

· 人工细胞智能设计再造 ·

鲁洪中 上海交通大学生命科学技术学院院长聘教轨副教授, 细胞数字化建模课题组组长、独立 PI、博导。入选 2022 年上海市浦江人才计划(A 类)。长期从事细胞代谢网络数字化建模和表征, 并将最新的多维细胞代谢模型应用于菌株智能设计和生物大数据分析等前沿领域。相关成果以一作或通讯作者发表于 *Nature Communication*、*Proceedings of the National Academy of Sciences of the United States of America*、*Molecular Systems Biology*、*Metabolic Engineering*、*Briefings in Bioinformatics*、*Trends in Biotechnology* 等国际期刊。详细内容可参见网页 <https://life.sjtu.edu.cn/teacher/luhongzhong>。课题组致力于高精度数字细胞模型构建, 破译基因型-表型关系, 解析生命运作规律; 开发细胞工厂精准设计算法, 实现 BT 和 IT 技术的深度结合, 以期解决细胞代谢全局可预测、菌株理性改造等合成生物学领域的关键问题与挑战, 加速绿色生物产品“智”造, 为我国工业菌株的系统开发和可持续升级提供原创性解决方案。



细胞生理代谢网络模型重构和优化的进展

肖陆驰, 鲁洪中*

上海交通大学 生命科学技术学院 微生物代谢国家重点实验室 上海 200240

肖陆驰, 鲁洪中. 细胞生理代谢网络模型重构和优化的进展[J]. 生物工程学报, 2025, 41(3): 1112-1132.

XIAO Luchi, LU Hongzhong. Advances in reconstruction and optimization of cellular physiological metabolic network models[J]. Chinese Journal of Biotechnology, 2025, 41(3): 1112-1132.

摘要: 细胞中自发或由酶催化的代谢反应组成了高度复杂的代谢网络, 其与细胞生理代谢活动运作密切相关。细胞生理代谢网络模型的重构有助于从系统层面上解析基因型与生长表型之间的关联, 为细胞生理代谢活动精准刻画与生物绿色制造等研究提供重要的计算生物学工具。本文系统介绍了全基因组尺度代谢网络模型(genome-scale metabolic models, GEMs)、动力学模型、酶约束代谢模型(enzyme-constrained genome-scale metabolic models, ecGEMs)等不同类型细胞生理代谢网络模型发展与应用的最新研究进展; 同时还介绍了 GEMs 自动化构建研究进展以及条件特异性 GEMs 建模策略。人工智能技术为高精度细胞生理代谢网络模型构建提供了全新机遇, 本文进一

资助项目: 国家重点研发计划(2022YFA0913000); 上海市浦江人才计划; 国家自然科学基金(22208211, 22378263)
This work was supported by the National Key Research and Development Program of China (2022YFA0913000), the Shanghai Pujiang Program, and the National Natural Science Foundation of China (22208211, 22378263).

*Corresponding author. E-mail: hongzhonglu@sjtu.edu.cn

Received: 2024-12-11; Accepted: 2025-03-11; Published online: 2025-03-13

步总结了人工智能技术在动力学模型和酶约束模型构建等领域的应用。各类细胞生理代谢网络模型的高质量重构将为今后的定量合成生物学与系统生物学等研究提供强大计算支撑。

关键词：基因组尺度代谢网络模型；动力学模型；酶约束代谢模型；深度学习

Advances in reconstruction and optimization of cellular physiological metabolic network models

XIAO Luchi, LU Hongzhong*

State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: The metabolic reactions in cells, whether spontaneous or enzyme-catalyzed, form a highly complex metabolic network closely related to cellular physiological metabolic activities. The reconstruction of cellular physiological metabolic network models aids in systematically elucidating the relationship between genotype and growth phenotype, providing important computational biology tools for precisely characterizing cellular physiological metabolic activities and green biomanufacturing. This paper systematically introduces the latest research progress in different types of cellular physiological metabolic network models, including genome-scale metabolic models (GEMs), kinetic models, and enzyme-constrained genome-scale metabolic models (ecGEMs). Additionally, our paper discusses the advancements in the automated construction of GEMs and strategies for condition-specific GEM modeling. Considering artificial intelligence offers new opportunities for the high-precision construction of cellular physiological metabolic network models, our paper summarizes the applications of artificial intelligence in the development of kinetic models and enzyme-constrained models. In summary, the high-quality reconstruction of the aforementioned cellular physiological metabolic network models will provide robust computational support for future research in quantitative synthetic biology and systems biology.

Keywords: genome-scale metabolic models (GEMs); kinetic models; enzyme-constrained genome-scale metabolic models (ecGEMs); deep learning

细胞生理代谢网络模型是开展系统生物学和合成生物学研究的重要工具之一，能够从系统层面解析基因型与表型之间的关系，并为生物制造、疾病治疗等领域提供模拟和分析支持。近年来，随着基因组测序技术、多组学数据整合以及人工智能技术的快速发展，细胞生理代谢网络模型在构建与应用等领域取得显著进展，其中基因组尺度代谢网络模型(genome-scale

metabolic models, GEMs)、动力学模型和酶约束代谢模型(enzyme-constrained genome-scale metabolic models, ecGEMs)等不同类型模型已广泛应用于代谢工程、精准医学等领域。本文按照模型类别分别总结这些模型的最新研究进展及其在生物制造和系统生物学中的典型应用，并展望了细胞生理代谢网络模型的未来发展方向。

1 基因组尺度代谢网络模型 (GEMs)

1.1 GEMs 的发展历程

在生物体的代谢活动中,大多数代谢反应由基因组编码的蛋白酶催化,这些生化反应形成了一个复杂的代谢网络^[1]。代谢网络代表了生物体与其生物和非生物环境相互作用并将养分转化为生物质的能力,因此代谢网络的构建与分析在系统生物学的研究中至关重要。数学的分析方法能够从拓扑、计量和动力学等不同水平上解析代谢网络的功能^[2],并提供一系列建模的方法^[3],其中基因组尺度代谢网络模型(GEMs)最为典型。GEMs以化学计量矩阵的形式整合生物体内的代谢物与反应信息,通过基因-蛋白-反应(gene-protein-reaction, GPR)规则将基因与反应相关联,并通过代谢通量平衡分析模拟细胞在不同环境和条件下的代谢行为。在实际应用中,GEMs不仅展现了其在预测细胞生长表型和基因敲除致死性方面的能力,还能够揭示基因型与代谢表型的潜在关联机制^[4]。由于这种预测潜力,GEMs已被广泛应用于研究宿主-微生物群体相互作用^[5]、预测药物靶点^[6],以及辅助工业生产或生物降解工程中的微生物菌株设计^[7]等。此外,DNA测序技术的最新进展使得数据库中基因组和宏基因组序列大幅增加^[8],为GEMs构建提供了充分的基因组数据来源。

大肠杆菌(*Escherichia coli*)是应用最为广泛的原核模式生物之一,研究人员针对常用的模式菌株如K-12 MG1655菌株,不断优化更新其代谢模型。1991年Edwards等^[9]发表了该菌株的首个大规模代谢网络模型,整合了已有的代谢路径数据,构建了代谢网络的初步框架。随后,Kauffman等^[10]利用基于约束的方法构建

了基因组尺度代谢网络模型(GEMs),这些模型能够预测细胞在不同环境条件下的代谢行为。著名的iJR904模型^[11]和iAF1260模型^[12]通过整合基因组序列数据、代谢途径数据以及文献中报道的数据,构建了较为完善的代谢网络模型。最新的模型如iJO1366^[13]和iML1515^[14]进一步优化了代谢网络的准确性和覆盖面,并结合了高通量数据和新兴的计算方法,在代谢路径的预测和合成生物学应用中表现出色。

在真核生物中,酿酒酵母(*Saccharomyces cerevisiae*) GEMs的构建也受到了广泛关注。随着*S. cerevisiae*的全基因组序列于1996年公布^[15],以此为基础构建的首个酿酒酵母的代谢模型iFF708于2003年发布^[16]。之后,研究人员发布了一系列不断更新的*S. cerevisiae* GEMs^[17],每个新版本都对之前的模型进行了逐步改进。在这些模型中,公共的酿酒酵母GEMs(Yeast1^[16]-Yeast9^[18])尤为值得关注。此系列模型由研究者社区不断完善,自2013年Yeast7.6发布以来^[19],这些模型在描述脂质、风味因子、辅因子和底物代谢等方面的完整度和质量得到了显著提升。研究人员在*S. cerevisiae*模型iWS902中校正并扩展了与风味因子形成相关的途径,阐释了酵母在工业应用过程中的香气形成机理^[20]。为了囊括辅因子代谢途径,酿酒酵母的铁元素代谢网络被整合到Yeast7.6中^[21],获得新版本模型Yeast7.Fe。在Yeast8中,研究人员进一步采用文献报道的辅因子浓度更新了模型的生物质组成,在Yeast8模型中,研究团队基于文献报道的辅因子浓度数据优化了生物质组分的表征,这一改进使得先前基因组尺度代谢模型(GEMs)中处于静默状态的相关代谢子途径^[22]被激活;Yeast8还基于微生物表型芯片技术所获得的胞内底物利用数据扩展了底物利用范围^[22]。此外,研究人员还向模型中添加了13种

额外的芳香族化合物及其相关反应, 以扩展 Yeast8 在工业葡萄酒生产中的应用^[23]。在最新的 Yeast9 模型中, 研究人员针对挥发性酯类和聚磷酸盐相关反应进行扩展, 进一步提升了代谢网络的完整度。

1.2 GEMs 自动化构建工具包介绍

尽管研究人员已经构建了如 *S. cerevisiae* 和 *E. coli* 等模式物种的高质量基因组尺度代谢网络模型(GEMs), 非模式物种 GEMs 的构建仍然面临较大挑战, 其主要原因在于非模式物种的实验数据匮乏导致基因结构与蛋白质功能注释信息不足, 以及手动构建代谢网络模型所耗费的时间与人力巨大。因此, 开发针对非模式物种的 GEMs 自动化构建工具成为了研究的重点。2018 年发布的 RAVEN 2.0 工具中提供了基于 KEGG 和 MetaCyc 数据库的 GEMs 快速构建工具^[24], 该工具主要采取自下而上的构建流程,

从多种数据库或模型来源中比对搜索代谢反应组成模型草稿, 再进一步进行模型优化。采用类似建模思路的还包括 ModelSEED^[25]和 Merlin^[26]等建模工具。同样于 2018 年发布的 CarveMe 工具^[27]重点应用了自上而下的模型构建方法(图 1), 该工具在详细校正过的通用模型(universal model)的基础上, 首先基于目标物种基因组与通用模型基因组信息的比对情况对模型中反应的存在性进行打分, 随后基于打分对反应进行筛选得到初始模型, 随后在给定的培养基上对模型进行缺口填补使得模型能够表征相应代谢特性; 与之类似的 gapseq 工具^[28]同样使用了一个来自 ModelSEED^[25]的反应和代谢物数据库来构建通用模型, 并采用 tblastn 来进行数据库中蛋白序列和目标物种基因组之间的比对, 并以比对得分作为权重采用简约酶通量平衡分析(parsimonious enzyme flux balance

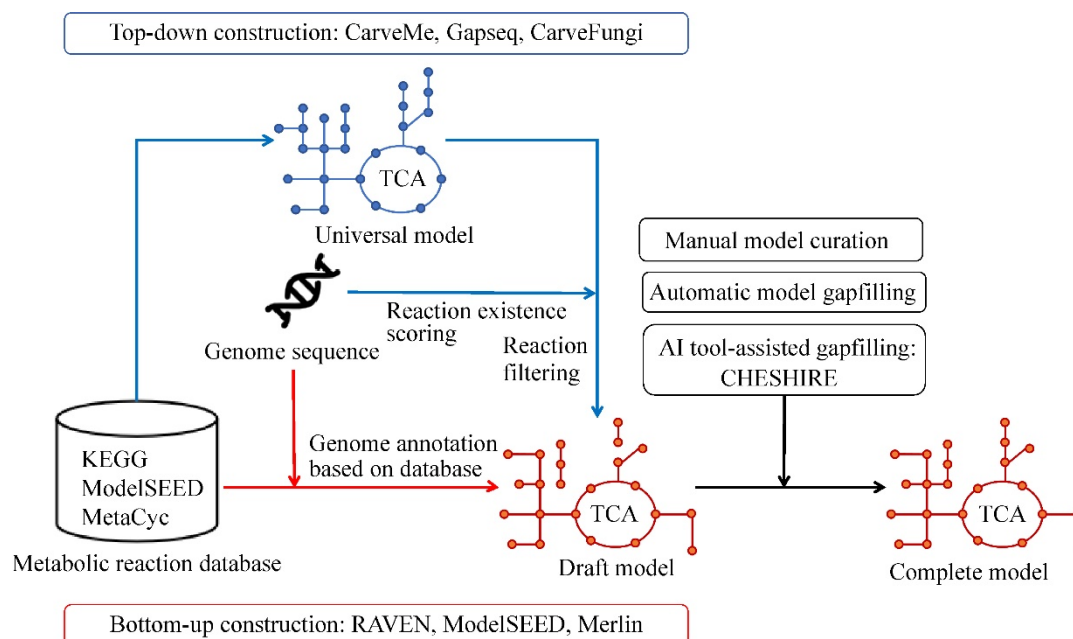


图 1 基因组尺度代谢网络模型建模思路 以 CarveMe 为代表的自上而下的建模方法和以 ModelSEED 为代表的自下而上的建模方法。

Figure 1 Modeling approaches for genome-scale metabolic network models. Top-down modeling methods represented by CarveMe and bottom-up modeling methods represented by ModelSEED.

analysis, pFBA)算法求解, 求解结果中通量大于 0 的反应被添加到目标物种的初始模型中, 随后基于生物量生产、碳源利用和代谢产物合成等条件进行缺口填补。

近年来, 随着深度学习领域的快速发展, 深度学习模型也被尝试应用到 GEMs 的自动构建流程中。CarveFungi 工具^[29]在 CarveMe^[27]的基础上, 引入深度学习模型预测反应发生区室, 进一步提高反应打分的可靠性, 进而提高建模质量。由于基因功能注释不准确、测序信息错误等原因, 非模式物种 GEMs 中往往存在一些缺口。针对这一问题, 基于图神经网络的深度学习模型工具 CHESHIRE^[30]通过学习代谢反应与代谢物之间的拓扑关系进行 GEMs 的缺口填补。在对 49 个物种进行代谢建模时, 采用 CHESHIRE 进行缺口填充的模型在预测发酵产

物和蛋白质的分泌等方面取得了显著提升, 证明 CHESHIRE 在代谢模型缺口填补方面具备重要应用潜力。上述自动化建模工具的优缺点比较以及适用范围在表 1 中进行了总结^[30]。

1.3 条件特异性 GEMs 构建进展

由于细胞内的蛋白分配和代谢调控是一个随时对外界条件做出适应调整的动态过程, 如何重建特定环境下的基因组尺度代谢网络模型 (GEMs), 即将实验所获取的组学数据整合到 GEMs 中从而提高 GEMs 的预测性能, 成为研究人员关注的重点。目前组学整合算法的原理可以分为 4 大类: 需要目标函数(objective function-required, OFR)、与表达数据兼容(expression data-compatible, EDC)、源自代谢任务(metabolic task-derived, MTD)和需要核心反应(core reaction-required, CRR)。OFR 算法将基

表 1 当前主流自动化建模工具优缺点与适用范围

Table 1 Summary of the advantages, disadvantages, and applicable scopes of current automated modeling tools

Tool name	Construction method	Advantages	Disadvantages	Applicability
RAVEN 2.0	Bottom-up	Integrates multiple databases, deeply explores metabolic potentials	Requires manual curation, which makes it time-consuming for large-scale modelling	Most species
ModelSEED	Bottom-up	Comprehensive database integration, supports phenotype optimization, high-quality core metabolism models	Only web version available (no open-source version), cannot deploy locally for large-scale modeling	Prokaryotes
CarveMe	Top-down	Template-based construction ensures relatively high-quality output models	Accuracy limited by templates, may omit complex reactions	Bacteria (Prokaryotes)
CarveFungi	Top-down (combined with deep learning)	Incorporates cellular compartment prediction and deep learning scoring to enhance accuracy	Primarily validated for fungi, effectiveness in other species needs verification	Fungi
Merlin	Bottom-up	Precise annotation of transport proteins; localized databases improve efficiency	Model quality depends on manual curation	Most species
Gapseq	Bottom-up	Uses tblastn alignment for deeper homology searches; optimizes gap-filling steps with pFBA algorithm	Long runtime (>1 hour per model)	Most species

因表达数据如代谢调节导致的基因失活(gene inactivation moderated by metabolism, GIMME)^[31]、代谢组学(metabolomics and expression, GIM3E)^[32]以及蛋白质组调节导致的基因失活(gene inactivity moderated by metabolism and expression by proteome, GIMMEp)^[33]等信息有机整合到 GEMs 中,通过最小化 GEMs 代谢通量和表达数据之间的不一致性来进行模型优化,同时保证目标函数(如代谢产物生成率或生长速率)维持正常状态。因此,得到的条件特异性 GEMs 具备相应代谢生产功能,同时能够尽可能与基因表达数据相吻合。由于 OFR 算法注重满足目标函数求解的可行性,GEMs 预测的代谢通量和实验数据中的基因表达水平之间仍存在不一致性,该问题尚未得到完全解决。另一方面,EDC 算法如整合代谢分析工具(integrative metabolic analysis tool, iMAT)^[34]、组织整合网络推断(integrative network inference for tissues, INIT)算法^[35]、E-Flux 算法^[36]和代谢概率调控(probabilistic regulation of metabolism, PROM)算法^[37]等在没有设定目标函数的情况下最大限度地提高了通量和基因表达水平之间的一致性。因此,虽然不一致性可以解决,但这些算法未考虑菌体生长或其他代谢产物生产的可行性,其所获得的条件特异性 GEMs 可能不具备功能性,无法用于精确模拟细胞生理代谢活动。为了分别解决 OFR 或 EDC 算法中的这 2 个问题,研究人员开发了 MTD 算法,如任务驱动的组织整合网络推断算法(task-driven integrative network inference for tissues, tINIT)^[38]、个性化代谢模型重建算法(personalized reconstruction of metabolic models, PRIME)^[39]、基于差异表达的代谢调整算法(metabolic adjustment by differential expression, MADE)^[40]和转录推断的代谢生物标志物响应算法(transcriptionally inferred metabolic biomarker

response, TIMBR)^[41]。MTD 算法在考虑目标函数或代谢任务的同时整合基因表达数据,最终重建的条件特异性 GEMs 与多组学数据之间的一致性比 OFR 算法更高。最新开发的 MTD 算法 TIMBR 首先使用转录组数据计算各种条件下的基因表达差异,然后在满足目标函数(如特定代谢物生产)的情况下向模型中增加反应,从而最小化多个条件下的通量权重的总和^[41]。CRR 算法主要包括模型构建算法(model-building algorithm, MBA)^[42]、基于确定性反应评估代谢环境特异性的算法(metabolic context-specificity assessed by deterministic reaction evaluation, mCADRE)^[43]、核心代谢网络快速重构算法(FASTCORE)^[44]、通过功能基因组学进行代谢重建算法(metabolic reconstruction via functional genomics, MIRAGE)^[45]和成本优化反应依赖性评估算法(cost optimization reaction dependency assessment, CORDA)^[46]等,这些算法将核心反应集作为输入数据之一,用来消除模型中不必要的反应或阻断反应。上述组学整合算法大多基于 MATLAB 计算框架 COBRA Toolbox v2.0^[47]和 RAVEN Toolbox 2.0^[24],或基于 Python 的 COBRAPy^[48]开发。

基于以上组学整合算法构建的条件特异性 GEMs 能够捕捉特定细胞或组织在特定环境下的代谢特性。如基于 tINIT 算法结合患者特异性蛋白质组数据构建的肝细胞癌(hepatocellular carcinoma, HCC)患者的个性化 GEM 能够预测特定代谢抑制剂对肿瘤的杀伤作用,为 HCC 的精准治疗提供潜在靶点^[38]。此外,研究人员采用 iMAT 算法,以人类 GEMs-Recon1^[49]为基础,整合多组学数据重建了肝细胞 GEMs。重建后的肝细胞 GEMs 描绘了肝脏代谢活动和功能图谱,并揭示了导致非酒精性脂肪肝病(nonalcoholic fatty liver disease, NAFLD)形成的关键分子线

索^[50]。条件特异性 GEMs 的应用不仅深化了对人类疾病代谢机制的理解,还为特定疾病的诊断和治疗提供了分子靶标,其在精准医学和代谢病研究中具有广阔的应用前景。

1.4 GEMs 在生物制造领域的应用

基于 GEMs 的通量平衡分析(flux balance analysis, FBA)可以鉴别生物制品合成的关键瓶颈。如在蛛丝蛋白合成中,研究人员基于 ¹³C 代谢通量分析获得的实验数据,分别对野生菌株和工程菌株的基因组尺度代谢模型(GEMs)施加通量约束并进行通量平衡分析,发现工程菌株在高代谢压力条件下存在 ATP 产量和三羧酸循环(tricarboxylic acid cycle, TCA)还原氢产量下降的现象,同时氨基酸合成过程所消耗 ATP 与还原氢显著增加。基于此研究人员采取了针对性策略,如向培养基中补充氨基酸以减少能量消耗,有效提高了工程菌株的蛋白产量^[51]。

GEMs 有助于生物制造领域中底盘菌株的理性设计和系统代谢工程研究。多个基于 GEMs 菌株的设计算法,如 OptKnock^[52]与 OptForce^[53]等,已成功用于提高目标产物合成效率。在此前的研究中,基于 OptKnock 预测实现了产物合成与生长耦合,成功提高了大肠杆菌中乳酸^[54]和 1,4-丁二醇的产量^[55],提升幅度分别达到 73%和 300%。基于 OptForce 预测,识别影响产物合成的关键反应,将大肠杆菌中的丙二酰-CoA 水平提升至 400%,进而实现柚皮素^[56]和脂肪酸^[57]的高效生产。此外,结合其他算法如代谢调整最小化(minimization of metabolic adjustment, MOMA)算法^[58]和 OptGene^[59],进一步提高了酿酒酵母中倍半萜^[60]和香草醛产量^[61]。这些算法的应用为工业生物制造提供了高效的菌株设计策略,推动了高附加值化学品的可持续生产。

1.5 跨物种代谢网络模型辅助异源途径探索

在生物合成研究中,细胞内单位底物生成产物的量,即途径产率 Y_p (pathway yield),是设计高效细胞工厂的关键指标。传统的单物种 GEM 可通过通量平衡分析(FBA)模拟计算目标产物的途径产率 Y_p ,但由于其仅包含种属特异性反应,限制了非天然合成产物的产率评估。此外,单物种 GEM 反应的多样性不足,无法探索通过异源反应提高 Y_p 的策略。因此研究人员尝试构建涵盖多物种生化反应的集成代谢网络模型,即跨物种代谢模型(cross-species metabolic networks, CSMNs)。

CSMNs 的基本构建思路与 GEMs 自动建模中的通用模型(universal model)类似,即通过整合跨物种生化反应数据库扩展单物种 GEM,以构建集成模型。例如,Chaturachai 等^[62]通过分步添加来自代谢数据库的异源反应整合了多个物种的 GEM,并基于集成模型评估非天然产物的合成可能性。然而高质量集成模型的获得需要进行严谨的质量控制和校正流程,否则可能由于模型中还原当量、能量或代谢物无限生成的错误使模型无法进行有效的产率评估(如模拟最大产率超出基于底物与产物还原度计算的理论最大产率)。Wei 等^[63]开发出了质量控制流程,通过配平反应分子式与电荷、校正反应方向以及基于 pFBA 算法筛选错误反应使得模型成为可靠的预测工具,并同时开发了定量异源途径设计算法(quantitative heterologous pathway design algorithm, QHEPath)用于探索能够突破产物 Y_p 的异源反应。在基于该算法设计的涉及 5 种工业微生物、4 种底物的 13 种能够突破产率极限的工程策略中,部分策略已得到实验验证,如在大肠杆菌中,丙酮的天然合成途径因丙酮酸脱羧生成乙酰辅酶 A 时的碳损失导致产

率较低, QHEPath 算法发现引入磷酸酮醇酶 (PKETF)反应可以将模拟所得 Y_p 得到由初始的 1.0 mol/mol 葡萄糖提升至 1.5 mol/mol 葡萄糖^[63]。该策略经 Yang 等^[64]的实验证明能够将工程菌株的丙酮产率从 0.38 mol/mol 葡萄糖 提升至 0.47 mol/mol 葡萄糖。

2 动力学模型

2.1 当前动力学模型的发展历程

与基于稳态假设的代谢网络模型(GEMs)不同, 代谢动力学模型基于细胞内代谢物的动态变化和酶促反应动力学, 采用常微分方程 (ordinary differential equations, ODEs)描述代谢物浓度随时间的变化。代谢物浓度和代谢反应速率被表征为随着时间动态变化的函数, 而其中的代谢反应速率的数学表达受到动力学理论(如过渡态理论)、热力学理论(如平衡理论、线性和非线性非平衡理论)和第一性原理(如代谢物的质量平衡)等一系列因素的驱动, 受到代谢物浓度以及酶动力学参数等一系列因素约束。

构建大规模动力学模型的主要困难在于动力学参数的获取。模型所需的动力学参数包括酶浓度、酶在正反 2 个反应方向上的转换速率、底物与酶结合的解离常数、抑制和激活常数以及平衡常数, 即酶促反应的标准吉布斯自由能。由此产生的高维度非凸参数分布空间使得当代代谢网络规模扩大时, 参数估计变得极为困难。研究人员往往采用优化算法, 通过最小化实验数据和模型模拟之间的差异来间接估计参数, 而数据库中提供的数据则往往用作参数的初始化和或用于对参数估计值施加约束^[65-66]。

由于动力学模型中的参数估计较为困难, 动力学模型的拓展往往依赖于新的实验数据的获取从而逐步进行。以酿酒酵母为例, 在早期

阶段, 其动力学模型主要涵盖特定的子途径, 例如海藻糖代谢^[67]、糖酵解^[68]和鞘脂生物合成^[69]。尽管这些模型已经可以在特定环境或遗传扰动下模拟细胞动态过程, 研究人员仍尝试构建更完整的模型以描述细胞整体的代谢活动。研究人员建立了一个包含了核心代谢途径的动力学模型, 囊括了糖酵解途径、戊糖磷酸途径和柠檬酸循环, 该模型能够定量预测酵母在葡萄糖和乙醇上的动态生长行为^[70]。此后, *S. cerevisiae* 代谢的动力学模型进一步扩展, 当前已经能够涵盖 102 个反应和 94 个细胞内代谢物^[71], 并且囊括了木糖同化途径, 使得模型能够在以葡萄糖和木糖作为混合底物的培养条件下取得较好的预测结果。

GEMs 的发展使得研究人员能够构建具有更大范围的动力学模型。较早期的酵母 GEMs, Yeast4^[72], 已被用作构建中等大小动力学模型的底盘。研究人员首先对 Yeast4 模型开展代谢通量分析, 筛选出通量值低于设定阈值的反应并将其剔除, 从而缩小模型规模; 在筛选后的包含 285 个反应的动力学模型中, 尽管研究人员仅使用了少量实验数据进行参数拟合, 模型仍然能够较为准确地预测细胞对外部底物浓度变化的动态响应^[73]。此外, 研究人员还在 iND750^[74]的基础上使用基于基因组规模多组学数据的贝叶斯推断方法估计动力学参数, 构建了一个具有 240 个反应和 203 个代谢物的中等大小的 *S. cerevisiae* 代谢动力学模型^[75]。该模型在表征酶与代谢物之间复杂的相互作用方面表现良好。而在大肠杆菌的建模方面, Khodayari 等^[76]采用了 iAF1260 模型^[12]作为基础构建了一个中型的大肠杆菌动力学模型——k-ecoli457, 该模型包含 457 个反应及其对应的底物-酶相互作用关系; 相较于传统模型, k-ecoli457 整合了更多的代谢流数据, 包括来

自 25 个突变株的实验数据, 涵盖了不同的生长条件和碳源利用表型; 此外, k-ecoli457 通过结合机器学习算法和集成模型(ensemble modeling, EM), 更精确地估算了反应中酶的动力学参数(如 K_m 和 k_{cat} 值), 使得该模型能够在多变的生长条件下更准确地预测代谢反应的行为; 与传统的基于通量平衡分析(FBA)的算法相比, k-ecoli457 能够更好地预测突变株中 24 种生物代谢产品的产量, 其实验测量值与预测值之间的 Pearson 相关系数达到 0.84, 展示了动力学模型预测的精准性。

2.2 动力学模型的自动化建模

当前研究人员聚焦于开发工具和算法来校准动态模型^[77], 特别是动力学模型的参数

拟合^[78-79], 但能够实现代谢动力学模型的自动构建的工具包仍然较少。2022 年发布的 GRASP 开源工具包^[80]采用 3 步工作流程构建动力学模型: (1) 输入准备阶段, 需定义代谢网络的稳态通量分布、相关反应的吉布斯自由能范围及酶调控机制等核心参数; (2) 参数采样与拟合阶段, 通过先验采样或近似贝叶斯计算拒绝采样方法, 生成符合热力学约束的动力学模型集合; (3) 后验分析阶段, 可对模型集合进行代谢控制分析、动态模拟及模型选择, 并支持 SBML 格式输出(图 2)。GRASP 基本实现了动力学模型的自动化构建, 能够探索细胞内可行的代谢动力学过程、基于贝叶斯进行参数拟合与模型筛选, 以及进行代谢活动的动态仿真。

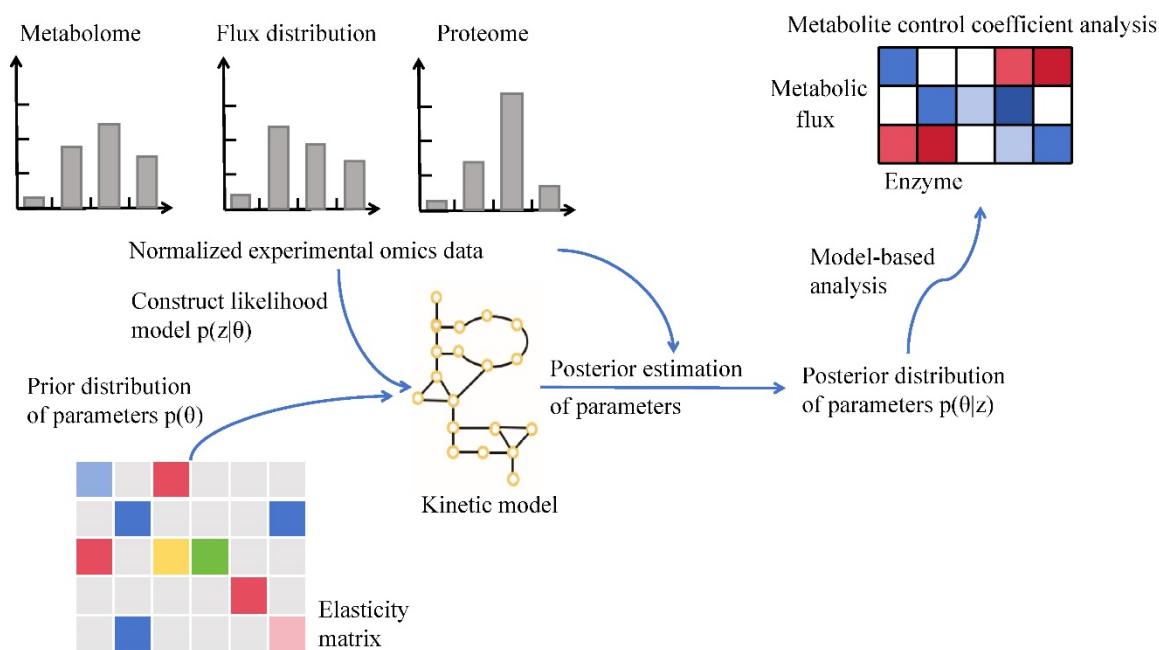


图 2 动力学模型的建模思路 基于代谢组学数据以及酶动力学参数设定模型参数的先验分布以及一定参数下模型通量分布的似然模型; 基于代谢组学数据, 通过采样的方法, 基于参数的后验分布对模型参数进行拟合; 基于模型运行代谢物控制系数分析等定量分析方法, 分析生物的动力学特征。
Figure 2 Modeling approaches for kinetic models. Model parameters are set with prior distributions based on metabolomics data and enzyme kinetics parameters, along with a likelihood model of flux distribution under certain parameters. Using sampling methods based on the posterior distribution of parameters, the model parameters are fitted with metabolomics data. Additionally, quantitative analyses such as metabolite control coefficient analysis are employed to assess the kinetic characteristics of the organism.

此外,深度学习领域的快速发展也为解决动力学模型构建中的参数调优问题提供了新的思路和方法。反应动力学参数的调整优化是动力学模型构建中的难点,而当前各种机器学习程序已被用于优化动力学模型或酶约束模型。例如,Choudhury 等^[81]开发了一个基于对抗学习网络的框架 REKINDLE,以快速重建能够反映细胞代谢动态特性的动力学模型。为了进一步减少动力学模型重建中的计算时间,Choudhury 等^[82]进一步建立了一种名为 RENAISSANCE 的新策略,该策略可以有效地完成动力学模型的参数化以表征目标生物的真实动态特性。

2.3 动力学模型在生物合成领域的应用

动力学模型相较于传统的 GEMs 模型,涵盖了更详细的酶动力学、反应动力学和热力学参数,能够更加精准地模拟代谢的动态变化过程,被广泛应用于指导菌株设计。

Chowdhury 等^[83]于 2014 年开发了基于动力学模型的菌株设计算法 k-OptForce,该算法参考了已发表的 OptForce 算法^[53]。与 OptForce 类似,k-OptForce 算法通过比较野生型和突变株的代谢通量来识别与基因上调、下调和基因敲除相关的代谢反应,然后进一步筛选和优化。相较于基于传统的静态计量模型的基因靶点筛选方法,k-OptForce 在发掘潜在的菌株改造靶点方面展现出了显著的优势。此外,Narayanan 等^[84]近期发表了一个基于动力学模型的新的菌株设计框架,称为基于非线性动态模型的理性代谢工程设计(nonlinear dynamic model assisted rational metabolic engineering design, NOMAD),该框架在保证菌株表型与参考菌株相近的条件下,通过非线性模拟的方法评估基因扰动在代谢网络中的响应,并基于此识别基因靶标以用于设计工程菌株。此框架的重点在于,在完成设计目标的同时使得大部分工程菌

株的稳健性保持与参考菌株相当的水平;为了进行验证,研究人员使用 NOMAD 预测提高大肠杆菌中蒽甲酸生产的潜在基因靶标,其能够预测到先前研究中已确定的 8 个基因靶标,包括香草酸途径(shikimate pathway)的关键酶(如 DDPA 和 SHKK)的上调,同时还识别了新的基因靶标以供进一步实验验证^[84]。

3 酶约束模型

3.1 ecGEMs 的原理与构建算法工具

由于准确的酶动力学参数难以获取,大规模的动力学模型构建仍然存在困难。为了避免对每个反应的酶动力学进行详细描述,研究人员开发出了新的建模策略,即酶约束代谢网络模型(ecGEMs),这种模型将决定反应通量的酶的属性作为约束条件加入模型(图 3)。ecGEMs 继承了传统 GEMs 的所有约束条件,包括代谢稳态平衡和代谢通量上下限。一个关键区别是 ecGEMs 中引入了一个新的约束:

$$v_j \leq k_{cat}^i * [E_i]$$

其中: v_j 是反应 j 的代谢通量, $[E_i]$ 是酶 i 的内部浓度, k_{cat}^i 是由酶 i 催化的反应 j 的酶转化数^[85-86]。

酶约束与模型的具体整合方式因采用的建模方法而异。基于是否在化学计量矩阵中显式引入转化数和酶质量平衡约束,可以将当前主流的建模工具分为 2 类。第一类包括分子拥挤效应通量平衡分析(flux balance analysis with molecular crowding, FBAwMC)^[87]、酶动力学代谢建模(metabolic modeling with enzyme kinetics, MOMENT)^[85]和 Python 中的酶约束代谢模型(enzyme-constrained metabolic models in Python, ECMpy)^[88],这些方法不改变原始的化学计量矩阵;另外一类包括基于动力学和组学数据的酶约束基因组代谢模型(GEM with enzymatic constraints using kinetic and omics data, GECKO)^[86]、

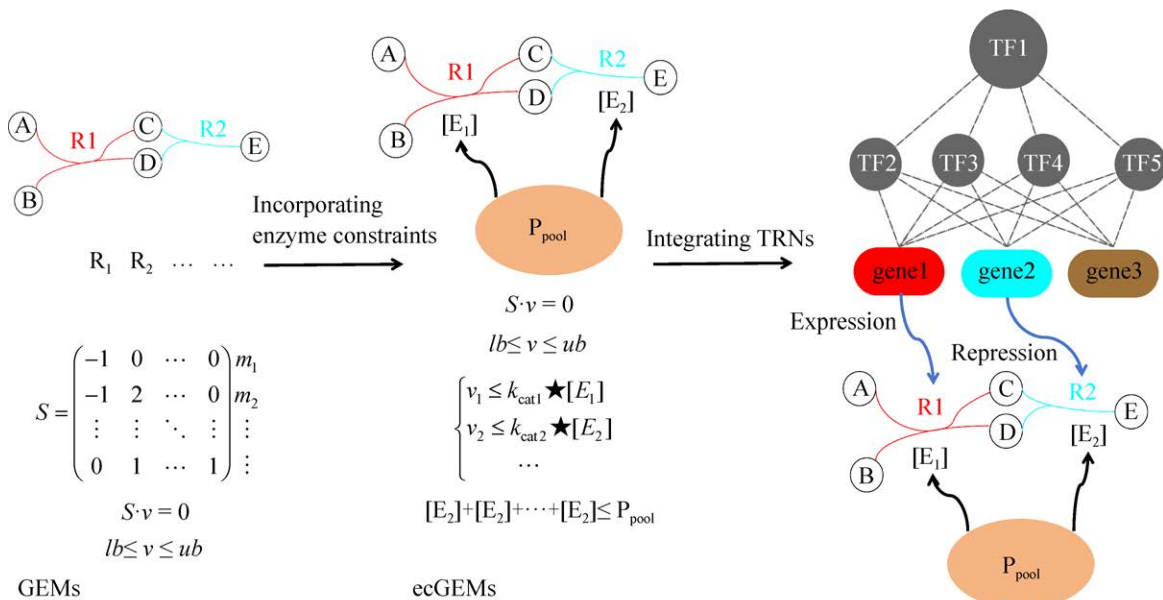


图 3 酶约束模型(ecGEMs)的构建与拓展 酶约束模型(ecGEMs)往往通过向 GEMs 中添加针对通量的酶约束条件来构建,其中酶的浓度往往受到胞内蛋白总量的约束;在 ecGEMs 的基础上进一步在模型中整合转录调控网络,能够模拟转录因子对基因和相关蛋白的调控作用,提升模型模拟效果。

Figure 3 Construction and expansion of enzyme-constrained models (ecGEMs). These models are often constructed by adding enzyme constraint conditions targeting flux to GEMs, where enzyme concentrations are typically limited by the total amount of intracellular proteins. Building upon ecGEMs, further integration of transcriptional regulatory networks into the model allows for the simulation of transcription factor regulation on genes and associated proteins, enhancing the model's simulation performance.

简化版酶动力学代谢建模(short MOMENT, sMOMENT)^[89]、蛋白质分配模型(protein allocation model, PAM)^[90]和 OVERLAY^[91],这些方法扩展化学计量矩阵以整合蛋白质约束。如果有蛋白质组学测量数据可用,可以用来指定酶的浓度 $[E_i]$ 。在没有这类测量的情况下,一个常见的假设是酶的使用受到总酶池 E_{pool} 的限制,该变量被用于表征模型中整合的所有代谢酶的总催化能力^[86,89]。ecGEMs 的另一个特点是将每个可逆反应分为 2 个不可逆的正向和反向反应。这样能够保证模型中的反应通量值为正从而简化了算法,同时也可以为可逆反应的 2 个方向赋予不同 k_{cat} 值,因为不同方向上的底物亲和力可能有所不同^[85]。

首个将酶参数整合进 GEMs 的工具是结合了

分子拥挤效应的通量平衡分析算法(FBAwMC),该方法利用可实验获取的细胞质密度、酶活、底物浓度和底物浓度等信息^[87],引入拥挤系数将代谢通量和酶的空间占用相联系,通过涉及细胞体积(V)的约束条件来实现对代谢通量的酶约束;2012 年发表的 MOMENT^[85]算法同样基于酶的总量来对酶浓度设立约束条件,并由酶浓度限制代谢通量的上下限,并且改进了异构酶和酶复合体的处理方法;sMOMENT 工具则在 MOMENT 的基础上消去了约束条件中的酶浓度变量,直接设立基于代谢通量的约束^[89];受 sMOMENT 中重新表达酶约束关系来减少模型复杂性的方法启发, Mao 等^[88]报道了一种被称为 ECMpy 的方法,其与 sMOMENT 工具类似,但进一步优化了酶约束条件,向其中引入

了酶的饱和系数; GECKO 工具与 MOMENT 的约束条件设置方法类似, 并在异构酶、多功能酶和酶复合体的处理方面做出了改进^[86]; PAM 工具在 GECKO 的基础上做进一步拓展, 将总酶使用量划分为 3 个部分: 转录蛋白、未使用的酶和活跃的酶, 并针对不同的酶设定不同的约束条件^[90]; OVERLAY 工具则用酶的有效转化率替代酶催化效率, 对于酶复合体则将其视为单一实体关注整体的转化率, 而不同于其他工具中分别处理复合体内的各个亚基的做法^[91]。

3.2 酶约束代谢模型参数的预测与优化

当前的酶约束模型建立工具建模原理大致类似, 其核心都在于如何基于有限的实验数据, 如蛋白质丰度和酶活性数据等尽可能准确地优化模型参数。缺乏高质量的酶丰度和催化效率是限制高质量酶约束模型构建的关键因素。因此针对这 2 种酶动力学参数的预测和调优方法就显得至关重要。

首先是针对酶的催化效率的预测和优化。Valgepea 等^[92]首先提出了一种方法, 通过对指数生长期大肠杆菌的绝对蛋白组和代谢通量的量化分析计算酶的表现催化效率作为催化效率的预测值; Davidi 等^[93]则采用蛋白质组数据估算酶浓度, 随后基于 pFBA 算法估算代谢通量, 进而估算在体水平酶催化效率; 在此基础上, Wilken 等^[94]报道了一种基于全局优化的参数估算方法, 其在满足 ecGEMs 的约束条件的情况下设定目标函数来最小化预测和实验测量得到的通量及酶浓度之间的误差, 通过将酶催化效率参数的预测转化为 ecGEMs 中的全局优化问题, 有效提升了参数预测的准确性; 基于机器学习和深度学习模型的数据驱动方法如基于深度学习的 k_{cat} 预测 (deep learning-based k_{cat} prediction, DLKcat)^[95]、预训练的 k_{cat} 预测模型 (unified framework based on pretrained language

models, UniKP)^[96]、DeepEnzyme^[97]和酶周转数预测器 (turnover number predictor, TurNuP)^[98]等工具能够基于蛋白序列和底物结构等模型以外的信息对催化效率做出预测。针对现有的 ecGEMs, Wendering 等^[99]提出了基于蛋白质丰度的周转数校正 (protein-abundance-based correction of turnover numbers, PRESTO) 算法, 通过为酶的催化效率 k_{cat} 添加校正因子, 并且在优化过程中最小化生长速率误差与 k_{cat} 校正量的加权线性组合来实现对模型内 k_{cat} 的全局优化。

针对蛋白质丰度的预测同样受到了广泛关注。基于数据驱动的预测模型能够基于转录组数据对蛋白质丰度做出较为准确的预测, 如 Ferreira 等^[100]于 2021 年报道的 AdaBoost 回归模型; 在酶约束模型中, 可以采用上述数据驱动的催化效率预测值, 直接基于 k_{cat} 值、蛋白质丰度和代谢通量这 3 者之间的定量关系来预测蛋白质丰度; 同样, 基于 ecGEMs 的全局优化算法能够对蛋白质丰度分布做出全局优化, 其中基于蛋白质约束代谢模型的酶丰度预测 (prediction of enzyme abundances using protein-constrained metabolic models, PARROT) 工具基于代谢分配调整最小化的原则在优化过程中最小化不同生长条件下酶浓度分布的变化量^[101], 另一种优化工具作为 OVERLAY 工具的一部分基于 RNA-seq 获得的转录组数据来优化酶浓度分布。

虽然当前的 ecGEMs 扩展了研究人员对酶使用和分配的理解, 并为多组学数据分析和代谢工程策略设计提供了平台和工具, 但在提升模型预测性能方面仍有很大的改进空间。酶约束模型在代谢工程设计的实际应用过程中仍然存在较大缺陷, 需要开发新的算法来进一步整合更多类型实验数据以提升模型质量。

3.3 酶约束模型整合调控网络

细胞代谢活动在多个层面通过各种异质分

子网络的相互作用受到严格调控,很多情况下仅靠 ecGEMs 无法模拟复杂调控的效果。因此如何将包括信号转导网络、转录调控网络(transcriptional regulatory networks, TRNs)和蛋白质分泌途径在内的异质细胞网络与 GEMs 整合来捕捉复杂代谢活动至关重要(图 3)。

其中转录调控网络(TRNs)在对代谢网络的调控中至关重要,因而受到了广泛关注。以酿酒酵母为例,基于 SGD 数据库构建的 TRN 模型包括 186 个转录因子(TF)和 5 727 个靶基因,涉及 28 260 个调控相互作用^[102];另外一个基于单细胞 RNA-seq 测量数据构建的 TRN 模型包含 12 228 个相互作用^[103]。将 TRN 与代谢网络模型整合的方法基本可以分为 2 类:采用布尔规则表征 TF 与其靶基因之间的相互作用,从而可以根据 TRN 或基因表达数据将 GEMs 中的反应打开或关闭^[104];另外一种概率调控(PROM)的方法,则引入了表征基因-TF 相互作用情况的概率,并根据这些概率,在 GEMs 中调整特定反应的最大通量以表达 TF 调控的效果^[37]。

3.4 酶约束模型在生物制造领域的应用

由于模型计算快捷且具有良好的预测性能,ecGEMs 也被广泛用于辅助细胞工厂的工程设计。为了方便基于 ecGEMs 的菌株设计,研究人员开发了多种工具包。例如, Yao 等^[105]开发了一个名为 PROSO 的工具箱,该工具箱可以利用 ecGEMs 进行计算菌株设计。PROSO 包含了如 PC-OptKnock 和最小化蛋白质组调整等多种算法,能够辅助合成生物学中的系统菌株设计。此外, Domenzain 等^[17]开发的 ecFactory 平台将经典的 FSEOF 算法与 ecGEMs 的预测能力相结合,可以轻松进行蛋白质成本的计算和蛋白质使用的变异性分析,以筛选出不合理的基因靶点。更重要的是, ecFactory 能够设计出最小化的基因操作(包括基因过表达、敲除和下

调)以用于实验验证。中国科学院天津工业生物技术研究所开发的 ECMpy2.0 工具^[106]则在建模过程中整合了 TurNuP^[98]等多种参数预测方法,显著提升模型质量的同时整合了包含 FSEOF 算法在内的 3 种菌株设计算法,实现了模型构建和菌株设计流程的一体化。

在具体应用方面,首先, ecGEMs 能够精准捕捉代谢活动的内在特征,常被用于模拟预测工程菌株的代谢表型从而指导代谢工程改造。研究人员将酿酒酵母的乙醇发酵途径替换为 2,3-丁二醇和甘油的共生产途径,并构建该工程菌株(GSY014)的 ecGEMs 进行模拟预测。结果显示模型预测的生长速率下降(0.360/h→0.175/h)与实验结果(0.150/h)高度吻合,且蛋白质组学数据也验证了 ecGEMs 所展现的酶资源重新分配现象,如糖酵解酶的上调^[107]。Wang 等^[108]于 2024 年基于 ECMpy 工具包成功构建了嗜热菌丝体(*Myceliophthora thermophila*)的酶约束模型 ecMTM,准确模拟了嗜热菌丝体对植物生物质水解产物(葡萄糖、木糖、阿拉伯糖等)的层级利用模式。模型显示,葡萄糖优先被快速摄取,而木糖和阿拉伯糖的代谢依赖于葡萄糖的消耗速率,这与实验观察的碳源利用顺序一致;模型还揭示了葡萄糖摄取速率与生物量产量之间的负相关关系,高葡萄糖摄取率会降低酶使用效率,导致生物量积累受限,这一发现为优化碳通量分配提供了理论依据。酿酒酵母的酶约束模型 ecYeast8^[22]则在动态模拟中展现了优秀的预测能力:其精准捕捉了连续培养中酿酒酵母在高稀释率下呼吸-发酵代谢的切换,在分批培养中还原了乙醇动态代谢过程,并通过底物的蛋白成本理论解释了在混合碳源培养中底物的消耗顺序(如葡萄糖先于果糖),与实验数据高度吻合,为工业菌株设计与工艺优化提供了高精度动态预测框架^[109]。

另外, ecGEMs 能够定量预测基因操作(如基因敲除、抑制和增强)对代谢物产量和菌株表型的影响, 从而被广泛应用于代谢改造靶点的筛选。Li 等^[110]使用 ec_iML1515^[111]模型模拟了大肠杆菌正常生长以及莽草酸生产最大化条件下的蛋白质需求, 确定了 7 种需要过表达的高需求蛋白质和 5 种需要下调的低需求蛋白质; 实验验证发现, 这 12 个基因靶标中的 11 个成功增强了菌株在莽草酸生产中的表现, 展示了 ecGEMs 的预测性能。随着与酶相关约束的逐步完善, ecGEMs 已经能够实现多种基因操作组合的模拟预测。例如, Ishchuk 等^[112]借助酵母的 ecGEM 模型(ecYeast8), 成功预测了能够增强血红素生产的 11 种基因靶点组合。基于盐单胞菌 TD01 菌株的基因组数据构建的酶约束代谢模型 Halo-ecGEM 则采用通量变异性分析(flux variability analysis, FVA)结合转录组与蛋白组数据筛选出 8 个对代谢扰动最大的高活性转座酶, 利用 CRISPR-Cas9 基因编辑技术精准敲除这 8 个高活性转座酶基因所构建的菌株表现出显著增强的稳定性, 证明了活跃转座酶是菌株不稳定的关键因素, 同时展现了 ecGEMs 指导的靶向编辑的应用潜力^[113]。

4 全细胞模型

4.1 全细胞模型(whole-cell model, WCM)的发展历程

此前讨论的各种代谢模型已经能够将转录调控和酶分配等细胞生理活动纳入代谢网络, 但仍有许多亚细胞过程难以通过添加约束的方式整合到模型中。针对这一问题, 研究人员提出了全细胞模型(WCM)。WCM 是最复杂的细胞生理代谢模型之一, 其将包括细胞代谢活动等各种亚细胞过程(如蛋白质的激活和折叠、细胞的压力响应和细胞复制周期等)整合在一起,

可以用于在更广泛的基因组水平上模拟细胞的生命活动。

目前, 研究人员已构建了针对支原体(*Mycoplasma genitalium*)^[114]、酵母(*S. cerevisiae*)^[115]和大肠杆菌^[116]的全细胞模型。以酿酒酵母的全细胞模型(WM_S288C)^[115]为例, 其构建方法与支原体(*M. genitalium*)全细胞模型的构建思路类似^[114], 主要基于酵母的 GEMs-iT0977 进行拓展, 将 15 种细胞状态和 26 种细胞过程整合到一个单一的计算模型中。WM_S288C 的子模型描述了多尺度的细胞过程, 包括运输与代谢、DNA 复制与修复、RNA 合成与修饰、蛋白质合成与折叠、细胞分裂^[115]。在使用 WM_S288C 进行模拟时, 这些代表不同细胞过程的子模型被解耦并以 1 s 的时间尺度独立进行模拟, 随后进行整合。通过这种子模型整合的方法, WM_S288C 模拟验证了非必需基因能够调控核苷酸浓度, 从而对细胞生长施加控制的现象。WM_S288C 进一步揭示了酵母基因型与表型之间的复杂关系, 展现了全细胞模型的发展潜力。

4.2 全细胞模型的应用

获取生物的最小基因组对于理解其细胞基础代谢活动以及进化过程极为重要, 这涉及生物体必需基因的筛选鉴别。由于完全依赖实验手段识别所有必需基因十分困难, WCM 往往能够凭借其强大的预测性能在最小基因组的识别与设计发挥独特价值。例如, Rees-Garbutt 等^[117-118]依托支原体全细胞模型开发了 2 种针对支原体最小基因组的设计算法 Minesweeper 和 Guess/Add/Mate Algorithm (GAMA), 重新设计了支原体的最小基因组; Rees-Garbutt 等^[118]通过使用全细胞模型(WCMs)进行设计-模拟-测试循环发现了 10 个低必需性基因, 这些基因相比于已报道的最小基因组(JCVI-Syn3.0^[119])

进一步减少了基因组的大小。

除了对最小基因组的识别与设计,WCM 还被用于对不同类型或不同来源的数据集进行深度整理。由于 WCM 可以在整体的层面上表征细胞表型, Macklin 等^[116]利用 WCM 对不同来源的大规模异质性实验数据集进行了整理和交叉验证,发现了数据中存在不一致性,如实验数据所描述的核糖体和 RNA 聚合酶的总产出不足以支持细胞在倍增时间内完成复制;所测量的代谢参数彼此之间的不完全兼容;细胞周期中缺失关键蛋白质时,细胞依然表现出强大的鲁棒性。这些矛盾现象的发现为实验研究提供了新的突破方向,有助于解释数据间的分歧并揭示新的生物学机制。此外,通过整合大规模的实验数据,模型成功地预测了胞内不同蛋白的半衰期,并且与最新的实验结果^[116]相符。

5 结论与展望

当前的细胞生理代谢模型构建的总体发展思路为通过整合多组学数据构建多尺度的细胞模型,而不仅仅局限于生物中心法则中的某一层面,从而能够有效模拟细胞生理活动的复杂性,例如酶约束模型与转录调控网络的整合。然而,当前的细胞代谢模型构建研究仍存在一些挑战。首先,非模式化物种的高精度 GEMs 的快速构建仍然面临巨大阻碍,当前随着高通量的基因测序技术的发展,研究人员往往能够大规模地获取新菌株的基因组信息^[120],但由于当前自动化建模工具的建模精度不足,无法在系统层次上全面表征非模式物种的代谢多样性分析;其次,当前数据库中可用的高质量酶动力学参数数据十分有限,酶约束模型与动力学模型的建模仍然很大程度依赖于参数的估计与优化;最后,现有模式生物的成熟 GEMs 模型

作为构建多尺度细胞模型的基石仍需要进一步迭代完善,以更完整地覆盖各种代谢子途径。以上困难与挑战都促使研究者不断整合各类组学实验数据以及新的算法工具来开发更高效的代谢网络发掘与重构算法以及酶动力学参数推断方案。

随着新兴技术手段的突破以及科研范式的不断创新,细胞生理代谢建模领域也面临诸多新的挑战与机遇。随着近年来深度学习领域的快速发展,人工智能(artificial intelligence, AI)愈发深入地被用于解决细胞代谢建模所遇到的阻碍。在蛋白质序列与结构分析领域,AlphaFold^[121]已经能够实现蛋白质结构的高质量预测,同时人工智能正持续不断地在蛋白质功能预测、蛋白质同源比对等方面取得进展,如 CLEAN^[122]、PLMSearch^[123]等工具的使用直接提升了细胞建模的便捷性与准确性;而在细胞生理代谢网络模型重构和优化领域,AI在酶动力学参数预测、模型网络优化以及模型分析等领域的尝试都取得了许多成果,展现了广阔的应用前景。此外,AI也为细胞建模领域的研究提出了新的科研范式。与传统的基于约束的数学模型不同,深度学习所具有的学习能力和黑箱模型为研究复杂生命系统的内在机理提出了新的思路。研究人员尝试将深度学习与 GEMs 耦合,提出了如人工智能代谢网络(artificial metabolic network, AMN)混合模型^[124]和 FlowGAT^[125]等新的代谢模型架构,这类模型在深度学习的总体架构下基于 GEMs 包含的机理信息构建模型,使得模型具备代谢网络基本结构的同时具备强大的学习能力,为拓展细胞模型预测能力边界提供了宝贵思路。

此外,随着单细胞测序和高通量分析技术的出现,单细胞建模在生物学和医学中的应用越来越广泛。通过将深度学习的方法应用到单

细胞组学技术产生的多组学、多模态数据的数据挖掘与分析中, 研究人员已经能够捕捉复杂分子机制并揭示生物体内不同细胞类型的生物网络多样性^[126]。如何利用单细胞 RNA 测序 (single-cell RNA sequencing, scRNA-seq) 和单细胞蛋白质组学等数据, 建立单细胞的高精度模型来有效表征单细胞的生理特征, 将是细胞生理模型构建的重要发展方向之一。

作者贡献声明

肖陆驰: 文献调研、初稿写作、绘图; 鲁洪中: 监督指导、论文审阅、稿件润色修改。

作者利益冲突公开声明

作者声明没有任何可能会影响本文所报告工作的已知经济利益或个人关系。

REFERENCES

- [1] FELL DA. Systems properties of metabolic networks[M]//Unifying Themes In Complex Systems, Volume 1. CRC Press, 2018: 163-178.
- [2] STEUER R. Computational approaches to the topology, stability and dynamics of metabolic networks[J]. *Phytochemistry*, 2007, 68(16/17/18): 2139-2151.
- [3] 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, KÖNIG R, SMITH RD, PALSSON BØ. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models[J]. *Molecular Systems Biology*, 2010, 6(1): 390.
- [4] HARCOTBE WR, DELANEY NF, LEIBY N, KLITGORD N, MARX CJ. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum[J]. *PLoS Computational Biology*, 2013, 9(6): e1003091.
- [5] STOLYAR S, van DIEN S, HILLESLAND KL, PINEL N, LIE TJ, LEIGH JA, STAHL DA. Metabolic modeling of a mutualistic microbial community[J]. *Molecular Systems Biology*, 2007, 3(1): 92.
- [6] OBERHARDT MA, YIZHAK K, RUPPIN E. Metabolically re-modeling the drug pipeline[J]. *Current Opinion in Pharmacology*, 2013, 13(5): 778-785.
- [7] RAU MH, ZEIDAN AA. Constraint-based modeling in microbial food biotechnology[J]. *Biochemical Society Transactions*, 2018, 46(2): 249-260.
- [8] LOMAN NJ, PALLEEN MJ. Twenty years of bacterial genome sequencing[J]. *Nature Reviews Microbiology*, 2015, 13(12): 787-794.
- [9] EDWARDS JS, PALSSON BO. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities[J]. *PNAS*, 2000, 97(10): 5528-5533.
- [10] KAUFFMAN KJ, PRAKASH P, EDWARDS JS. Advances in flux balance analysis[J]. *Current Opinion in Biotechnology*, 2003, 14(5): 491-496.
- [11] REED JL, VO TD, SCHILLING CH, PALSSON BO. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR)[J]. *Genome Biology*, 2003, 4(9): R54.
- [12] FEIST AM. iAF1260: A genome-scale metabolic model of *E. coli* as a platform for metabolic engineering[C]//The Annual Meeting and Exhibition 2007, 2007.
- [13] ORTH JD, CONRAD TM, NA J, LERMAN JA, NAM H, FEIST AM, PALSSON BØ. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism: 2011[J]. *Molecular Systems Biology*, 2011, 7(1): 535.
- [14] MONK JM, LLOYD CJ, BRUNK E, MIH N, SASTRY A, KING Z, TAKEUCHI R, NOMURA W, ZHANG Z, MORI H, FEIST AM, PALSSON BO. iML1515, a knowledgebase that computes *Escherichia coli* traits[J]. *Nature Biotechnology*, 2017, 35(10): 904-908.
- [15] JOHNSTON M. Genome sequencing: The complete code for a eukaryotic cell[J]. *Current Biology*, 1996, 6(5): 500-503.
- [16] FÖRSTER J, FAMILI I, FU P, PALSSON BØ, NIELSEN J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network[J]. *Genome Research*, 2003, 13(2): 244-253.
- [17] DOMENZAIN I, LI FR, KERKHOVEN EJ, SIEWERS V. Evaluating accessibility, usability and interoperability of genome-scale metabolic models for diverse yeasts species[J]. *FEMS Yeast Research*, 2021, 21(1): foab002.
- [18] ZHANG C, SÁNCHEZ BJ, LI F, EIDEN CWQ, SCOTT WT, LIEBAL UW, BLANK LM, MENGERS HG, ANTON M, RANGEL AT, MENDOZA SN, ZHANG L, NIELSEN J, LU H, KERKHOVEN EJ. Yeast9: a consensus genome-scale metabolic model for *S. cerevisiae* curated by the community[J]. *Molecular Systems Biology*, 2024, 20(10): 1134-1150.
- [19] AUNG H W, HENRY S A, WALKER L P. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism[J]. *Industrial Biotechnology*, 2013, 9(4): 215-228.
- [20] SCOTT WT Jr, SMID EJ, NOTEBAART RA, BLOCK DE. Curation and analysis of a *Saccharomyces cerevisiae* genome-scale metabolic model for predicting production of sensory impact molecules under enological conditions[J]. *Processes*, 2020, 8(9):

- 1195.
- [21] DIKICIOGLU D, OLIVER SG. Extension of the yeast metabolic model to include iron metabolism and its use to estimate global levels of iron-recruiting enzyme abundance from cofactor requirements[J]. *Biotechnology and Bioengineering*, 2019, 116(3): 610-621.
- [22] LU HZ, LI FR, SÁNCHEZ BJ, ZHU ZM, LI G, DOMENZAIN I, MARCIŠAUSKAS S, ANTON PM, LAPP A D, LIEVEN C, BEBER ME, SONNENSCHN N, KERKHOVEN EJ, NIELSEN J. A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism[J]. *Nature Communications*, 2019, 10: 3586.
- [23] HENRIQUES D, MINEBOIS R, MENDOZA S, MACÍAS L G, PÉREZ-TORRADO R, BARRIO E, TEUSINK B, QUEROL A, BALSACANTO E. A multi-phase multi-objective genome-scale model shows diverse redox balance strategies in yeasts[J]. *bioRxiv*, 2021. <https://www.biorxiv.org/content/10.1101/2021.02.11.430755v1>
- [24] WANG H, MARCIŠAUSKAS S, SÁNCHEZ BJ, DOMENZAIN I, HERMANSSON D, AGREN R, NIELSEN J, KERKHOVEN EJ. RAVEN 2.0: a versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*[J]. *PLoS Computational Biology*, 2018, 14(10): e1006541.
- [25] HENRY CS, DeJONGH M, BEST AA, FRYBARGER PM, LINSAY B, STEVENS RL. High-throughput generation, optimization and analysis of genome-scale metabolic models[J]. *Nature Biotechnology*, 2010, 28(9): 977-982.
- [26] DIAS O, ROCHA M, FERREIRA EC, ROCHA I. Reconstructing genome-scale metabolic models with merlin[J]. *Nucleic Acids Research*, 2015, 43(8): 3899-3910.
- [27] MACHADO D, ANDREJEV S, TRAMONTANO M, PATIL KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities[J]. *Nucleic Acids Research*, 2018, 46(15): 7542-7553.
- [28] ZIMMERMANN J, KALETA C, WASCHINA S. Gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models[J]. *Genome Biology*, 2021, 22(1): 81.
- [29] CASTILLO S, PEDDINTI G, BLOMBERG P, JOUHTEN P. Reconstruction of compartmentalized genome-scale metabolic models using deep learning for over 800 fungi[J]. *bioRxiv*, 2023. DOI:10.1101/2023.08.23.554328.
- [30] CHEN C, LIAO C, LIU YY. Teasing out missing reactions in genome-scale metabolic networks through hypergraph learning[J]. *Nature Communications*, 2023, 14: 2375.
- [31] BECKER SA, PALSSON BO. Context-specific metabolic networks are consistent with experiments[J]. *PLoS Computational Biology*, 2008, 4(5): e1000082.
- [32] SCHMIDT BJ, EBRAHIM A, METZ TO, ADKINS JN, PALSSON BØ, HYDUKE DR. GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data[J]. *Bioinformatics*, 2013, 29(22): 2900-2908.
- [33] BORDBAR A, MO ML, NAKAYASU ES, SCHRIMPE-RUTLEDGE AC, KIM YM, METZ TO, JONES MB, FRANK BC, SMITH RD, PETERSON SN, HYDUKE DR, ADKINS JN, PALSSON BO. Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation[J]. *Molecular Systems Biology*, 2012, 8(1): 558.
- [34] ZUR H, RUPPIN E, SHLOMI T. iMAT: an integrative metabolic analysis tool[J]. *Bioinformatics*, 2010, 26(24): 3140-3142.
- [35] AGREN R, BORDEL S, MARDINOGLU A, PORNPUTTAPONG N, NOOKAEW I, NIELSEN J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT[J]. *PLoS Computational Biology*, 2012, 8(5): e1002518.
- [36] COLIJN C, BRANDES A, ZUCKER J, LUN DS, WEINER B, FARHAT MR, CHENG TY, BRANCH MOODY D, MURRAY M, GALAGAN JE. Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production[J]. *PLoS Computational Biology*, 2009, 5(8): e1000489.
- [37] CHANDRASEKARAN S, PRICE ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107(41): 17845-17850.
- [38] AGREN R, MARDINOGLU A, ASPLUND A, KAMPF C, UHLEN M, NIELSEN J. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling[J]. *Molecular Systems Biology*, 2014, 10(3): 721.
- [39] YIZHAK K, GAUDE E, Le DÉVÉDEC S, WALDMAN YY, STEIN GY, van de WATER B, FREZZA C, RUPPIN E. Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer[J]. *eLife*, 2014, 3: e03641.
- [40] JENSEN PA, PAPIN JA. Functional integration of a metabolic network model and expression data without arbitrary thresholding[J]. *Bioinformatics*, 2011, 27(4): 541-547.
- [41] BLAIS EM, RAWLS KD, DOUGHERTY BV, LI ZI, KOLLING GL, YE P, WALLQVIST A, PAPIN JA. Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions[J]. *Nature Communications*, 2017, 8: 14250.
- [42] JERBY L, SHLOMI T, RUPPIN E. Computational reconstruction of tissue-specific metabolic models:

- application to human liver metabolism[J]. *Molecular Systems Biology*, 2010, 6(1): 401.
- [43] WANG YL, EDDY JA, PRICE ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE[J]. *BMC Systems Biology*, 2012, 6: 153.
- [44] VLASSIS N, PACHECO MP, SAUTER T. Fast reconstruction of compact context-specific metabolic network models[J]. *PLoS Computational Biology*, 2014, 10(1): e1003424.
- [45] VITKIN E, SHLOMI T. MIRAGE: a functional genomics-based approach for metabolic network model reconstruction and its application to cyanobacteria networks[J]. *Genome Biology*, 2012, 13(11): R111.
- [46] SCHULTZ A, QUTUB AA. Reconstruction of tissue-specific metabolic networks using CORDA[J]. *PLoS Computational Biology*, 2016, 12(3): e1004808.
- [47] SCHELLENBERGER J, QUE R, FLEMING RMT, THIELE I, ORTH JD, FEIST AM, ZIELINSKI DC, BORDBAR A, LEWIS NE, RAHMANIAN S, KANG J, HYDUKE DR, PALSSON BØ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0[J]. *Nature Protocols*, 2011, 6(9): 1290-1307.
- [48] EBRAHIM A, LERMAN JA, PALSSON BO, HYDUKE DR. COBRApy: CONstraints-based reconstruction and analysis for Python[J]. *BMC Systems Biology*, 2013, 7: 74.
- [49] DUARTE NC, BECKER SA, JAMSHIDI N, THIELE I, MO ML, VO TD, SRIVAS R, PALSSON BØ. Global reconstruction of the human metabolic network based on genomic and bibliomic data[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(6): 1777-1782.
- [50] HYÖTYLÄINEN T, JERBY L, PETÄJÄ EM, MATTILA I, JÄNTTI S, AUVINEN P, GASTALDELLI A, YKI-JÄRVINEN H, RUPPIN E, OREŠIĆ M. Genome-scale study reveals reduced metabolic adaptability in patients with non-alcoholic fatty liver disease[J]. *Nature Communications*, 2016, 7: 8994.
- [51] XIAO ZY, CONNOR AJ, WORLAND AM, TANG YJ, ZHA RH, KOFFAS M. Silk fibroin production in *Escherichia coli* is limited by a positive feedback loop between metabolic burden and toxicity stress[J]. *Metabolic Engineering*, 2023, 77: 231-241.
- [52] BURGARD AP, PHARKYA P, MARANAS CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization[J]. *Biotechnology and Bioengineering*, 2003, 84(6): 647-657.
- [53] RANGANATHAN S, SUTHERS PF, MARANAS CD. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions[J]. *PLoS Computational Biology*, 2010, 6(4): e1000744.
- [54] FONG SS, BURGARD AP, HERRING CD, KNIGHT EM, BLATTNER FR, MARANAS CD, PALSSON BO. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid[J]. *Biotechnology and Bioengineering*, 2005, 91(5): 643-648.
- [55] YIM H, HASELBECK R, NIU W, PUJOL-BAXLEY C, BURGARD A, BOLDT J, KHANDURINA J, TRAWICK JD, OSTERHOUT RE, STEPHEN R, ESTADILLA J, TEISAN S, SCHREYER HB, ANDRAE S, YANG TH, LEE SY, BURK MJ, van DIEN S. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol[J]. *Nature Chemical Biology*, 2011, 7(7): 445-452.
- [56] XU P, RANGANATHAN S, FOWLER ZL, MARANAS CD, KOFFAS MAG. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA[J]. *Metabolic Engineering*, 2011, 13(5): 578-587.
- [57] RANGANATHAN S, TEE TW, CHOWDHURY A, ZOMORRODI AR, YOON JM, FU YF, SHANKS JV, MARANAS CD. An integrated computational and experimental study for overproducing fatty acids in *Escherichia coli*[J]. *Metabolic Engineering*, 2012, 14(6): 687-704.
- [58] SEGRÈ D, VITKUP D, CHURCH GM. Analysis of optimality in natural and perturbed metabolic networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(23): 15112-15117.
- [59] PATIL KR, ROCHA I, FÖRSTER J, NIELSEN J. Evolutionary programming as a platform for in silico metabolic engineering[J]. *BMC Bioinformatics*, 2005, 6: 308.
- [60] ASADOLLAHI MA, MAURY J, PATIL KR, SCHALK M, CLARK A, NIELSEN J. Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering[J]. *Metabolic Engineering*, 2009, 11(6): 328-334.
- [61] BROCHADO AR, MATOS C, MØLLER BL, HANSEN J, MORTENSEN UH, PATIL KR. Improved vanillin production in baker's yeast through in silico design[J]. *Microbial Cell Factories*, 2010, 9: 84.
- [62] CHATSURACHAI S, FURUSAWA C, SHIMIZU H. An in silico platform for the design of heterologous pathways in nonnative metabolite production[J]. *BMC Bioinformatics*, 2012, 13: 93.
- [63] WEI F, CAI JY, MAO YF, WANG RY, LI HR, MAO ZT, LIAO XP, LI AN, DENG XG, LI FR, YUAN QQ, MA HW. Unveiling metabolic engineering strategies by quantitative heterologous pathway design[J]. *Advanced Science*, 2024, 11(45): 2404632.
- [64] YANG XY, YUAN QQ, ZHENG YY, MA HW, CHEN T, ZHAO XM. An engineered non-oxidative glycolysis pathway for acetone production in *Escherichia coli*[J]. *Biotechnology Letters*, 2016, 38(8): 1359-1365.
- [65] SRIYUDTHSAK K, SHIRAIISHI F, HIRAI MY. Mathematical modeling and dynamic simulation of

- metabolic reaction systems using metabolome time series data[J]. *Frontiers in Molecular Biosciences*, 2016, 3: 15.
- [66] CHONG C, MOHAMAD M, DERIS S, SHAMSIR M, CHOON Y, CHAI L. A review on modelling methods, pathway simulation software and recent development on differential evolution algorithms for metabolic pathways in systems biology[J]. *Current Bioinformatics*, 2014, 9(5): 509-521.
- [67] SMALLBONE K, MALYS N, MESSIHA HL, WISHART JA, SIMEONIDIS E. Building a kinetic model of trehalose biosynthesis in *Saccharomyces cerevisiae*[M]//*Methods in Systems Biology*. Amsterdam: Elsevier, 2011: 355-370.
- [68] SMALLBONE K, MESSIHA HL, CARROLL KM, WINDER CL, MALYS N, DUNN WB, MURABITO E, SWAINSTON N, DADA JO, KHAN F, PIR P, SIMEONIDIS E, SPASIĆ I, WISHART J, WEICHART D, HAYES NW, JAMESON D, BROOMHEAD DS, OLIVER SG, GASKELL SJ, et al. A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes[J]. *FEBS Letters*, 2013, 587(17): 2832-2841.
- [69] SAVOGLIDIS G, Da SILVEIRA dos SANTOS AX, RIEZMAN I, ANGELINO P, RIEZMAN H, HATZIMANIKATIS V. A method for analysis and design of metabolism using metabolomics data and kinetic models: Application on lipidomics using a novel kinetic model of sphingolipid metabolism[J]. *Metabolic Engineering*, 2016, 37: 46-62.
- [70] MOISSET P, VAISMAN D, CINTOLESI A, URRUTIA J, RAPAPORT I, ANDREWS BA, ASENJO JA. Continuous modeling of metabolic networks with gene regulation in yeast and *in vivo* determination of rate parameters[J]. *Biotechnology and Bioengineering*, 2012, 109(9): 2325-2339.
- [71] MISKOVIC L, ALFF-TUOMALA S, SOH KC, BARTH D, SALUSJÄRVI L, PITKÄNEN JP, RUOHONEN L, PENTTILÄ M, HATZIMANIKATIS V. A design-build-test cycle using modeling and experiments reveals interdependencies between upper glycolysis and xylose uptake in recombinant *S. cerevisiae* and improves predictive capabilities of large-scale kinetic models[J]. *Biotechnology for Biofuels*, 2017, 10: 166.
- [72] DOBSON PD, SMALLBONE K, JAMESON D, SIMEONIDIS E, LANTHALER K, PIR P, LU C, SWAINSTON N, DUNN WB, FISHER P, HULL D, BROWN M, OSHOTA O, STANFORD NJ, KELL DB, KING RD, OLIVER SG, STEVENS RD, MENDES P. Further developments towards a genome-scale metabolic model of yeast[J]. *BMC Systems Biology*, 2010, 4: 145.
- [73] CHEN Y, LI FR, MAO JW, CHEN Y, NIELSEN J. Yeast optimizes metal utilization based on metabolic network and enzyme kinetics[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2021, 118(12): e2020154118.
- [74] MIH N, PALSSON BO. Expanding the uses of genome-scale models with protein structures[J]. *Molecular Systems Biology*, 2019, 15(11): e8601.
- [75] SÁNCHEZ R, SALI A. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome[J]. *PNAS*, 1998, 95(23): 13597-13602.
- [76] KHODAYARI A, MARANAS CD. A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains[J]. *Nature Communications*, 2016, 7: 13806.
- [77] VILLAVERDE AF, PATHIRANA D, FRÖHLICH F, HASENAUER J, BANGA JR. A protocol for dynamic model calibration[J]. *Briefings in Bioinformatics*, 2022, 23(1): bbab387.
- [78] CHAKRABARTI A, MISKOVIC L, SOH KC, HATZIMANIKATIS V. Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints[J]. *Biotechnology Journal*, 2013, 8(9): 1043-1057.
- [79] GOPALAKRISHNAN S, DASH S, MARANAS C. K-FIT: an accelerated kinetic parameterization algorithm using steady-state fluxomic data[J]. *Metabolic Engineering*, 2020, 61: 197-205.
- [80] MATOS MRA, SAA PA, COWIE N, VOLKOVA S, de LEEUW M, NIELSEN LK. GRASP: a computational platform for building kinetic models of cellular metabolism[J]. *Bioinformatics Advances*, 2022, 2(1): vbac066.
- [81] CHOUDHURY S, MORET M, SALVY P, WEILANDT D, HATZIMANIKATIS V, MISKOVIC L. Reconstructing kinetic models for dynamical studies of metabolism using generative adversarial networks[J]. *Nature Machine Intelligence*, 2022, 4(8): 710-719.
- [82] CHOUDHURY S, NARAYANAN B, MORET M, HATZIMANIKATIS V, MISKOVIC L. Generative machine learning produces kinetic models that accurately characterize intracellular metabolic states[J]. *Nature Catalysis*, 2024, 7(10): 1086-1098.
- [83] CHOWDHURY A, ZOMORRODI AR, MARANAS CD. K-OptForce: integrating kinetics with flux balance analysis for strain design[J]. *PLoS Computational Biology*, 2014, 10(2): e1003487.
- [84] NARAYANAN B, WEILANDT D, MASID M, MISKOVIC L, HATZIMANIKATIS V. Rational strain design with minimal phenotype perturbation[J]. *Nature Communications*, 2024, 15: 723.
- [85] ADADI R, VOLKMER B, MILO R, HEINEMANN M, SHLOMI T. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters[J]. *PLoS Computational Biology*, 2012, 8(7): e1002575.
- [86] SÁNCHEZ BJ, ZHANG C, NILSSON A, LAHTVEE PJ, KERKHOVEN EJ, NIELSEN J. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints[J].

- Molecular Systems Biology, 2017, 13(8): 935.
- [87] BEG QK, VAZQUEZ A, ERNST J, de MENEZES MA, BAR-JOSEPH Z, BARABÁSI AL, OLTVAI ZN. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity[J]. PNAS, 2007, 104(31): 12663-12668.
- [88] MAO ZT, ZHAO X, YANG X, ZHANG PJ, DU JW, YUAN QQ, MA HW. ECMpy, a simplified workflow for constructing enzymatic constrained metabolic network model[J]. Biomolecules, 2022, 12(1): 65.
- [89] BEKIARIS PS, KLAMT S. Automatic construction of metabolic models with enzyme constraints[J]. BMC Bioinformatics, 2020, 21(1): 19.
- [90] ALTER TB, BLANK LM, EBERT BE. Proteome regulation patterns determine *Escherichia coli* wild-type and mutant phenotypes[J]. mSystems, 2021, 6(2): e00625-20.
- [91] YAO HY, DAHAL S, YANG L. Novel context-specific genome-scale modelling explores the potential of triacylglycerol production by *Chlamydomonas reinhardtii*[J]. Microbial Cell Factories, 2023, 22(1): 13.
- [92] VALGEPEA K, ADAMBERG K, SEIMAN A, VILU R. *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins[J]. Molecular BioSystems, 2013, 9(9): 2344.
- [93] DAVIDI D, NOOR E, LIEBERMEISTER W, BAR-EVEN A, FLAMHOLZ A, TUMMLER K, BARENHOLZ U, GOLDENFELD M, SHLOMI T, MILO R. Global characterization of *in vivo* enzyme catalytic rates and their correspondence to *in vitro* k_{cat} measurements[J]. Proceedings of the National Academy of Sciences of the United States of America, 2016, 113(12): 3401-3406.
- [94] WILKEN SE, BESANÇON M, KRATOCHVÍL M, FOKO KUATE CA, TREFOIS C, GU W, EBENHÖH O. Interrogating the effect of enzyme kinetics on metabolism using differentiable constraint-based models[J]. Metabolic Engineering, 2022, 74: 72-82.
- [95] LI FR, YUAN L, LU HZ, LI G, CHEN Y, ENGQVIST MKM, KERKHOVEN EJ, NIELSEN J. Deep learning-based k_{cat} prediction enables improved enzyme-constrained model reconstruction[J]. Nature Catalysis, 2022, 5(8): 662-672.
- [96] YU H, DENG H, HE J, KEASLING J, LUO X. UniKP: a unified framework for the prediction of enzyme kinetic parameters[J]. Nature Communications, 2023, 14(1): 8211.
- [97] WANG T, XIANG GM, HE SW, SU LY, WANG YG, YAN XF, LU HZ. DeepEnzyme: a robust deep learning model for improved enzyme turnover number prediction by utilizing features of protein 3D-structures[J]. Briefings in Bioinformatics, 2024, 25(5): bbae409.
- [98] KROLL A, ROUSSET Y, HU XP, LIEBRAND NA, LERCHER MJ. Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning[J]. Nature Communications, 2023, 14: 4139.
- [99] WENDERING P, AREND M, RAZAGHI-MOGHADAM Z, NIKOLOSKI Z. Data integration across conditions improves turnover number estimates and metabolic predictions[J]. Nature Communications, 2023, 14: 1485.
- [100] FERREIRA M, VENTORIM R, ALMEIDA E, SILVEIRA S, SILVEIRA W. Protein abundance prediction through machine learning methods[J]. Journal of Molecular Biology, 2021, 433(22): 167267.
- [101] FERREIRA MAM, SILVEIRA WBD, NIKOLOSKI Z. PARROT: Prediction of enzyme abundances using protein-constrained metabolic models[J]. PLoS Computational Biology, 2023, 19(10): e1011549.
- [102] SUN DQ, TIAN L, MA BG. Spatial organization of the transcriptional regulatory network of *Saccharomyces cerevisiae*[J]. FEBS Letters, 2019, 593(8): 876-884.
- [103] JACKSON CA, CASTRO DM, SALDI GA, BONNEAU R, GRESHAM D. Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments[J]. eLife, 2020, 9: e51254.
- [104] HERRGÅRD MJ, LEE BS, PORTNOY V, PALSSON BØ. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*[J]. Genome Research, 2006, 16(5): 627-635.
- [105] YAO HY, YANG L. PROSO Toolbox: a unified protein-constrained genome-scale modelling framework for strain designing and optimization[J]. arXiv Preprint arXiv:2308.14869, 2023.
- [106] MAO ZT, NIU JH, ZHAO JX, HUANG YY, WU K, YUN LY, GUAN JR, YUAN QQ, LIAO XP, WANG ZW, MA HW. ECMpy 2.0: a Python package for automated construction and analysis of enzyme-constrained models[J]. Synthetic and Systems Biotechnology, 2024, 9(3): 494-502.
- [107] SJÖBERG G, REKĚNA A, FORNSTAD M, LAHTVEE PJ, van MARIS AJA. Evaluation of enzyme-constrained genome-scale model through metabolic engineering of anaerobic co-production of 2,3-butanediol and glycerol by *Saccharomyces cerevisiae*[J]. Metabolic Engineering, 2024, 82: 49-59.
- [108] WANG YT, MAO ZT, DONG JC, ZHANG PJ, GAO Q, LIU DF, TIAN CG, MA HW. Construction of an enzyme-constrained metabolic network model for *Myceliophthora thermophila* using machine learning-based k_{cat} data[J]. Microbial Cell Factories, 2024, 23(1): 138.
- [109] MORENO-PAZ S, SCHMITZ J, MARTINS dos SANTOS VAP, SUAREZ-DIEZ M. Enzyme-constrained models predict the dynamics of *Saccharomyces cerevisiae* growth in continuous, batch and fed-batch bioreactors[J]. Microbial Biotechnology, 2022, 15(5): 1434-1445.
- [110] LI ZD, GAO C, YE C, GUO L, LIU J, CHEN XL,

- SONG W, WU J, LIU LM. Systems engineering of *Escherichia coli* for high-level shikimate production[J]. *Metabolic Engineering*, 2023, 75: 1-11.
- [111] YE C, LUO QL, GUO L, GAO C, XU N, ZHANG L, LIU LM, CHEN XL. Improving lysine production through construction of an *Escherichia coli* enzyme-constrained model[J]. *Biotechnology and Bioengineering*, 2020, 117(11): 3533-3544.
- [112] ISHCHUK OP, DOMENZAIN I, SÁNCHEZ BJ, PETRANOVIC D. Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae*[J]. *Proceedings of the National Academy of Sciences*, 2022, 119(30): e2108245119.
- [113] ZHANG LZ, YE JW, LI G, PARK H, LUO H, LIN YN, LI SW, YANG WN, GUAN YY, WU FQ, HUANG WZ, WU Q, SCRUTTON NS, NIELSEN J, CHEN GQ. A long-term growth stable *Halomonas* sp. deleted with multiple transposases guided by its metabolic network model Halo-ccGEM[J]. *Metabolic Engineering*, 2024, 84: 95-108.
- [114] KARR JR, SANGHVI JC, MACKLIN DN, GUTSCHOW MV, JACOBS JM, BOLIVAL B, ASSAD-GARCIA N, GLASS JI, COVERT MW. A whole-cell computational model predicts phenotype from genotype[J]. *Cell*, 2012, 150(2): 389-401.
- [115] YE C, XU N, GAO C, LIU GQ, XU JZ, ZHANG WG, CHEN XL, NIELSEN J, LIU LM. Comprehensive understanding of *Saccharomyces cerevisiae* phenotypes with whole-cell model WM_S288C[J]. *Biotechnology and Bioengineering*, 2020, 117(5): 1562-1574.
- [116] MACKLIN DN, AHN-HORST TA, CHOI H, RUGGERO NA, CARRERA J, MASON JC, SUN G, AGMON E, DeFELICE MM, MAAYAN I, LANE K, SPANGLER RK, GILLIES TE, PAULL ML, AKHTER S, BRAY SR, WEAVER DS, KESELER IM, KARP PD, MORRISON JH, COVERT MW. Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation[J]. *Science*, 2020, 369(6502): eaav3751.
- [117] REES-GARBUTT J, CHALKLEY O, GRIERSON C, MARUCCI L. Minimal genome design algorithms using whole-cell models[J]. *Computational Methods in Synthetic Biology*, 2021: 183-198.
- [118] REES-GARBUTT J, CHALKLEY O, LANDON S, PURCELL O, MARUCCI L, GRIERSON C. Designing minimal genomes using whole-cell models[J]. *Nature communications*, 2020, 11(1): 836.
- [119] HUTCHISON CA, CHUANG RY, NOSKOV VN, ASSAD-GARCIA N, DEERINCK TJ, ELLISMAN MH, GILL J, KANNAN K, KARAS BJ, MA L, PELLETIER JF, QI ZQ, ALEXANDER RICHTER R, STRYCHALSKI EA, SUN LJ, SUZUKI Y, TSVETANOVA B, WISE KS, SMITH HO, et al. Design and synthesis of a minimal bacterial genome[J]. *Science*, 2016, 351(6280): aad6253.
- [120] YAN QL, LI SH, YAN QS, HUO XK, WANG C, WANG XF, SUN Y, ZHAO WY, YU ZL, ZHANG Y, GUO RC, LV QB, HE X, YAO CL, LI ZM, CHEN F, JI QR, ZHANG AQ, JIN H, WANG GY, et al. A genomic compendium of cultivated human gut fungi characterizes the gut mycobiome and its relevance to common diseases[J]. *Cell*, 2024, 187(12): 2969-2989.e24.
- [121] JUMPER J, EVANS R, PRITZEL A, GREEN T, FIGURNOV M, RONNEBERGER O, TUNYASUVUNAKOOL K, BATES R, ŽÍDEK A, POTAPENKO A, BRIDGLAND A, MEYER C, KOHL SAA, BALLARD AJ, COWIE A, ROMERA-PAREDES B, NIKOLOV S, JAIN R, ADLER J, BACK T, et al. Highly accurate protein structure prediction with AlphaFold[J]. *Nature*, 2021, 596(7873): 583-589.
- [122] YU TH, CUI HY, LI JC, LUO YN, JIANG GD, ZHAO HM. Enzyme function prediction using contrastive learning[J]. *Science*, 2023, 379(6639): 1358-1363.
- [123] LIU W, WANG ZY, YOU RH, XIE CH, WEI H, XIONG Y, YANG JY, ZHU SF. Author Correction: PLMSearch: Protein language model powers accurate and fast sequence search for remote homology[J]. *Nature Communications*, 2024, 15: 7766.
- [124] FAURE L, MOLLET B, LIEBERMEISTER W, FAULON JL. A neural-mechanistic hybrid approach improving the predictive power of genome-scale metabolic models[J]. *Nature Communications*, 2023, 14: 4669.
- [125] HASIBI R, MICHOEL T, OYARZÚN DA. Integration of graph neural networks and genome-scale metabolic models for predicting gene essentiality[J]. *NPJ Systems Biology and Applications*, 2024, 10: 24.
- [126] MA AJ, WANG XY, LI JX, WANG CK, XIAO T, LIU YT, CHENG H, WANG JX, LI Y, CHANG YZ, LI JP, WANG DL, JIANG YX, SU L, XIN G, GU SP, LI ZH, LIU BQ, XU D, MA Q. Single-cell biological network inference using a heterogeneous graph transformer[J]. *Nature Communications*, 2023, 14: 964.