

· 工业细胞科学 ·

刘立明 博士、教授、博士生导师，教育部长江学者特聘教授。研究方向为工业生物技术、代谢工程、生物转化和生物催化。在 *Chemical Reviews*, *Nature Communications*, *ACS Catalysis*, *Metabolic Engineering*, *Biotechnology and Bioengineering*, *Applied and Environmental Microbiology*、《生物工程学报》《微生物学报》等国内外生物工程类主流学术期刊上以第一或责任作者身份发表学术论文 120 余篇，其中 SCI 论文 100 余篇。出版科技著作 2 部。获得授权发明专利 28 项，申请或公开发明专利 20 余项。研究成果近 5 年来获国家技术发明二等奖、教育部科技进步一等奖、中国石油与化学工业联合会科技进步一等奖等 5 项科技奖励。



应用代谢网络模型解析工业微生物胞内代谢

叶超^{1,2}, 徐楠³, 陈修来^{1,2}, 刘立明^{1,2}

1 江南大学 食品科学与技术国家重点实验室, 江苏 无锡 214122

2 江南大学 工业生物技术教育部重点实验室, 江苏 无锡 214122

3 扬州大学 生物科学与技术学院, 江苏 扬州 225009

叶超, 徐楠, 陈修来, 等. 应用代谢网络模型解析工业微生物胞内代谢. 生物工程学报, 2019, 35(10): 1901–1913.

Ye C, Xu N, Chen XL, et al. Application of metabolic network model to analyze intracellular metabolism of industrial microorganisms. Chin J Biotech, 2019, 35(10): 1901–1913.

摘要: 为了快速、高效地理解工业微生物胞内代谢特征, 寻找潜在的代谢工程改造靶点, 基因组规模代谢网络模型 (GSMM) 作为一种系统生物学工具越来越受到人们的关注。文中在回顾 GSMM 20 年发展历程的基础上, 分析了当前 GSMM 的研究现状, 总结了 GSMM 的构建及分析方法, 从预测细胞表型和指导代谢工程两个方面阐述了 GSMM 在解析工业微生物胞内代谢中的应用, 并展望了 GSMM 未来的发展趋势。

关键词: 微生物胞内代谢, 生产性能, 基因组规模代谢网络模型, 系统生物学

Received: June 16, 2019; **Accepted:** August 19, 2019

Supported by: National Key Research and Development Program of China (No. 2018YFA0901401), the Scientific and Technological Innovation Leading Talents of National “Ten Thousand Talent Program” (2018), the National Natural Science Foundation of China (No. 21808083), the Key Technologies Research and Development Program of Jiangsu Province (Nos. BE2018623, BE2017622), the Key Field Research and Development Program of Guangdong Province (No. 2019B020218001).

Corresponding author: Liming Liu. Tel/Fax: +86-510-85197875; E-mail: mingli@jiangnan.edu.cn

国家重点研发计划 (No. 2018YFA0901401), 2018 年“万人计划”科技创新领军人才支持经费, 国家自然科学基金 (No. 21808083), 江苏省科技支撑计划社会发展项目 (Nos. BE2018623, BE2017622), 广东省重点领域研发计划项目 (No. 2019B020218001) 资助。

Application of metabolic network model to analyze intracellular metabolism of industrial microorganisms

Chao Ye^{1,2}, Nan Xu³, Xiulai Chen^{1,2}, and Liming Liu^{1,2}

¹ State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi 214122, Jiangsu, China

² Key Laboratory of Industrial Biotechnology, Ministry of Education, Jiangnan University, Wuxi 214122, Jiangsu, China

³ College of Bioscience and Biotechnology, Yangzhou University, Yangzhou 225009, Jiangsu, China

Abstract: To quickly and efficiently understand the intracellular metabolic characteristics of industrial microorganisms, and to find potential metabolic engineering targets, genome-scale metabolic network models (GSMMs) as a systems biology tool, are attracting more and more attention. We review here the 20-year history of metabolic network model, analyze the research status and development of GSMMs, summarize the methods for model construction and analysis, and emphasize the applications of metabolic network model for analyzing intracellular metabolic activity of microorganisms from cellular phenotypes, and metabolic engineering. Furthermore, we indicate future development trend of metabolic network model.

Keywords: intracellular metabolic activity of microorganisms, production performance, genome-scale metabolic model, systems biology

基因组规模代谢网络模型 (Genome-scale metabolic model, GSMM) 作为一种数学模型, 其本质是用于表征基因-蛋白-反应 (GPR) 三者之间的关系。GSMM 包括一系列化学计量平衡的生化反应, 通过形成一个矩阵 S , 将其转化为一个数学模型。矩阵 S 的每行代表代谢物, 每列代表反应。GSMM 已经广泛应用在分析网络特性、预测细胞表型、指导菌株设计、驱动模型发现、研究进化过程和分析相互作用等 6 个方面 (图 1)^[1-2]。

工业微生物胞内代谢活动具有较高的复杂性, 单一的研究手段难以系统地理解其调控机制, 且无法高效地获取所需表型。目前利用 GSMM 分析工业微生物胞内的代谢调控, 尤其是筛选代谢工程改造靶点, 提高目标产物产量方面已经取得了一定的进展。例如, Nocon 等利用 MOMA 算法鉴定出乙醇脱氢酶 *adh2* 是毕赤酵母中过量生成胞质人超氧化物歧化酶 (hSOD) 基因敲除靶点, 实验证实 *adh2* 的敲除使得 hSOD 产量增加了 20%^[3]。Song 等利用流量响应分析方法, 筛选出扩增磷酸烯醇丙酮酸羧化酶 (*pcc*) 使富马酸产量提高了 2.8 倍^[4]。这些基于 GSMM 预测和实验结合的方法, 在乙醇^[5-6]、丁醇^[7-8]、琥

珀酸^[9-10]、乳酸^[11]、番茄红素^[12]、氨基酸^[13-14]、香兰素^[15]和 1,4-丁二醇^[16]等产品生产中展现了巨大的应用前景。

本文综述了目前 GSMM 的研究进展, 比较了目前常见的模型构建及评价标准, 总结出流量平衡类和遗传扰动类模型分析方法的应用范围。进一步结合这些模型分析工具, 系统地阐述 GSMM 在工业微生物表型预测、代谢改造等胞内代谢研究中的应用。同时展望了 GSMM 的未来发展趋势, 以期能够利用系统生物学工具, 为“湿实验”提供理论指导。

1 代谢网络模型的发展

1.1 代谢网络模型的研究现状

自 1999 年完成流感嗜血杆菌 *Haemophilus influenzae* GSMM 的构建以来, GSMM 经历了 20 年的发展, 截至 2019 年 3 月, 已经有超过 153 种微生物, 348 个 GSMMs 完成了构建 (图 2)^[1-17]。其中一些模式菌株的 GSMMs, 如大肠杆菌 *Escherichia coli* K12 和酿酒酵母 *Saccharomyces cerevisiae* S288c, 经过不断的修正和完善, 分别构建了 6 个和 12 个 GSMMs^[17-18]。改进后的 GSMMs,

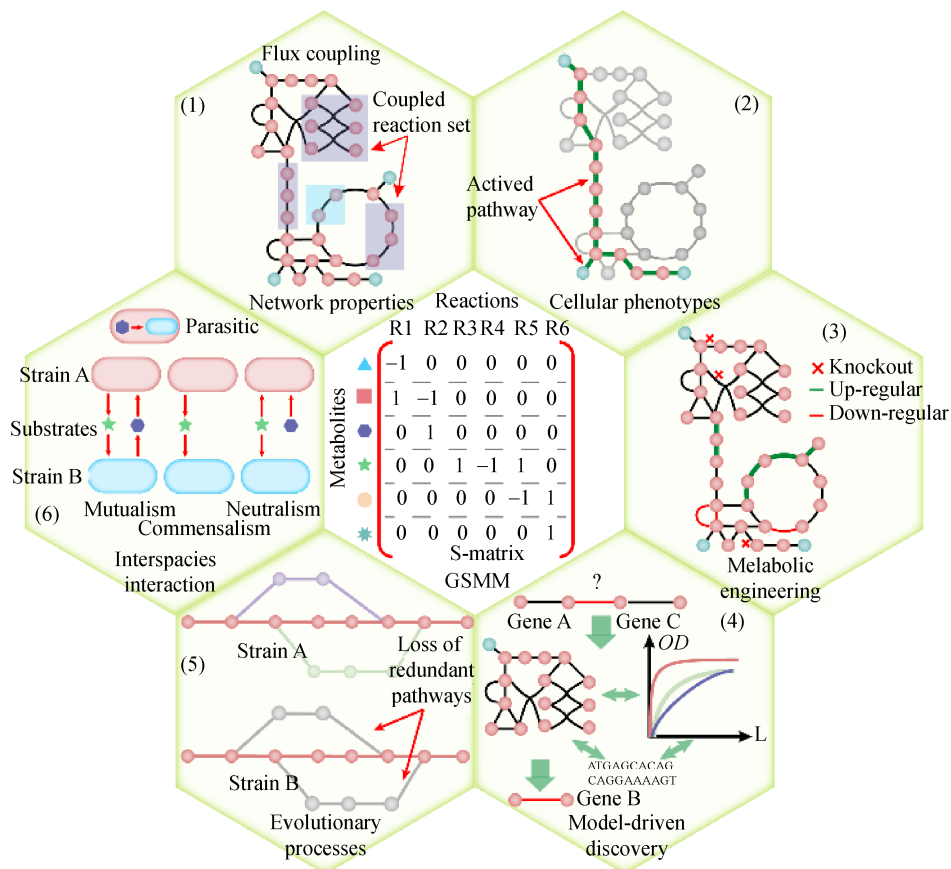


图1 GSMM在工业微生物中的应用

Fig. 1 Application of GSMM in industrial strain.

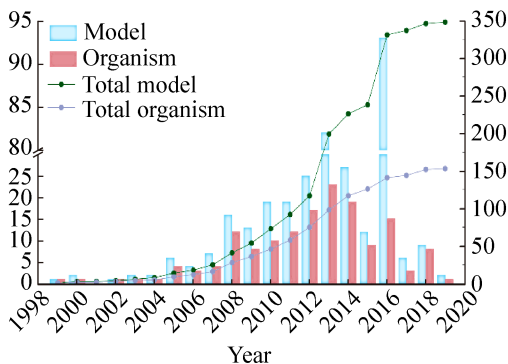


图2 已发表代谢网络模型的统计

Fig. 2 The number of published GSMM (As of March 2019).

不仅增加了模型规模 (基因、反应、代谢物), 而且提高了模型预测结果的准确性。如最新 *E. coli* 模型 *iML1515* 预测基因敲除表型的准确率为 93.4%, 相较 *iJO1366* 准确率 89.8% 提高了 3.6%^[19]。

由于传统的 GSMM 主要是通过对底物的利用作为约束条件进行模拟分析, 鲁棒性分析碳源吸收速率与细胞生长之间的关系通常呈线性相关。表明随着碳源吸收速率的增加, 细胞比生长速率可以一直增加, 最终远大于菌株的最大比生长速率。因此, 为了提高模型预测结果的准确性, 需要在传统 GSMM 的基础上整合其他方法。目前的改进方法主要有 4 个方面: 1) 整合组学数据^[20-22]: 如通过整合转录组学数据, 得到基因表达模型 (ME-model), 预测结果的准确性从 91.2% 提高至 92.3%^[20]; 2) 添加各种约束条件^[23-26]: 将动力学参数整合至 *E. coli* 核心 GSMM 中, 模拟结果与实验数据的皮尔森相关系数达到了 0.84^[23]; 3) 整合多种生物模型: 在 GSMM 基础上整合基因表达数据、转录调控模型、信号转导

网络, 得到 *E. coli* 整合模型, 皮尔森相关系数提高了 2 倍^[27]; 4) 构建微生物全细胞模型: 利用生殖支原体 *Mycoplasma genitalium* 全细胞模型不仅可以从单细胞水平描述一个细胞周期内个体分子及其相互作用, 而且实现了一系列可观察到的细胞行为的准确预测^[28]。

1.2 代谢网络模型的发展趋势

1.2.1 互作网络模型

自然界中的微生物相互作用关系主要存在互生、共生、竞争和寄生。通过构建微生物群体互作模型, 模拟不同物种之间的代谢物交换, 已经被用于鉴定微生物之间的相互作用关系。通过构建维生素 C 混菌发酵模型 *iWZ-KV-663-BM-1055*, 共培养条件下, 生酮基古龙酸菌 *Ketogulonicigenium vulgare* 和巨大芽孢杆菌 *Bacillus megaterium* 的最大比生长速率分别比单独培养提高了 1.5 倍和 6.6 倍, 表明两菌之间存在共生关系, 并且是通过代谢物交换实现的: *B. megaterium* 能为 *K. vulgare* 提供 6 种氨基酸、3 种核苷酸、6 种维生素和辅因

子、3 种有机酸以及甘油等营养物质; 而 *K. vulgare* 通过分泌苯丙氨酸、富马酸和甲酸促进 *B. megaterium* 生长^[29]。类似地, 通过分别构建木质纤维素生产丁醇菌株丙酮丁醇梭菌 *Clostridium acetobutylicum* 和解纤维梭菌 *Clostridium cellulolyticum* 的 GSMs, 模拟共培养条件下丁醇发酵过程, 发现 *C. acetobutylicum* 和 *C. cellulolyticum* 之间也是代谢物交换而实现共生 (图 3)^[30]。

1.2.2 泛基因组模型

泛基因组 (Pan-genome) 是指某一物种全部基因的总称, 包括核心基因组 (Core genome) 和附加基因组 (Dispensable genome)。通过构建泛基因组模型, 对核心基因和附加基因进行组合分析, 比较同一物种不同菌株之间的表型差异, 从而分析菌株特异性。目前已经完成构建的泛基因组模型主要有: 包含了 55 株 *E. coli* 的泛基因组模型^[31]、64 株金黄色葡萄球菌 *Staphylococcus aureus* 泛基因组模型^[32]以及 410 株沙门氏菌属 *Salmonella*^[33]的泛基因组模型。由于涉及的菌株

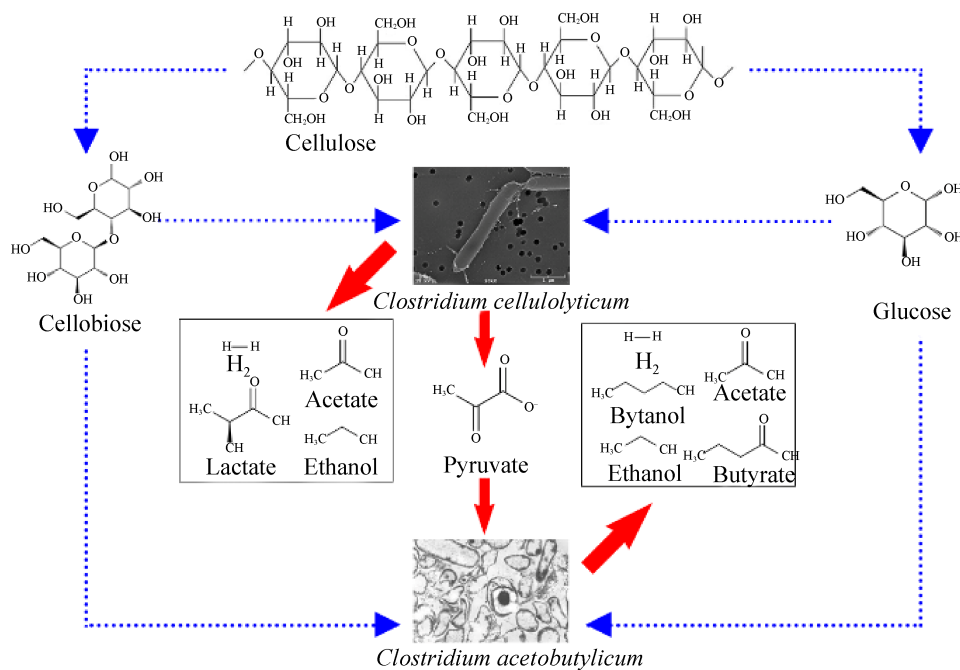


图 3 *C. acetobutylicum* 和 *C. cellulolyticum* 之间的代谢物交换

Fig. 3 Metabolites exchange between *C. acetobutylicum* and *C. cellulolyticum*.

都是致病性菌株, 主要是利用泛基因组模型预测不同菌株之间的代谢能力差异, 从而鉴定致病性。例如, 在 *Salmonella* 泛基因组模型中, 通过预测 410 株 *Salmonella* 在 530 种环境中的生长能力, 鉴定出这些 *Salmonella* 存在显著差异的代谢途径包括碳代谢和细胞壁合成, 并且代谢特异性与每个菌株的血清型和分离宿主相对应^[33]。

1.2.3 宏基因组模型

宏基因组 (Meta-genome) 是指特定环境中全部微生物的总 DNA。物种丰富度 (用来描述和量化微生物群落, 反映特定区域物种的数量), 是宏基因组数据的重要评价指标。在单个微生物 GSMM 的基础上, 通过构建宏基因组网络模型, 一方面可以结合模型分析菌群在不同环境中的代谢特征, 另一方面可以预测不同微生物之间的相互作用关系。目前已经构建完成的宏基因组模型是包含 773 个肠道微生物的宏基因组模型^[34]和包含 1 562 个人类相关微生物的宏基因组模型^[35]。考虑到宏基因组数据的物种丰富度, 需要批量构建成百上千个 GSMMs, 因此在构建宏基因组模型的过程中需要开发特定的方法, 如 AGORA^[34]和 MOMBO^[35]。这些算法的特点在于实现模型批量构建的过程中, 建立了特定的模型质量评价体系, 从而满足模拟分析。

2 代谢网络模型的分析

2.1 模型构建

根据 Palsson 实验室发表在 *Nature Protocols* 上的 GSMM 构建教程^[36-37], GSMM 的构建主要包括 4 个部分: 粗模型的构建、模型的精炼、模型的数学转换、模型的调试和验证, 涉及 102 个具体步骤。而 GSMM 的构建方式有 3 种: 手动构建、自动构建和半自动构建。手动构建方法依赖于基因组注释结果, 结合 KEGG 中的 pathway map, 对每个代谢途径涉及的基因-蛋白-反应进行

收集与整理, 这种方法的优势在于模型的准确性较高, 缺点则是模型构建过程耗时。自动构建方法则是通过已有的工具, 提交相关的信息, 如基因组序列, 从而自动得到目的菌株的 GSMM, 目前已经开发了许多工具, 如 COBRA^[37]、RAVEN^[38]、Model SEED^[39]和 IMGMD^[17], 用于 GSMM 的自动化构建。这种方式的优势在于可以实现模型的快速、批量构建, 缺点则是构建得到的模型准确性低。半自动构建的方法则是先通过一些自动化构建工具得到 GSMM, 然后在该模型的基础上进行手动精炼, 得到精确的模型。

在完成模型的构建后, 需要进行一系列的验证, 从而判断其覆盖范围是否完整, 以及与现有的模型相比是否具有更好的预测能力。常见的模型评价指标主要有: 通用指标、连通度指标、生长指标和基因敲除指标 (表 1)。前两个属于描述模型的指标 (定量和定性), 后两个是预测生物行为的指标^[40]。通过将不同培养条件下的模拟结果与实验值进行比较, 从而判断模型模拟结果的准确性。

2.2 模型分析

2.2.1 流量平衡分析类算法

代谢流量 (Flux) 是胞内分子通过代谢途径的转换率, 受到代谢途径中涉及的酶调控。流量平衡分析 (FBA) 作为最基本的数学算法, 目前已经被广泛应用于计算最优条件下的生长速率或者代谢产物合成速率^[41]。然而由于 FBA 算法局限性是只能模拟稳态条件下的流量分布, 并且通过底物的吸收速率作为约束条件, 限制了 FBA 预测流量分布的准确性。因此, 在 FBA 的基础上开发了一系列的算法对代谢流量进行额外的约束 (表 2)。这些算法一方面提高了模型预测流量分布结果的准确性, 例如, 通过整合转录调控信息, 开发了 rFBA 算法, 根据培养基中的代谢物浓度对反应流量添加布尔约束, 能够解释、分析和预测转录调

表 1 用于评价代谢网络模型的主要指标

Table 1 Main indicators used to evaluate GSMM

Descriptive metric		Predictive metric	
General metrics	Detailed parameters	Growth metrics	Detailed parameters
Model size	<ul style="list-style-type: none"> • Reactions, Metabolites, Genes 	Growth metrics	<ul style="list-style-type: none"> • Growth rates under different carbon/nitrogen sources, or aerobic/anaerobic
Model size (without compartments)	<ul style="list-style-type: none"> • Reactions, Metabolites, Genes 	Substrate consumption	<ul style="list-style-type: none"> • Consumption rate for different carbon/nitrogen sources
Similarity	<ul style="list-style-type: none"> • Jaccard distance^[34] 	Product excretion	<ul style="list-style-type: none"> • Predicting product generation rate
Additional information	<ul style="list-style-type: none"> • New pathways • Standardization • Other levels of regulation 	Flux distribution	<ul style="list-style-type: none"> • Consuming carbon and nitrogen sources • Getting energy • Core pathway flux distribution
Connectivity Metrics	Detailed parameters	Gene deletion metrics	Detailed parameters
Graph metrics	<ul style="list-style-type: none"> • Metrics for any pair of nodes (M_i, M_j): Shortest path, Shortest path diversity • Metrics for any node M_i: Node degree, Local clustering coefficient, Betweenness centrality • Global metrics: Global clustering coefficient, Characteristic path length, Network diameter 	Single knockout – minimal media	<ul style="list-style-type: none"> • Identifying essential genes under minimal media
Dead-end metabolites	<ul style="list-style-type: none"> • Metabolites that are not consumed 	Single knockout – complex media	<ul style="list-style-type: none"> • Identifying essential genes under complex media
		Single knockout – other media	<ul style="list-style-type: none"> • Identifying essential genes under other media
Blocked reactions	<ul style="list-style-type: none"> • Reactions always have zero flux 	Double knockout	<ul style="list-style-type: none"> • Detecting synthetic lethal genes

表 2 常见的流量平衡分析方法

Table 2 Common algorithms for flux balance analysis

Algorithm	Description	General concept	Reference
FBA	Flux balance analysis	Simplex linear programming optimization can be used with a metabolic map described in matrix format.	[41]
rFBA	Regulatory FBA	Transcription regulation can be incorporated into constraint-based models of metabolism.	[42]
iFBA	Integrated FBA	To model metabolism, regulation, and signaling, rFBA is integrated with ordinary differential equations.	[43]
pFBA	Parsimonious enzyme usage FBA	pFBA uses a bi-level LP to minimize enzyme-associated flux, subject to optimal biomass.	[46]
CoupledFBA	Couple non-metabolic networks to metabolism	Constraints, such as the transcription and translation machinery are provided for coupling with metabolism.	[47]
MD-FBA	FBA account for metabolite dilution	MD-FBA is an MILP problem that accounts for the dilution of internal metabolites.	[48]
CoPE-FBA	Comprehensive polyhedral enumeration FBA	CoPE-FBA indicates that the thousands to millions of optimal flux patterns result from a combinatorial explosion of flux patterns in just a few metabolic sub-networks.	[49]
gFBA	Geometric FBA	Compared with FBA solution, geometric FBA provides a standard, central, reproducible solution.	[50]
dFBA	Dynamic FBA	Nonlinear or linear programming on a series of short time intervals was applied to constrain the fluxes changing with media conditions.	[44]
cFBA	Community FBA	cFBA considers constraints deriving from reaction stoichiometry, reaction thermodynamics, and the ecosystem to study the metabolic behavior of microbial communities.	[45]

控在系统水平上对细胞代谢的影响^[42]。在 rFBA 的基础上,进一步整合常微分方程 (ODEs),开发了 iFBA 算法^[43],模拟结果表明,相较于 rFBA, iFBA 对 334 个单基因扰动和野生型的预测更准确。另一方面拓展了模型的应用,如 dFBA 算法将 FBA 扩展至能够考虑动力学因素,实现代谢流量的动态模拟。应用该算法,准确预测了 *E. coli* 在葡萄糖分批补料条件下的二次生长情况^[44]。

cFBA 则通过将反应化学计量数、热力学和生态系统作为约束条件,用于研究微生物群体的代谢行为^[45]。

2.2.2 遗传扰动类算法

微生物遗传改造主要是通过对微生物基因组进行修饰,改变基因型,从而获得不同表型微生物。结合 GSMM,开发相应的算法可以实现对这些扰动的模拟(表 3)。这些算法可以分为基因敲

表 3 常见的模拟菌株设计方法

Table 3 Common *in silico* algorithms for strain design

Algorithm	Description	Classification	Reference
OptKnock	Bi-level optimization to find gene knockout targets leads to product formation under optimal growth conditions	Gene deletion	[51]
ReacKnock	Inspired by OptKnock, screening for up to 20 gene knockout targets	Gene deletion	[52]
MOMA	Use quadratic programming to minimize changes in metabolic flux levels after gene deletion	Gene deletion	[58]
MOMAKnock	A kind of gene knockout optimization algorithm for target products overproduction	Gene deletion	[59]
RobustKnock	Prediction of gene knockout targets for overproduction by considering the existence of competitive pathways	Gene deletion	[60]
OptCouple	Combination of gene knockout, insertion and medium modification to predict growth-coupled strain design	Gene deletion	[61]
OptGene	Use genetic algorithms to explore viable solution regions to identify gene knockout targets that meet the desired phenotype	Gene deletion	[62]
OptORF	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	Gene deletion/expression	[63]
OptSwap	A method for determining the optimal modification of cofactor-specific (NAD(H) and NADP(H)).	Gene addition	[53]
OptForce	Identification of response genes with significant flux changes by flux differential analysis of wild-type and mutant (with ideal phenotype)	Gene up/down regulation	[54]
k-OptForce	An extension of OptForce, integrated with enzyme kinetic constants to allow optimal solutions for metabolic and/or enzymatic engineering	Gene up/down regulation	[55]
IdealKnock	A top-down framework, that first scans the mutant of interest, then determines the knockout strategy	Gene up/down regulation	[64]
OptReg	An extension of OptKnock that predicts the up- and down-regulation of reactions to achieve the desired phenotype	Gene up/down regulation	[65]
Redirector	Iteratively identify all reactions with flux changes to accommodate the gradual changes in biomass and desired products	Gene up/down regulation	[66]
FSEOF	Identification of response genes for flux increase during maximum product formation	Gene up regulation	[12]
OptStrain	Using a general database of known enzymatic reactions to determine the minimal modification path required for maximizing product formation	Heterologous pathway addition	[56]
SimOptStrain	Simultaneous identification of deleted reactions and added heterologous reactions in host metabolism	Heterologous pathway addition	[57]

除、添加、上调/下调或者异源途径导入等。OptKnock 是典型的基因敲除方法,是基于双目标方程,即分别对生物量方程和目标产物求解,筛选出能够提高目标产物的基因敲除靶点^[51]。在 OptKnock 的基础上进行改进,得到了 ReacKnock 算法,能够实现多达 20 个基因敲除靶点的筛选^[52]。OptSwap 作为基因添加类唯一的算法,主要是一种确定氧化还原酶辅因子特异性 (NAD(H)和 NADP(H)) 的最佳修饰的方法,用于微生物生产菌株的设计^[53]。在基因上调、下调方面,OptForce 通过比较突变菌株和野生型的反应流量差异,从而确定上调或下调的靶点,实现目标产物的过量合成^[54]。而 k-OptForce 则是 OptForce 的扩展,通过整合酶动力学常数,预测酶编码基因的上调或下调^[55]。异源表达相关的算法目前有 OptStrain 和 SimOptStrain。前者通过在包含一系列酶催化反应的反应库中检索,筛选能够满足最大产物合成速率下的最小修饰路径添加至宿主细胞中^[56]。后者则是不仅能够实现异源途径的添加,同时鉴定出宿主细胞中需要删除的反应^[57]。

3 代谢网络模型在解析胞内代谢中的应用

3.1 预测细胞表型

在工业菌株生产过程中,微生物的表型是基因型和外部环境共同作用的结果。微生物的表型主要涉及细胞生长、能量利用、底物利用、产物合成和基因必需性等方面。利用 GSMM 对细胞表型进行预测,从而理解微生物的表型潜力。

为了预测微生物的细胞生长,利用 FBA 算法,计算稳态条件下的各个反应的流量分布。通过设定不同底物的吸收速率,以生物量方程为目标方程求解,从而计算出 *E. coli* 的最大比生长速率 (μ)^[41]。

通过改变目标方程,探讨微生物代谢在不同目标方程下的流量分布,以检验满足细胞功能背后的驱动力,进而理解能量利用与细胞次优行为

之间的关系。通过评价 11 种目标方程下的流量分布,模拟不同条件下的最大 ATP 得率,与 6 种环境下的 C13 标记实验进行比较,证实模型能够准确预测微生物胞内能量利用^[67]。

通过改变培养条件,模拟不同培养基中底物利用对微生物生长的影响,从而鉴定最优培养基。利用维生素 C 生产菌株 *K. vulgare* WSH001 模型 iWZ663,通过 FBA 模拟,发现从完全培养基中分别去除甘氨酸、半胱氨酸、甲硫氨酸、色氨酸、腺嘌呤、胸腺嘧啶、硫胺素和泛酸会导致细胞生长分别下降 1%、21%、16%、1%、26%、57%、73%和 24%。基于这些结果开发了 *K. vulgare* 最小合成培养基,不仅满足了 *K. vulgare* 的生长,而且 2-酮基-L-古龙酸产量达到了完全培养基条件下的 96.5%^[68]。

利用鲁棒性分析算法,对生物量方程进行约束,以产物合成反应作为目标方程求解,能够得到一个解空间,反映了不同生长速率下的产物合成能力。结合 OptKnock 算法,筛选 *E. coli* 过量生成乳酸的靶点,结合适应性进化策略,最终得到能够在 M9 培养基中同时满足快速生长和乳酸大量分泌的 *E. coli* 菌株^[11]。

利用单基因敲除程序,对微生物的必需基因鉴定。通过比较不同培养条件下的必需基因差异,从而理解基因的绝对必需性和相对必需性。在最小培养基中,高山被孢霉 *Mortierella alpina* 的模型 iCY1106 中有 86 个基因被鉴定为必需基因;而在酵母提取物培养基中,只有 49 个基因是必需基因。进一步分析,前者有 36.05%的必需基因属于氨基酸代谢途径,而后者有 23.26%的必需基因参与核苷酸代谢^[69]。表明在营养丰富的培养条件下,氨基酸可以通过从培养基中直接摄入,不需要从头合成,因此氨基酸代谢相关的基因是相对必需的。

综上所述,在细胞表型预测方面,GSMM 展现了巨大的应用前景。利用 GSMM 不仅可以模拟

微生物在不同培养条件下的生长情况,为发酵过程优化提供参考,而且实现了胞内代谢流分布的定量分析,为代谢瓶颈的鉴定提供了理论依据。

3.2 指导代谢工程

代谢工程改造目标菌株,提高产量是一种有效策略,已经被广泛地应用^[70]。但是由于缺乏对微生物整个代谢活动的系统认识,单一水平上的基因改造策略难以高效地实现目标。利用 GSMM,结合不同的遗传改造方法,可以从全局水平模拟:1) 竞争路径的消除;2) 合成路径关键基因的强化;3) 反馈抑制的消除;4) 外源路径的导入;5) 辅因子优化等改造策略对目标产物的影响,为系统代谢工程提供指导^[71]。

为了消除竞争路径,通过一系列的算法筛选基因敲除靶点,对基因敲除结果的组合优化,从而提高目标产品的产量。利用 OptCouple,鉴定出大肠杆菌分别过量生产丙酸、衣康酸的一系列基因敲除靶点,这些靶点大多存在于其竞争途径,此外还鉴定出甲基化修饰提高产量的靶点^[61]。

在基因敲除菌株的基础上,表达特定基因,模拟合成路径关键基因的强化,分析基因表达对产物合成的影响。利用 OptORF 算法鉴定出敲除转录因子 *fnr*,同时敲除基因 *plfB*、*tdcE* 和 *pgi*,在此基础上过量表达 *edd*,使得乙醇得率从 39.3% 提高至 86.2%^[63]。

反馈抑制的消除主要是通过不同的算法模拟代谢路径中某个基因的上调或下调,从而使碳流更多地流向目标产物。利用 OptForce 算法,对不同链长脂肪酸合成过程中的上调、下调、敲除靶点进行了鉴定。发现上调 *fabZ* 和 *acyl-ACP* 硫酯酶,同时敲除 *fadD*,最小 M9 培养基中 C14-16 脂肪酸的产量达到了 1.7 g/L,得率为 0.14 g/g 葡萄糖(最大理论得率的 39%)^[72]。同样地,利用该方法鉴定出敲除 *fumC* 和 *sucC*,同时上调 ACC、PGK、GAPD 和 PDH 的表达,能够使胞内丙酰辅

酶 A 的含量提高 3.1 倍^[73]。

对于非天然产物的合成,通常涉及异源途径的导入。利用 SimOptStrain 鉴定出在 *E. coli* 中敲除 *sdhC*、*gnd* 和 *glyA* 的同时,引入氧代戊二酸脱氢酶 (EC: 1.2.1.52) 催化的反应 (2-酮戊二酸+CoA+NADP⁺=琥珀酰辅酶 A+CO₂+NADPH),使得琥珀酸得率达到了最大理论得率的 32.5%^[57]。

通过改善产物合成路径中关键酶的辅因子的偏好性,消除辅因子失衡所造成的负面效应,有助于构建高效的产物合成路径。基于 OptSwap 算法设计和优化光滑球拟酵母 *Candida glabrata* 生产丙酮酸,得到了 12 种辅因子交换和敲除策略,导致丙酮酸合成速率从 0 最高增加至 20.42 mmol/(g DW·h)^[74]。

综上所述,基于 GSMM,筛选不同代谢改造靶点并组合,用于指导代谢工程,进而通过实验进行验证。应用这种策略,已经成功地应用在有机酸、氨基酸、醇类等化学品的生产中。模型与实验相结合,不仅实现了菌株改造的理性设计,提高了代谢工程改造效率,而且能够从全局水平考虑遗传扰动对整个微生物胞内代谢的影响,有助于实现代谢流的精准调控。

4 结论与展望

针对工业微生物胞内代谢的研究,国内外研究人员采用生化工程和代谢工程等策略,开展了卓有成效的研究。结合 GSMM,可以将理性设计与实验操作结合,从而高效地获得工业生产菌株。GSMM 经过 20 年的发展,一方面通过不断添加约束条件,提高了模型预测结果的准确性,另一方面从单一菌株的模型发展至宏基因组模型。随着一系列模型构建工具和分析方法的开发,不仅实现了微生物 GSMM 的快速构建,而且能够从预测细胞表型、指导代谢工程这两个方面对微生物胞内代谢活动进行系统分析。

随着 GSMM 构建工具的不断开发和完善,目前已经可以实现 GSMM 的快速、批量构建,但是受限于模型构建方法及构建团队自身水平的限制,导致构建的 GSMMs 质量参差不齐,此外,这些 GSMMs 中的反应和代谢物列表中的展现形式也缺乏统一,限制了 GSMM 的进一步应用。虽然目前已有一些数据库,如 BIGG Models 实现了已发表 GSMMs 的标准化,但是该数据库目前仅收录了 85 个模型^[18]。因此亟需建立相应的标准对模型进行统一,并且搭建功能更全面的数据库平台,提供 GSMMs 的检索和下载。

虽然 GSMM 已经被广泛地应用在多个方面,但是这些模型的约束参数都是固定的。而最新发展的机器学习 (Machine learning) 方法,基于黑箱理论 (Black-box),不需要对约束参数进行设定,只需要提供一系列的训练集 (Training data),就能够对参数进行自动修正,从而实现模型的精确预测。此外,为了克服黑箱机器学习方法无法直接揭示生物分子与细胞表型之间的相互作用关系的不足, Yang 等进一步开发了白箱理论 (White-box) 用于揭示抗生素作用机制^[75]。机器学习方法,作为一种新兴的技术,在现代社会展现了巨大的应用潜力,已经广泛地应用在语言识别、视觉对象识别、物体检测、药物发现和基因组分析等方面^[76-78]。因此,采用机器学习方法对 GSMM 进行分析,除了进一步实现代谢流调控的精准预测,还可以用于复杂环境下相互作用关系的表征,则是未来 GSMM 的发展趋势。

REFERENCES

- [1] Kim WJ, Kim HU, Lee SY. Current state and applications of microbial genome-scale metabolic models. *Curr Opin Syst Biol*, 2017, 2: 10–18.
- [2] O'Brien EJ, Monk JM, Palsson BO. Using genome-scale models to predict biological capabilities. *Cell*, 2015, 161(5): 971–987.
- [3] Nocon J, Steiger MG, Pfeffer M, et al. Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metab Eng*, 2014, 24: 129–138.
- [4] Song CW, Kim DI, Choi S, et al. Metabolic engineering of *Escherichia coli* for the production of fumaric acid. *Biotechnol Bioeng*, 2013, 110(7): 2025–2034.
- [5] Dikicioglu D, Pir P, Onsan ZI, et al. Integration of metabolic modeling and phenotypic data in evaluation and improvement of ethanol production using respiration-deficient mutants of *Saccharomyces cerevisiae*. *Appl Environ Microbiol*, 2008, 74(18): 5809–5816.
- [6] Lee KY, Park JM, Kim TY, et al. The genome-scale metabolic network analysis of *Zymomonas mobilis* ZM4 explains physiological features and suggests ethanol and succinic acid production strategies. *Microb Cell Fact*, 2010, 9: 94.
- [7] Lee J, Yun H, Feist AM, et al. Genome-scale reconstruction and *in silico* analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. *Appl Microbiol Biotechnol*, 2008, 80(5): 849–862.
- [8] Ranganathan S, Maranas CD. Microbial 1-butanol production: identification of non-native production routes and *in silico* engineering interventions. *Biotechnol J*, 2010, 5(7): 716–725.
- [9] Lee SJ, Lee DY, Kim TY, et al. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation. *Appl Environ Microbiol*, 2005, 71(12): 7880–7887.
- [10] Mienda BS, Shamsir MS, Ilias RM. Model-guided metabolic gene knockout of *gnd* for enhanced succinate production in *Escherichia coli* from glucose and glycerol substrates. *Comput Biol Chem*, 2016, 61: 130–137.
- [11] Fong SS, Burgard AP, Herring CD, et al. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng*, 2005, 91(5): 643–648.
- [12] Choi HS, Lee SY, Kim TY, et al. *In silico* identification of gene amplification targets for improvement of lycopene production. *Appl Environ Microbiol*, 2010, 76(10): 3097–3105.

- [13] Lee KH, Park JH, Kim TY, et al. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol Syst Biol*, 2007, 3(1): 149.
- [14] Park JH, Lee KH, Kim TY, et al. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. *Proc Natl Acad Sci USA*, 2007, 104(19): 7797–7802.
- [15] Brochado AR, Matos C, Møller BL, et al. Improved vanillin production in baker's yeast through *in silico* design. *Microb Cell Fact*, 2010, 9: 84.
- [16] Yim H, Haselbeck R, Niu W, et al. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat Chem Biol*, 2011, 7(7): 445–452.
- [17] Ye C, Xu N, Dong C, et al. IMGMD: a platform for the integration and standardisation of *in silico* microbial genome-scale metabolic models. *Sci Rep*, 2017, 7(1): 727.
- [18] King ZA, Lu J, Dräger A, et al. BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res*, 2016, 44(D1): D515–D522.
- [19] Monk JM, Lloyd CJ, Brunk E, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol*, 2017, 35(10): 904–908.
- [20] O'Brien EJ, Lerman JA, Chang RL, et al. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol*, 2013, 9(1): 693.
- [21] Yizhak K, Benyamini T, Liebermeister W, et al. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*, 2010, 26(12): i255–i260.
- [22] Feng XY, Zhao HM. Investigating xylose metabolism in recombinant *Saccharomyces cerevisiae* via ¹³C metabolic flux analysis. *Microb Cell Fact*, 2013, 12: 114.
- [23] Khodayari A, Maranas CD. A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nat Commun*, 2016, 7: 13806.
- [24] Krumholz EW, Libourel IGL. Thermodynamic constraints improve metabolic networks. *Biophys J*, 2017, 113(3): 679–689.
- [25] Sánchez BJ, Zhang C, Nilsson A, et al. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol*, 2017, 13(8): 935.
- [26] Brunk E, Sahoo S, Zielinski DC, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol*, 2018, 36(3): 272–281.
- [27] Carrera J, Estrela R, Luo J, et al. An integrative, multi-scale, genome-wide model reveals the phenotypic landscape of *Escherichia coli*. *Mol Syst Biol*, 2014, 10(7): 735.
- [28] Karr JR, Sanghvi JC, Macklin DN, et al. A whole-cell computational model predicts phenotype from genotype. *Cell*, 2012, 150(2): 389–401.
- [29] Ye C, Zou W, Xu N, et al. Metabolic model reconstruction and analysis of an artificial microbial ecosystem for vitamin C production. *J Biotechnol*, 2014, 182–183: 61–67.
- [30] Salimi F, Zhuang K, Mahadevan R. Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. *Biotechnol J*, 2010, 5(7): 726–738.
- [31] Monk JM, Charusanti P, Aziz RK, et al. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci USA*, 2013, 110(50): 20338–20343.
- [32] Bosi E, Monk JM, Aziz RK, et al. Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc Natl Acad Sci USA*, 2016, 113(26): E3801–E3809.
- [33] Seif Y, Kavvas E, Lachance JC, et al. Genome-scale metabolic reconstructions of multiple *Salmonella* strains reveal serovar-specific metabolic traits. *Nat Commun*, 2018, 9(1): 3771.
- [34] Magnúsdóttir S, Heinken A, Kutt L, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol*, 2017, 35(1): 81–89.
- [35] Garza DR, Van Verk MC, Huynen MA, et al. Towards predicting the environmental metabolome from metagenomics with a mechanistic model. *Nat Microbiol*, 2018, 3(4): 456–460.

- [36] Schellenberger J, Que R, Fleming RMT, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc*, 2011, 6(9): 1290–1307.
- [37] Heirendt L, Arreckx S, Pfau T, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc*, 2019, 14(3): 639–702.
- [38] Wang H, Marcišauskas S, Sánchez BJ, et al. RAVEN 2.0: a versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Comput Biol*, 2018, 14(10): e1006541.
- [39] Alper HS. *Systems Metabolic Engineering*. New York: Humana Press, 2013: 17–45.
- [40] Sanchez BJ, Nielsen J. Genome scale models of yeast: towards standardized evaluation and consistent omic integration. *Integr Biol (Camb)*, 2015, 7(8): 846–858.
- [41] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*, 2010, 28(3): 245–248.
- [42] Covert MW, Schilling CH, Palsson B. Regulation of gene expression in flux balance models of metabolism. *J Theor Biol*, 2001, 213(1): 73–88.
- [43] Covert MW, Xiao N, Chen TJ, et al. Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics*, 2008, 24(18): 2044–2050.
- [44] Mahadevan R, Edwards JS, Doyle III FJ. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J*, 2002, 83(3): 1331–1340.
- [45] Khandelwal RA, Olivier BG, Roling WFM, et al. Community flux balance analysis for microbial consortia at balanced growth. *PLoS ONE*, 2013, 8(5): e64567.
- [46] Lewis NE, Hixson KK, Conrad TM, et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol*, 2010, 6(1): 390.
- [47] Thiele I, Fleming RMT, Bordbar A, et al. Functional characterization of alternate optimal solutions of *Escherichia coli*'s transcriptional and translational machinery. *Biophys J*, 2010, 98(10): 2072–2081.
- [48] Benyamini T, Folger O, Ruppin E, et al. Flux balance analysis accounting for metabolite dilution. *Genome Biol*, 2010, 11(4): R43.
- [49] Kelk SM, Olivier BG, Stougie L, et al. Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Sci Rep*, 2012, 2: 580.
- [50] Smallbone K, Simeonidis E. Flux balance analysis: a geometric perspective. *J Theor Biol*, 2009, 258(2): 311–315.
- [51] Burgard AP, Pharkya P, Maranas CD. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*, 2003, 84(6): 647–657.
- [52] Xu ZX, Zheng P, Sun JB, et al. ReacKnock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. *PLoS ONE*, 2013, 8(12): e72150.
- [53] King ZA, Feist AM. Optimizing cofactor specificity of oxidoreductase enzymes for the generation of microbial production strains—OptSwap. *Ind Biotechnol*, 2013, 9(4): 236–246.
- [54] Ranganathan S, Suthers PF, Maranas CD. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol*, 2010, 6(4): e1000744.
- [55] Chowdhury A, Zomorodi AR, Maranas CD. k-OptForce: integrating kinetics with flux balance analysis for strain design. *PLoS Comput Biol*, 2014, 10(2): e1003487.
- [56] Pharkya P, Burgard AP, Maranas CD. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res*, 2004, 14(11): 2367–2376.
- [57] Kim J, Reed JL, Maravelias CT. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *PLoS ONE*, 2011, 6(9): e24162.
- [58] Segrè D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA*, 2002, 99(23): 15112–15117.
- [59] Ren SG, Zeng B, Qian XN. Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. *BMC Bioinformatics*, 2013, 14(S2): S17.
- [60] Tepper N, Shlomi T. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics*, 2010, 26(4):

- 536–543.
- [61] Jensen K, Broeken V, Hansen ASL, et al. OptCouple: joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs. *Metab Eng Commun*, 2019, 8: e00087.
- [62] Patil KR, Rocha I, Förster J, et al. Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics*, 2005, 6: 308.
- [63] Kim J, Reed JL. OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst Biol*, 2010, 4: 53.
- [64] Gu DQ, Zhang C, Zhou SG, et al. IdealKnock: a framework for efficiently identifying knockout strategies leading to targeted overproduction. *Comput Biol Chem*, 2016, 61: 229–237.
- [65] Pharkya P, Maranas CD. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab Eng*, 2006, 8(1): 1–13.
- [66] Rockwell G, Guido NJ, Church GM. Redirector: designing cell factories by reconstructing the metabolic objective. *PLoS Comput Biol*, 2013, 9(1): e1002882.
- [67] Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol*, 2007, 3(1): 119.
- [68] Fan SC, Zhang ZY, Zou W, et al. Development of a minimal chemically defined medium for *Ketogulonicigenium vulgare* WSH001 based on its genome-scale metabolic model. *J Biotechnol*, 2014, 169: 15–22.
- [69] Ye C, Xu N, Chen HQ, et al. Reconstruction and analysis of a genome-scale metabolic model of the oleaginous fungus *Mortierella alpina*. *BMC Syst Biol*, 2015, 9: 1.
- [70] Chen XL, Gao C, Guo L, et al. DCEO Biotechnology: tools to design, construct, evaluate, and optimize the metabolic pathway for biosynthesis of chemicals. *Chem Rev*, 2018, 118(1): 4–72.
- [71] Choi KR, Jang WD, Yang D, et al. Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. *Trends Biotechnol*, 2019, 37(8): 817–837.
- [72] Ranganathan S, Tee TW, Chowdhury A, et al. An integrated computational and experimental study for overproducing fatty acids in *Escherichia coli*. *Metab Eng*, 2012, 14(6): 687–704.
- [73] Xu P, Ranganathan S, Fowler ZL, et al. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab Eng*, 2011, 13(5): 578–587.
- [74] Xu N, Ye C, Chen XL, et al. Genome-scale metabolic modelling common cofactors metabolism in microorganisms. *J Biotechnol*, 2017, 251: 1–13.
- [75] Yang JH, Wright SN, Hamblin M, et al. A white-box machine learning approach for revealing antibiotic mechanisms of action. *Cell*, 2019, 177(6): 1649–1661.e9.
- [76] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521(7553): 436–444.
- [77] Moen E, Bannon D, Kudo T, et al. Deep learning for cellular image analysis. *Nat Methods*, 2019, doi: 10.1038/s41592-019-0403-1.
- [78] Gazestani VH, Lewis NE. From genotype to phenotype: augmenting deep learning with networks and systems biology. *Curr Opin Syst Biol*, 2019, 15: 68–73.

(本文责编 陈宏宇)