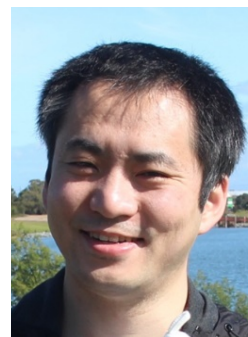


· 智能生物过程控制优化 ·

孙际宾 研究员，中国科学院天津工业生物技术研究所系统生物学中心主任，中国科学院引才计划入选者。长期从事工业微生物系统生物学研究，发展和完善系统生物学理论与研究方法，解析工业微生物的高产抗逆分子基础和代谢瓶颈，设计和创造符合工业化需求的微生物。发表 SCI 论文 150 篇，申请发明专利 110 项，PCT 专利 16 项。获内蒙古自治区科学技术进步一等奖等省部级科技奖项 6 项。担任中国微生物学会常务理事、中国生物发酵产业协会常务理事、中国生物工程学会合成生物学专业委员会副主任委员等。



朱岩 中国科学院天津工业生物技术研究所研究员，中国科学院引才计划入选者。聚焦微生物系统生物学研究，整合多组学与数字细胞方法，“干”“湿”结合，深入探索微生物生长代谢与胁迫耐受的分子机制。发表 SCI 论文 96 篇。担任 *International Journal of Antimicrobial Agents* 等期刊编委。



数据驱动的生物制造过程多组学分析与建模

朱岩^{1,3*}，张志丹^{1,3}，覃培斌¹，申杰¹，孙际宾^{1,2,3*}

1 中国科学院天津工业生物技术研究所系统生物学中心，天津 300308

2 国家合成生物技术创新中心，天津 300308

3 低碳合成工程生物学全国重点实验室，天津 300308

朱岩, 张志丹, 覃培斌, 申杰, 孙际宾. 数据驱动的生物制造过程多组学分析与建模[J]. 生物工程学报, 2025, 41(3): 1152-1178.

ZHU Yan, ZHANG Zhidan, QIN Peibin, SHEN Jie, SUN Jibin. Data-driven multi-omics analyses and modelling for bioprocesses[J]. Chinese Journal of Biotechnology, 2025, 41(3): 1152-1178.

摘要: 生物制造是新质生产力的重要代表，是通过工程细胞或无细胞体系实现物质高效转化的创新生产方式。其生产过程具有时空异质性、复杂性及动态性，给过程系统认知、优化调控带来了显著挑战。本文总结了生物过程多组学数据获取及分析的关键技术，归纳了基于多组学数据的生物过程建模

资助项目：国家重点研发计划(2020YFA0908300)；天津市合成生物技术创新能力提升行动(TSBICIP-CXRC-073, TSBICIP-PTJJ-012)

This work was supported by the National Key Research and Development Program of China (2020YFA0908300) and the Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project (TSBICIP-CXRC-073, TSBICIP-PTJJ-012).

*Corresponding authors. E-mail: ZHU Yan, zhuyan@tib.cas.cn; SUN Jibin, sun_jb@tib.cas.cn

Received: 2025-01-22; Accepted: 2025-03-20

方法,深入探讨了多组学和建模在过程参数调整、发酵控制、环境应激机制解析、营养供给优化及生产实时监测等关键环节的实际应用,详细阐释了多组学数据深度整合及模型构建在提升生物过程操作精度方面的巨大潜力。此外,本文讨论了目前生物过程优化面临的若干挑战,并探索了可能的解决策略。未来通过克服这些挑战,可以更好地理解和控制复杂的生物过程,推动生物制造领域的快速发展。

关键词: 生物过程; 多组学; 数据科学; 数字细胞模型; 整合分析

Data-driven multi-omics analyses and modelling for bioprocesses

ZHU Yan^{1,3*}, ZHANG Zhidan^{1,3}, QIN Peibin¹, SHEN Jie¹, SUN Jibin^{1,2,3*}

1 Systems Biology Center, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

2 National Center of Technology Innovation for Synthetic Biology, Tianjin 300308, China

3 National Key Laboratory of Engineering Biology for Low-carbon Manufacturing, Tianjin 300308, China

Abstract: Biomanufacturing has emerged as a crucial driving force for efficient material conversion through engineered cells or cell-free systems. However, the intrinsic spatiotemporal heterogeneity, complexity, and dynamic characteristics of these processes pose significant challenges to systematic understanding, optimization, and regulation. This review summarizes essential methodologies for multi-omics data acquisition and analyses for bioprocesses and outlines modelling approaches based on multi-omics data. Furthermore, we explore practical applications of multi-omics and modelling in fine-tuning process parameters, improving fermentation control, elucidating stress response mechanisms, optimizing nutrient supplementation, and enabling real-time monitoring and adaptive adjustment. The substantial potential offered by integrating multi-omics with computational modelling for precision bioprocessing is also discussed. Finally, we identify current challenges in bioprocess optimization and propose the possible solutions, the implementation of which will significantly deepen understanding and enhance control of complex bioprocesses, ultimately driving the rapid advancement of biomanufacturing.

Keywords: bioprocess; multi-omics; data science; digital cell model; integrative analysis

生物制造(biomanufacturing)利用生物系统转化原料生产化学品、医药、食品、饲料、生物材料和生物燃料,是保障供应链安全、推动绿色可持续发展的新质生产力。研究表明,到2050年全球生物经济规模预计突破30万亿美元,其中生物基化学品与材料在化学工业中的占比将显著提升^[1]。合成生物学“学习-设计-构建-测试”(learn-design-build-test)方法论广泛应用于工程细胞开发,但其生产效能最终取决于与

之适配的生物过程或称生物工艺(bioprocessing)体系。完整的生物工艺链涵盖原料前处理、规模化发酵、产品分离纯化等关键环节,其中规模化发酵作为核心阶段,直接影响产物产量、质量及后续分离纯化效率。当前生物过程研究面临若干关键挑战:(1) 细胞生理代谢经历从生长期到生产期、从成熟至衰老的动态演变,关键调控的分子机制尚未完全阐明^[2]。(2) 在发酵过程中,工程细胞需应对高温、溶氧限制、极

端 pH、高渗透压、饥饿、氧化压力、剪切力、静水压力以及毒性化合物等多重环境胁迫，其应激响应网络仍有待解析。(3)生物过程中状态变量如光密度(optical density, OD)、溶解氧、pH、温度、氧化还原电位、黏度，以及发酵液中代谢物、分泌蛋白、无机离子组分的动态变化，共同构成了环境组(图 1)。环境组与细胞基因组、转录组、蛋白组、代谢组的相互作用机制尚未被深入揭示。(4) 尽管基于实验设计(design of experiments, DOE)、多参数相关分析的优化理论已取得突破，且人工智能(artificial intelligence, AI)与大数据分析在生物过程中的应用案例显著增多，但实际工业场景中仍面临模型泛化能力受限、精准调控不足、智能控制标准化体系缺失等挑战。(5) 生物过程逐级放大(scale-up)导致细胞的生理代谢状态异质性增强，生产性能降低。揭示异质性原因，并实现小型化(scale-down)动态模拟以优化规模放大，仍是亟待解决的科学问题。总之，尽管生命科学领域的基础研究不断深入，基因编辑、分子生物学等底层技术取得显著进展，但其在生物过程优化与控制中的应用仍面临瓶颈。生命科学聚焦分子机制解析，而过程工程强调系统优化，二者在研究方法(微观相对于宏观)与技术标准(实验室相对于生产车间)上存在显著差异，导致基础研究成果难以高效转化为工业级解决方案。

系统生物学通过整合基因组、转录组、蛋白组、代谢组、互作组学等多维组学技术，定量检测生物过程中细胞内外组分与互作的动态变化，解析生物系统与环境间的动态互动原理，为弥合生命科学基础研究与工业过程工程应用之间的鸿沟提供了重要工具。通过挖掘海量组学数据，构建机理或数据驱动的数学模型，系统生物学能够预测环境波动、基因调控或外源

干扰对细胞功能的影响，实现生命科学与过程科学的有机融合，优化生物过程与控制策略，推动生物制造过程的智能化与自动化，显著提高工业生产效能与稳定性^[2]。

1 生物过程多组学数据获取

生物过程多组学数据获取需克服工程细胞生长代谢动态异质性与反应器环境梯度的双重挑战。采样需设计时空特异性策略；样品制备则需解决代谢物、RNA 等生物分子稳定性及高通量标准化处理等难点，精准捕获细胞转录、翻译、代谢等多维动态过程；数据采集需突破多组学原位同步监测分析技术瓶颈，最终构建生物过程中工程细胞内部运作、工程细胞与环境互作的全息解析体系。

1.1 组学采样

生物过程的组学采样需兼顾细胞代谢动态异质性(如生长、生产不同阶段)与宏观反应器环

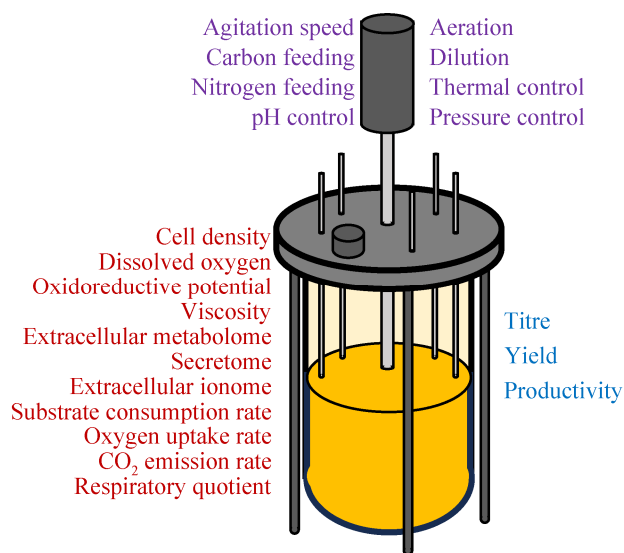


图 1 生物过程中的控制变量(紫色)、状态变量(红色)和生产性能(蓝色)

Figure 1 Control variables (purple), state variables (red) and production performance (blue) in bioprocesses.

境异质性(如传质梯度、热力学分布及流体剪切场差异)的多尺度耦合效应。基于发酵动力学特征,应规划时序采样策略,在代谢表型转换关键节点如指数生长期或产物合成高峰期密集取样,捕捉细胞代谢调控网络的快速变化。另外需要避免在补料、pH调节和搅拌速率变化后立即采样,需等待系统稳定后执行,避免因混合不完全引发的空间异质性采样偏差。传统手动采样存在显著局限性,例如操作复杂度高、采样频率低、染菌风险高,难以满足细胞代谢动态变化的检测需求。近年来,基于物联网与微流控技术的在线自动化采样系统如 Numera[®]、Seg-Flow[®] S3、MAST[®]、bioPROBE 等逐步出现,通过封闭式无菌采样通道与集成式多参数传感器(监测溶氧、温度、底物浓度等),实现了高频次、高时空分辨率的连续监测,显著提升了生物过程动态数据的捕获能力^[3]。

1.2 样品制备

生物过程多组学样品制备技术正朝着高通量、自动化与高保真方向迭代升级,以适配复杂生物系统多维度动态解析需求。基因组学样品制备的核心环节是 DNA 提取与文库构建(图 2),基于磁珠的固相可逆固定技术(solid phase reversible immobilisation, SPRI)实现了微生物 DNA 的高通量纯化。转录组学样品制备包括 RNA 提取、反转录和文库构建。RNA 标签测序(RNAtag sequencing, RNAtag-seq)通过 3'端寡核苷酸标签多重标记实现单次建库混合测序,大幅简化了转录组文库制备流程,提高了处理通量^[4-5]。基因组学和转录组学样品制备的多个环节,如核酸提取^[6]、文库构建^[7]等,也已不同程度地实现了自动化。蛋白质组学样品制备主要包括蛋白质提取、变性、酶解与除盐等多个步骤,面对大批量样品,目前一些整合磁珠或滤膜处理技术(如 SP3、S-Trap)的自动化样品前处

理平台(如 AutoSP3)能够实现稳定的高通量样品前处理需求^[8-11]。胞外代谢组学样品制备相对简单,已实现从样品制备到上机检测的高通量、自动化^[12]。胞内代谢组学样品制备的关键在于快速淬灭代谢活动,以防止代谢物浓度在样品处理过程中发生变化^[13]。基于同位素标记的组学样品制备与代谢组学基本一致,但需开展预实验确定稳态时间。相互作用组学样品制备的核心在于维持分子间相互作用原态,需在采集时确保互作分子存在,可通过交联固定避免 pH 波动及有机溶剂等变性因素干扰。捕获的互作组分(DNA/蛋白质/小分子等)需适配对应组学分析流程进行特异性制备与检测^[14-15]。各组学技术路线的自动化进程有着显著差异,但在高通量与标准化需求驱动下,全流程自动化正加速渗透。结合无人机采样与自动引导车物流衔接,已形成了从样本采集到制备的全链条自动化解决方案。

1.3 数据采集

数据采集是多组学研究的基础环节,涵盖从核酸到代谢物的多维信息获取,其方法因研究目标和技术平台呈现高度多样性。基因组通过测定 DNA 序列解析遗传信息^[16]。重测序可用于发现碱基突变和结构变异^[17]。三代测序技术提供了长读长,适合微生物基因组的高保真组装、结构变异研究^[18]以及 DNA 修饰分析^[19]。三代测序技术的进步,如 PacBio HiFi,达到了 99.8%准确度^[20],打破了传统技术中长读长与高精度不可兼得的局限,在高度重复序列解析和结构变异检测等复杂基因组研究中展现出技术优势。

转录组学用于研究细胞 RNA 的种类、修饰和表达水平,帮助理解基因调控机制^[21]。二代测序技术因能够提供高质量数据且成本较低,常用于基因差异表达分析。三代测序的长读长

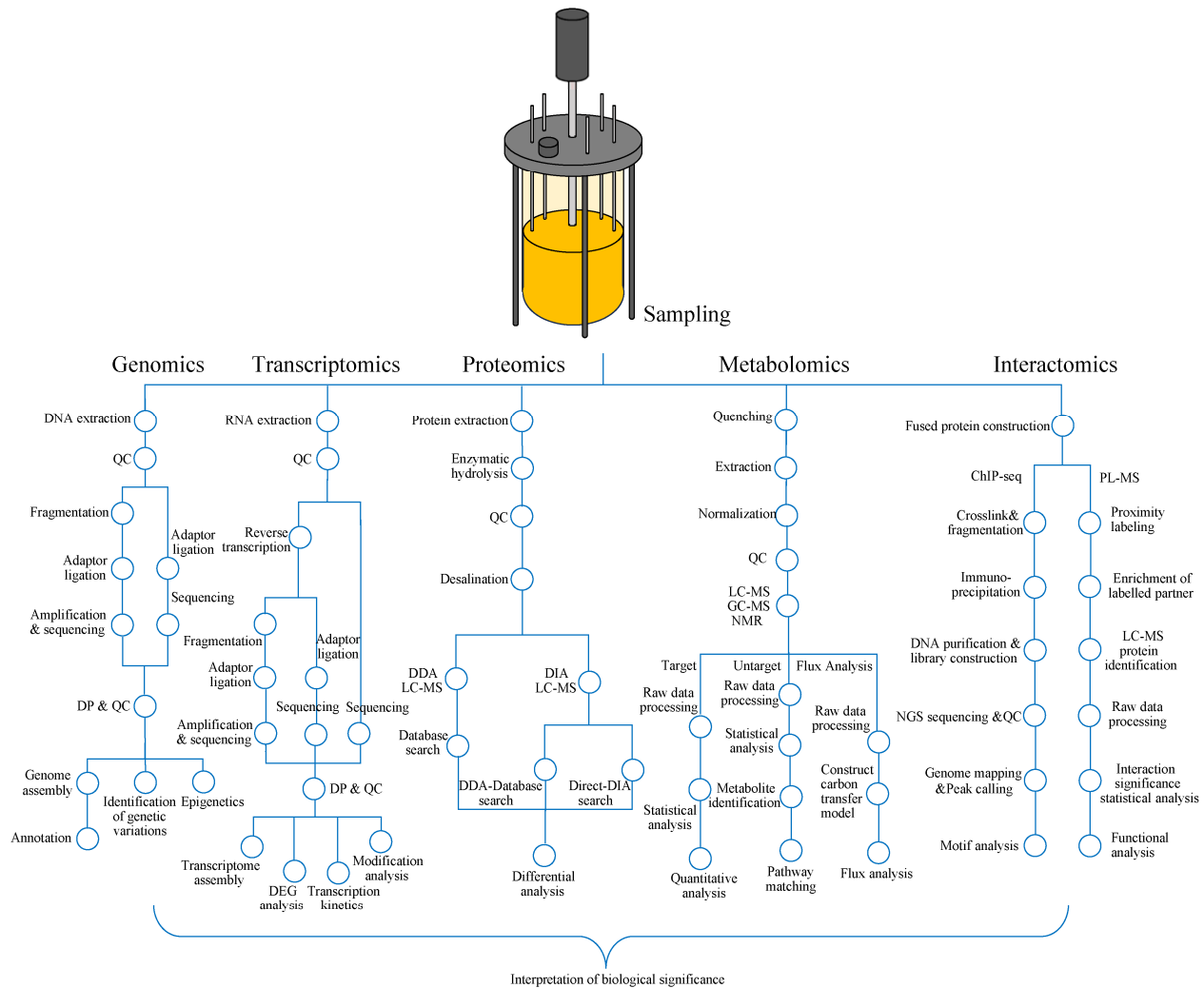


图2 生物过程多组学样品制备、数据采集及分析流程

Figure 2 Sample preparation, data acquisition, and analysis for bioprocess multi-omics. DP: Data processing; QC: Quality control; DEG: Differentially expressed gene; DDA: Data-dependent acquisition; DIA: Data-independent acquisition; LC-MS: Liquid chromatography-mass spectrometry; GC-MS: Gas chromatography-mass spectrometry; NMR: Nuclear magnetic resonance; CHIP-seq: Chromatin immunoprecipitation sequencing; PL-MS: Proximity labelling-mass spectrometry.

技术更擅长识别转录本异构体，为研究复杂转录调控提供了新工具^[22]。此外，牛津纳米孔测序(Oxford nanopore sequencing)技术可完成RNA直接测序，实现对RNA序列和碱基修饰信息的读取^[23-24]。

蛋白质组学数据采集依赖于高分辨质谱(high-resolution mass spectrometry)，经酶解产生

的多肽通过液相色谱分离后，依次进入质谱进行离子化和质量分析。一级质谱通过检测完整肽段的质荷比(mass charge ratio, m/z)确定分子量及丰度信息，二级质谱则通过碰撞能量诱导肽段碎裂，生成碎片离子谱以解析氨基酸序列及翻译后修饰。数据采集模式包括数据依赖性采集(data-dependent acquisition, DDA)和数据非

依赖性采集(data-independent acquisition, DIA)。DDA 优先选择高峰肽段碎裂,适用于小规模样本的定性分析;DIA 将全扫描范围划分为连续窗口,无差别采集所有肽段的碎裂信息,结合谱图库可实现复杂样本的高精度定量^[25]。随着技术的不断进步^[26-28],DIA 模式已成为蛋白质组分析的首选^[29],复杂样品的快速分析可压缩至数分钟内完成^[30]。

代谢组学数据采集的核心技术选择遵循非靶向覆盖、靶向高精度原则。高分辨质谱作为非靶向代谢组学的主流技术,依托 Orbitrap、Q-TOF 等平台,通过 DDA 或 DIA 模式实现全扫描范围内代谢物的高灵敏度检测^[31]。靶向代谢组学则以三重四极杆质谱(triple quadrupole mass spectrometry, QQQ)为主,构建多反应监测(multiple reaction monitoring, MRM)离子对库,通过优化碰撞能参数与保留时间窗口,实现目标代谢物的超灵敏定量^[31]。核磁共振波谱(nuclear magnetic resonance, NMR)技术则以非破坏性检测和结构解析能力见长^[32]。代谢物的化学稳定性受温度等环境因素影响显著,样本采集处理后需尽快上机分析。进样量可通过总蛋白定量或内源参比代谢物校准实现生物学归一化。数据采集序列设计需遵循间隔插入质控(quality control, QC)样本策略,用于评估仪器稳定性并辅助后期数据处理^[33]。同位素标记通量组学的数据采集需权衡方法学特性,靶向采集需优化同位素标记后的目标代谢物离子对;非靶向采集无需预设目标,但需通过 QC 样本校正基质效应与同位素峰簇干扰。

相互作用组学研究蛋白-DNA 相互作用(protein-DNA interaction, PDI)、蛋白-蛋白相互作用(protein-protein interaction, PPI)、蛋白-代谢物相互作用(protein-metabolite interaction, PMI)等细胞组分相互作用关系。PDI 技术主要包括

基于免疫沉淀的染色质免疫沉淀测序技术(chromatin immunoprecipitation sequencing, ChIP-seq)、核酸酶切割与核酸释放技术(cleavage under targets and release using nuclease, CUT-RUN)/靶向切割与转座酶标记技术(cleavage under targets and tagmentation, CUT-Tag)及依赖甲基化酶或脱氨酶的结合位点标记技术^[34-36]等,使用特异性酶或条件切割、标记或释放目标 DNA 片段并进行文库构建与测序分析,鉴定蛋白与 DNA 作用关系。PPI 分为靶向(局部)PPI 和非靶向(全局)PPI: 靶向技术如亲和纯化-质谱联用技术(affinity purification-mass spectrometry, AP-MS)和邻近标记-质谱联用技术(proximity labelling-mass spectrometry, PL-MS),需通过标签融合或工具富集互作蛋白;非靶向技术则基于化学交联质谱,通过交联多肽酶解后质谱分析全局互作网络。PMI 检测方法包括基于蛋白质稳定性变化特征检测的热蛋白质组分析^[37]、溶剂诱导蛋白沉淀^[38]、蛋白质氧化速率的稳定性分析、基于空间构象变化的靶点响应可及性分析^[39]、限制性酶解-质谱分析技术^[40]、肽中心局部稳定性探测技术^[41]等,通过检测蛋白质结合小分子后的稳定性或酶解/标记干扰效应解析互作信息。

组学实验中,生物学重复与技术重复的合理设置是保障数据可靠性的核心要素。生物学重复使用同一菌种,检测生物过程操作过程中的波动情况。技术重复则针对同一样本进行多次测量,用于评估实验操作中移液偏差、仪器波动等系统误差。重复数目的设置需综合考虑统计效力及技术稳定性需求。

2 组学数据分析

生物过程的组学分析首先围绕单组学数据展开,包括数据质量控制、数据清洗、去噪、标准化归一化,以及批次效应校正;随后利用

差异分析、类别及通路富集分析、功能注释、网络构建、关键节点识别等方法,挖掘重要生物分子及调控机制。在单组学分析基础上,多组学整合分析则通过因子分析、协同表达分析、机器学习建模等手段,实现跨组学数据的系统性关联与机制深度解析,精准识别影响性能的关键因子,为工程细胞设计改造及过程优化提供可靠依据。

2.1 单组学数据分析方法

不同组学的数据分析方法因研究对象及数据类型而有所差异,但其分析思路相近,通常包括数据预处理、差异分析、功能富集、网络构建,以及关键代谢途径或调控机制分析等核心步骤。以基因组数据为例,首先使用 FastQC 等工具^[42]去除测序接头和低质量 reads,选用有参组装、从头组装等合适的组装策略,重构基因组^[43],常用工具包括 Canu^[44]、Falcon^[45]、flye^[46]和 SOAPdenovo2^[47]等;随后使用 Prokka^[48]、RAST^[49]、MAKER^[50]、PGAP^[51]、Glimmer^[52]、AUGUSTUS^[53]和 GeneMark^[54]等工具对组装好的基因组序列进行结构和功能注释;变异分析通过 GATK^[55]、Snippy 等工具比对测序数据与参考基因组,检测单核苷酸多态性(single nucleotide polymorphism, SNP)以及插入或缺失(insertion or deletion, InDel),帮助明确菌株遗传变异并指导菌株改造策略。此外,比较基因组学方法通过多基因组比对分析,揭示菌株间的进化关系和特定功能基因,进一步支持工业应用中的菌株筛选与优化。例如,基因组学分析揭示了工业乙醇发酵中微生物群落动态与产量的关联机制。De Oliveira Lino 等^[56]对巴西 2 家甘蔗乙醇精炼厂进行跨季节采样,利用鸟枪法宏基因组测序,通过 Canu 组装获得高质量基因组,并利用 Prokka 注释基因功能,发现脂多糖合成和磷酸转移酶系统等通路与发

酵性能显著相关;变异分析揭示了关键菌株[如发酵乳杆菌(*Limosilactobacillus fermentum*)]的 SNP/InDel 突变对代谢表型有显著影响;进一步分离 *L. fermentum* 菌株并开展静态共培养实验,发现特定菌株因代谢差异(如乳酸/乙酸分泌比例改变)导致酵母生长抑制,证实了基因型-表型关联;最终提出基于温度调控和菌株特异性干预的发酵优化策略,为工业微生物群落管理提供了基因组驱动的解决方案。

转录组测序技术(RNA sequencing, RNA-seq)是当前广泛使用的转录组学技术,能够通过测定基因表达水平揭示生物过程中菌株的基因表达调控机制,并提供菌株改造的潜在靶点。其基本分析流程包括数据质量控制、去除接头序列与低质量读段、数据标准化及归一化,以确保数据可靠性。针对有参考基因组的数据,通过 BWA^[57]、STAR^[58]、HISAT2^[59]、Bowtie2^[60]等工具将 reads 映射到参考基因组,利用 FeatureCounts^[61]、HTSeq^[62]、Salmon^[63]、Kallisto^[64]等工具获得表达定量信息,单位通常采用 RPKM (reads per kilobase of transcript per million mapped reads)、FPKM (fragments per kilobase of transcript per million mapped reads)或 TPM (transcripts per million)^[65]。随后,利用 DESeq2^[66]、edgeR^[67]和 limma-voom^[68]等工具开展差异表达分析,以识别不同条件下显著变化的基因。进一步构建共表达网络(如 weighted gene co-expression network analysis, WGCNA),系统挖掘协同作用的基因模块,推断其功能关联。此外,利用 KEGG (Kyoto encyclopedia of genes and genomes)和 GO (gene ontology)数据库开展通路富集分析,帮助识别关键代谢途径或调控通路,为菌株的代谢工程改造提供潜在靶点。在近期一项构建大肠杆菌工程细胞以生产游离脂肪酸(free fatty acids,

FFAs)的研究中,研究者使用转录组学和代谢组学分析比较 FFA 高产菌株与对照菌株,发现 26 个与 FFAs 代谢不直接相关但能提升产量的基因,如 *ihfA*、*aidB*、*ryfA* 和 *gadA*,通过组合调整这 4 个基因创制的重组菌株在分批发酵中实现了 30.0 g/L 的 FFAs 产量^[69]。

蛋白质组学通过解析生物过程中细胞蛋白质动态调控规律,为发酵工艺优化提供依据。其分析思路通常包含数据预处理、肽段序列鉴定、蛋白定量以及差异表达蛋白分析与功能解析等核心步骤。在数据预处理阶段,需对原始质谱数据进行噪声滤除、基线校正和碎片离子筛选。针对蛋白质组 DDA 数据,采用 MaxQuant 或 Mascot 精准比对质谱二级谱图与理论谱图实现肽段鉴定,进一步筛选转换完成蛋白质组学结果的定量解析^[70]。对于 DIA 数据,可使用基于预建 DDA 谱图库的比对方法或 direct-DIA 数据库直比法,后者因其高效性和便利性占据主导地位^[71]。差异蛋白质统计分析后,结合 GO 和 KEGG 等数据库进行功能富集分析,明确差异蛋白的生物学功能;将蛋白质组学数据映射到 STRING、KEGG 等已知的生物通路和网络中,识别关键的调控节点与功能模块。最近一项研究通过多组学分析揭示了酿酒酵母蛋白组随生长速率的线性分配规律,发现核糖体蛋白与生长速率呈正相关关系,糖酵解蛋白与分子伴侣蛋白呈现显著负相关,其活性调控通过磷酸化修饰实现;结合代谢建模发现酶饱和度驱动氨基酸合成通量提升,为工业菌种代谢工程提供靶点,并通过优化发酵条件,最终实现异丁醇等产物的高效合成,降低能耗与成本^[72]。

代谢组学数据分析是揭示发酵过程中代谢调控机制的重要方法,其分析流程主要包括数据质量控制、峰提取、归一化、去噪、对齐等

数据预处理、代谢物鉴定、统计分析及生物学意义的挖掘。靶向代谢组学数据分析较为简单,可直接进行目标代谢物的定量分析。非靶向代谢组学较为复杂^[73],通常采用单变量(*t* 检验、方差分析、相关分析等)和多变量分析(主成分分析、偏最小二乘法判别分析、聚类分析、自组织映射、随机森林和支持向量机等)筛选差异代谢物。代谢物鉴定通常利用标准品比对与数据库检索,如 Metlin、Massbank、Chemspider 以及一些专属数据库如 HMDB (Human Metabolome Database)、ECMDB (*Escherichia coli* Metabolome Database)、MMCD (Madison Metabolomics Consortium Database)等。此外,新兴的辅助谱图解析工具如 MetDNA^[74]和 MS2Query^[75]显著提高了代谢物鉴定效率。在功能分析方面,利用 MetaboAnalyst、KEGG 等工具或数据库,进行代谢物功能注释与通路映射分析,以明确差异代谢物所涉及的具体代谢路径和潜在调控机制,进一步指导菌株的代谢工程改造。白仲虎团队^[76]采用非靶向代谢组学策略探究了谷氨酸棒杆菌在分泌表达外源蛋白时的代谢差异;首先使用高分辨质谱仪分别采集了正负离子模式下的无偏向代谢物信息,然后采用正交偏最小二乘判别分析模型对分泌表达外源蛋白组和对照组样本进行区分,筛选出具有显著差异的 176 个代谢物,鉴定出关键代谢物为 D-甘露糖-6-磷酸、叶酸、肌苷、L-色氨酸、L-苯丙氨酸等;继而采用 MetaboAnalyst 进行 KEGG 通路分析,发现分泌表达外源蛋白会导致中心代谢加剧和多种氨基酸特别是芳香族氨基酸代谢的变化,以满足蛋白合成和分泌对能量的需求,同时保护胞内蛋白以维持菌体正常生长。

基于 ¹³C 同位素标记的通量组学分析是精确解析生物发酵过程代谢流分布的关键方法。在实验设计阶段,需要明确代谢稳态与同位素

稳态条件。数据处理阶段涉及质谱原始数据预处理、同位素标记峰的识别与定量,随后采用 INCA^[77]等工具进行同位素标记模式分析及代谢通量计算。基于获得的代谢通量数据,进一步整合生物过程与代谢网络模型,系统解析菌株代谢网络的关键节点与瓶颈反应,从而为菌株的代谢工程改造和发酵过程优化提供明确的指导。研究人员通过 ¹³C 标记通量分析,分析了解脂耶式酵母(*Yarrowia lipolytica*)在脂质生物燃料生产中的代谢通量;他们使用 ¹³C 标记葡萄糖追踪碳流,结合气相色谱-质谱联用仪(gas chromatography-mass spectrometry, GC-MS),使用 INCA 工具计算了脂肪酸合成途径和三羧酸循环的通量分布;结果显示,乙酰辅酶 A 的供应是限制脂质产量的关键瓶颈;基于此,研究人员通过过表达脂肪酸合成酶基因,优化菌株,使脂质产量提升了 40%^[78]。

由于相互作用组学数据因策略不同而且多样,这里以基于质谱检测的 PPI 互作组学数据为例对分析方法进行概述。AP-MS 数据分析时,由于不同的诱饵蛋白会共纯化出大量的猎物蛋白,其中包括相当一部分的非特异性结合蛋白,因此,为了得到真正特异性结合的猎物蛋白,往往需要结合显著性分析软件对所有诱饵蛋白的质谱数据进行统计性分析,包括 SAINT、CompPASS 等软件^[79-80]。不同的软件算法有所不同,但核心都会考虑猎物蛋白的丰度、在对照中出现的频率以及在重复实验中的一致性。首先需要将不同诱饵蛋白所得样品的质谱相对定量数据整理为列表,其中行为每个猎物蛋白在不同实验即不同诱饵蛋白、对照及重复实验中的相对定量强度,列为不同实验样品,然后将该文件输入到上述显著性分析软件进行统计分析和打分,根据分值分布、错误发现率(false discovery rate, FDR)及阳性得分

等选择合适的阈值,从而筛选出较高可信度的相互作用蛋白。对于最具全局性 PPI 检测的交联质谱技术,交联试剂的选择与针对交联多肽复杂谱图的分析软件至关重要,交联剂的选择考虑因素有交联反应基团、交联臂长、能否质谱断裂、能否穿透细胞膜、是否含富集功能团等,选择具有富集功能的交联剂如 PhoX^[81]、tBu-PhoX^[82]、DSBSO^[83]等可以捕捉到低丰度蛋白的 PPI 信息。交联质谱分析软件应用较多的有 pLink^[84]、XlinkX^[85]、MeroX^[86]等,通过设置交联试剂类型、断裂规则、特征离子、酶解规则等信息,与蛋白质标准序列数据库进行比对,识别鉴定出交联多肽以及交联位点信息,经 FDR 过滤后得到蛋白-蛋白相互作用结果,PPI 的结果图可以采用 xiView^[87]、Cytoscape^[88]等工具来展示。

2.2 多组学数据整合分析

多组学数据整合主要包括基于统计的、基于生物学知识的、基于约束模型的和基于 AI 的方法。统计驱动的方法通过算法挖掘跨组学内在联系:相似性/核方法(similarity/kernel-method)通过计算样本间的相似性分数构建多组学相似性网络,具有处理缺失数据的优势,运行时间主要取决于样本数量而非组学特征数,且较少依赖特征选择,适用于高通量数据集;多核学习(multi-kernel learning)通过核函数将多组学数据映射到更高维的特征空间并计算相似性矩阵,进而组合各数据视图的核函数生成整合空间,用于模式分析和聚类,使得此方法适用于高维数据的子群识别;因子分析(factor analysis)通过矩阵分解降维多组学数据,提取潜在生物因子以实现样本聚类与模式识别,MOFA (multi-omics factor analysis)等工具结合变分推理,在低维空间中保留生物学解释因子,利用稀疏正则化抑制噪声与缺失值干扰,

显著提升分子亚型识别效率；基于相关性/协方差的联合降维方法(correlation/covariance-based joint dimension reduction)通过最大化转化空间中的相关性或协方差实现多组学数据整合，并用于模式分析和样本聚类。

基于生物学知识的组学整合方法，根据数据库知识库，构建多层网络模型，如基因-蛋白质-代谢物网络，再将多组学数据映射到网络模型上，或基于知识鉴别关键元件，或使用基于图的算法或聚类方法来揭示复杂的多组学数据集中的新模式^[89]。约束优化模型引入多组学数据，可计算特定条件下的代谢流分布(详见 3.1)。

人工智能驱动的技术则通过机器学习与深度学习(deep learning, DL)挖掘非线性模式：监督学习[如支持向量机(supportive vector machine)、随机森林(random forest)]用于分类预测，非监督学习[如聚类分析(K-means, hierarchical clustering)、关联规则挖掘]揭示隐含结构，深度学习[如卷积神经网络(convolutional neural network, CNN)、图神经网络(graph neural network, GNN)]处理高维稀疏数据并捕捉跨组学交互效应。当前技术演进正朝着智能化、动态化与可解释化方向发展，深度学习与因果推断结合增强跨组学因果关系解析，而基于约束的生成模型则为虚拟生物系统模拟开辟新路径，推动从数据整合到知识生成的范式跨越。

Culley 等^[90]提出了一种结合代谢模型与多组学机器学习的方法，用于预测酵母细胞生长；研究者首先构建了 1 143 种单基因敲除酵母菌株的特异代谢模型，利用基因表达数据调整代谢网络反应边界，并通过简约代谢平衡分析(parsimony flux balance analysis, pFBA)计算代谢通量分布；随后，采用多种机器学习方法比较单组学与多组学数据的预测性能；结果显示，整合了转录组与代谢通量数据的混合模型，如

多模态神经网络(multimodal artificial neural network, MMANN)显著优于单一组学模型，其中位数绝对误差低至 0.043，Pearson 相关系数高达 0.906；进一步验证表明，该模型在独立测试集(含单基因和双基因敲除菌株)中仍保持高泛化能力，揭示了代谢通量数据对非代谢基因(如 DNA 修复相关基因)的间接调控作用；并发现核心预测特征主要涉及蛋白质合成、RNA 加工及脂质代谢通路，为理解细胞生长的分子机制提供了新视角。此研究不仅推动了多模态机器学习在生物系统中的应用，还为工程酵母优化及疾病模型开发提供了数据驱动的工具框架。Roy 等^[91]针对合成生物学中实验数据量大但预测能力不足的挑战，提出了一套集成工具链与模拟方法的解决方案，开发了 Omics Mock Generator (OMG)合成数据生成器模拟多组学时间序列数据，构建了包含基因组规模代谢模型、自动推荐工具(automated recommendation tool, ART)和数据存储平台(inventory of composable elements/experiment data depot, ICE/EDD)的完整 workflow；基于关键代谢反应通量预测异戊烯醇产量，最终推荐出工程改造策略，使模拟产量提升 23%；实验验证表明，ART 的预测结果与 OMG 模拟数据高度吻合(误差<5%)，证明了该方法在合成生物学中的实用价值；该研究不仅为多模态数据整合与机器学习应用提供了标准化流程，还为工程酵母优化及生物燃料开发奠定了数据驱动的工具基础。

3 生物过程基于多组学数据的建模

生物过程涉及生物系统与其所在环境复杂且动态的相互作用，这些过程涵盖了从分子、细胞到群体层次的广泛联系。为深入理解这些复杂过程并准确预测生物系统的行为，基于组

学数据建立精确的数理模型至关重要。这些模型不仅能揭示生物过程内在调控的分子机制,还为生物过程的优化控制与放大提供了理论基础和决策依据。

3.1 生物过程模型

生物过程模型的常用架构包括约束计量模型、动力学模型、多尺度模型和 AI 模型。细胞代谢网络包含代谢物浓度和反应速率这 2 类变量,微分方程组表示为 $dx/dt=S \cdot v$, 其中 x 、 S 和 v 分别为代谢物浓度、化学计量系数与代谢反应速率。细胞处于代谢稳态时,胞内每一个代谢物的生成速率等于消耗速率,浓度不变,即 $S \cdot v=0$ 。引入反应方向、可逆性、酶量等反应约束,使用基于线性规划的 FBA^[92]方法计算代谢流分布,就得到了基因组尺度代谢模型(genome-scale metabolic model, GSMM or M model)。进一步扩展引入转录、翻译、大分子组装等其他模块,并引入更多约束,就得到了基因组尺度代谢与表达模型[genome-scale model of metabolism and gene expression (or macromolecular expression), ME model]^[93]。除此以外,还有 ETFL (expression and thermodynamics flux model)、RBA (resource balance analysis)等多种类似 ME 模型的架构。动态流平衡分析(dynamic flux balance analysis, dFBA)通过向 M 模型和 ME 模型中引入动力学模块,近似计算生物过程中的胞内外动态代谢过程^[94](图 3A)。

动力学模型种类和规模更为多样(图 3B)。简单的动力学模型如 Monod 模型、logistic 模型、Haldane 模型、Gompertz 模型等主要包括生物量、底物浓度、产物浓度等状态变量和比生长速率、底物利用速率等速率变量,描述生长速率、底物消耗及产物生成的时变规律,变量少、预测能力有限且无法与组学数据结合^[95]。复杂的动力学模型使用上述的微分方程组,根

据对反应机理的近似程度可分为规范模型(canonical model)、近似模型(approximate model)和机制模型(mechanistic model)。其中规范模型一般仅在特定范围(operating point)有效,例如基于幂律的 Mass action 模型或 S-系统模型^[96],基于线性泰勒展开的 Log-lin 模型和 Lin-log 模型等^[97]。Mass action 或 S-系统使用参照标准化,等式两边 log 转换,即可将代谢流的相对变化写为酶量变化以及底物/效应物相对变化的加和^[96]。Log-lin 模型和 Lin-log 模型的原理也类似,但在参考点泰勒一阶展开^[97]。这些模型对参考点的选择依赖性很高,实际应用相对不多。机制模型遵循物质平衡和热力学原则,对于可逆反应 $S \rightleftharpoons P$, 反应速率 $v=E_{\text{total}} \cdot (k_{\text{cat}+}x_S/K_S - k_{\text{cat}-}x_P/K_P)/(1+x_S/K_S+x_P/K_P)$, 其中 E_{total} 为酶量, x_S 与 x_P 分别为底物、产物的浓度, $k_{\text{cat}+}$ 和 $k_{\text{cat}-}$ 分别为 ES 复合物形成与分解的正向反应速率常数, K_S 与 K_P 分别为底物、产物的解离常数。酶活性存在别构效应(allosteric effect), MWC 模型(Monod-Wyman-Changeux model)和 KNF 模型(Koshland-Nemethy-Filmer)都基于一个假设,即酶存在松弛态和紧张态之间的平衡,但状态之间的转换处理方式不同。MWC 模型强调协同对称,KNF 模型突出顺序结合,这 2 种模型都强调了构象变化在变构调节终态结构中的作用^[97]。近似模型是在机制模型基础上进一步简化。基于拟稳态和快平衡假设,Michaelis-Menten 模型简化单底物酶动力学,得到 $v=(v_{\text{max}} \cdot x_S)/(K_m+x_S)$ ^[98]。使用 Hill 方程近似描述单底物和多底物反应酶修饰和别构效应^[98]。基因组规模的动力学建模架构包括 SKM (structured kinetic modelling)^[96]、ORACLE (optimisation and risk analysis of complex living entities)^[99]、MASS framework^[100]等。

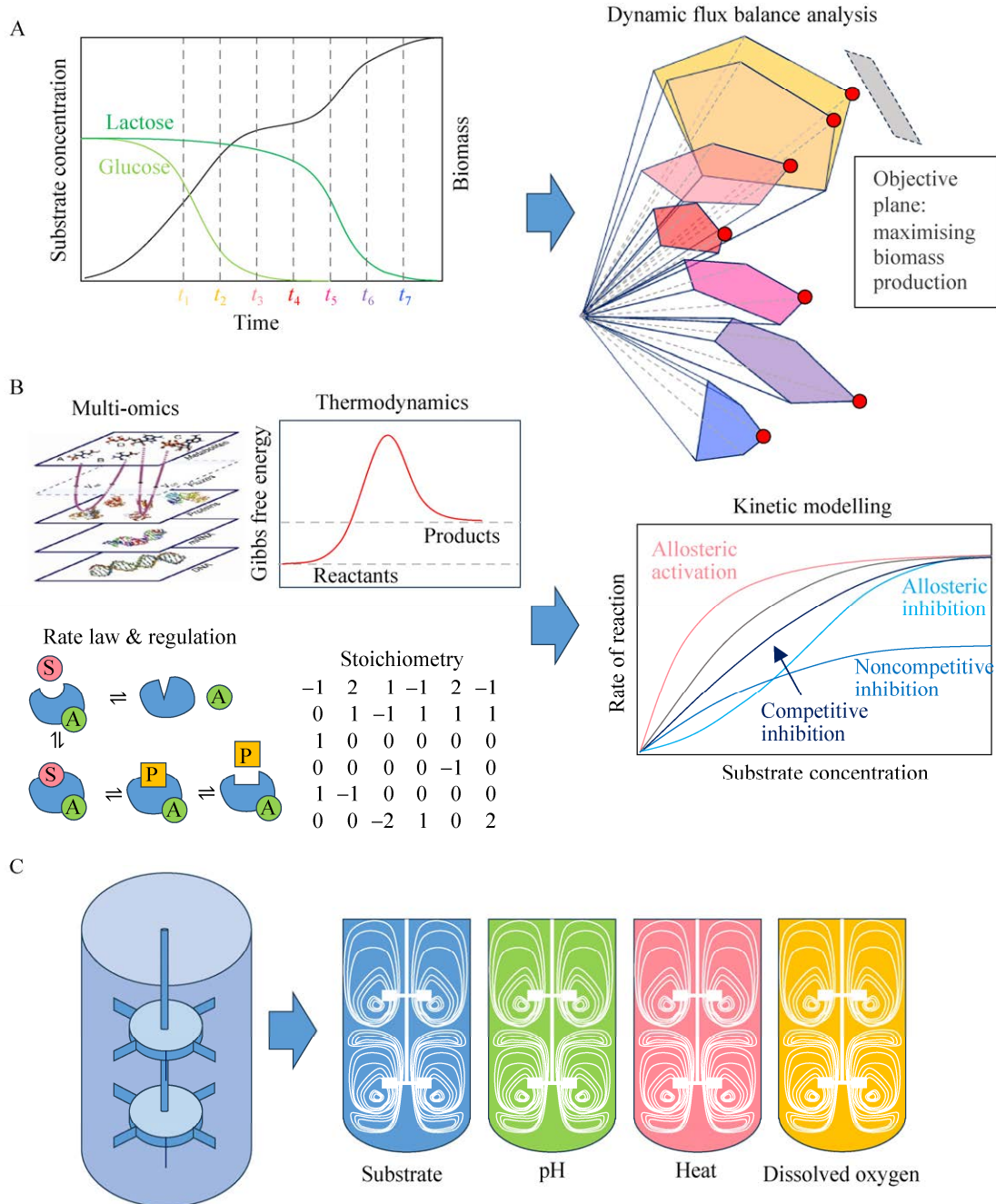


图3 生物过程机理模型^[101] A: 使用动态通量平衡分析计算代谢物抑制效应; B: 整合多组学、热力学、酶动力学、化学计量信息构建动力学模型; C: 结合流场动力学模拟生物过程中底物、pH、传热、溶氧等条件的异质性。

Figure 3 Mechanistic model of bioprocess^[101]. A: Conduct dynamic residue balance analysis based on the constrained model to simulate the metabolite repression effect in the bioprocess of co-utilisation of glucose and lactose; B: Integrate multi-omics, thermodynamics, reaction kinetics, and stoichiometry to construct a metabolic kinetic model; C: Model the bioprocess based on computational fluid dynamics, taking into account the spatial gradients of substrate concentration, pH, temperature, and dissolved oxygen in the reactor.

生物过程从实验室小试、中试、示范到规模化生产逐级放大。大规模生物过程通常在流场不均匀的生物反应器中进行,反应器内存在温度、底物浓度和溶氧梯度,这些因素对工程细胞的生理代谢产生显著影响,且反应器体积越大,不均一程度越高,细胞微环境波动增大,从而加剧代谢异质性。计算流体动力学(computational fluidic dynamics, CFD)使用 Euler-Lagrangian 方法计算模拟反应器内流场变化,结合动态区室模型(dynamic compartment model)预测营养梯度和细胞密度分布(图 3C)。集成单细胞代谢动力学(如 GSMM 或动力学模型)、生物过程动力学和计算流体动力学就形成多尺度模型研究方法,可深入揭示过程放大时反应器内传质传热导致细胞代谢异质性的分子机制。

大数据和 AI 技术快速发展颠覆了生物过程研究范式。深度学习、迁移学习(transfer learning)、强化学习(reinforcement learning)、集成学习(ensemble learning)和生成对抗网络(generative adversarial networks)等算法正在重塑数据处理和分析方式,特别是在处理和整合生物过程在线测量数据与多组学数据时, AI 技术展现出了巨大的潜力。这些技术不仅帮助构建复杂的系统生物学模型以模拟和理解生物过程,还催生了如过程分析技术(process analytical technology, PAT)、数字孪生(digital twins)等过程优化控制技术。机器学习也能够与 GSMM 机制模型整合提升预测分析能力,例如人工代谢网络(artificial metabolic network, AMN)方法整合 FBA 与机器学习,实现了根据胞外营养物浓度精确预测代谢流,并能够进行机制解析^[102](图 4)。AI 应用于生物过程研究中仍面临数据异质性、模型解释性等挑战。随着计算能力的提升和海量组学数据的积累以及高级算法的涌现,

未来有望进一步解析生物过程的复杂动态机制,实现对生物过程的精准在线控制,从而大幅提升生物制造的技术水平。

3.2 多组学数据与模型的整合

多组学海量数据为生物过程建模提供了丰富的数据支撑。基因组注释结合 KEGG、BioCyc、BRENDA、Reactome 等生化代谢数据库,构建 GSMM 约束计量模型。转录组、蛋白组及代谢组数据提供了基因表达和代谢动态的关键信息,可用于定性或定量地约束模型,提升其预测精度。在定性整合方面,转录组学推断基因表达的开关状态,并以此引入模型中。MADE、iMAT、INIT、RIPTiDe 等算法基于表达信息进行定性约束^[103-107]。MADE 方法通过基因表达变化的显著性来评估基因状态,预测代谢反应的活化或失活^[104]。iMAT 方法基于基因表达均值和方差,将表达水平分为高、中、低这 3 种状态,利用混合整数线性优化对代谢反应上下限施加约束,使预测结果与基因表达数据及热力学约束相符^[105]。INIT^[106]和 RIPTiDe^[107]算法通过整合转录组数据,筛选符合表达数据的代谢反应网络,不以细胞生长作为目标函数,而是最大化基因表达和代谢流的拟合度,以获得与组学数据一致的代谢网络模型。非靶向代谢组学数据则为代谢物及相关反应的存在性提供证据,辅助模型的进一步修正。

在定量约束方面, E-Flux^[108]是一种较粗糙的算法,根据特定函数关系将转录组数据转化为代谢反应约束,以求解特定条件下的代谢流分布。PROM^[109]算法则根据转录组学数据估算基因表达的概率,并引入模型计算代谢流。定量蛋白质组学提供了酶蛋白丰度数据, GECKO 方法^[110]通过 $v < k_{cat}[E]$ (其中 v 为反应速率, k_{cat} 为酶转换数, $[E]$ 为酶量)将定量蛋白组学数据引入模型中,作为代谢反应上限的约束条件。而

时间序列的定量代谢组学数据则估算了代谢物的净消耗或净生成速率 b ，通过方程 $S \cdot v = b$ (其中 S 为化学计量系数, v 为反应速率), 构建非稳态 FBA 算法实现对过程代谢流变化的计算。

动力学模型通常包含较多参数, 构建精确模型依赖实验数据进行参数估计。定量蛋白质组学、定量代谢组学、通量组学分别提供酶浓度、代谢物浓度及反应速率数据, 用于参数估计。然而, 这些数据往往数量有限, 导致参数估计的不确定性较大, 如何量化参数的不确定性成为构建大规模动力学模型的核心挑战之一。Ensemble 建模通过寻找能重现观测数据的参数值分布来应对这一挑战。最近研究采用 Bayesian 推断计算 Ensemble 建模的参数后验分布, 为整合多组学构建大规模动力学模型提供了方法^[111]。

多组学为生物过程 AI 模型提供了海量数据。整合时序转录组学、蛋白质组学和代谢组

学数据, 并耦合过程参数与生长表型监测, 为 AI 模型提供了多层次跨尺度的动态观测信息。数据的异质性和高维度性带来了整合挑战。基于图卷积网络(graph convolutional network)构建包含基因-蛋白-代谢物网络的系统图谱, 可在组学数据之间建立关联。多模态学习(multi-modal learning)技术也广泛应用于组学数据的融合, 通过共享隐层结构, 使得不同组学数据在同一模型中高效协同, 从而提升生物过程 AI 模型的准确性。此外, 迁移学习和自监督学习等策略可有效克服组学数据的异质性, 使模型在多源数据上具有良好的泛化能力。多组学数据的时序特征通过递归神经网络或长短期记忆网络(long short-term memory, LSTM)纳入模型中, 允许捕捉生物过程的时序动态变化, 为实时过程控制提供支持。通过这些方法, AI 模型能够整合多组学数据生成精确的预测和优化方案, 为生物过程中的智能化调控决策支持。

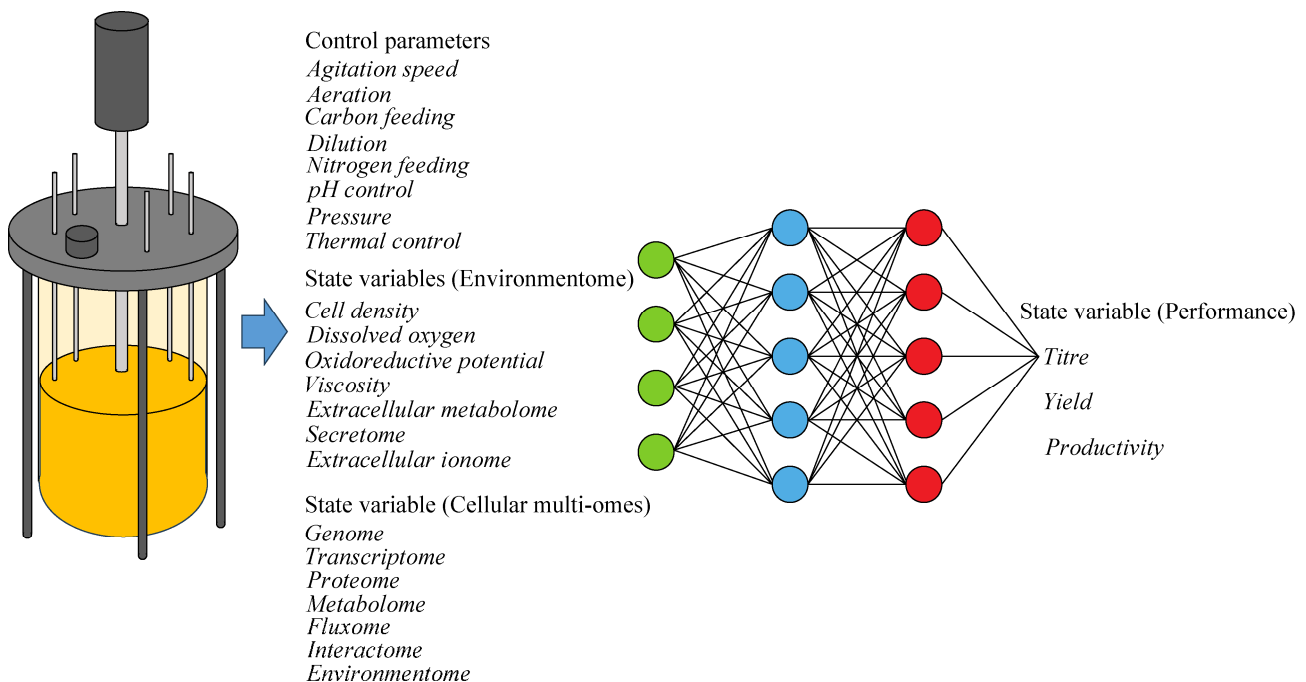


图 4 使用生物过程的控制变量与状态变量数据训练 AI 模型

Figure 4 AI model training using control variables and state variables of bioprocesses.

4 多组学分析与建模在生物制造过程中的应用

数据驱动的多组学分析与建模正重塑生物制造过程的优化范式。通过整合基因组、转录组、蛋白质组、代谢组及表型数据,构建生物系统的多维度数字模型,从而揭示工程细胞代谢调控与生物过程优化的底层逻辑。生物制造中,工程细胞需兼具环境适应性(如耐受极端 pH、高温或毒性代谢物)与代谢鲁棒性(如维持 NADH/ATP 平衡);而生物反应器需协同细胞特性与工程参数实现高效转化目标产物。多组学技术与计算建模的深度融合为此提供了创新解决方案:一方面,基于动态转录组、代谢组时序数据与代谢通量分析(如 FBA),可系统解析工程细胞在生长、胁迫及产物合成不同阶段的代谢网络重构机制;另一方面,结合三传一反(质量传递、热量传递、动量传递与化学反应工程)的工程学原理,揭示生物反应器中混合效率、传质限制及局部代谢异质性对产物合成的影响规律。这种多尺度建模方法不仅为定向进化筛选高产菌株、优化发酵工艺参数提供理论指导,更通过建立“基因-代谢-表型”的全局关联模型,推动生物制造从经验试错向理性设计的范式转变。在生物制造(尤其是发酵工程)中,过程变量、环境胁迫、营养底物、发酵阶段是影响产物合成与工艺优化的 4 个核心要素。通过系统性调控这些因素,可实现高效的过程优化与放大,本文将重点阐述多组学分析与建模在这 4 个方面的应用。

4.1 过程变量

过程变量又称过程参数,包括影响生物过程的各种操作变量以及描述过程状态的状态变量。常见的过程变量包括温度、压力、搅拌、通气、光照、补料速率等物理变量, pH、溶氧、

氧化还原电位、气体分压、底物浓度、生长因子浓度等化学变量,以及与生物相关的细胞密度、细胞活性、氧气摄取速率(oxygen uptake rate, OUR)、CO₂ 释放速率(carbon-dioxide evolution rate, CER)、呼吸熵(respiratory quotient, RQ)等生物变量。多组学分析和数理建模过程深入分析过程参数如何影响细胞生理代谢的分子机制,为优化生物过程提供理论根据。生物过程中温度显著影响细胞膜、蛋白表达及活性。通过转录组学和代谢组学比较低温(16 °C)和高温(32 °C)对产油红酵母类胡萝卜素、脂质和胞外多糖合成的影响发现,低温显著增加了胞外多糖的产量,而高温则通过影响三羧酸循环(citric acid cycle, TCA)循环关键酶的活性以及类胡萝卜素、脂质和胞外多糖合成基因的表达,显著提高了脂质和类胡萝卜素的产量,通过控温可以实现发酵中不同阶段的生产表型^[112]。氧化还原水平对于胞内生化代谢至关重要。生物过程中保持较低的发醇液氧化还原电位(oxidoreductive potential, ORP)有利于还原性产物的合成,Xia 等^[113]通过控制 ORP 在-350 mV,使用酶解汽爆玉米秸秆生产丁醇,发酵周期缩短了 6 h,总溶剂增加了 27.5%,丁醇产量提高了 34.2%至 10.2 g/L;使用 GSMM 比较不同条件下的代谢流分布,发现糖酵解和 TCA 循环代谢流增强,产生了充足的 NAD(P)H 促进丁醇合成。pH 是链霉菌聚赖氨酸发酵生产的关键参数,通过转录组学分析发现酸性 pH 冲击上调了聚赖氨酸合成酶的转录水平,上调了脂肪酸去饱和酶,增强了呼吸相关基因表达,最终显著提升了聚赖氨酸的产量。溶氧水平是生物过程中调控工程细胞代谢的关键。整合代谢组学、通量组学和转录组学与 GSMM 模型,Lu 等^[114]研究了黑曲霉生产工业酶过程中细胞如何响应限氧压力;发现在氧气受限的条件下,黑曲霉

的代谢转向更高的氧化还原状态，脂肪酸合成相关基因表达下调，能量供给减少，生长速率也因此下降；进一步发现，乙醛酸循环的上调减少了 TCA 循环中 NADH 的生成，从而帮助细胞在缺氧环境下维持氧化还原平衡；此外，糖酵解途径的通量增加可能有助于缓解氧气限制下细胞的能量需求；通过这种多组学整合的方法，研究揭示了黑曲霉如何通过代谢调控适应氧气限制，为微生物细胞工厂的系统设计与优化提供了新的见解。分批、补料-分批、连续是经常使用的几种发酵方式。Henriques 等^[115]结合 dFBA 及动力学模拟描述酵母在批式发酵过程中利用碳源氮源产出一系列代谢产物。通过模拟揭示了耐低温酵母可能利用 γ -氨基丁酸通路产生还原产物，从而实现还原力平衡，模型为模拟酵母批式发酵提供了理论架构。

4.2 环境胁迫

工业生物过程中，工程细胞持续面临多重胁迫压力。温度的升高或降低可能导致代谢酶活性的改变；pH 偏移削弱质子梯度驱动的物质转运；高浓度产物引起渗透压胁迫，影响底物和产物运输；原料纤维素水解液中的毒性化合物也会显著影响细胞生长和代谢。此类多因素协同作用导致细胞生长受限、代谢网络异常重构及产物合成效率衰减，其根本机制涉及从蛋白质变性、代谢通路阻塞到全局调控网络失活的跨尺度级联效应。然而，从分子损伤到系统代谢重构跨尺度响应机制仍存在认知盲区，亟需整合多组学动态监测与计算建模，解析胁迫应答的信号传导网络与代谢弹性调控节点，为定向进化或合成回路设计抗逆性强化策略提供理论框架。营养限制也是生物过程的常见胁迫之一，研究人员通过逐步降低酵母培养基氮浓度并结合多组学分析与代谢建模，揭示了细胞在碳/氮限制下的资源分配策略；通过多组学分

析解释在碳源受限条件下，细胞通过维持 75% 转录组活性和 50% 蛋白质组储备实现弹性调控，其中 74% 的基因翻译潜能可在环境变化时被快速激活以应对代谢需求；同时构建的酶约束代谢模型(ecYeast8.1)首次在恒定生长速率下量化了核心代谢通路储备中 80% 的酶容量储备，并通过 FBA 阐明了代谢储备的动态调节机制；进一步解析发现核糖体复合体存在不完全形式，氮限制时选择性上调 17 种核糖体蛋白，通过动态重组核糖体复合体实现翻译能力的快速重编程；该研究揭示了代谢和翻译储备如何使酵母能够迅速适应环境变化，重新分配资源，从而为代谢工程和合成生物学的应用提供了重要的理论支持^[116]。使用生物质水解液生产乙醇时，水解液中的香豆酸(p-hydroxycinnamic acid, pCA)有一定的细胞毒性，研究人员整合转录组学、代谢组学分析工业酵母 SA-1 在香豆酸胁迫下通过动态调控线粒体代谢、核糖体重构及氧化还原平衡等关键通路实现耐受的分子机制，并鉴定出 20 个调控乙醇生产和生物质积累的核心枢纽基因^[117]。马克思克鲁维酵母(*Kluyveromyces marxianus*)可发酵糖产乙醇，且能够在高于 37 °C 的条件下生长，但乙醇耐受性较低。整合转录组学以及 GSMM，深入研究乙醇胁迫响应机制；代谢建模预测生长速率增加时会出现呼吸-发酵代谢共存的现象，基因共表达网络分析揭示蛋白质错误折叠修复、蛋白降解、分子伴侣相关基因可能是响应乙醇胁迫的主要机制，HAP3、MET4 和 SNF2 可能是关键调控因子，模型计算表明，氨基酸代谢、膜脂代谢和麦角固醇合成在乙醇胁迫条件下代谢通量增加，相关酶的利用率上升；该研究为构建乙醇耐受 *K. marxianus* 细胞工厂提供了关键线索^[118]。中国仓鼠卵巢细胞(Chinese hamster ovary cells, CHO)是重组蛋白生产常用的细胞

工厂。蛋白质组学分析高产 CHO 细胞的蛋白合成、能量代谢、氧化还原途径都上调, 细胞生长减弱, 细胞衰亡途径下调, 比较生长和静止细胞, 意外发现了 ROS 清除相关的蛋白如 peroxidases 和 SOD2 上调表达^[119]。¹³C 代谢通量分析表明 CHO 细胞生产处于高产率时氧化代谢增强, ROS 产生较多, 谷胱甘肽氧化、磷酸戊糖途径氧化分支都比较活跃, 说明使用高产蛋白时 CHO 处于氧化胁迫响应状态^[120]。

4.3 营养底物

起始培养基组成及生物过程中营养成分的添加能够显著影响工程细胞的生产性能。多组学分析生物过程, 深入理解细胞营养需求, 进而指导工程细胞设计构建、培养基与过程优化。例如, 谷氨酸棒杆菌不能同时利用高浓度葡萄糖(15 g/L)和阿拉伯糖(15 g/L), 使用代谢组学分析发现丙酮酸激酶可能是限制因素; 敲除阿拉伯糖利用阻遏转录因子 *araR* 并过表达丙酮酸激酶, 成功构建了高效共利用葡萄糖-阿拉伯糖的细胞工厂^[121]。底物中碳氮比关系到化合物、脂质、菌体蛋白产量。研究人员通过转录组学和代谢组学分析, 发现琥珀酸和氨基酸合成受阻可能影响吡哆醇合成瓶颈, 发酵时添加琥珀酸、组氨酸和丝氨酸并优化碳氮比显著提高了吡哆醇的产量^[122]。高碳氮比条件下, 解脂酵母脂质产量较高, 代谢组学分析显示甘油相比葡萄糖更能促进脂肪酸的合成^[123]。通过脂质组学分析不同氮源(如豆粕、酵母提取物、硝酸钾和酒石酸铵)对脂质合成的影响, 发现豆粕和酒石酸铵有助于提高甘油三酯、花生四烯酸和 C16-C18 长链脂肪酸的比例^[124]。营养限制可促进圆红酵母脂质合成, 整合转录组、蛋白质组和代谢组学分析, 发现限磷条件下磷酸代谢和甘油三酯合成提高, TCA 循环下调, 碳流流向脂质合成^[125]。此外, 金属离子对发酵过程的影

响亦不容忽视。酿酒酵母以不同培养基生产类胡萝卜素差异显著, 比较转录组分析发现锌和铜转运蛋白的表达水平显著变化, 提示培养基中可能缺乏这 2 种离子, 外源单独添加锌、铜离子或同时添加两种离子分别将产量提高了 2.3、9.2 和 9.7 倍^[126]。在两阶段补料分批发酵中, 大肠杆菌工程菌能够合成泛酸, 但其生长表现较弱。通过结合转录组学和代谢组学分析, 发现乙醛酸循环和 TCA 循环基因的表达水平显著下降, 代谢物的浓度降低; 在生长阶段外源添加 1 g/L 柠檬酸, 结果使得生长速率提高了 1.54 倍, 泛酸产量增加 33%, 达到 45.4 g/L, 总时空收率提高了 22%^[127]。在大肠杆菌中过表达梭菌丁醇合成途径并删减其他发酵支路实现了正丁醇高效生产。进一步代谢组学分析发现, 由于删除了磷酸乙酰转移酶基因(*pta*), 胞内 CoA 不平衡造成了丙酮酸、丁酸及其他 CoA 衍生物的胞内积累, 另外催化丁酰-CoA 还原为丁醛的醇脱氢酶 AdhE2 活性不高, 可能限制了丁醇合成; 添加半胱氨酸提升游离 CoA 水平并提高 AdhE2 活性, 最终将正丁醇产量提升至 18.3 g/L^[128]。整合转录组学、代谢组学以及 GSMM 分析, 系统揭示了产氢梭菌(*Clostridium autoethanogenum*) 中聚-3-羟基丁酸酯 (poly-3-hydroxybutyrate, PHB) 合成受限于 NADH/NAD⁺ 氧化还原失衡及乙酰辅酶 A 浓度不足这两个制约因素。通过上调 *phbA* 启动子活性和重构细胞膜脂质组成实现碳流高效重定向, PHB 产量增加了 12 倍^[129]。

结合 dFBA 和非靶向代谢组学分析, 发现在葡萄糖和氨基酸共培养下, 大肠杆菌利用氨基酸, 减少葡萄糖摄取并伴随乙酸生成, 关键代谢物丙酮酸和草酰乙酸积累, 可能通过抑制磷酸转移酶系统调节营养摄取与代谢途径选择, 该研究为大肠杆菌在营养供应变化条件下的动态代谢调节机制提供了深入见解^[130]。此

外,使用蛋白质组和通量组数据校准,构建了酿酒酵母资源依赖的代谢动力学模型,成功预测了不同稀释率恒化培养下的特定蛋白质组,揭示了 Crabtree 效应并非由总蛋白质组的空间限制引起,而是由线粒体限制导致,蛋白质组的“过量容量”可能用于增强动态营养条件下的细胞鲁棒性。模型对动态条件下蛋白质组的预测与实验结果一致,该模型可用于工程细胞的代谢适应研究和生物过程的优化设计^[131]。

4.4 发酵阶段

生物过程的阶段化控制策略广泛用于优化生产性能。双阶段发酵是其中的典型,例如毕赤酵母前期使用甘油、葡萄糖快速生长菌体,后期使用甲醇作为底物和诱导剂生产蛋白^[132];溶剂发酵中梭菌前期生长菌体快速产酸,pH 快速下降诱导产溶剂^[133]。有机酸、氨基酸生产中前期通过有氧生长促进细胞快速增殖,后期转入无氧发酵提高产物合成效率^[134-135]。甾醇化学品生产前期长菌体耗尽碳源,后期添加底物启动甾醇合成^[136-137]。透明质酸发酵前期生长菌体,后期添加透明质酸酶降低发酵液黏度提高通气量,提高菌体生长和产物合成^[138]。使用黑曲霉分批发酵葡萄糖酸钠,前期菌体生长时供氧水平高、OUR 高、底物消耗快,后期菌丝停止生长时,限制供氧使 OUR 水平降低,这种两阶段 OUR 控制策略提高得率至 94%;在这些过程中,底物限制、诱导物添加、温度、pH 和通气的调控至关重要;通过 GSMM 计算分析过表达 ATPase-F₁ 亚基导致糖酵解速率加快,与对照相比,产酸期提前结束,并向产醇代谢转变,提高了溶剂总产量和丁醇生产速率^[139]。生物过程中出现代谢振荡现象,降低代谢鲁棒性,影响生产性能。研究表明代谢物浓度的周期变化可以调控代谢过程,引起代谢振荡。最近研究发现自产乙醇梭菌(*Clostridium autoethanogenum*)

进行自养连续气体发酵过程中表现出生长代谢振荡;该过程包括单独利用 CO 和共利用 CO/H₂ 的交替过程;整合高精度时序代谢组学和蛋白质组学分析发现,转氢酶反应热力学驱动力(NADH/NAD⁺×NADP⁺/NADPH)动态变化与代谢振荡高度相关,但蛋白组无明显变化;此外,采用热力学代谢流分析(thermodynamic metabolic flux analysis, tMFA)表明底物利用交替和代谢振荡确实在热力学水平上受到调控^[140]。在连续发酵甘油生产 1,3-丙二醇(1,3-propanediol, 1,3-PDO)的过程中,观察到丁酸梭菌的自发振荡现象,甘油代谢表现出周期性变化,乳酸、甲酸和氢气等产物滞后于生物质、1,3-PDO 和丁酸的生成;胞外氧化还原电位和胞内 NAD⁺/NADH 比值分析显示细胞在振荡过程中经历显著的氧化还原状态转换,随生长速率下降由氧化状态转向还原状态;转录组分析表明胁迫条件下丙酮酸代谢紊乱、乙酰辅酶 A 转化为乙醛导致乙醛过量积累可能是振荡产生的原因^[141]。

4.5 基于模型的过程优化放大

实验室小试到工业规模生产需要经过逐级放大,效率低、周期长,严重制约工程细胞的产业化应用。大体积生物反应器物料混合和传质效率相对较低,物料完全混合时间显著延长,通气补料过程中会导致溶氧、底物、pH 值不均一分布^[142]。例如,大肠杆菌在 900 L 反应器中发酵,靠近进料点的顶部区域葡萄糖浓度较高,而在反应器底部的葡萄糖浓度接近零^[143]。相对而言,细胞内代谢平衡时间为秒级。因此反应器不同区域的细胞代谢呈现出异质性。细胞模型的应用为生物过程的优化提供了强有力的支持。整合代谢建模、定量代谢组学和计算流体动力学分析,计算模拟溶氧、营养和压力的异质性可为过程放大提供理论基础,并发展基于脉冲、振荡、斜坡的动态补料模式。通过计算

Lagrangian 模拟反应器流体动力学发现,在青霉素发酵过程中,反应器体积 57%的底物耗尽,存在营养限制,微生物在底物过剩和底物耗竭的区域之间交替变化,时间尺度为秒级^[144]。将代谢模型与计算流体动力学模拟结合,评估了底物异质性对生物反应器中青霉素链霉菌代谢响应的影响,使用 scale-down 方法模拟 54 m³ 工业规模的青霉素发酵动力学,并基于模拟优化设计反应器,改变进料位置减少青霉素生产的下降幅度^[145]。使用 CFD 计算食气永达尔梭菌(*Clostridium ljungdahlii*)在 125 m³ 反应器进行的 CO 气体发酵,分析细胞运动模式、循环时间和时序响应;生命线分析表明 97%细胞营养受限,84%细胞经历胁迫而发生转录变化^[146]。

在生物过程的优化中,AI 模型在检测、过程优化、放大和过程控制方面展现出巨大的潜力。使用人工神经网络化了尼泊尔德巴利酵母(*Debaryomyces nepalensis*)木糖醇生产的温度、发酵时间、pH、氧气传质系数 k_{La} 、生物量^[147]。使用耐高温的地芽孢杆菌生产脂肪酶,测试了不同的培养基,经过神经网络和响应面模型优化培养参数,获得了最佳的生长温度、培养基体积、接种量、搅拌速率、培养时间和初始 pH^[148]。将通过代谢流平衡计算得到的中央碳代谢途径的代谢通量作为输入,结合集成学习预测解脂酵母的产品产量,对产量 >1 g/L 的产品的预测准确度达到 $R^2=0.87$ ^[149]。同样地,整合大肠杆菌细胞工厂的改造策略、生物过程变量与计算得到的代谢流作为输入,通过数据增强和集成学习解决数据稀疏、非标准化和不完整的问题,主成分分析帮助识别出影响生产的关键因素,从而实现对不同生物过程产量、产率和转化率的高精度预测($R^2=0.80-0.93$)^[150]。总之,整合代谢建模、计算流体动力学分析及人工智能模型的应用,成为优化生物过程、提高

生产效率及可持续性的关键手段。

5 未来展望

生物过程使用工程细胞或无细胞体系在生物反应器中利用淀粉糖、木质纤维素生物质、一碳化合物等原料来生产化学品、医药、食品、材料、生物燃料等产品,是连接基础生物科学与工业应用之间的桥梁。通过综合运用多组学与模型解析,可以深入探索生物过程中细胞的动态响应机制,从而为优化和提升生物过程生产效能提供理论基础并创新研究思路。然而,多组学和模型研究生物过程存在诸多挑战:(1) 不同于实验室的实验器材。工业生物反应器的大容积带来传质、传热、传动的不均匀性,导致温度、pH、溶氧、压力、水活度、黏度、底物浓度、代谢物浓度、毒性化合物浓度、离子强度等物理化学参数存在显著的时空异质性。在组学取样的过程中,这些异质性常常被忽视,从而影响数据的准确性和可靠性,进而阻碍对生物过程的深入理解。(2) 控制参数的设定。包括搅拌、补料、排料、控温、通气、酸碱调节、光照等,直接作用于反应器,显著影响工程细胞的生长代谢环境。这些参数的调整会影响细胞的生长分裂、基因表达和生化代谢等各层次生命活动。然而,大多数组学仅针对工程细胞取样,研究细胞组分的变化和相互作用,而忽略了环境中的代谢物、离子、蛋白等组分(即环境组)的变化,导致组学分析与实际应用之间存在脱节。(3) 生物过程的动态变化性给研究带来了挑战。工程细胞经历从延滞期、对数期、稳定期到衰亡期等不同的群体生长阶段,也经历了从菌丝萌发、生长到产孢等不同的生理阶段。另外从细胞层次上,代谢、转录、翻译、降解等细胞过程在时间尺度上存在显著差异。如何选取代表性的时间点,既能捕捉到关键的变化

拐点, 又能符合经济原则, 是研究中的一个重要挑战。(4) 生物分子之间广泛的动态的相互作用是生命活动的基础, 也是生物系统与传统化工过程的重要区别, 但是目前研究生物分子间全部相互作用的高通量自动化手段还比较缺乏。虽然传统研究大多关注代谢, 但代谢物水平的变化不仅是蛋白组层面酶量变化的结果, 也是酶活性变化的结果, 而后者与蛋白修饰和蛋白-小分子互作相关。转录组学和蛋白组学常用于解析细胞对环境中的营养和胁迫的响应机制, 但细胞如何感知这些外界信号尚不明确。同时, 细胞在转录、翻译等过程间的复杂协同调控机制也未被充分理解。(5) 在开展整合多组学研究时, 多数情况下多层次组学数据通常不是同时获取, 关联性不强, 数据的整合过程中, 数据的不完整性和质量问题非常突出。另外组学取样分析往往在线下非原位进行, 存在显著滞后性, 缺乏在线质谱和测序设备使得实时分析面临巨大挑战。(6) 生物过程组学数据分析涵盖多种方法, 其中多组学整合分析大多依赖于统计模型, 约束模型基于代谢稳态假设, 代谢动力学模型需要丰富的生化细节, 代谢动力学模型结合 CFD 能够在反应器尺度上进行精确的过程优化, 但需要大量的计算和实验资源。AI 模型在应用中也面临挑战, 参数设置不当可能导致模型欠拟合或过拟合, 影响泛化能力。因此, 针对具体问题选择合适的模型架构是一大挑战。

为了克服上述挑战, 首先需要充分考虑生物过程的时空异质性。组学取样时, 应确保样本能够代表性地反映工程细胞在不同生理阶段的生长和代谢状态, 同时也要精确记录反应器内如温度、pH、溶氧等物理化学参数的时空变化。这一步骤是确保样本具有代表性及数据准确性的基础, 也是深入理解和优化生物过程的

关键起点。其次, 多组学数据的整合与标准化亦是关键。应建立统一的数据标准和技术手段, 加强不同数据源之间的整合与对齐, 以增强数据的可比性和一致性, 同时实现多组学联合取样分析。再次, 发展和应用拉曼光谱、联合质谱和实时测序等在线组学技术, 实现生物过程的实时监测, 以提高数据的时效性和分析的准确性。此外, 需要发展离子组学、多糖组学、环境组学技术, 定量分析生物过程中细胞外环境代谢物、离子、蛋白组分的动态变化, 并将这些数据与离线细胞多组学数据协同分析, 也是提高生物过程理解深度的重要方向。在生物过程组学分析中, AI 的应用至关重要。积极发展相互作用组的高通量自动化研究方法, 针对工业底盘细胞组织大科学工程, 构建面向 AI 的系统化、标准化、规模化的工业生物大数据。利用 AI 技术高效处理和挖掘海量数据, 通过先进的数据分析方法, 显著提升数据解析的精度和深度。同时, 构建数字细胞模型, 结合组学数据和生物建模, 可以深入探讨生物过程的复杂机制, 为生物制造过程的优化提供精确的理论基础和数据支持。

展望未来, 随着高通量测序、高性能质谱、光谱等先进测量技术的快速发展, 大数据科学、人工智能技术与生物科学深度融合, 研究人员对生物制造过程中生物学与工程学的理解将不断深化, 数字孪生技术和智能反应器技术加速演进, 未来的生物过程检测与控制将迎来自动化、数据化、智能化的变革, 有力支撑生物制造产业和生物经济的发展壮大。

6 结语

生物过程是生物制造的核心环节, 具备时空异质性、多尺度复杂性和非线性动态演变等特征。本文详细总结了多组学技术探究生物过

程中复杂机制的创新途径,通过整合跨组学动态监测数据,构建机理模型或人工智能模型,实现分子-细胞-反应器多尺度的全面解析。同时,文章也探讨了组学分析和模型仿真在过程参数、控制参数、发酵阶段、环境压力、实时在线监测和优化等关键方面的应用。此外,本文还归纳了当前面临的挑战和未来可能的解决方案,这些方法工具的革新将推动生物制造从经验试错向可预测设计范式转型,为工业生物过程的开发和应用提供支持。

作者贡献声明

朱岩:方案设计、初稿写作;张志丹:初稿写作;覃培斌、申杰:文献收集整理;孙际宾:方案设计、监督指导、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] The Global Bioeconomy[EB/OL]. [2025-03-20]. https://www.naturefinance.net/wp-content/uploads/2024/05/ENG-TheGlobalBioeconomy_FINAL.pdf.
- [2] SHABESTARY K, KLAMT S, LINK H, MAHADEVAN R, STEUER R, HUDSON EP. Design of microbial catalysts for two-stage processes[J]. *Nature Reviews Bioengineering*, 2024, 2: 1039-1055.
- [3] DAN L, YING WY, PRABHU AV, BIN ABDUL RAHIM AA, SHENG ZLJ. Device for automated aseptic sampling: automated sampling solution for future cell and gene manufacturing[J]. *Frontiers in Bioengineering and Biotechnology*, 2024, 12: 1452674.
- [4] SHISHKIN AA, GIANNOUKOS G, KUCUKURAL A, CIULLA D, BUSBY M, SURKA C, CHEN J, BHATTACHARYA RP, RUDY RF, PATEL MM, NOVOD N, HUNG DT, GNIRKE A, GARBER M, GUTTMAN M, LIVNY J. Simultaneous generation of many RNA-seq libraries in a single reaction[J]. *Nature Methods*, 2015, 12: 323-325.
- [5] ZHENG GXY, TERRY JM, BELGRADER P, RYVKIN P, BENT ZW, WILSON R, ZIRALDO SB, WHEELER TD, McDERMOTT GP, ZHU JJ, GREGORY MT, SHUGA J, MONTESCLAROS L, UNDERWOOD JG, MASQUELIER DA, NISHIMURA SY, SCHNALL-LEVIN M, WYATT PW, HINDSON CM, BHARADWAJ R, et al. Massively parallel digital transcriptional profiling of single cells[J]. *Nature Communications*, 2017, 8: 14049.
- [6] ZHANG R, GONG HQ, ZENG XD, LOU CP, SZE C. A microfluidic liquid phase nucleic acid purification chip to selectively isolate DNA or RNA from low copy/single bacterial cells in minute sample volume followed by direct on-chip quantitative PCR assay[J]. *Analytical Chemistry*, 2013, 85(3): 1484-1491.
- [7] MILDRUM S, HENDRICKS A, STORTCHEVOI A, KAMELAMELA N, BUTTY VL, LEVINE SS. High-throughput minaturized RNA-seq library preparation[J]. *Journal of Biomolecular Techniques*, 2020, 31(4): 151-156.
- [8] SIELAFF M, KUHAREV J, BOHN T, HAHLBROCK J, BOPP T, TENZER S, DISTLER U. Evaluation of FASP, SP3, and iST protocols for proteomic sample preparation in the low microgram range[J]. *Journal of Proteome Research*, 2017, 16(11): 4060-4072.
- [9] CHEN WD, WANG S, ADHIKARI S, DENG ZH, WANG LJ, CHEN L, KE M, YANG PY, TIAN RJ. Simple and integrated spintip-based technology applied for deep proteome profiling[J]. *Analytical Chemistry*, 2016, 88(9): 4864-4871.
- [10] ZOUGMAN A, SELBY PJ, BANKS RE. Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis[J]. *Proteomics*, 2014, 14(9): 1006-1010.
- [11] 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.
- [12] ZHANG ZD, ZHANG C, ZHANG X, CHEN JZ, CAI NY, ZHONG SS, HAN ZB, ZHU Y, ZHENG P, SUN JB, LIU CX. Automated and integrated ultrahigh throughput screening for industrial strain enabled by acoustic-droplet-ejection mass spectrometry[J]. *Journal of Pharmaceutical Analysis*, 2024, 14(9): 100949.
- [13] HEINLE L, SULAIMAN K, OLSON A, RUTERBORIES K. A homologous series of internal standards for near universal application in the discovery LC-MS/MS bioanalytical laboratory[J]. *Journal of Pharmaceutical and Biomedical Analysis*, 2020, 190: 113578.
- [14] FROMMELT F, FOSSATI A, ULIANA F, WENDT F, XUE P, HEUSEL M, WOLLSCHIED B, AEBERSOLD R, CIUFFA R, GSTAIGER M. DIP-MS: ultra-deep interaction proteomics for the deconvolution of protein complexes[J]. *Nature Methods*, 2024, 21(4): 635-647.
- [15] CLASEN MA, RUWOLT M, WANG C, RUTA JL, BOGDANOW B, KURT LU, ZHANG ZH, WANG S, GOZZO FC, CHEN T, CARVALHO PC, LIMA DB, LIU F. Proteome-scale recombinant standards and a robust high-speed search engine to advance cross-linking MS-based interactomics[J]. *Nature Methods*, 2024, 21(12): 2327-2335.
- [16] COSTESSI A, van den BOGERT B, MAY A, VER LOREN van THEMAAT E, ROUBOS JA, KOLKMAN MAB, BUTLER D, PIROVANO W. Novel sequencing technologies to support industrial biotechnology[J]. *FEMS Microbiology Letters*, 2018, 365(16): fny103.
- [17] van DIJK EL, AUGER H, JASZCZYSSZYN Y,

- THERMES C. Ten years of next-generation sequencing technology[J]. *Trends in Genetics*, 2014, 30(9): 418-426.
- [18] ZHAO WX, ZENG W, PANG B, LUO M, PENG Y, XU JL, KAN B, LI ZP, LU X. Oxford nanopore long-read sequencing enables the generation of complete bacterial and plasmid genomes without short-read sequencing[J]. *Frontiers in Microbiology*, 2023, 14: 1179966.
- [19] NYE TM, van GIJTENBEEK LA, STEVENS AG, SCHROEDER JW, RANDALL JR, MATTHEWS LA, SIMMONS LA. Methyltransferase DnmA is responsible for genome-wide N6-methyladenosine modifications at non-palindromic recognition sites in *Bacillus subtilis*[J]. *Nucleic Acids Research*, 2020, 48(10): 5332-5348.
- [20] WENGER AM, PELUSO P, ROWELL WJ, CHANG PC, HALL RJ, CONCEPCION GT, EBLER J, FUNGTAMMASAN A, KOLESNIKOV A, OLSON ND, TOPFER A, ALONGE M, MAHMOUD M, QIAN YF, CHIN CS, PHILLIPPY AM, SCHATZ MC, MYERS G, DEPRISTO MA, RUAN J, HUNKAPILLER MW. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome[J]. *Nature Biotechnology*, 2019, 37(10): 1155-1162.
- [21] STARK R, GRZELAK M, HADFIELD J. RNA sequencing: the teenage years[J]. *Nature Reviews Genetics*, 2019, 20(11): 631-656.
- [22] JENJAROENPUN P, WONGSURAWAT T, PEREIRA R, PATUMCHAROENPOL P, USSERY DW, NIELSEN J, NOOKAEW I. Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN.PK113-7D[J]. *Nucleic Acids Research*, 2018, 46(7): e38.
- [23] PUST MM, DAVENPORT CF, WIEHLMANN L, TÜMMLER B. Direct RNA nanopore sequencing of *Pseudomonas aeruginosa* clone C transcriptomes[J]. *Journal of Bacteriology*, 2022, 204(1): e0041821.
- [24] TAN L, GUO ZH, SHAO YW, YE LW, WANG MM, DENG X, CHEN S, LI RS. Analysis of bacterial transcriptome and epitranscriptome using nanopore direct RNA sequencing[J]. *Nucleic Acids Research*, 2024, 52(15): 8746-8762.
- [25] FERNÁNDEZ-COSTA C, MARTÍNEZ-BARTOLOMÉ S, McCLATCHY DB, SAVIOLA AJ, YU NK, YATES JR 3rd. Impact of the identification strategy on the reproducibility of the DDA and DIA results[J]. *Journal of Proteome Research*, 2020, 19(8): 3153-3161.
- [26] DELAFIELD DG, LU GY, KAMINSKY CJ, LI LJ. High-end ion mobility mass spectrometry: a current review of analytical capacity in omics applications and structural investigations[J]. *TrAC Trends in Analytical Chemistry*, 2022, 157: 116761.
- [27] XU ST, YANG C, YAN XP, LIU HW. Towards high throughput and high information coverage: advanced single-cell mass spectrometric techniques[J]. *Analytical and Bioanalytical Chemistry*, 2022, 414(1): 219-233.
- [28] RIDGEWAY ME, BLEIHOLDER C, MANN M, PARK MA. Trends in trapped ion mobility–Mass spectrometry instrumentation[J]. *TrAC Trends in Analytical Chemistry*, 2019, 116: 324-331.
- [29] HEBERT AS, RICHARDS AL, BAILEY DJ, ULBRICH A, COUGHLIN EE, WESTPHALL MS, COON JJ. The one hour yeast proteome[J]. *Molecular & Cellular Proteomics*, 2014, 13(1): 339-347.
- [30] GUZMAN UH, MARTINEZ-VAL A, YE ZL, DAMOC E, ARREY TN, PASHKOVA A, RENUSE S, DENISOV E, PETZOLDT J, PETERSON AC, HARKING F, ØSTERGAARD O, RYDBIRK R, AZNAR S, STEWART H, XUAN Y, HERMANSON D, HORNING S, HOCK C, MAKAROV A, et al. Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition[J]. *Nature Biotechnology*, 2024, 42(12): 1855-1866.
- [31] JOHNSON CH, IVANISEVIC J, SIUZDAK G. Metabolomics: beyond biomarkers and towards mechanisms[J]. *Nature Review Molecular Cell Biology*, 2016, 17(7): 451-459.
- [32] MARKLEY JL, BRUSCHWEILER R, EDISON AS, EGHBALNIA HR, POWERS R, RAFTERY D, WISHART DS. The future of NMR-based metabolomics[J]. *Current Opinion in Biotechnology*, 2017, 43: 34-40.
- [33] LU WY, SU XY, KLEIN MS, LEWIS IA, FIEHN O, RABINOWITZ JD. Metabolite measurement: pitfalls to avoid and practices to follow[J]. *Annual Review of Biochemistry*, 2017, 86: 277-304.
- [34] SKENE PJ, HENIKOFF S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites[J]. *eLife*, 2017, 6: e21856.
- [35] KAYA-OKUR HS, WU SJ, CODOMO CA, PLEDGER ES, BRYSON TD, HENIKOFF JG, AHMAD K, HENIKOFF S. CUT&Tag for efficient epigenomic profiling of small samples and single cells[J]. *Nature Communications*, 2019, 10(1): 1930.
- [36] GALLAGHER LA, VELAZQUEZ E, PETERSON SB, CHARITY JC, RADEY MC, GEBHARDT MJ, HSU F, SHULL LM, CUTLER KJ, MACARENO K, de MORAES MH, PENEWIT KM, KIM J, ANDRADE PA, LaFRAMBOISE T, SALIPANTE SJ, RENIERE ML, de LORENZO V, WIGGINS PA, DOVE SL, et al. Genome-wide protein–DNA interaction site mapping in bacteria using a double-stranded DNA-specific cytosine deaminase[J]. *Nature Microbiology*, 2022, 7(6): 844-855.
- [37] SAVITSKI MM, REINHARD FBM, FRANKEN H, WERNER T, SAVITSKI MF, EBERHARD D, MARTINEZ MOLINA D, JAFARI R, DOVEGA RB, KLAEGER S, KUSTER B, NORDLUND P, BANTSCHIEFF M, DREWES G. Tracking cancer drugs in living cells by thermal profiling of the proteome[J]. *Science*, 2014, 346(6205): 1255784.
- [38] ZHANG XL, WANG Q, LI YN, RUAN CF, WANG SY, HU LH, YE ML. Solvent-induced protein precipitation for drug target discovery on the proteomic scale[J]. *Analytical Chemistry*, 2020, 92(1): 1363-1371.
- [39] TIAN Y, WAN N, ZHANG HQ, SHAO C, DING M, BAO QY, HU HY, SUN HY, LIU CG, ZHOU K, CHEN

- S, WANG GJ, YE H, HAO HP. Chemoproteomic mapping of the glycolytic targetome in cancer cells[J]. *Nature Chemical Biology*, 2023, 19(12): 1480-1491.
- [40] FENG YH, de FRANCESCO G, KAHRAMAN A, SOSTE M, MELNIK A, BOERSEMA PJ, de LAURETO PP, NIKOLAEV Y, OLIVEIRA AP, PICOTTI P. Global analysis of protein structural changes in complex proteomes[J]. *Nature Biotechnology*, 2014, 32(10): 1036-1044.
- [41] LI KJ, CHEN SJ, WANG KY, WANG Y, XUE LJ, YE YY, FANG Z, LYU JW, ZHU HY, LI YN, YU T, YANG F, ZHANG XL, GUO SQ, RUAN CF, ZHOU JH, WANG Q, DONG MM, LUO C, YE ML. A peptide-centric local stability assay enables proteome-scale identification of the protein targets and binding regions of diverse ligands[J]. *Nature Methods*, 2025, 22(2): 278-282.
- [42] CHEN SF, ZHOU YQ, CHEN YR, GU J. Fastp: an ultra-fast all-in-one FASTQ preprocessor[J]. *Bioinformatics*, 2018, 34(17): i884-i890.
- [43] AYLING M, CLARK MD, LEGGETT RM. New approaches for metagenome assembly with short reads[J]. *Briefings in Bioinformatics*, 2020, 21(2): 584-594.
- [44] KOREN S, WALENZ BP, BERLIN K, MILLER JR, BERGMAN NH, PHILLIPPY AM. Canu: scalable and accurate long-read assembly *via* adaptive *k*-mer weighting and repeat separation[J]. *Genome Research*, 2017, 27(5): 722-736.
- [45] CHIN CS, PELUSO P, SEDLAZECK FJ, NATTESTAD M, CONCEPCION GT, CLUM A, DUNN C, O'MALLEY R, FIGUEROA-BALDERAS R, MORALES-CRUZ A, CRAMER GR, DELLEDONNE M, LUO CY, ECKER JR, CANTU D, RANK DR, SCHATZ MC. Phased diploid genome assembly with single-molecule real-time sequencing[J]. *Nature Methods*, 2016, 13(12): 1050-1054.
- [46] FREIRE B, LADRA S, PARAMA JR. Memory-efficient assembly using flye[J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2022, 19(6): 3564-3577.
- [47] LUO RB, LIU BH, XIE YL, LI ZY, HUANG WH, YUAN JY, HE GZ, CHEN YX, PAN Q, LIU YJ, TANG JB, WU GX, ZHANG H, SHI YJ, LIU Y, YU C, WANG B, LU Y, HAN CL, CHEUNG DW, et al. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler[J]. *GigaScience*, 2012, 1(1): 18.
- [48] SEEMANN T. Prokka: rapid prokaryotic genome annotation[J]. *Bioinformatics*, 2014, 30(14): 2068-2069.
- [49] BRETTIN T, DAVIS JJ, DISZ T, EDWARDS RA, GERDES S, OLSEN GJ, OLSON R, OVERBEEK R, PARRELLO B, PUSCH GD, SHUKLA M, THOMASON JA 3rd, STEVENS R, VONSTEIN V, WATTAM AR, XIA FF. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes[J]. *Scientific Reports*, 2015, 5: 8365.
- [50] CANTAREL BL, KORF I, ROBB SMC, PARRA G, ROSS E, MOORE B, HOLT C, ALVARADO AS, YANDELL M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes[J]. *Genome Research*, 2008, 18(1): 188-196.
- [51] TATUSOVA T, DiCUCCIO M, BADRETDIN A, CHETVERNIN V, NAWROCKI EP, ZASLAVSKY L, LOMSADZE A, PRUITT KD, BORODOVSKY M, OSTELL J. NCBI prokaryotic genome annotation pipeline[J]. *Nucleic Acids Research*, 2016, 44(14): 6614-6624.
- [52] AGGARWAL G, RAMASWAMY R. Ab initio gene identification: prokaryote genome annotation with GeneScan and GLIMMER[J]. *Journal of Biosciences*, 2002, 27(1 Suppl 1): 7-14.
- [53] NACHTWEIDE S, STANKE M. Multi-genome annotation with AUGUSTUS[M]//Gene Prediction. New York, NY: Springer New York, 2019: 139-160.
- [54] BESEMER J, BORODOVSKY M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses[J]. *Nucleic Acids Research*, 2005, 33(Web Server issue): W451-W454.
- [55] McKENNA A, HANNA M, BANKS E, SIVACHENKO A, CIBULSKIS K, KERNYTSKY A, GARIMELLA K, ALTSHULER D, GABRIEL S, DALY M, DePRISTO MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data[J]. *Genome Research*, 2010, 20(9): 1297-1303.
- [56] de OLIVEIRA LINO FS, GARG S, LI SS, MISIAKOU MA, KANG K, VALE Da COSTA BL, BEYER-PEDERSEN TS, GIACON TG, BASSO TO, PANAGIOTOU G, SOMMER MOA. Strain dynamics of contaminating bacteria modulate the yield of ethanol biorefineries[J]. *Nature Communications*, 2024, 15(1): 5323.
- [57] LI H, DURBIN R. Fast and accurate short read alignment with Burrows-Wheeler transform[J]. *Bioinformatics*, 2009, 25(14): 1754-1760.
- [58] DOBIN A, DAVIS CA, SCHLESINGER F, DRENKOW J, ZALESKI C, JHA S, BATUT P, CHAISSON M, GINGERAS TR. STAR: ultrafast universal RNA-seq aligner[J]. *Bioinformatics*, 2013, 29(1): 15-21.
- [59] KIM D, LANGMEAD B, SALZBERG SL. HISAT: a fast spliced aligner with low memory requirements[J]. *Nature Methods*, 2015, 12(4): 357-360.
- [60] LANGMEAD B, SALZBERG SL. Fast gapped-read alignment with bowtie 2[J]. *Nature Methods*, 2012, 9(4): 357-359.
- [61] LIAO Y, SMYTH GK, SHI W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features[J]. *Bioinformatics*, 2014, 30(7): 923-930.
- [62] ANDERS S, PYL PT, HUBER W. HTSeq: a Python framework to work with high-throughput sequencing data[J]. *Bioinformatics*, 2015, 31(2): 166-169.
- [63] PATRO R, DUGGAL G, LOVE MI, IRIZARRY RA, KINGSFORD C. Salmon provides fast and bias-aware quantification of transcript expression[J]. *Nature Methods*, 2017, 14(4): 417-419.

- [64] BRAY NL, PIMENTEL H, MELSTED P, PACHTER L. Near-optimal probabilistic RNA-seq quantification[J]. *Nature Biotechnology*, 2016, 34(5): 525-527.
- [65] ZHAO SR, YE Z, STANTON R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols[J]. *RNA*, 2020, 26(8): 903-909.
- [66] LOVE MI, HUBER W, ANDERS S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2[J]. *Genome Biology*, 2014, 15(12): 550.
- [67] ROBINSON MD, McCARTHY DJ, SMYTH GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data[J]. *Bioinformatics*, 2010, 26(1): 139-140.
- [68] LAW CW, CHEN YS, SHI W, SMYTH GK. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts[J]. *Genome Biology*, 2014, 15(2): R29.
- [69] FANG LX, FAN J, LUO SL, CHEN YR, WANG CY, CAO YX, SONG H. Genome-scale target identification in *Escherichia coli* for high-titer production of free fatty acids[J]. *Nature Communications*, 2021, 12(1): 4976.
- [70] TYANOVA S, TEMU T, COX J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics[J]. *Nature Protocols*, 2016, 11(12): 2301-2319.
- [71] LOU RH, CAO Y, LI SS, LANG XY, LI YX, ZHANG YY, SHUI WQ. Benchmarking commonly used software suites and analysis workflows for DIA proteomics and phosphoproteomics[J]. *Nature Communications*, 2023, 14(1): 94.
- [72] XIA JY, SÁNCHEZ BJ, CHEN Y, CAMPBELL K, KASVANDIK S, NIELSEN J. Proteome allocations change linearly with the specific growth rate of *Saccharomyces cerevisiae* under glucose limitation[J]. *Nature Communications*, 2022, 13(1): 2819.
- [73] SMITH MJ, IVANOV DP, WEBER RJM, WINGFIELD J, VIANT MR. Acoustic mist ionization mass spectrometry for ultrahigh-throughput metabolomics screening[J]. *Analytical Chemistry*, 2021, 93(26): 9258-9266.
- [74] ZHOU ZW, LUO MD, ZHANG HS, YIN YD, CAI YP, ZHU ZJ. Metabolite annotation from knowns to unknowns through knowledge-guided multi-layer metabolic networking[J]. *Nature Communications*, 2022, 13(1): 6656.
- [75] de JONGE NF, LOUWEN JJR, CHEKMENEVA E, CAMUZEUX S, VERMEIR FJ, JANSEN RS, HUBER F, van der HOOFT JJJ. MS2Query: reliable and scalable MS² mass spectra-based analogue search[J]. *Nature Communications*, 2023, 14(1): 1752.
- [76] 黄丹妮, 邹宇, 刘秀霞, 杨艳坤, 白仲虎. 基于非靶向代谢组学分析谷氨酸棒杆菌分泌表达外源蛋白代谢差异[J]. *食品与发酵工业*, 2023, 49(20): 1-10. HUANG DN, ZOU Y, LIU XX, YANG YK, BAI ZH. Analysis for metabolite changes of *Corynebacterium glutamicum* secreting and expressing exogenous proteins based on untargeted metabolomics[J]. *Food and Fermentation Industries*, 2023, 49(20): 1-10 (in Chinese).
- [77] RAHIM M, RAGAVAN M, DEJA S, MERRITT ME, BURGESS SC, YOUNG JD. INCA 2.0: A tool for integrated, dynamic modeling of NMR- and MS-based isotopomer measurements and rigorous metabolic flux analysis[J]. *Metabolic engineering*, 2022, 69: 275-285.
- [78] CUI ZY, ZHONG YT, SUN ZJ, JIANG ZN, DENG JY, WANG Q, NIELSEN J, HOU J, QI QS. Reconfiguration of the reductive TCA cycle enables high-level succinic acid production by *Yarrowia lipolytica*[J]. *Nature Communications*, 2023, 14(1): 8480.
- [79] TEO G, KOH H, FERMIN D, LAMBERT JP, KNIGHT JDR, GINGRAS AC, CHOI H. SAINTq: Scoring protein-protein interactions in affinity purification - mass spectrometry experiments with fragment or peptide intensity data[J]. *Proteomics*, 2016, 16(15/16): 2238-2245.
- [80] HUTTLIN EL, BRUCKNER RJ, PAULO JA, CANNON JR, TING L, BALTIER K, COLBY G, GEBREAB F, GYGI MP, PARZEN H, SZPYT J, TAM S, ZARRAGA G, PONTANO-VAITES L, SWARUP S, WHITE AE, SCHWEPPE DK, RAD R, ERICKSON BK, OBAR RA, et al. Architecture of the human interactome defines protein communities and disease networks[J]. *Nature*, 2017, 545(7655): 505-509.
- [81] STEIGENBERGER B, PIETERS RJ, HECK AJR, SCHELTEMA RA. PhoX: an IMAC-enrichable cross-linking reagent[J]. *ACS Central Science*, 2019, 5(9): 1514-1522.
- [82] JIANG PL, WANG C, DIEHL A, VINER R, ETIENNE C, NANDHIKONDA P, FOSTER L, BOMGARDEN RD, LIU F. A membrane-permeable and immobilized metal affinity chromatography (IMAC) enrichable cross-linking reagent to advance *in vivo* cross-linking mass spectrometry[J]. *Angewandte Chemie (International Ed)*, 2022, 61(12): e202113937.
- [83] KAAKE RM, WANG XR, BURKE A, YU C, KANDUR W, YANG YY, NOVTISKY EJ, SECOND T, DUAN JC, KAO A, GUAN SH, VELLUCCI D, RYCHNOVSKY SD, HUANG L. A new *in vivo* cross-linking mass spectrometry platform to define protein-protein interactions in living cells[J]. *Molecular & Cellular Proteomics*, 2014, 13(12): 3533-3543.
- [84] CHEN ZL, MENG JM, CAO Y, YIN JL, FANG RQ, FAN SB, LIU C, ZENG WF, DING YH, TAN D, WU L, ZHOU WJ, CHI H, SUN RX, DONG MQ, HE SM. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides[J]. *Nature Communications*, 2019, 10(1): 3404.
- [85] LIU F, LÖSSL P, SCHELTEMA R, VINER R, HECK AJR. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification[J]. *Nature Communications*, 2017, 8: 15473.
- [86] IACOBUCCI C, GÖTZE M, IHLING CH, PIOTROWSKI C, ARLT C, SCHÄFER M, HAGE C, SCHMIDT R, SINZ A. A cross-linking/mass spectrometry workflow based on MS-cleavable

- cross-linkers and the MeroX software for studying protein structures and protein-protein interactions[J]. *Nature Protocols*, 2018, 13(12): 2864-2889.
- [87] COMBE CW, GRAHAM M, KOLBOWSKI L, FISCHER L, RAPPSILBER J. xiVIEW: visualisation of crosslinking mass spectrometry data[J]. *Journal of Molecular Biology*, 2024, 436(17): 168656.
- [88] SHANNON P, MARKIEL A, OZIER O, BALIGA NS, WANG JT, RAMAGE D, AMIN ND, SCHWIKOWSKI B, IDEKER T. Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. *Genome Research*, 2003, 13(11): 2498-2504.
- [89] LOVINO M, RANDAZZO V, CIRAVEGNA G, BARBIERO P, FICARRA E, CIRRINCIONE G. A survey on data integration for multi-omics sample clustering[J]. *Neurocomputing*, 2022, 488: 494-508.
- [90] CULLEY C, VIJAYAKUMAR S, ZAMPIERI G, ANGIONE C. A mechanism-aware and multiomic machine-learning pipeline characterizes yeast cell growth[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, 117(31): 18869-18879.
- [91] ROY S, RADIVOJEVIC T, FORRER M, MARTI JM, JONNALAGADDA V, BACKMAN T, MORRELL W, PLA HAR H, KIM J, HILLSON N, GARCIA MARTIN H. Multiomics data collection, visualization, and utilization for guiding metabolic engineering[J]. *Frontiers in Bioengineering and Biotechnology*, 2021, 9: 612893.
- [92] EDWARDS JS, IBARRA RU, PALSSON BO. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data[J]. *Nature Biotechnology*, 2001, 19(2): 125-130.
- [93] THIELE I, FLEMING RMT, QUE R, BORDBAR A, DIEP D, PALSSON BO. Multiscale modeling of metabolism and macromolecular synthesis in *E. coli* and its application to the evolution of *Codon* usage[J]. *PLoS One*, 2012, 7(9): e45635.
- [94] PAN DT, WANG P, WANG XL, SUN YQ, XIU ZL. Dynamic flux balance analysis of 1, 3-propanediol production by *Clostridium butyricum* fermentation[J]. *Biotechnology Progress*, 2024, 40(1): e3411.
- [95] GHENU AH, MARREC L, BANK C. Challenges and pitfalls of inferring microbial growth rates from lab cultures[J]. *Frontiers in Ecology and Evolution*, 2024, 11: 1313500.
- [96] STEUER R, GROSS T, SELBIG J, and BLASIUS B. Structural kinetic modeling of metabolic networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103(32): 11868-73.
- [97] SAA PA and NIELSEN LK. Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks[J]. *Biotechnology Advances*, 2017, 35(8): 981-1003.
- [98] CORNISH-BOWDEN A. *Fundamentals of enzyme kinetics*[M]. 3rd ed. London: Portland Press, 2004.
- [99] MISKOVIC L and HATZIMANIKATIS V. Production of biofuels and biochemicals: in need of an ORACLE[J]. *Trends in Biotechnology*, 2010, 28(8): 391-7.
- [100] JAMSHIDI N and PALSSON BO. Formulating genome-scale kinetic models in the post-genome era[J]. *Molecular Systems Biology*, 2008, 4: 171.
- [101] WANG G, HARINGA C, TANG WJ, NOORMAN H, CHU J, ZHUANG YP, ZHANG SL. Coupled metabolic-hydrodynamic modeling enabling rational scale-up of industrial bioprocesses[J]. *Biotechnology and Bioengineering*, 2020, 117(3): 844-867.
- [102] 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(1): 4669.
- [103] KERKHOVEN EJ. Advances in constraint-based models: methods for improved predictive power based on resource allocation constraints[J]. *Current Opinion in Microbiology*, 2022, 68: 102168.
- [104] JENSEN PA and PAPIN JA. Functional integration of a metabolic network model and expression data without arbitrary thresholding[J]. *Bioinformatics*, 2011, 27(4): 541-547.
- [105] JERBY L, SHLOMI T, and RUPPIN E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism[J]. *Molecular Systems Biology*, 2010, 6: 401.
- [106] AGREN R, BORDEL S, MARDINOGLU A, PORNPUTTAPONG N, NOOKAEW I, and 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.
- [107] JENIOR ML, MOUTINHO TJ, JR., DOUGHERTY BV, and PAPIN JA. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments[J]. *PLoS Computational Biology*, 2020, 16(4): e1007099.
- [108] COLIJN C, BRANDES A, ZUCKER J, LUN DS, WEINER B, FARHAT MR, CHENG TY, MOODY DB, MURRAY M, and GALAGAN JE. Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production[J]. *PLoS Computational Biology*, 2009, 5(8): e1000489.
- [109] CHANDRASEKARAN S and 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.
- [110] DOMENZAIN I, SANCHEZ B, ANTON M, KERKHOVEN EJ, MILLAN-OROPEZA A, HENRY C, SIEWERS V, MORRISSEY JP, SONNENSCHIEIN N, and NIELSEN J. Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0[J]. *Nature Communications*, 2022, 13(1): 3766.
- [111] TRAN LM, RIZK ML, and LIAO JC. Ensemble modeling of metabolic networks[J]. *Biophysics Journal*, 2008, 95(12): 5606-17.
- [112] ZHAO D, LI CJ. Multi-omics profiling reveals

- potential mechanisms of culture temperature modulating biosynthesis of carotenoids, lipids, and exopolysaccharides in oleaginous red yeast *Rhodotorula glutinis* ZHK[J]. LWT, 2022, 171: 114103.
- [113] XIA ML, WANG D, XIA YM, SHI HJ, TIAN ZY, ZHENG Y, WANG M. Oxidoreduction potential controlling for increasing the fermentability of enzymatically hydrolyzed steam-exploded corn stover for butanol production[J]. Microbial Cell Factories, 2022, 21(1): 130.
- [114] LU HZ, CAO WQ, LIU XY, SUI YF, OUYANG LM, XIA JY, HUANG MZ, ZHUANG YP, ZHANG SL, NOORMAN H, CHU J. Multi-omics integrative analysis with genome-scale metabolic model simulation reveals global cellular adaptation of *Aspergillus niger* under industrial enzyme production condition[J]. Scientific Reports, 2018, 8(1): 14404.
- [115] HENRIQUES D, MINEBOIS R, MENDOZA SN, MACÍAS LG, PÉREZ-TORRADO R, BARRIO E, TEUSINK B, QUEROL A, BALSACANTO E. A multiphase multiobjective dynamic genome-scale model shows different redox balancing among yeast species of the *Saccharomyces* genus in fermentation[J]. mSystems, 2021, 6(4): e0026021.
- [116] YU R, CAMPBELL K, PEREIRA R, BJÖRKEROTH J, QI Q, VORONTSOV E, SIHLBOM C, NIELSEN J. Nitrogen limitation reveals large reserves in metabolic and translational capacities of yeast[J]. Nature Communications, 2020, 11(1): 1881.
- [117] CIAMPONI FE, PROCÓPIO DP, MURAD NF, FRANCO TT, BASSO TO, BRANDÃO MM. Multi-omics network model reveals key genes associated with p-coumaric acid stress response in an industrial yeast strain[J]. Scientific Reports, 2022, 12(1): 22466.
- [118] de MOURA FERREIRA MA, Da SILVEIRA WB. Multi-omics data and model integration reveal the main mechanisms associated with respiro-fermentative metabolism and ethanol stress responses in *Kluyveromyces marxianus*[J]. Biochemical Engineering Journal, 2024, 211: 109471.
- [119] ALI AS, CHEN R, RAJU R, KSHIRSAGAR R, GILBERT A, ZANG L, KARGER BL, IVANOV AR. Multi-omics reveals impact of cysteine feed concentration and resulting redox imbalance on cellular energy metabolism and specific productivity in CHO cell bioprocessing[J]. Biotechnology Journal, 2020, 15(8): e1900565.
- [120] TEMPLETON N, SMITH KD, MCATEE-PEREIRA AG, DORAI H, BETENBAUGH MJ, LANG SE, YOUNG JD. Application of ¹³C flux analysis to identify high-productivity CHO metabolic phenotypes[J]. Metabolic Engineering, 2017, 43: 218-225.
- [121] KAWAGUCHI H, YOSHIHARA K, HARA KY, HASUNUMA T, OGINO C, KONDO A. Metabolome analysis-based design and engineering of a metabolic pathway in *Corynebacterium glutamicum* to match rates of simultaneous utilization of D-glucose and L-arabinose[J]. Microbial Cell Factories, 2018, 17(1): 76.
- [122] TIAN ZZ, LIU LX, WU LJ, YANG ZX, ZHANG YH, DU LP, ZHANG DW. Enhancement of vitamin B₆ production driven by omics analysis combined with fermentation optimization[J]. Microbial Cell Factories, 2024, 23(1): 137.
- [123] YUN EJ, LEE J, KIM DH, KIM J, KIM S, JIN YS, KIM KH. Metabolomic elucidation of the effects of media and carbon sources on fatty acid production by *Yarrowia lipolytica*[J]. Journal of Biotechnology, 2018, 272: 7-13.
- [124] LU HQ, CHEN HQ, TANG X, YANG Q, ZHANG H, CHEN YQ, CHEN W. Ultra performance liquid chromatography-Q exactive orbitrap/mass spectrometry-based lipidomics reveals the influence of nitrogen sources on lipid biosynthesis of *Mortierella alpina*[J]. Journal of Agricultural and Food Chemistry, 2019, 67(39): 10984-10993.
- [125] WANG YN, ZHANG SF, ZHU ZW, SHEN HW, LIN XP, JIN X, JIAO X, ZHAO ZK. Systems analysis of phosphate-limitation-induced lipid accumulation by the oleaginous yeast *Rhodospiridium toruloides*[J]. Biotechnology for Biofuels, 2018, 11: 148.
- [126] SU BL, LI AZ, DENG MR, ZHU HH. Transcriptome analysis reveals a promotion of carotenoid production by copper ions in recombinant *Saccharomyces cerevisiae*[J]. Microorganisms, 2021, 9(2): 233.
- [127] ZOU SP, LIU JL, ZHAO K, ZHU XT, ZHANG B, LIU ZQ, ZHENG YG. Metabolic engineering of *Escherichia coli* for enhanced production of D-pantothenic acid[J]. Bioresource Technology, 2024, 412: 131352.
- [128] OHTAKE T, PONTRELLI S, LAVIÑA WA, LIAO JC, PUTRI SP, FUKUSAKI E. Metabolomics-driven approach to solving a CoA imbalance for improved 1-butanol production in *Escherichia coli*[J]. Metabolic Engineering, 2017, 41: 135-143.
- [129] De SOUZA PINTO LEMGRUBER R, VALGEPEA K, TAPPEL R, BEHRENDORFF JB, PALFREYMAN RW, PLAN M, HODSON MP, SIMPSON SD, NIELSEN LK, KÖPKE M, MARCELLIN E. Systems-level engineering and characterisation of *Clostridium autoethanogenum* through heterologous production of poly-3-hydroxybutyrate (PHB)[J]. Metabolic Engineering, 2019, 53: 14-23.
- [130] ZAMPIERI M, HÖRL M, HOTZ F, MÜLLER NF, SAUER U. Regulatory mechanisms underlying coordination of amino acid and glucose catabolism in *Escherichia coli*[J]. Nature Communications, 2019, 10(1): 3354.
- [131] VERHAGEN KA, EERDEN SA, SIKKEMA BJ, WAHL SA. Predicting metabolic adaptation under dynamic substrate conditions using a resource-dependent kinetic model: a case study using *Saccharomyces cerevisiae*[J]. Frontiers in Molecular Biosciences, 2022, 9: 863470.
- [132] YANG ZL, ZHANG ZS. Engineering strategies for enhanced production of protein and bio-products in *Pichia pastoris*: a review[J]. Biotechnology Advances, 2018, 36(1): 182-195.
- [133] RIAZ S, MAZHAR S, ABIDI SH, SYED Q, ABBAS

- N, SALEEM Y, NADEEM AA, MARYAM M, ESSA R, ASHFAQ S. Biobutanol production from sustainable biomass process of anaerobic ABE fermentation for industrial applications[J]. Archives of Microbiology, 2022, 204(11): 672.
- [134] HAO YN, MA Q, LIU XQ, FAN XG, MEN JX, WU HY, JIANG S, TIAN DG, XIONG B, XIE XX. High-yield production of L-valine in engineered *Escherichia coli* by a novel two-stage fermentation[J]. Metabolic Engineering, 2020, 62: 198-206.
- [135] HARDER BJ, BETTENBROCK K, KLAMT S. Temperature-dependent dynamic control of the TCA cycle increases volumetric productivity of itaconic acid production by *Escherichia coli*[J]. Biotechnology and Bioengineering, 2018, 115(1): 156-164.
- [136] WANG JH, JIANG W, LIANG CJ, ZHU LH, LI YR, MO Q, XU S, CHU A, ZHANG L, DING ZY, SHI GY. Overproduction of α -farnesene in *Saccharomyces cerevisiae* by farnesene synthase screening and metabolic engineering[J]. Journal of Agricultural and Food Chemistry, 2021, 69(10): 3103-3113.
- [137] YE ZL, HUANG YL, SHI B, XIANG ZL, TIAN Z, HUANG M, WU LL, DENG ZX, SHEN K, LIU TG. Coupling cell growth and biochemical pathway induction in *Saccharomyces cerevisiae* for production of (+)-valencene and its chemical conversion to (+)-nootkatone[J]. Metabolic Engineering, 2022, 72: 107-115.
- [138] ZHANG YJ, DONG JJ, XU GC, HAN RZ, ZHOU JY, NI Y. Efficient production of hyaluronic acid by *Streptococcus zooepidemicus* using two-stage semi-continuous fermentation[J]. Bioresource Technology, 2023, 377: 128896.
- [139] DAI ZJ, ZHU Y, DONG HJ, ZHAO CH, ZHANG YP, LI Y. Enforcing ATP hydrolysis enhanced anaerobic glycolysis and promoted solvent production in *Clostridium acetobutylicum*[J]. Microbial Cell Factories, 2021, 20(1): 149.
- [140] MAHAMKALI V, VALGEPEA K, de SOUZA PINTO LEMGRUBER R, PLAN M, TAPPEL R, KÖPKE M, SIMPSON SD, NIELSEN LK, MARCELLIN E. Redox controls metabolic robustness in the gas-fermenting acetogen *Clostridium autoethanogenum*[J]. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117(23): 13168-13175.
- [141] ZHOU JJ, SHEN JT, WANG XL, SUN YQ, XIU ZL. Metabolism, morphology and transcriptome analysis of oscillatory behavior of *Clostridium butyricum* during long-term continuous fermentation for 1, 3-propanediol production[J]. Biotechnology for Biofuels, 2020, 13(1): 191.
- [142] ENFORS SO, JAHIC M, ROZKOV A, XU B, HECKER M, JURGEN B, KRUGER E, SCHWEDER T, HAMER G, O'BEIRNE D, NOISOMMIT-RIZZI N, REUSS M, BOONE L, HEWITT C, MCFARLANE C, NIENOW A, KOVACS T, TRAGARDH C, FUCHS L, REVSTEDT J, et al. Physiological responses to mixing in large scale bioreactors[J]. Journal of Biotechnology, 2001, 85(2): 175-85.
- [143] LAPIN A, SCHMID J, REUSS M. Modeling the dynamics of *E. coli* populations in the three-dimensional turbulent field of a stirred-tank bioreactor: a structured-segregated approach[J]. Chemical Engineering Science, 2006, 61(14): 4783-4797.
- [144] HARINGA C, TANG WJ, DESHMUKH AT, XIA JY, REUSS M, HEIJNEN JJ, MUDDER RF, NOORMAN HJ. Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: an analysis of organism lifelines[J]. Engineering in Life Sciences, 2016, 16(7): 652-663.
- [145] HARINGA C, TANG WJ, WANG G, DESHMUKH AT, van WINDEN WA, CHU J, van GULIK WM, HEIJNEN JJ, MUDDER RF, NOORMAN HJ. Computational fluid dynamics simulation of an industrial *P. chrysogenum* fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization[J]. Chemical Engineering Science, 2018, 175: 12-24.
- [146] SIEBLER F, LAPIN A, HERMANN M, TAKORS R. The impact of CO gradients on *C. ljungdahlii* in a 125 m³ bubble column: Mass transfer, circulation time and lifeline analysis[J]. Chemical Engineering Science, 2019, 207: 410-423.
- [147] PAPPU JSM, GUMMADI SN. Modeling and simulation of xylitol production in bioreactor by *Debaryomyces nepalensis* NCYC 3413 using unstructured and artificial neural network models[J]. Bioresource Technology, 2016, 220: 490-499.
- [148] EBRAHIMPOUR A, ABD RAHMAN RN, EAN CH'NG DH, BASRI M, SALLEH AB. A modeling study by response surface methodology and artificial neural network on culture parameters optimization for thermostable lipase production from a newly isolated thermophilic *Geobacillus* sp. strain ARM[J]. BMC Biotechnology, 2008, 8: 96.
- [149] CZAJKA JJ, OYETUNDE T, TANG YJ. Integrated knowledge mining, genome-scale modeling, and machine learning for predicting *Yarrowia lipolytica* bioproduction[J]. Metabolic Engineering, 2021, 67: 227-236.
- [150] OYETUNDE T, LIU D, MARTIN HG, TANG YJ. Machine learning framework for assessment of microbial factory performance[J]. PLoS One, 2019, 14(1): e0210558.