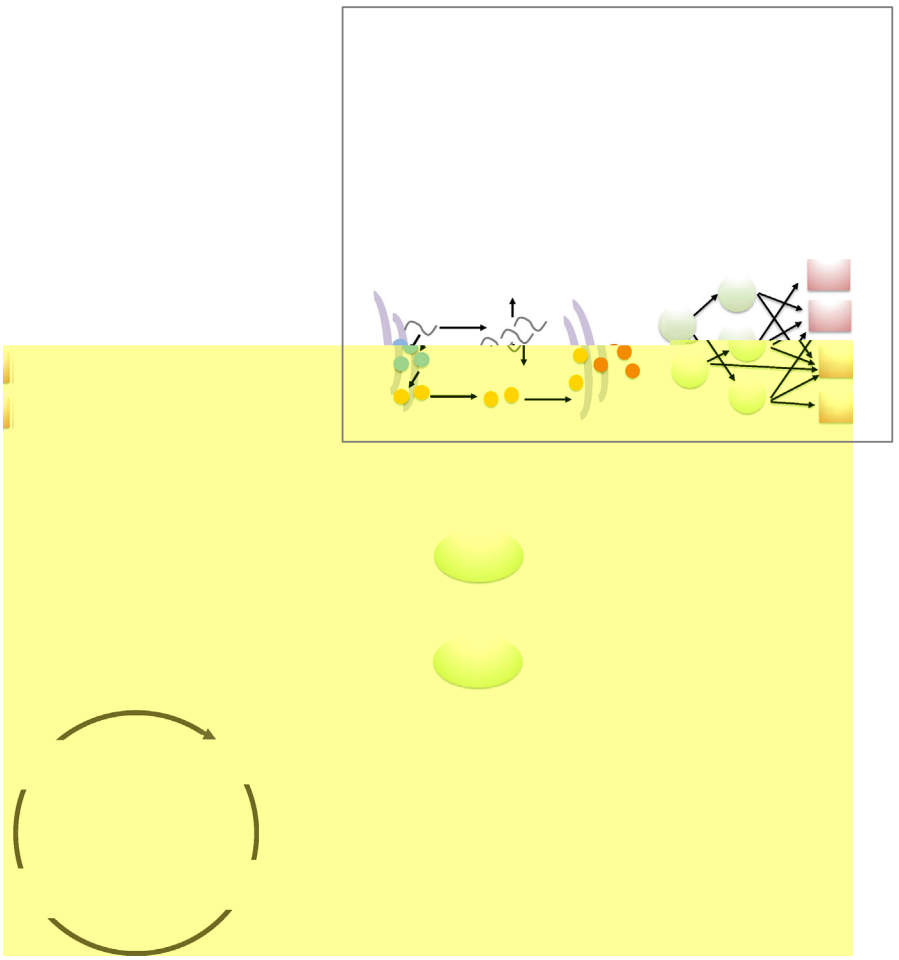
Cellular metabolic activities are tightly regulated at multiple levels through the interaction of various heterogeneous molecular networks, and even ecGEMs alone are not able to simulate the effects of complex regulation. Thus, the multiscale models are essential to capture the intricate metabolism by integrating heterogeneous cellular networks including signal transduction networks, transcriptional regulatory networks (TRNs), and protein secretion pathways with GEMs (Figure 2).

Stress- and nutrition-related signal transduction networks

The inclusion of signal transduction networks in a multiscale model would allow it to predict cellular responses to external stimuli or stress. As one of the earliest examples, a model of osmoregulation was integrated with a metabolic model to describe the cellular response to hyperosmotic shock [56]. Since then, computational models for many more signal transduction networks have been constructed, including the signaling networks for MAPK [57], Snf1 [58], and ion regulation [59]. Moreover, the complexity of signaling network models has increased by considering multiple stress and nutrition stimuli simultaneously. As one attempt, all six major stress-response pathways related to ion homeostasis, nutrient adaptation, osmotic stress, oxidative, heat shock, and pheromone stress response were merged into a holistic molecular interaction map [60]. Interestingly, this comprehensive map showed that yeast stress-response pathways are organized into bow-tie structures, and complex-mediated reversible reactions obtained through network motif analysis play a unique role in the regulation of stress responses. An integrated nutrient signaling network was recently built for yeast, and this could be used to predict nutrient-responsive transcription factor (TF) activities in mutant strains under nutrient shifts [61]. However, all these newly developed stress-related signal transduction network models were not coupled with yeast GEMs, and they were therefore not able to explore how cellular metabolism was quantitatively regulated in response to these external stresses. By contrast, Boolean modeling of a glucose-sensing regulatory pathway has recently been successfully integrated with a small enzyme-constrained metabolic model for yeast, and this could elucidate how dynamic regulation through a signaling pathway affects cellular metabolism and results in improved enzyme utilization predictions for both respiratory and mixed metabolism [62].

Transcriptional regulatory networks

Regulation of gene transcription influences metabolism on a global scale [63], rendering it important to develop TRNs that accompany yeast GEMs to allow comprehensive simulations of metabolic network regulation. The functional annotation of *S. cerevisiae* TFs is ever increasing and is catalogued in several public databases, such as the *Saccharomyces* Genome Database (SGD) [64] and YEASTRACT [65], which are instrumental for building high-quality TRN models. As an example, a comprehensive TRN model was built for *S. cerevisiae* based on the SGD database [64], consisting of 186 TFs and 5727 target genes, involving 28 260 regulatory interactions [66]. Large-scale RNA-seq data from divergent conditions is another important source to infer TRNs for *S. cerevisiae*. With the aid of machine learning, a global TRN comprising 12 228 interactions was built based on single-cell RNA-seq measurements on 38 285 individual cells under 11 different environmental conditions [67]. Timecourse gene expression data has also been used to develop a whole-cell transcriptional model which could predict and validate new transcriptional interactions [68]. The high-confidence TRN map of yeast could be expanded based on multiple datasets by using dual threshold optimization and network inference algorithms, resulting with a high-confidence yeast TRN made up of 96 TFs, 1686 target genes, and 3268 regulatory interactions [69].



High-quality yeast TRNs set a solid base to formulize integrated models, even though combining TRNs with GEMs remains challenging. Multiple novel methods that couple TRNs with GEMs have recently been reviewed [70]. Several of these novel algorithms have been used for yeast, and they can roughly be divided into two main approaches. (i) Based on experimental data, Boolean rules reflect the interactions between TFs and their target genes, and thereby the reactions in the GEM

can be switched on or off based either on the TRN or on gene expression data [71]. (ii) Using the probabilistic regulation of metabolism (PROM) [72] approach, the probabilities to characterize gene states and gene–TF interactions are introduced. According to these probabilities, the maximum fluxes through specific reactions are tuned to represent the effects of TF regulation. This probabilistic framework successfully combined TRNs with GEMs, leading to more accurate growth prediction for *S. cerevisiae* [73].

Protein secretion pathway

In yeast, the protein secretion pathway encompasses numerous distinct steps that are catalyzed by >100 cellular proteins [74], which together can determine not only cellular phenotypes but also the production yields of heterologous proteins [75]. By describing 16 subsystems that cover all the secretory machinery processes from translocation to sorting, a stoichiometric model of the *S. cerevisiae* protein secretion pathway was reconstructed [76

with RNA-seq analysis showed that mitochondrial fluxes are positively associated with changes at the transcript level, suggesting transcriptional regulation [87]. Similarly, combining dynamic flux simulations with mRNA measurements identified Zwf1p (glucose 6-phosphate dehydrogenase) as a bottleneck in the production of ethanol when xylose is the substrate [88]. Enzyme-constrained models such as ecYeast7.6 can furthermore predict both intracellular enzyme usages [45] and fluxes through each reaction, thereby helping to illuminate the consistency and discrepancy between measured and predicted protein abundances (or fluxes). More recently, gene expression and fluxomic data for 1143 *S. cerevisiae* mutants were combined with advanced machine-learning procedures to improve predictions of yeast cell growth, thereby outperforming the predictions that could be made from single datasets [89].

In turn, integration of omics constraints can enhance predictions with both classical GEMs and multiscale models. For instance, constraining the reaction bounds based on abundances from proteomics and transcriptomics measurements improved flux predictions with iMM904 [90]. Quantitative proteomics could be used directly to constrain an enzyme-constrained model of yeast (i.e., ecYeast7) to increase its prediction abilities [45], in which the maximal flux of one reaction could be constrained by both enzyme abundance and its turnover number (Box 2). As one application, simulations with ecYeast7.6 under a series of growth rates helped to identify key enzymes controlling fluxes towards amino acid synthesis [91]. With kinetic models, multi-omics datasets including metabolomics, proteomics, and fluxomics can be fully mined and exploited to tune the model structure and parameters, thereby significantly improving model predictions [40,92]. Omics data have been instrumental for building condition-specific GEMs, in particular of human tissues and cells [93] as part of investigations into disease. However, this approach has so far rarely been used to build condition-specific GEMs for yeast.

Despite the early stage of their development, GEMs enhanced with protein 3D structures have shown potential to connect structural bioinformatics with the systems biology paradigm, and this approach has been able to generate new biological insights into the role of specific residue mutations in cellular fitness [49,94,95]. Protein 3D structures enable mutation cluster analysis based on mutation and structural information (Figure 3B), and such an analysis with proYeast8^{DB} [19] was able to relate potential mutation signatures to specific traits of wine production and substrate utilization. In studies of human disease, mutation cluster analysis with protein 3D structures is more widely used, and for instance can help to identify driver mutations for the formation of various types of cancers [96,97].

Cell factory design and optimization

Various computational methods, including MOMA [98], OptKnock [99], OptForce [100], and FSEOF [101], that have been reviewed previously [102], have been developed for using classical GEMs in the design and optimization of microbial cell factories (Figure 3C). Indeed, *in silico* strain design with yeast GEMs has been used for a wide range of products, including succinic acid [30], dicarboxylic acid [103], L-phenylacetylcarbinol [104], lipid [105], and human superoxide dismutase [106], which have also been summarized in detail [32].

Gradually, multiscale models of yeast are beginning to show their value in identifying rational targets for systematic metabolic engineering. First, a multiscale model including regulation could be employed with novel methods such as IDREAM [73] and OptRAM [107] to predict TF targets for genetic engineering to improve the production of multiple products because the roles of TF in regulation could be evaluated using the integrated models. Furthermore, ecYeast7.6 has been used to rank gene targets for synthetic biology [45] by flexibly adjusting protein abundances or k_{cat} values and evaluating the response of these interventions. Similarly, the enzyme- and

temperature-constrained GEM etcYeast7.6 has identified key enzymes associated with heat-tolerance, and one of the candidates, ERG1, was experimentally verified to affect heat-tolerance [47]. In addition, by using a kinetic model of yeast, the potential enzyme targets for improving the flux of desired products can be prioritized via flux control coefficient analysis or correlation analysis between the predicted enzyme usages and the relevant product formation rates [40].

Concluding remarks and future perspectives

Multiscale models enable interrogation of biological complexity at multidimensional levels instead of limiting them to only one level in the central dogma of biology. Despite significant advances in yeast multiscale models, there are still several challenges that could hinder progress in the near future, in particular when expanding multiscale model reconstruction to non-model yeast species (see Outstanding questions). First, the number of enzymes for which high-quality experimentally determined kinetic parameters are available is still limited [108], particularly for non-model yeast species. However, developments in machine learning and novel parameter inference procedures may pave the way for the prediction of unknown parameters of enzyme/protein for less well studied yeast species [109]. Second, model reconstruction needs large numbers of standardized datasets, such as growth data obtained from continuous cultivation and absolute quantitative protein abundance data from mass spectrometry measurements. However, the availability of such datasets is limited for most yeast species beyond *S. cerevisiae*. The collection of high-quality growth and omics datasets for non-model yeast species under standard conditions would therefore be very valuable and alleviate the shortage of essential data. Lastly, the metabolic models of *S. cerevisiae* and other yeast strains still require additional curation to yield a more complete coverage of metabolic sub-pathways, and the resulting high-quality models will act as new cornerstones to build comprehensive multiscale models for yeast. To solve the issue, automatic protein function prediction [110] together with evidence from omics measurements and molecular experiments could help to increase the metabolic coverage. In addition, inconsistencies between model predictions and *in vitro* experimental results will provide clues for further iterative improvement of model quality.

Even though a whole-cell model WM_S288C has been developed based on *S. cerevisiae* GEM iTO977 [12], challenges remain in developing a fully functional whole-cell model for yeast from the aforementioned multiscale models. Combining high-quality TRNs with metabolic models will certainly help to illustrate how transcriptional regulation affects cellular metabolism through resource allocation under various genetic or environmental perturbations. However, TRNs and stress-response networks have not yet been integrated with yeast ME- or whole-cell models because computational toolboxes to couple multiple types of heterogeneous networks are still lacking. Some novel coupling algorithms and simulation strategies, such as Bayesian metamodeling [111] and multi-algorithmic simulators [112], have recently been updated and evaluated, and these approaches could be used to integrate more interconnected cellular processes with yeast whole-cell models. Overall, we anticipate that progress in measurements and algorithms will promote yeast ME-modeling and whole-cell models to provide a more powerful computation platform that will play a prominent role both in fundamental studies of yeast and in cell factory design.

Outstanding questions

How can we efficiently develop a new generation of multiscale models for non-model yeasts?

How can we further develop more advanced yeast whole-cell models by integrating more heterogeneous cellular networks?

How can we develop genome-scale kinetic models for yeasts?

How can we use the various omics data to tune model parameters for better prediction?

How can we exploit complex models to carry out intelligent cell factory design?

How can we integrate different types of regulatory networks with GEMs during model simulations?

How can we predict the activity of enzymes with specific mutations and use these data as input for the model simulation?

Acknowledgments

This project received funding from the Novo Nordisk Foundation (grant NNF10CC1016517), the Knut and Alice Wallenberg Foundation, and the EU Horizon 2020 research and innovation projects DD-DeCaF and CHASSY (grant agreements 686070 and 720824).

Declaration of interests

The authors declare no conflicts of interest.

References

1. Domenzain, I. *et al.* (2021) Evaluating accessibility, usability and interoperability of genome-scale metabolic models for diverse yeasts species. *FEMS Yeast Res.* 21, foab002
2. Lopes, H. and Rocha, I. (2017) Genome-scale modeling of yeast: chronology, applications and critical perspectives. *FEMS Yeast Res.* 17, fox050
3. Zhang, J.D. *et al.* (2019) Multiscale modelling of drug mechanism and safety. *Drug Discov. Today* 25, 519–534
4. O'Brien, E.J. *et al.* (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* 9, 693
5.

48. Chen, Y. *et al.* (2021) Yeast optimizes metal utilization based on metabolic network and enzyme kinetics. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2020154118
49. Mih, N. and Palsson, B.O. (2019) Expanding the uses of genome-scale models with protein structures. *Mol. Syst. Biol.* 15, e8601
50. Sánchez, R. and Sali, A. (1998) Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc. Natl. Acad. Sci. U. S. A.* 95, 13597–13602
51. Malmstrom, L. *et al.* (2007) Superfamily assignments for the yeast proteome through integration of structure prediction with the gene ontology. *PLoS Biol.* 5, e76
52. Rose, P.W. *et al.* (2017) The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* 45, 271–281
53. Waterhouse, A. *et al.* (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, 296–303
54. Pieper, U. *et al.* (2011) ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.* 39, D465–D474
55. Mih, N. *et al.* (2018) ssbio: a Python framework for structural systems biology. *Bioinformatics* 34, 2155–2157
56. Klipp, E. *et al.* (2005) Integrative model of the response of yeast to osmotic shock. *Nat. Biotechnol.* 23, 975–982
57. Zou, X. *et al.* (2008) Modeling specificity in the yeast MAPK signaling networks. *J. Theor. Biol.* 250, 139–155
58. Lubitz, T. *et al.* (2015) Network reconstruction and validation of the Snf1/AMPK pathway in baker's yeast based on a comprehensive literature review. *NPJ Syst. Biol. Appl.* 1, 15007
59. Ke, R. *et al.* (2013) An integrative model of ion regulation in yeast. *PLoS Comput. Biol.* 9, e1002879
60. Kawakami, E. *et al.* (2016) Network analyses based on comprehensive molecular interaction maps reveal robust control structures in yeast stress response pathways. *NPJ Syst. Biol. Appl.* 2, 15018
61. Jalihal, A.P. *et al.* (2020) Modeling and analysis of the macronutrient signaling network in budding yeast. *bioRxiv* Published online February 15, 2020. https://doi.org/10.1101/2020.02.15.950881
62. Österberg, L. *et al.* (2021) A novel yeast hybrid modeling framework integrating Boolean and enzyme-constrained networks enables exploration of the interplay between signaling and metabolism. *PLoS Comput. Biol.* 17, e1008891
63. Chechik, G. *et al.* (2008) Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nat. Biotechnol.* 26, 1251–1259
64. Cherry, J.M. *et al.* (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* 40, D700–D705
65. Monteiro, P.T. *et al.* (2019) YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. *Nucleic Acids Res.* 48, D642–D649
66. Sun, D.Q. *et al.* (2019) Spatial organization of the transcriptional regulatory network of *Saccharomyces cerevisiae*. *FEBS Lett.* 593, 876–884
67. Jackson, C.A. *et al.* (2020) Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. *eLife* 9, e51254
68. Hackett, S.R. *et al.* (2020) Learning causal networks using inducible transcription factors and transcriptome-wide time series. *Mol. Syst. Biol.* 16, e9174
69. Kang, Y. *et al.* (2020) Dual threshold optimization and network inference reveal convergent evidence from TF binding locations and TF perturbation responses. *Genome Res.* 30, 459–471
70. Cruz, F. *et al.* (2020) A review of methods for the reconstruction and analysis of integrated genome-scale models of metabolism and regulation. *Biochem. Soc. Trans.* 48, 1889–1903
71. Herrgard, M.J. *et al.* (2006) Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res.* 16, 627–635
72. Chandrasekaran, S. and Price, N.D. (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* 107, 17845–17850
73. Wang, Z. *et al.* (2017) Combining inferred regulatory and reconstructed metabolic networks enhances phenotype prediction in yeast. *PLoS Comput. Biol.* 13, e1005489
74. Delic, M. *et al.* (2013) The secretory pathway: exploring yeast diversity. *FEMS Microbiol. Rev.* 37, 872–914
75. Huang, M. *et al.* (2018) Engineering the protein secretory pathway of *Saccharomyces cerevisiae* enables improved protein production. *Proc. Natl. Acad. Sci. U. S. A.* 115, E11025–E11032
76. Feizi, A. *et al.* (2013) Genome-scale modeling of the protein secretory machinery in yeast. *PLoS One* 8, e63284
77. Irani, Z.A. *et al.* (2016) Genome-scale metabolic model of *Pichia pastoris* with native and humanized glycosylation of recombinant proteins. *Biotechnol. Bioeng.* 113, 961–969
78. Gu, C. *et al.* (2019) Current status and applications of genome-scale metabolic models. *Genome Biol.* 20, 121
79. O'Brien, E.J. *et al.* (2015) Using genome-scale models to predict biological capabilities. *Cell* 161, 971–987
80. McCloskey, D. *et al.* (2013) Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol. Syst. Biol.* 9, 661
81. Bordbar, A. *et al.* (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* 15, 107–120
82. Kerkhoven, E.J. *et al.* (2016) Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. *NPJ Syst. Biol. Appl.* 2, 16005
83. Tiukova, I.A. *et al.* (2019) Genome-scale model of *Rhodotorula toruloides* metabolism. *Biotechnol. Bioeng.* 116, 3396–3408
84. Marcišauskas, S. *et al.* (2019) Reconstruction and analysis of a *Kluyveromyces marxianus* genome-scale metabolic model. *BMC Bioinforma.* 20, 551
85. Correia, K. and Mahadevan, R. (2020) Pan-genome-scale network reconstruction: harnessing phylogenomics increases the quantity and quality of metabolic models. *Biotechnol. J.* 15, 1900519
86. Hongzhong, F. *et al.* (2020) SysBioChalmers/Yeast-Species-GEMs: Yeast_species_GEM v0.3.1 (Version v0.3.1). *Zenodo* Published online December 4, 2020. https://doi.org/10.5281/zenodo.4305960
87. Lahtvee, P.J. *et al.* (2017) Absolute quantification of protein and mRNA abundances demonstrate variability in gene-specific translation efficiency in yeast. *Cell Syst.* 4, 495–504
88. Hohenschuh, W. *et al.* (2020) Using high-throughput data and dynamic flux balance modeling techniques to identify points of constraint in xylose utilization in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 1, 58–75
89. Culley, C. *et al.* (2020) A mechanism-aware and multiomic machine-learning pipeline characterizes yeast cell growth. *Proc. Natl. Acad. Sci. U. S. A.* 117, 18869–18879
90. Tian, M. and Reed, J.L. (2018) Integrating proteomic or transcriptomic data into metabolic models using linear bound flux balance analysis. *Bioinformatics* 34, 3882–3888
91. Xia, J. *et al.* (2021) Proteome allocations change linearly with specific growth rate of *Saccharomyces cerevisiae* under glucose-limitation. *ResearchSquare* Published online May 18, 2021. http://dx.doi.org/10.21203/rs.3.rs-464207/v1
92. Khodayari, A. and Maranas, C.D. (2016) A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nat. Commun.* 7, 13806
93. Cho, J.S. *et al.* (2019) Reconstruction of context-specific genome-scale metabolic models using multiomics data to study metabolic rewiring. *Curr. Opin. Syst. Biol.* 15, 1–11
94. Brunk, E. *et al.* (2018) Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* 36, 272–281
95. Brunk, E. *et al.* (2016) Systems biology of the structural proteome. *BMC Syst. Biol.* 10, 26
96. Wang, X. *et al.* (2012) Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat. Biotechnol.* 30, 159–164
97. Kamburov, A. *et al.* (2015) Comprehensive assessment of cancer missense mutation clustering in protein structures. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5486–5495

98. Segrè, D. *et al.* (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15112–15117
99. Burgard, A.P. *et al.* (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84, 647–657
100. Ranganathan, S. *et al.* (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.* 6, e1000744
101. Choi, H.S. *et al.* (2010) In silico identification of gene amplification targets for improvement of lycopene production. *Appl. Environ. Microbiol.* 76, 3097–3105
102. King, Z.A. *et al.* (2015) Next-generation genome-scale models for metabolic engineering. *Curr. Opin. Biotechnol.* 35, 23–29
103. Mishra, P. *et al.* (2018) Genome-scale model-driven strain design for dicarboxylic acid production in *Yarrowia lipolytica*. *BMC Syst. Biol.* 12, 12
104. Iranmanesh, E. *et al.* (2020) Improving l-phenylacetylcarbinol production in *Saccharomyces cerevisiae* by in silico aided metabolic engineering. *J. Biotechnol.* 308, 27–34
105. Kim, M. *et al.* (2019) In silico identification of metabolic engineering strategies for improved lipid production in *Yarrowia lipolytica* by genome-scale metabolic modeling. *Biotechnol. Biofuels* 12, 187
106. Nocon, J. *et al.* (2014) Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metab. Eng.* 24, 129–138
107. Shen, F. *et al.* (2019) OptRAM: in-silico strain design via integrative regulatory-metabolic network modeling. *PLoS Comput. Biol.* 15, e1006835
108. Nilsson, A. *et al.* (2017) Metabolic models of protein allocation call for the kinetome. *Cell Syst.* 5, 538–541
109. Kim, G.B. *et al.* (2020) Machine learning applications in systems metabolic engineering. *Curr. Opin. Biotechnol.* 64, 1–9
110. Mendoza, S.N. *et al.* (2019) A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biol.* 20, 158
111. Raveh, B. *et al.* (2021) Bayesian metamodeling of complex biological systems across varying representations. *bioRxiv* Published online March 29, 2021. https://doi.org/10.1101/2021.03.29.437574
112. Goldberg, A.P. *et al.* (2016) Toward scalable whole-cell modeling of human cells. In *Proceedings of the 2016 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*, pp. 259–262, Association for Computing Machinery
113. Thiele, I. and Palsson, B.O. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121
114. Orth, J.D. *et al.* (2010) What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248
115. Caspi, R. *et al.* (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 44, 471–480
116. Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30
117. Domenzain, I. *et al.* (2021) Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *bioRxiv* Published online March 5, 2021. https://doi.org/10.1101/2021.03.05.433259
118. Lerman, J.A. *et al.* (2012) In silico method for modelling metabolism and gene product expression at genome scale. *Nat. Commun.* 3, 929
119. Porta-Pardo, E. *et al.* (2017) Comparison of algorithms for the detection of cancer drivers at subgene resolution. *Nat. Methods* 14, 782–788