Applied Microbiology and Biotechnology (2022) 106:4907–4920
https://doi.org/10.1007/s00253-022-12066-y

MINI-REVIEW

Genome-scale metabolic network models: from first-generation to next-generation

Chao Ye[1] · Xinyu Wei[1] · Tianqiong Shi[1] · Xiaoman Sun[1] · Nan Xu[2] · Cong Gao[3] · Wei Zou[4]

Received: 2 September 2021 / Revised: 24 June 2022 / Accepted: 2 July 2022 / Published online: 13 July 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Over the last two decades, thousands of genome-scale metabolic network models (GSMMs) have been constructed. These GSMMs have been widely applied in various fields, ranging from network interaction analysis, to cell phenotype prediction. However, due to the lack of constraints, the prediction accuracy of first-generation GSMMs was limited. To overcome these limitations, the next-generation GSMMs were developed by integrating omics data, adding constrain condition, integrating different biological models, and constructing whole-cell models. Here, we review recent advances of GSMMs from the first generation to the next generation. Then, we discuss the major application of GSMMs in industrial biotechnology, such as predicting phenotypes and guiding metabolic engineering. In addition, human health applications, including understanding biological mechanisms, discovering biomarkers and drug targets, are also summarized. Finally, we address the challenges and propose new trend of GSMMs.

Key points

- *This mini-review updates the literature on almost all published GSMMs since 1999.*
- *Detailed insights into the development of the first- and next-generation GSMMs.*
- *The application of GSMMs is summarized, and the prospects of integrating machine learning are emphasized.*

Keywords Genome-scale metabolic model · Phenotype prediction · Metabolic engineering · Biological mechanisms · Biomarkers · Drug targets

Introduction

Genome-scale metabolic network models (GSMMs) are mathematical models that have become crucial systems biology tools. The core of a GSMM is a representation of the relationships among genes-proteins-reactions (GPRs). The construction of GSMMs included four steps: (1) draft model construction, (2) model refinement, (3) model mathematization, and (4) model verification, which involve 103 detailed steps (Heirendt et al. 2019; Schellenberger et al. 2011). With the development of model construction tools, such as the RAVEN Toolbox (Wang et al. 2018), Model SEED (Henry et al. 2010), and Merlin (Dias et al. 2015), GSMMs can be constructed automatically, thus avoiding labor-intensive manual construction. These tools can build a GSMM quickly in a user-friendly manner, but the model still needs to be further refined, including the replacement of the common biomass function with a unique function.

In addition to model construction, the analysis algorithms also play a key role in the model application.

✉ Chao Ye
chaoye09@njnu.edu.cn

✉ Wei Zou
weizou1985@163.com

[1] School of Food Science and Pharmaceutical Engineering, Nanjing Normal University, Nanjing 210023, China

[2] College of Bioscience and Biotechnology, Yangzhou University, Yangzhou 225009, China

[3] State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi 214122, China

[4] College of Bioengineering, Sichuan University of Science & Engineering, Yibin 644005, China

Since the flux reflects the conversion rate of intracellular molecules through metabolic pathways, which are regulated by specific enzymes, different algorithms were developed to calculate the flux distribution. FBA (flux balance analysis) is the basic algorithm for calculating the flux distribution, which was used to calculate optimal growth rates or product synthesis rates (Orth et al. 2010). However, FBA could only simulate the flux distribution in the steady state and was constrained by the substrate uptake rate, which limited the prediction accuracy of GSMMs. Based on FBA, other flux balance analysis algorithms were developed, such as rFBA (regulatory FBA, incorporating transcriptional regulation) (Covert et al. 2001), MD-FBA (FBA accounting for metabolite dilution) (Benyamini et al. 2010), dFBA (dynamic FBA, considering enzyme dynamics) (Mahadevan et al. 2002), and cFBA (Community FBA, involving reaction thermodynamics) (Khandelwal et al. 2013). Furthermore, the in silico strain design algorithms, which are used to identify genetic modification targets, can be classified into four groups, including gene deletion (Burgard et al. 2003; Segre et al. 2002), gene addition (King and Feist 2013), gene up-/down-regulation (Chowdhury et al. 2014; Ranganathan et al. 2010), and addition of heterologous pathways (Kim et al. 2011; Pharkya et al. 2004) (Table 1). Using these algorithms, the simulation results could be combined with experimental results, which have been applied to produce ethanol, butanol, succinic acid, lactic acid, lycopene, amino acids, vanillin, and 1,4-butanediol, showing great application prospects.

In this review, we first collected almost all the GSMMs published to date by literature mining, and we analyzed the characteristic of GSMMs among bacteria, eukaryotic microorganisms, plants, and animals. According to their different constraints, these published GSMMs could be classified as first-generation and next-generation GSMMs. The first-generation GSMMs were further classified based on the species number. For the next-generation GSMMs, we summarized the constraints that were added to improve the prediction accuracy of GSMMs. In addition, we summarize the applications of GSMMs in industrial biotechnology and human health. Finally, we discuss the current challenges and research trends of GSMMs.

First-generation GSMMs

Typical GSMMs

Since the first GSMM of *Haemophilus influenzae* RD (Edwards 1999), GSMMs have been developed for over 20 years. By August 2021, 2002 GSMMs have been published for 1021 organisms (based on the literature survey results from web of science, Fig. 1A). Of these models, bacteria account for 85.43%, and eukaryotic microorganism accounts for 7.14%. The rest are plants and animals. The GSMM of *Zea mays* has both the highest number of genes and metabolites, 5824 and 9153, respectively (Simons et al. 2014). In addition, the GSMM of human cells has the highest number of reactions—33,090 (Schultz and Qutub 2016). The GSMMs of plants have the largest average number of genes, while those of animals have both the largest average numbers of reactions and metabolites (Fig. 1B). Notably, some typical model organisms have more than one GSMM. For example, there are 6 GSMMs for *Escherichia coli* K12 and 12 GSMMs for *Saccharomyces cerevisiae* S288c (Fig. 2) (King et al. 2016; Ye et al. 2017). These upgraded GSMMs cannot only increase the scale of the model (number of genes, reactions, and metabolites), but also improve the prediction accuracy. For example, the newest *E. coli* GSMM *i*ML1515 model achieved 93.4% accuracy for phenotype prediction, while *i*JO1366 only reached an accuracy of 89.8% (Monk et al. 2017).

Interaction models

Microbial interactions in nature mainly include mutualism, commensalism, parasitism, and competition. To investigate the relationship between two species, interaction models were constructed. Interaction models were developed based on two existing GSMMs and treated each GSMM as a separate compartment (Fig. 2). The space representing the community of both organisms was defined as a new compartment. In addition, another compartment represented the environment, through which the metabolites can be exchanged in the medium and the community space. Interaction models have been widely applied to simulate the exchange of metabolites between two species. For example, the *i*WZ-KV-663-BM-1055 model was constructed to guide the mixed fermentation of vitamin C (Ye et al. 2014). Under co-culture conditions, the maximum growth rates of *Ketogulonicigenium vulgare* and *Bacillus megaterium* were improved 1.5- and 6.6-fold, compared with mono-culture condition, respectively. This result proved there was symbiosis between these two strains, which was based on the exchange of metabolites. For example, *B. megaterium* could supply nutrients, including six amino acids, three organic acids, and glycerol for *K. vulgare*. In addition, *K. vulgare* could secrete aspartate, fumarate, and formate for *B. megaterium* (Ye et al. 2014). Similarly, GSMMs of *Clostridium acetobutylicum* and *Clostridium cellulolyticum* were used to simulate the fermentation process of butanol and have also proved the exchange of metabolites, indicating the symbiotic relationship (Salimi et al. 2010).

Table 1 Common algorithms used for the simulation of GSMMs

Algorithm	Description	General concept	Reference
Flux balance analysis			
FBA	Flux balance analysis	Simplex linear programming optimization can be used with a metabolic map described in matrix format	(Orth et al. 2010)
rFBA	Regulatory FBA	Transcriptional regulation can be incorporated into constraint-based metabolic models	(Covert et al. 2001)
iFBA	Integrated FBA	To model metabolism, regulation, and signaling, rFBA is integrated with ordinary differential equations	(Covert et al. 2008)
pFBA	Parsimonious enzyme usage FBA	pFBA uses a bi-level LP to minimize enzyme-associated flux, subject to optimal biomass	(Lewis et al. 2010)
CoupledFBA	Coupling of non-metabolic networks to metabolism	Constraints, such as the transcription and translation machinery, are provided for coupling with metabolism	(Thiele et al. 2010)
MD-FBA	FBA accounting for metabolite dilution	MD-FBA is a MILP problem that accounts for the dilution of internal metabolites	(Benyamini et al. 2010)
CoPE-FBA	Comprehensive polyhedral enumeration FBA	CoPE-FBA indicates that thousands to millions of optimal flux patterns result from a combinatorial explosion of flux patterns in just a few metabolic sub-networks	(Kelk et al. 2012)
gFBA	Geometric FBA	Compared with FBA solution, geometric FBA provides a standard, central, reproducible solution	(Smallbone and Simeonidis 2009)
dFBA	Dynamic FBA	Nonlinear or linear programming on a series of short time intervals was applied to constrain the fluxes changing with media conditions	(Mahadevan et al. 2002)
cFBA	Community FBA	cFBA considers constraints derived from reaction stoichiometry, reaction thermodynamics, and the ecosystem to study the metabolic behavior of microbial communities	(Khandelwal et al. 2013)
Strain design			
OptKnock	Gene deletion	Bi-level optimization to find gene knockout targets leads to product formation under optimal growth conditions	(Burgard et al. 2003)
ReacKnock	Gene deletion	Inspired by OptKnock, screening of up to 20 gene knockout targets	(Xu et al. 2013)
MOMA	Gene deletion	Use quadratic programming to minimize changes in metabolic flux levels after gene deletion	(Segre et al. 2002)
MOMAKnock	Gene deletion	A kind of gene knockout optimization algorithm for target products overproduction	(Ren et al. 2013)
RobustKnock	Gene deletion	Prediction of gene knockout targets for overproduction by considering the existence of competitive pathways	(Tepper and Shlomi 2010)
OptCouple	Gene deletion	Combination of gene knockout, insertion, and medium modification to predict growth-coupled strain design	(Jensen et al. 2019)
OptGene	Gene deletion	Use genetic algorithms to explore viable solution regions to identify gene knockout targets that meet the desired phenotype	(Patil et al. 2005)

Table 1 (continued)

Algorithm	Description	General concept	Reference
OptORF	Gene deletion/expression	Search for strain designs that maximize growth and target products, screening for targets that are affected by a specific number of gene deletions and that meet known transcriptional regulatory rules	(Kim and Reed 2010)
OptSwap	Gene addition	A method for determining the optimal modification of cofactor-specific (NAD (H) and NADP (H))	(King and Feist 2013)
OptForce	Gene up-/down-regulation	Identification of response genes with significant flux changes by flux differential analysis of wild-type and mutant (with ideal phenotype)	(Ranganathan et al. 2010)
k-OptForce	Gene up-/down-regulation	An extension of OptForce, integrated with enzyme kinetic constants to allow optimal solutions for metabolic and/or enzymatic engineering	(Chowdhury et al. 2014)
IdealKnock	Gene up-/down-regulation	A top-down framework, which first scans the mutant of interest, then determines the knockout strategy	(Gu et al. 2016)
OptReg	Gene up-/down-regulation	An extension of OptKnock that predicts the up- and down-regulation of reactions to achieve the desired phenotype	(Pharkya and Maranas 2006)
Redirector	Gene up-/down-regulation	Iteratively identify all reactions with flux changes to accommodate the gradual changes in biomass and desired products	(Rockwell et al. 2013)
FSEOF	Gene up regulation	Identification of response genes for flux increase during maximum product formation	(Choi et al. 2010)
OptStrain	Heterologous pathway addition	Using a general database of known enzymatic reactions to determine the minimal modification path required for maximizing product formation	(Pharkya et al. 2004)
SimOptStrain	Heterologous pathway addition	Simultaneous identification of deleted reactions and addition of heterologous reactions to host metabolism	(Kim et al. 2011)

Pan-genome models

A pan-genome is the entire set of genes for all strains within a clade, including the core genome and the dispensable genome (Fig. 2). To analyze the genome characteristic and identify phenotypic differences among different strains of a species, pan-genome models were developed. The pan-genome models were constructed using the microorganisms which belong to the same genus. Until now, only three pan-genome models have been published, such as the *E. coli* pan-genome model (55 strains) (Monk et al. 2013), *Staphylococcus aureus* pan-genome model (64 strains) (Bosi et al. 2016), and *Salmonella* pan-genome model (410 strains) (Seif et al. 2018). Since most of these species are pathogenic, the application of pan-genome models aimed at about the pathogenicity analysis by comparing metabolic capacity. During the pan-genome model construction for *Salmonella*, the growth ability of 410 *Salmonella* stains, which were cultured under 530 conditions, was predicted. The significantly different metabolic pathways were related to carbohydrate and cell wall synthesis. Metabolic specificity corresponds to the serovar of each strain and the host from which it was isolated (Seif et al. 2018).

Meta-genome models

Meta-genome includes all the genetic material present in an environmental sample, consisting of the genomes of many individual organisms. Species richness is an important evaluation indicator for metagenomic data, which is used to describe and quantify a microbial community, and reflects the number of species in a particular region. Based on the GSMM for single microorganisms, meta-genome models were also developed (Fig. 2). Using meta-genome models, the metabolic characteristics of strains under different environments can be evaluated.

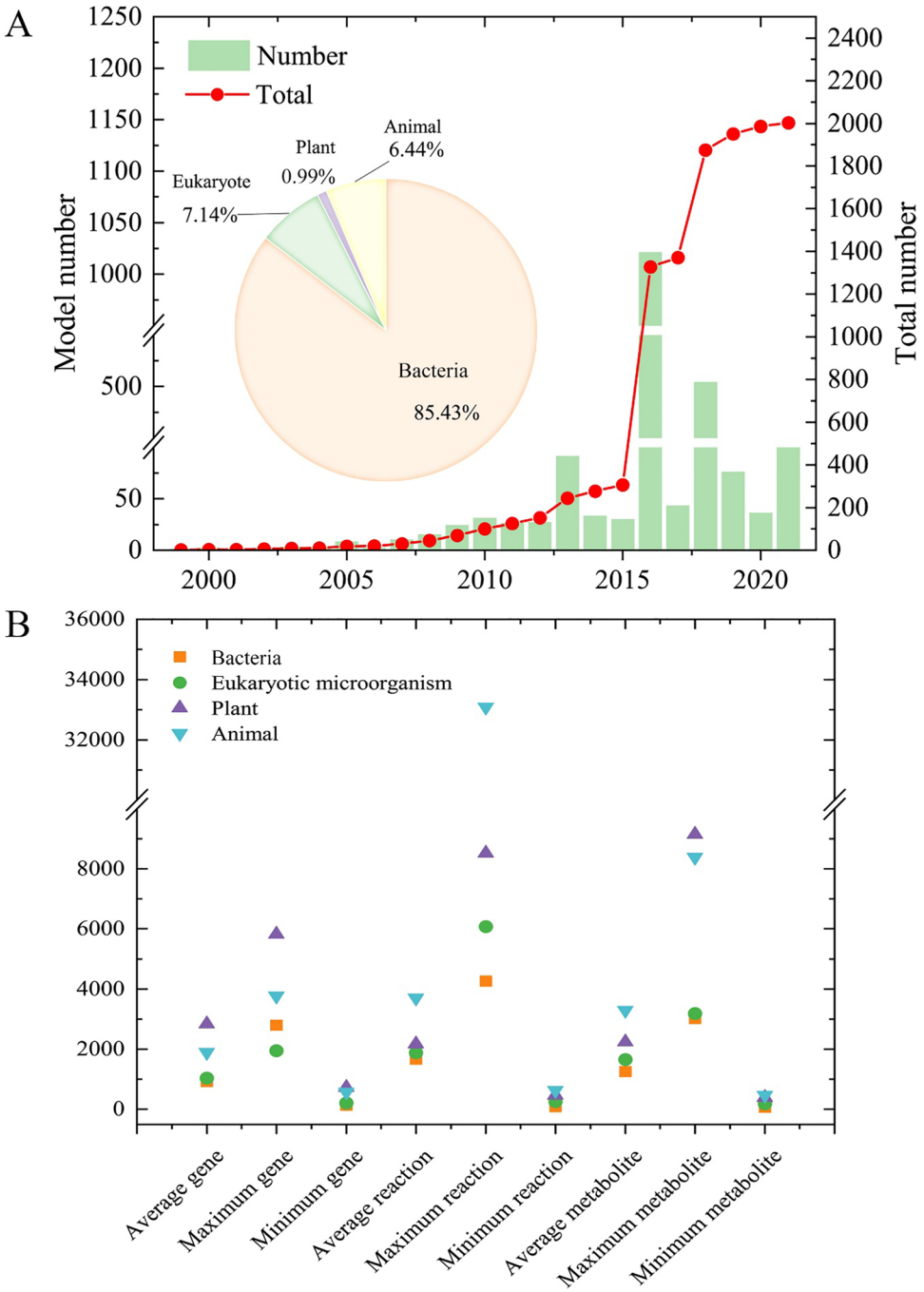

Fig. 1 The statistical result of GSMMs since 1999. A: the distribution of all GSMMs, and the increase in each year

In addition, interactions among different organisms can also be predicted. However, only two meta-genome models have been constructed to date, including a model of the 773 gut microbes and a model of 1562 human-related microbes, respectively (Garza et al. 2018). Considering the complexity of species richness, hundreds of GSMMs need to be constructed. To enable accurate model construction, methods such as AGORA (Magnusdottir et al. 2017) and MOMBO (Garza et al. 2018) have been developed. The advantages of these methods lie in the establishment of a specific quality evaluation system for the simulation and analysis of each model.

Next-generation GSMMs

The first-generation GSMMs relied on the substrate uptake rate to constrain the simulation process. For example, when the effect of the glucose uptake rate (GUR) on the cell growth rate (μ) was simulated using the Robustness Analysis

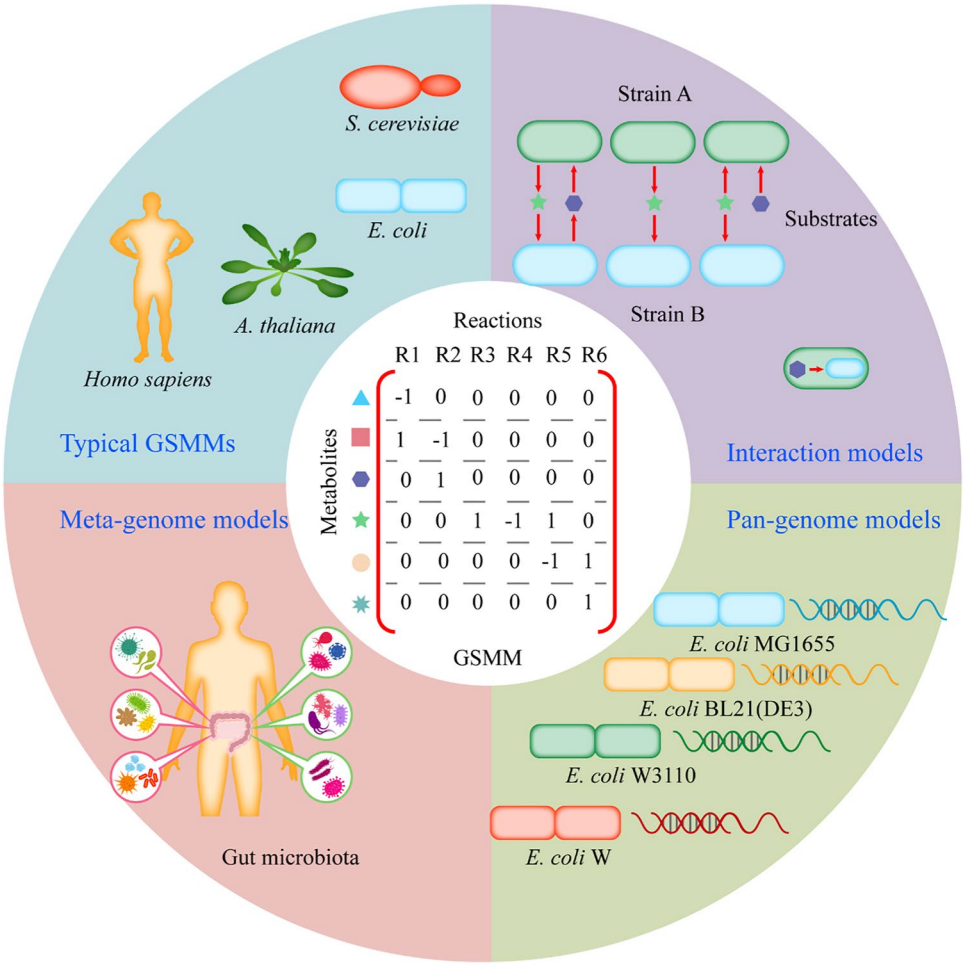

Fig. 2 The development of the first-generation GSMMs

algorithm, which is an algorithm used to simulate the relationship between a reaction of interest and an objective of interest, the results showed that with the increase of GUR, μ also kept increasing. Thus, the simulated growth rate value was much higher than the experimental results. To improve the prediction accuracy, novel methods were developed based on the first-generation GSMMs. These improvements were focused on about four aspects, including the integration of omics data, adding constraint conditions, integrating different biological networks, and constructing a whole-cell model (Fig. 3).

Integrating omics data

Genomics was the basis of GSMMs, and the genome annotation results were used to predict the function of each gene. Transcriptomics produces expression data of RNA transcripts that are produced by the genome, under specific circumstances. Similarly, proteomics reflects the abundance of each protein that is expressed under the investigated conditions, while metabolomics provides data on the concentrations of substrates and metabolic products. Fluxomics was defined as the complete set of metabolic fluxes in a cell. To improve the prediction accuracy of GSMMs, different omics data were integrated in recent studies (Kamsen et al. 2021; Lin et al. 2021; Pries et al. 2021). With the introduction of transcriptomics, a metabolism and gene expression model (ME-model) for *E. coli* was constructed. Using this ME-model, parameters such as the maximal growth rate on M9 medium, substrate uptake rate, extracellular secretion rate, flux distribution, and gene expression levels were calculated. Compared with the experimental results obtained under 24 culture conditions, the Pearson correlation coefficient (PCC), which is a measure of linear correlation between two sets of data of the ME-model was 2 times higher than that of the GSMM *i*JO1366 model, increasing from 0.20 (p value = 0.49) to 0.50 (p value = 0.07) (Ye et al. 2020a). Notably, a PCC between 0.5 and 0.7 indicates moderate correlation, while the initial PCC was below 0.4 and, therefore, indicated a weak or no correlation. Other tools, such as GIMME (Gene Inactivity Moderated by Metabolism and Expression), could also improve the accuracy of phenotype prediction by introducing gene expression data

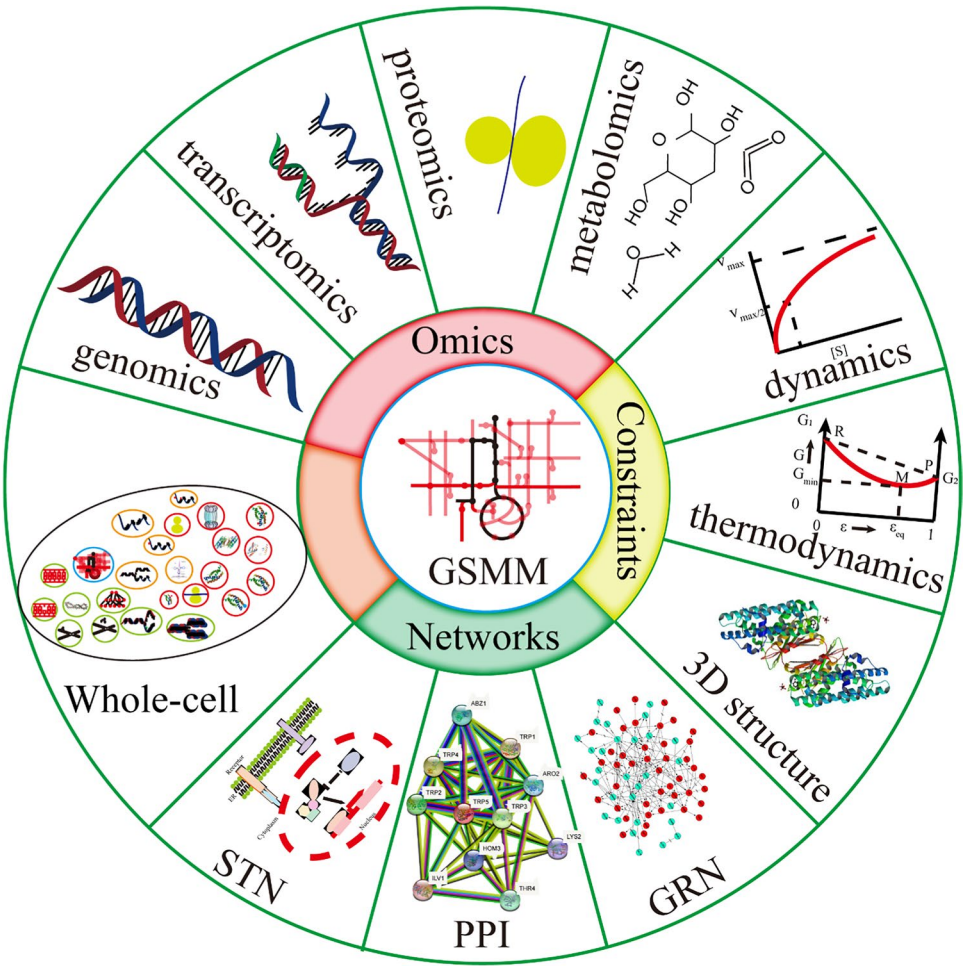

Fig. 3 The development of the second-generation GSMMs. GRN gene transcriptional regulatory network, PPI protein–protein interaction network, STN signal transduction network

(Becker and Palsson 2008). Furthermore, iMAT (an Integrative Metabolic Analysis Tool) was developed by integrating transcriptomic and proteomic data with GSMMs (Zur et al. 2010). The IOMA (Integrative Omics-Metabolic Analysis) algorithm can quantitatively integrate proteomic and metabolomic data with GSMMs, allowing more accurate prediction of metabolic flux distributions. Using IOMA, the PCC for the simulation results of wild-type and 23 single-gene deletions was 0.54, higher than the PCC predicted by FBA (0.44) and MOMA (minimization of metabolic adjustment, 0.38). MOMA can predict the effects of gene deletion using quadratic programming to identify a point in flux space, which is closest to the wild-type point, and is compatible with the gene deletion constraint (Yizhak et al. 2010). In addition, to explain the xylose utilization capacity of *S. cerevisiae*, ^{13}C fluxomics data was integrated into a GSMM. Using this method, the high cell maintenance energy was identified as one of the key factors constraining xylose utilization (Feng and Zhao 2013).

Adding constraint conditions

Based on traditional GSMMs, some algorithms have been developed to constrain metabolic fluxes by integrating kinetics (Henriques et al. 2021; Khodayari and Maranas 2016), thermodynamics (Krumholz and Libourel 2017; Oftadeh et al. 2021; Yang et al. 2021), enzymatic properties (Massaiu et al. 2019; Sanchez et al. 2017), or 3D structures (Brunk et al. 2018; Monk et al. 2017). A kinetics model, named k-ecoli457, was constructed by introducing 185 km values, and 49 k_{cat} values, based on the core *E. coli* GSMM. Using k-ecoli457, the PCC value of 24 product yields reached 0.84, which as higher than with other algorithms (Khodayari and Maranas 2016). Based on the GSMMs of *Streptococcus pneumoniae*, *Bacillus subtilis*, *E. coli* MG1655, and *Acinetobacter baylyi* ADP1, thermodynamics constraints ($\Delta_r G'^o$) were integrated to modify the reaction direction in order to improve model prediction performance (Krumholz and Libourel 2017). GECKO is a MATLAB/Python package for enhancing a genome-scale model to account

for enzymological constraints, using enzyme kinetics and omics data. It can enhance a GEM to account for enzymes as part of reactions, thereby ensuring that each metabolic flux does not exceed its maximum capacity, equal to the product of the enzyme's abundance and turnover number. Using the GECKO toolbox, enzymes were introduced as metabolic reactions in the GSMMs of *S. cerevisiae* and *E. coli*. Compared with the traditional model Yeast7, the upgraded model ecYeast7 achieved a significant improvement of phenotype predictions (Sanchez et al. 2017). For example, the growth rates under 24 culture conditions predicted using the ec_*i*ML1515 model achieved a PCC of 0.57, which was higher than that of the *i*ML1515 model (Ye et al. 2020a). In *E. coli*, the 3D structures of proteins that participate in metabolic processes were incorporated into *i*JO1366, resulting in the parametrization of the metabolic model based on structural information (Chang et al. 2013).

Integrating different biological networks

In addition to GSMMs, there are other biology networks, such as gene regulatory networks (GRNs), protein interaction networks (PPIs), and signal transduction networks (STNs). Based on GSMMs, other biological networks have been integrated to improve the ability of phenotype prediction (Cruz et al. 2020). With the integration of GSMMs and GRNs, a new model of *Mycobacterium tuberculosis*, named MTBPROM2.0, was constructed, which covered 2555 interactions linked to 104 transcription factors (TFs). Compared with GSMM, MTBPROM2.0 offers more accurate simulation of gene deletion effects but also predicts the growth defects associated with TF overexpression. Moreover, condition-specific models based on MTBPROM2.0 successfully predicted synergistic growth consequences of overexpressing the TF gene *whiB4* in the presence of two standard anti-TB drugs (Ma et al. 2015). Furthermore, the GSMM, GRN, and STN of *E. coli* were integrated using the iFBA algorithm. The algorithm is based on an approach called integrated flux balance analysis, which is used to model the dynamic behavior of metabolic, regulatory, and signaling networks by combining FBA with regulatory Boolean logic and ordinary differential equations(Covert et al. 2008). Using this model, *E. coli* wild-type and single gene perturbation phenotypes for diauxic growth on glucose/lactose and glucose/glucose-6phosphate were predicted, which proved that iFBA is a significant improvement over the pure GSMM (Covert et al. 2008). In addition, based on *E. coli* GSMM, gene expression data, GRN, and STN were incorporated. The integrated model was used to evaluate the growth ability under 14 different conditions, as well as the effects of single-gene deletion. The PCC between experimentally measured and predicted values was 0.6, while the PCC of the basic GSMM was only 0.2 (Carrera et al. 2014).

Constructing a whole-cell model

The whole-cell model is the most complex biological network, which is used to describe the formation process of cellular macromolecular, such as DNA, RNA, proteins, and small molecule metabolites. It was also applied to explore interaction mechanisms among these macromolecular and metabolic systems. In a whole-cell model, all cell activities, including metabolism, are modularized, and the interaction between modules can be explored systemically. Eventually, whole-cell activities can be digitized. Following seven fundamental principles and four practical principles, Karr et al. constructed the first whole-cell model (Karr et al. 2015). With the whole-cell model, cell activities of *Mycoplasma genitalium* can be simulated dynamically. It covers 16 cell states (such as geometry and cell mass) and 28 cell processes (such as DNA replication, RNA transcription, and protein translation). Using the *M. genitalium* whole-cell model, (1) the life cycle of a single cell can be described from the level of individual molecules and their interactions; (2) the specific function of every annotated gene product can be accounted for; and (3) a wide range of observable cellular behaviors can be predicted accurately (Karr et al. 2012). Recently, a whole-cell model of *S. cerevisiae* was also constructed, named WM_S288C, which consists of 15 cellular states and 26 cellular processes (Ye et al. 2020b). Based on the WM_S288C model, (1) the relationship between genotype and phenotype was explored; (2) the resource allocation for *S. cerevisiae* during the cell cycle was calculated; and (3) the regulatory mechanisms of intracellular nucleotides were identified. The whole-cell model is a highly integrated network that can be used to simulate the cellular dynamic changes and predict phenotypes more accurately. There are 7 challenges that must be overcome to construct a whole-cell model, including experimental interrogation, data curation, model construction, and integration, accelerated computation, data analysis and visualization, model validation, as well as collaboration and community development, which limited the development of whole-cell models (Macklin et al. 2014). To overcome these challenges, team collaboration by interdisciplinary researchers is needed. Furthermore, a community should be established, so that experimental data and simulated results can be shared.

Application of biological networks in industrial biotechnology

Predicting cellular phenotypes

In the production of industrial strains, the phenotype of microorganisms is the result of the interaction of the genotype and external environment, involving cell growth, energy

usage, substrate consumption, product synthesis, and gene essentiality. Accordingly, GSMMs have been used to understand the phenotypic potential of microorganisms. For example, FBA was applied to calculate flux distribution under steady-state conditions. This was carried out by setting different substrate uptake rates, with biomass as an objective function, which allowed researchers to calculate the maximal growth rate of *E. coli* (Orth et al. 2010). To simulate the change of culture conditions, the uptake rates of metabolites were changed. Then, an optimized medium was identified according to the effect of different substrates on cell growth. Based on the *K. vulgare* model *i*WZ663, which was used for vitamin C production, the FBA results showed that removing L-glycine, L-cysteine, L-methionine, L-tryptophan, adenine, thymine, thiamine, and pantothenate from complete chemically defined medium (CDM) would decrease the biomass of *K. vulgare* to 1%, 21%, 16%, 1%, 26%, 57%, 73%, and 24%, respectively (Fan et al. 2014). Then, a minimal medium for *K. vulgare* was developed which not only supported the cell growth, but also resulted in 96.5% of the maximal production of 2-keto-l-gulonic acid with corn steep liquor powder medium.

In addition, essential genes could be identified using a single-gene deletion algorithm. A comparison of growth in different culture media helped the researchers understand the absolute and relative essentiality of genes. According to the *Mortierella alpina* model *i*CY1106, 86 genes were identified as essential in minimal medium, and 36.05% of these genes belong to the amino acid metabolism pathways. By contrast, 49 genes were essential in yeast extract medium, and 23.26% of these were related to the nucleotide metabolism (Ye et al. 2015). These results indicated that in a rich culture medium, amino acids could be directly absorbed from the medium, so that amino acid synthesis genes can be considered to have relative essentiality. GSMMs have great potential for the prediction of cellular phenotypes. Using GSMMs, the cell growth under different conditions could be simulated, providing a reference for the optimization of fermentation processes. Furthermore, the flux distribution can be quantitatively analyzed, providing a theoretical basis for the identification of metabolic bottleneck.

Guiding metabolic engineering

Metabolic engineering can improve the target strain to increase its growth and product yield (Chen et al. 2018). However, due to the lack of a systematic understanding of the whole metabolic activity of microorganisms, the engineering of individual genes cannot achieve the goal of optimization. By using GSMMs in combination with different strain design algorithms, different strategies could be simulated on a global level. These include (1) removal of competing pathways. OptCouple is a constraint-based modeling algorithm that can simultaneously identify combinations of gene knockouts, insertions, and medium supplements that lead to growth-coupled production of a target compound (Jensen et al. 2019). Using this algorithm, a series of genetic engineering targets for the over-production of propionate and itaconate were predicted. Most of these targets were related to competing pathways (Jensen et al. 2019); (2) enhancement of key genes in synthetic pathways. Using OptORF, the transcription factor gene *fnr*, as well as the *plfB*, *tdcE*, and *pgi* genes, were identified as knockout targets. By also over-expressing the *edd*, the ethanol yield was increased from 39.3 to 86.2% (Kim and Reed 2010); (3) eliminating feedback inhibition. OptForce can identify all possible engineering interventions by classifying reactions in the metabolic model based on the necessary change of flux values to meet a pre-specified overproduction target (Ranganathan et al. 2010). Using this strategy, the up-regulation, down-regulation, and knockout targets were identified for the production of fatty acids with different chain lengths. After upregulating *fabZ* and acyl-ACP thioesterase, as well as deleting *fadD*, the *E. coli* strain could produce 1.70 g/L and 0.14 g fatty acids/g glucose (39% of the maximal theoretical yield) with chain lengths of C14–16 in minimal M9 medium (Ranganathan et al. 2012); (4) introduction of heterologous pathways. SimOptStrain simultaneously considers gene deletions in a host organism and reaction additions from a universal database such as KEGG or MetaCyc (Kim et al. 2011). Using this approach, *sdhC*, *gnd*, and *glyA* were predicted as deletion targets, and a reaction catalyzed by oxoglutarate dehydrogenase (EC: 1.2.1.52, 2-Oxoglutarate + CoA + NADP = Succinyl-CoA + CO_2 + NADPH) was introduced, leading to succinate production reaching 32.5% of the maximal theoretical yield (Kim et al. 2011); and (5) optimization of cofactors. With the help of OptSwap (King and Feist 2013), 12 different strategies for cofactor exchange and knockout were developed, which increased the pyruvate synthesis rate from 0 to 20.42 mmol gDW^{-1} h^{-1} (Xu et al. 2017). Based on GSMMs, screening and combining different targets to guiding metabolic engineering, followed by experimental verification, have been successfully applied to the production of organic acids, amino acids, and alcohols. The combination of GSMMs and experiments can guide rational strain design and improve the efficiency of metabolic engineering. In addition, the influence of genetic disturbances on the metabolism of the whole microorganism is considered at the global level, which enables the precise regulation of metabolic fluxes.

Application of biological networks for human health

Since the publication of the first GSMM of human cells 15 years ago (Duarte et al. 2007), more than seven models have been constructed (Table 2). In addition, models for

more than 120 unique tissues and organs have also been constructed, including breast, liver, and stomach (Schultz and Qutub 2016; Wang et al. 2012). Using these human GSMMs, the biological mechanisms behind various physiological and pathological states could be investigated, such as the Warburg effect (Shlomi et al. 2011), aiding design of appropriate disease treatments by the identification of novel drug targets and biomarkers.

Understanding biological mechanisms

Human metabolism is much more complex than that of microorganisms and is regulated by more than 3000 genes. Cellular metabolism is highly related to human health, and abnormal metabolism can lead to various diseases, such as type 2 diabetes, fatty liver, and even cancer. GSMMs have been used to elucidate the metabolic malfunction of cells under chronic or acute disease conditions. Based on 917 primary tumor samples across 13 types, 917 cancer GSMMs were constructed (Gatto et al. 2020). Using these GSMMs, the reactions catalyzed by enzymes that are coded by *ARG2*, *RHAG*, *SLC6*, as well as *SLC16* family gene members, and *PTGS1* and *PTGS2* were found to be at the core of cancer. Since considering the stoichiometric and enzyme solvent capacity in human GSMM, the Warburg effect was proved to be the result of the metabolic adaptation of cancer cells to increase biomass production rate (Shlomi et al. 2011).

Discovering biomarkers and drug targets

Biomarkers can be chemical, physical, or biological and may be functional at the physiological, biochemical, and the cellular level, or at the level of molecular interactions. According to the definition of World Health Organization, "any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease could be classified as biomarkers" (Strimbu and Tavel 2010). Biomarkers have been widely applied in the field of risk assessment, disease diagnosis, and treatment prediction. However, the naturally occurring biomarkers are frequently found in low concentrations, so that the detection process is time-consuming and expensive (Wu and Qu 2015). With the help of GSMMs, concentration changes of thousands of metabolites under different conditions can be simulated. Using the hepatocyte GSMM *i*Hepatocytes2322, chondroitin and heparan sulphate were identified as biomarkers for diagnosing non-alcoholic steatohepatitis and staging of non-alcoholic fatty liver disease (Mardinoglu et al. 2014).

In addition to detect biomarkers, GSMMs have been applied to predict drug targets. Using *i*Hepatocytes2322, *PSPH*, *SHMT1*, and *BCAT1* were identified as potential therapeutic targets for the treatment of non-alcoholic steatohepatitis (Mardinoglu et al. 2014). By integrating of RNA-seq data with GSMMs, the mevalonate pathway was predicted as a potential pathway against cancer proliferation (Raskevicius et al. 2018).

Table 2 The development of human GSMMs

Model	Reactions	Metabolites	Genes	Reference
Recon 1	3311	2766	2004	(Duarte et al. 2007)
EHMN	2824	2715	2322	(Ma et al. 2007)
Recon 2	7440	5063	2194	(Thiele et al. 2013)
HMR	9311	8181	3668	(Mardinoglu et al. 2013)
HMR 2.0	8184	6007	3765	(Agren et al. 2014)
Recon 2.2	7785	5324	1675	(Swainston et al. 2016)
Recon 3D	13,534	4140	3288	(Brunk et al. 2018)

Concluding remarks and future prospects

After more than 20 years of development, GSMMs have been improved from the first generation to next generation. Since the model construction process has been standardized (Heirendt et al. 2019; Schellenberger et al. 2011), and many automated-construction tools were released, it has become convenient to construct large numbers of GSMMs in a single run, enabling the investigation of interactions at the pan-genome or meta-genome level. To overcome the deficiencies of first-generation GSMMs and improve the predicting accuracy, omics data, additional constrains, and other biological networks were integrated into GSMMs. In addition, whole-cell models were constructed. Combined with different algorithms, GSMMs have been widely applied in different fields, including industrial biotechnology and human health.

However, most parameters in GSMMs were fixed, which limits the accuracy of model prediction. Recently developed machine learning approaches, which do not need fixed constraint parameters, only require a series of training sets. These parameters can then be modified automatically, so that the model can be more accurate. In addition, as the "black-box" approach was unable to provide direct mechanistic insights into the relationship between biological molecules and cellular phenotypes, a "white-box" method was developed and successfully used to reveal the action mechanisms of antibiotics (Yang et al. 2019). Machine learning has shown great potential in speech recognition, visual object recognition, object detection, drug discovery, and genome analysis (Gazestani and Lewis 2019; LeCun et al. 2015; Moen et al. 2019). Thus, there is a tendency to integrate machine learning approaches with GSMMs, allowing precise prediction of the regulation of metabolic fluxes, as well as characterizing the interactions under complex conditions

(Czajka et al. 2021; Lewis and Kemp 2021; Pearcy et al. 2021; Saini et al. 2021; Schinn et al. 2021).

Author contribution CY, XYW, NX, and WZ wrote the manuscript. TQS and XMS generated the figures and tables. CG proofread the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by grants from the National Key R & D Program of China (2021YFC2100700) and the National Natural Science Foundation of China (31801522, 21808196).

Declarations

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors declare no competing interests.

References

Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J (2014) Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. Mol Syst Biol 10(3):721. https://doi.org/10.1002/msb.145122

Becker SA, Palsson BO (2008) Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol 4(5):e1000082. https://doi.org/10.1371/journal.pcbi.1000082

Benyamini T, Folger O, Ruppin E, Shlomi T (2010) Flux balance analysis accounting for metabolite dilution. Genome Biol 11(4):43. https://doi.org/10.1186/gb-2010-11-4-r43

Bosi E, Monk JM, Aziz RK, Fondi M, Nizet V, Palsson BO (2016) Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. P Natl Acad Sci USA 113(26):3801–3809. https://doi.org/10.1073/pnas.1523199113

Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Drager A, Mih N, Gatto F, Nilsson A, Preciat Gonzalez GA, Aurich MK, Prlic A, Sastry A, Danielsdottir AD, Heinken A, Noronha A, Rose PW, Burley SK, Fleming RMT, Nielsen J, Thiele I, Palsson BO (2018) Recon3D enables a three-dimensional view of gene variation in human metabolism. Nat Biotechnol 36(3):272–281. https://doi.org/10.1038/nbt.4072

Burgard AP, Pharkya P, Maranas CD (2003) OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84(6):647–657. https://doi.org/10.1002/bit.10803

Carrera J, Estrela R, Luo J, Rai N, Tsoukalas A, Tagkopoulos I (2014) An integrative, multi-scale, genome-wide model reveals the phenotypic landscape of *Escherichia coli*. Mol Syst Biol 10(7):735. https://doi.org/10.15252/msb.20145108

Chang RL, Andrews K, Kim D, Li Z, Godzik A, Palsson BO (2013) Structural systems biology evaluation of metabolic thermotolerance in *Escherichia coli*. Science 340(6137):1220–1223. https://doi.org/10.1126/science.1234012

Chen X, Gao C, Guo L, Hu G, Luo Q, Liu J, Nielsen J, Chen J, Liu L (2018) DCEO Biotechnology: tools to design, construct, evaluate, and optimize the metabolic pathway for biosynthesis of chemicals. Chem Rev 118(1):4–72. https://doi.org/10.1021/acs.chemrev.6b00804

Choi HS, Lee SY, Kim TY, Woo HM (2010) *In silico* identification of gene amplification targets for improvement of lycopene production. Appl Environ Microbiol 76(10):3097–3105. https://doi.org/10.1128/AEM.00115-10

Chowdhury A, Zomorrodi AR, Maranas CD (2014) k-OptForce: integrating kinetics with flux balance analysis for strain design. PLoS Comput Biol 10(2):e1003487. https://doi.org/10.1371/journal.pcbi.1003487

Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux balance models of metabolism. J Theor Biol 213(1):73–88. https://doi.org/10.1006/jtbi.2001.2405

Covert MW, Xiao N, Chen TJ, Karr JR (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. Bioinformatics 24(18):2044–2050. https://doi.org/10.1093/bioinformatics/btn352

Cruz F, Faria JP, Rocha M, Rocha I, Dias O (2020) A review of methods for the reconstruction and analysis of integrated genome-scale models of metabolism and regulation. Biochem Soc Trans 48(5):1889–1903. https://doi.org/10.1042/Bst20190840

Czajka JJ, Oyetunde T, Tang YJ (2021) Integrated knowledge mining, genome-scale modeling, and machine learning for predicting *Yarrowia lipolytica* bioproduction. Metab Eng 67:227–236. https://doi.org/10.1016/j.ymben.2021.07.003

Dias O, Rocha M, Ferreira EC, Rocha I (2015) Reconstructing genome-scale metabolic models with merlin. Nucleic Acids Res 43(8):3899–3910. https://doi.org/10.1093/nar/gkv294

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. P Natl Acad Sci USA 104(6):1777–1782. https://doi.org/10.1073/pnas.0610772104

Edwards IJ (1999) Systems Properties of the *Haemophilus influenzae* Rd metabolic genotype. J Biol Chem 274(25):17410–17416. https://doi.org/10.1074/jbc.274.25.17410

Fan SC, Zhang ZY, Zou W, Huang Z, Liu J, Liu LM (2014) Development of a minimal chemically defined medium for *Ketogulonicigenium vulgare* WSH001 based on its genome-scale metabolic model. J Biotechnol 169:15–22. https://doi.org/10.1016/j.jbiotec.2013.10.027

Feng XY, Zhao HM (2013) Investigating xylose metabolism in recombinant *Saccharomyces cerevisiae* via C-13 metabolic flux analysis. Microb Cell Fact 12:114. https://doi.org/10.1186/1475-2859-12-114

Garza DR, van Verk MC, Huynen MA, Dutilh BE (2018) Towards predicting the environmental metabolome from metagenomics with a mechanistic model. Nat Microbiol 3:456–460. https://doi.org/10.1038/s41564-018-0124-8

Gatto F, Ferreira R, Nielsen J (2020) Pan-cancer analysis of the metabolic reaction network. Metab Eng 57:51–62. https://doi.org/10.1016/j.ymben.2019.09.006

Gazestani VH, Lewis NE (2019) From genotype to phenotype: augmenting deep learning with networks and systems biology. Curr Opin Syst Biol 15:68–73. https://doi.org/10.1016/j.coisb.2019.04.001

Gu DQ, Zhang C, Zhou SG, Wei LJ, Hua Q (2016) IdealKnock: a framework for efficiently identifying knockout strategies leading to targeted overproduction. Comput Biol Chem 61:229–237. https://doi.org/10.1016/j.compbiolchem.2016.02.014

Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdottir HS, Wachowiak J, Keating SM, Vlasov V, Magnusdottir S, Ng CY, Preciat G, Zagare A, Chan SHJ, Aurich MK, Clancy CM, Modamio J, Sauls JT, Noronha A, Bordbar A, Cousins B, El Assal DC, Valcarcel LV, Apaolaza I, Ghaderi S, Ahookhosh M, Ben Guebila M, Kostromins A, Sompairac N, Le HM, Ma D, Sun Y, Wang L, Yurkovich JT, Oliveira MAP, Vuong PT, El Assal LP, Kuperstein I, Zinovyev A, Hinton HS, Bryant WA,

Aragon Artacho FJ, Planes FJ, Stalidzans E, Maass A, Vempala S, Hucka M, Saunders MA, Maranas CD, Lewis NE, Sauter T, Palsson BØ, Thiele I, Fleming RMT (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat Protoc 14(3):639–702. https://doi.org/10.1038/s41596-018-0098-2

Henriques D, Minebois R, Mendoza SN, Macias LG, Perez-Torrado R, Barrio E, Teusink B, Querol A, Balsa-Canto E (2021) A multiphase multiobjective dynamic genome-scale model shows different redox balancing among yeast species of the *Saccharomyces* genus in fermentation. mSystems 6(4):e0026021. https://doi.org/10.1128/mSystems.00260-21

Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28(9):977–982. https://doi.org/10.1038/nbt.1672

Jensen K, Broeken V, Hansen ASL, Sonnenschein N, Herrgard MJ (2019) OptCouple: joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs. Metab Eng Commun 8:e00087. https://doi.org/10.1016/j.mec.2019.e00087

Kamsen R, Kalapanulak S, Chiewchankaset P, Saithong T (2021) Transcriptome integrated metabolic modeling of carbon assimilation underlying storage root development in cassava. Sci Rep 11:8758. https://doi.org/10.1038/s41598-021-88129-3

Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B Jr, Assad-Garcia N, Glass JI, Covert MW (2012) A whole-cell computational model predicts phenotype from genotype. Cell 150(2):389–401. https://doi.org/10.1016/j.cell.2012.05.044

Karr JR, Takahashi K, Funahashi A (2015) The principles of whole-cell modeling. Curr Opin Microbiol 27:18–24. https://doi.org/10.1016/j.mib.2015.06.004

Kelk SM, Olivier BG, Stougie L, Bruggeman FJ (2012) Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. Sci Rep 2:580. https://doi.org/10.1038/srep00580

Khandelwal RA, Olivier BG, Roling WFM, Teusink B, Bruggeman FJ (2013) Community flux balance analysis for microbial consortia at balanced growth. PLoS ONE 8(5):e64567. https://doi.org/10.1371/journal.pone.0064567

Khodayari A, Maranas CD (2016) A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. Nat Commun 7:13806. https://doi.org/10.1038/ncomms13806

Kim J, Reed JL (2010) OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst Biol 4(53):53. https://doi.org/10.1186/1752-0509-4-53

Kim J, Reed JL, Maravelias CT (2011) Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. PLoS ONE 6(9):e24162. https://doi.org/10.1371/journal.pone.0024162

King ZA, Feist AM (2013) Optimizing cofactor specificity of oxidoreductase enzymes for the generation of microbial production strains—OptSwap. Ind Biotechnol 9(4):236–246. https://doi.org/10.1089/ind.2013.0005

King ZA, Lu J, Drager A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE (2016) BiGG models: a platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res 44(D1):D515–D522. https://doi.org/10.1093/nar/gkv1049

Krumholz EW, Libourel IGL (2017) Thermodynamic constraints improve metabolic networks. Biophys J 113(3):679–689. https://doi.org/10.1016/j.bpj.2017.06.018

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539

Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, Konig R, Smith RD, Palsson BO (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Mol Syst Biol 6:390. https://doi.org/10.1038/msb.2010.47

Lewis JE, Kemp ML (2021) Integration of machine learning and genome-scale metabolic modeling identifies multi-omics biomarkers for radiation resistance. Nat Commun 12:2700. https://doi.org/10.1038/s41467-021-22989-1

Lin YT, Lin YS, Cheng WL, Chang JC, Chao YC, Liu CS, Wei AC (2021) Transcriptomic and metabolic network analysis of metabolic reprogramming and IGF-1 modulation in SCA3 transgenic mice. Int J Mol Sci 22(15):7974. https://doi.org/10.3390/ijms22157974

Ma HW, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I (2007) The Edinburgh human metabolic network reconstruction and its functional analysis. Mol Syst Biol 3:135. https://doi.org/10.1038/msb4100177

Ma SY, Minch KJ, Rustad TR, Hobbs S, Zhou SL, Sherman DR, Price ND (2015) Integrated modeling of gene regulatory and metabolic networks in *Mycobacterium tuberculosis*. PLoS Comput Biol 11(11):e1004543. https://doi.org/10.1371/journal.pcbi.1004543

Macklin DN, Ruggero NA, Covert MW (2014) The future of whole-cell modeling. Curr Opin Biotechnol 28:111–115. https://doi.org/10.1016/j.copbio.2014.01.012

Magnusdottir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, Greenhalgh K, Jager C, Baginska J, Wilmes P (2017) Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol 35:81–89. https://doi.org/10.1038/nbt.3703

Mahadevan R, Edwards JS, Doyle FJ III (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys J 83(3):1331–1340. https://doi.org/10.1016/S0006-3495(02)73903-9

Mardinoglu A, Agren R, Kampf C, Asplund A, Nookaew I, Jacobson P, Walley AJ, Froguel P, Carlsson LM, Uhlen M, Nielsen J (2013) Integration of clinical data with a genome-scale metabolic model of the human adipocyte. Mol Syst Biol 9:649. https://doi.org/10.1038/msb.2013.5

Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J (2014) Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. Nat Commun 5:3083. https://doi.org/10.1038/ncomms4083

Massaiu I, Pasotti L, Sonnenschein N, Rama E, Cavaletti M, Magni P, Calvio C, Herrgard MJ (2019) Integration of enzymatic data in *Bacillus subtilis* genome-scale metabolic model improves phenotype predictions and enables *in silico* design of poly-γ-glutamic acid production strains. Microb Cell Fact 18(1):3. https://doi.org/10.1186/s12934-018-1052-2

Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D (2019) Deep learning for cellular image analysis. Nat Methods 16(12):1233–1246. https://doi.org/10.1038/s41592-019-0403-1

Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Palsson BO (2013) Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. P Natl Acad Sci USA 110(50):20338–20343. https://doi.org/10.1073/pnas.1307797110

Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R, Nomura W, Zhang Z, Mori H, Feist AM, Palsson BO (2017) *i*ML1515, a knowledgebase that computes *Escherichia coli* traits. Nat Biotechnol 35(10):904–908. https://doi.org/10.1038/nbt.3956

Oftadeh O, Salvy P, Masid M, Curvat M, Miskovic L, Hatzimanikatis V (2021) A genome-scale metabolic model of *Saccharomyces cerevisiae* that integrates expression constraints and reaction

thermodynamics. Nat Commun 12(1):4790. https://doi.org/10.1038/s41467-021-25158-6
Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? Nat Biotechnol 28(3):245–248. https://doi.org/10.1038/nbt.1614
Patil KR, Rocha I, Forster J, Nielsen J (2005) Evolutionary programming as a platform for *in silico* metabolic engineering. BMC Bioinformatics 6:308. https://doi.org/10.1186/1471-2105-6-308
Pearcy N, Hu Y, Baker M, Maciel-Guerra A, Xue N, Wang W, Kaler J, Peng Z, Li F, Dottorini T (2021) Genome-scale metabolic models and machine learning reveal genetic determinants of antibiotic resistance in *Escherichia coli* and unravel the underlying metabolic adaptation mechanisms. mSystems 6(4):e0091320. https://doi.org/10.1128/mSystems.00913-20
Pharkya P, Maranas CD (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng 8(1):1–13. https://doi.org/10.1016/j.ymben.2005.08.003
Pharkya P, Burgard AP, Maranas CD (2004) OptStrain: a computational framework for redesign of microbial production systems. Genome Res 14(11):2367–2376. https://doi.org/10.1101/Gr.2872004
Pries C, Razaghi-Moghadam Z, Kopka J, Nikoloski Z (2021) Integration of relative metabolomics and transcriptomics time-course data in a metabolic model pinpoints effects of ribosome biogenesis defects on *Arabidopsis thaliana* metabolism. Sci Rep 11(1):4787. https://doi.org/10.1038/s41598-021-84114-y
Ranganathan S, Suthers PF, Maranas CD (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol 6(4):e1000744. https://doi.org/10.1371/journal.pcbi.1000744
Ranganathan S, Tee TW, Chowdhury A, Zomorrodi AR, Yoon JM, Fu Y, Shanks JV, Maranas CD (2012) An integrated computational and experimental study for overproducing fatty acids in *Escherichia coli*. Metab Eng 14(6):687–704. https://doi.org/10.1016/j.ymben.2012.08.008
Raskevicius V, Mikalayeva V, Antanaviciute I, Cesleviciene I, Skeberdis VA, Kairys V, Bordel S (2018) Genome scale metabolic models as tools for drug design and personalized medicine. PLoS ONE 13(1):e0190636. https://doi.org/10.1371/journal.pone.0190636
Ren SG, Zeng B, Qian XN (2013) Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. BMC Bioinformatics 14:17. https://doi.org/10.1186/1471-2105-14-S2-S17
Rockwell G, Guido NJ, Church GM (2013) Redirector: designing cell factories by reconstructing the metabolic objective. PLoS Comput Biol 9(1):e1002882. https://doi.org/10.1371/journal.pcbi.1002882
Saini DK, Rai A, Devi A, Pabbi S, Chhabra D, Chang JS, Shukla P (2021) A multi-objective hybrid machine learning approach-based optimization for enhanced biomass and bioactive phycobiliproteins production in *Nostoc* sp. CCC-403. Bioresour Technol 329:124908. https://doi.org/10.1016/j.biortech.2021.124908
Salimi F, Zhuang K, Mahadevan R (2010) Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. Biotechnol J 5(7):726–738. https://doi.org/10.1002/biot.201000159
Sanchez BJ, Zhang C, Nilsson A, Lahtvee PJ, Kerkhoven EJ, Nielsen J (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol Syst Biol 13(8):935. https://doi.org/10.15252/msb.20167411
Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. Nat Protoc 6(9):1290–1307. https://doi.org/10.1038/nprot.2011.308
Schinn SM, Morrison C, Wei W, Zhang L, Lewis NE (2021) A genome-scale metabolic network model and machine learning predict amino acid concentrations in Chinese Hamster Ovary cell cultures. Biotechnol Bioeng 118(5):2118–2123. https://doi.org/10.1002/bit.27714
Schultz A, Qutub AA (2016) Reconstruction of tissue-specific metabolic networks using CORDA. PLoS Comput Biol 12(3):e1004808. https://doi.org/10.1371/journal.pcbi.1004808
Segre D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. P Natl Acad Sci USA 99(23):15112–15117. https://doi.org/10.1073/pnas.232349399
Seif Y, Kavvas E, Lachance JC, Yurkovich JT, Nuccio SP, Fang X, Catoiu E, Raffatellu M, Palsson BO, Monk JM (2018) Genome-scale metabolic reconstructions of multiple *Salmonella* strains reveal serovar-specific metabolic traits. Nat Commun 9(1):3771. https://doi.org/10.1038/s41467-018-06112-5
Shlomi T, Benyamini T, Gottlieb E, Sharan R, Ruppin E (2011) Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect. PLoS Comput Biol 7(3):e1002018. https://doi.org/10.1371/journal.pcbi.1002018
Simons M, Saha R, Amiour N, Kumar A, Guillard L, Clement G, Miquel M, Li ZN, Mouille G, Lea PJ, Hirel B, Maranas CD (2014) Assessing the metabolic impact of nitrogen availability using a compartmentalized maize leaf genome-scale model. Plant Physiol 166(3):1659–1674. https://doi.org/10.1104/pp.114.245787
Smallbone K, Simeonidis E (2009) Flux balance analysis: a geometric perspective. J Theor Biol 258(2):311–315. https://doi.org/10.1016/j.jtbi.2009.01.027
Strimbu K, Tavel JA (2010) What are biomarkers? Curr Opin HIV AIDS 5(6):463–466. https://doi.org/10.1097/COH.0b013e32833ed177
Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hanscho M, Zielinski DC, Ang KS, Gardiner NJ, Gutierrez JM, Kyriakopoulos S, Lakshmanan M, Li SZ, Liu JK, Martinez VS, Orellana CA, Quek LE, Thomas A, Zanghellini J, Borth N, Lee DY, Nielsen LK, Kell DB, Lewis NE, Mendes P (2016) Recon 2.2: from reconstruction to model of human metabolism. Metabolomics 12(7):109. https://doi.org/10.1007/s11306-016-1051-4
Tepper N, Shlomi T (2010) Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics 26(4):536–543. https://doi.org/10.1093/bioinformatics/btp704
Thiele I, Fleming RM, Bordbar A, Schellenberger J, Palsson BØ (2010) Functional characterization of alternate optimal solutions of *Escherichia coli*'s transcriptional and translational machinery. Biophys J 98(10):2072–2081. https://doi.org/10.1016/j.bpj.2010.01.060
Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bolling C, Bordel S, Chavali AK, Dobson P, Dunn WB, Endler L, Hala D, Hucka M, Hull D, Jameson D, Jamshidi N, Jonsson JJ, Juty N, Keating S, Nookaew I, Le Novere N, Malys N, Mazein A, Papin JA, Price ND, Selkov E, Sigurdsson MI, Simeonidis E, Sonnenschein N, Smallbone K, Sorokin A, van Beek JHGM, Weichart D, Goryanin I, Nielsen J, Westerhoff HV, Kell DB, Mendes P, Palsson BO (2013) A community-driven global reconstruction of human metabolism. Nat Biotechnol 31(5):419–427. https://doi.org/10.1038/nbt.2488
Wang YL, Eddy JA, Price ND (2012) Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC Syst Biol 6:153. https://doi.org/10.1186/1752-0509-6-153
Wang H, Marcisauskas S, Sanchez BJ, Domenzain I, Hermansson D, Agren R, Nielsen J, Kerkhoven EJ (2018) RAVEN 2.0: a versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. PLoS Comput Biol 14(10):e1006541. https://doi.org/10.1371/journal.pcbi.1006541
Wu L, Qu XG (2015) Cancer biomarker detection: recent achievements and challenges. Chem Soc Rev 44(10):2963–2997. https://doi.org/10.1039/c4cs00370e

Xu Z, Zheng P, Sun J, Ma Y (2013) ReacKnock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. PLoS ONE 8(12):e72150. https://doi.org/10.1371/journal.pone.0072150

Xu N, Ye C, Chen X, Liu J, Liu L (2017) Genome-scale metabolic modelling common cofactors metabolism in microorganisms. J Biotechnol 251:1–13. https://doi.org/10.1016/j.jbiotec.2017.04.004

Yang JH, Wright SN, Hamblin M, McCloskey D, Alcantar MA, Schrubbers L, Lopatkin AJ, Satish S, Nili A, Palsson BO, Walker GC, Collins JJ (2019) A White-Box machine learning approach for revealing antibiotic mechanisms of action. Cell 177(6):1649–1661. https://doi.org/10.1016/j.cell.2019.04.016

Yang X, Mao Z, Zhao X, Wang R, Zhang P, Cai J, Xue C, Ma H (2021) Integrating thermodynamic and enzymatic constraints into genome-scale metabolic models. Metab Eng 67:133–144. https://doi.org/10.1016/j.ymben.2021.06.005

Ye C, Zou W, Xu N, Liu L (2014) Metabolic model reconstruction and analysis of an artificial microbial ecosystem for vitamin C production. J Biotechnol 182–183:61–67. https://doi.org/10.1016/j.jbiotec.2014.04.027

Ye C, Xu N, Chen H, Chen YQ, Chen W, Liu L (2015) Reconstruction and analysis of a genome-scale metabolic model of the oleaginous fungus *Mortierella alpina*. BMC Syst Biol 9:1. https://doi.org/10.1186/s12918-014-0137-8

Ye C, Xu N, Dong C, Ye Y, Zou X, Chen X, Guo F, Liu L (2017) IMGMD: a platform for the integration and standardisation of *in silico* microbial genome-scale metabolic models. Sci Rep 7(1):727. https://doi.org/10.1038/s41598-017-00820-6

Ye C, Luo Q, Guo L, Gao C, Xu N, Zhang L, Liu L, Chen X (2020) Improving lysine production through construction of an *Escherichia coli* enzyme-constrained model. Biotechnol Bioeng 117(11):3533–3544. https://doi.org/10.1002/bit.27485

Ye C, Xu N, Gao C, Liu G, Xu J, Zhang W, Chen X, Nielsen J, Liu L (2020) Comprehensive understanding of *Saccharomyces cerevisiae* phenotypes with whole-cell model WM_S288C. Biotechnol Bioeng 117(5):1562–1574. https://doi.org/10.1002/bit.27298

Yizhak K, Benyamini T, Liebermeister W, Ruppin E, Shlomi T (2010) Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. Bioinformatics 26(12):255–260. https://doi.org/10.1093/bioinformatics/btq183

Zur H, Ruppin E, Shlomi T (2010) iMAT: an integrative metabolic analysis tool. Bioinformatics 26(24):3140–3142. https://doi.org/10.1093/bioinformatics/btq602

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