

· 智能生物过程控制优化 ·

张庆华 中国科学院过程工程研究所副研究员、硕士生导师，中国科学院大学教授、中国化工学会过程强化委员会青年委员，中国化工学会混合与搅拌专业委员会委员。长期从事化学反应工程、多相流反应器设计和放大的基础和应用研究。主持或参与多项国家自然科学基金、863 计划项目、国家重点研发计划、中国科学院先导 C、中国科学院科技服务网络计划(STS 计划)等项目。发表 SCI 论文 40 余篇，授权发明专利 13 项，实用新型专利 8 项，撰写专著 1 章。获 2023 年度国家科技进步二等奖及中国石油和化学工业联合会科技进步一等奖各 1 项。



杨超 中国科学院过程工程研究所所长、研究员。国家杰出青年科学基金获得者，教育部长江学者特聘教授，北京学者。主要从事反应器工程、航天固体动力材料和电子级化学品研究。主持基金委创新研究群体、国家重点研发计划、国家重点基金、国家重大科研仪器研制等项目。获国家科技进步二等奖 2 项、国家技术发明二等奖 1 项、国家自然科学基金二等奖 1 项，获何梁何利基金科学与技术奖、中国青年科技奖等。



生物过程的介尺度模拟仿真与 AI 优化

王智慧¹，王聪²，张庆华^{1,3*}，夏建业⁴，丛威¹，杨超^{1,3*}

1 石油化工分子转化与反应工程全国重点实验室 中国科学院过程工程研究所，北京 100190

2 中国矿业大学(北京)化学与环境工程学院，北京 100083

3 中国科学院大学 化工学院，北京 100049

4 中国科学院天津工业生物技术研究所，天津 300308

王智慧，王聪，张庆华，夏建业，丛威，杨超. 生物过程的介尺度模拟仿真与 AI 优化[J]. 生物工程学报, 2025, 41(3): 1197-1218.

WANG Zhihui, WANG Cong, ZHANG Qinghua, XIA Jianye, CONG Wei, YANG Chao. Mesoscale simulation and AI optimization of bioprocesses[J]. Chinese Journal of Biotechnology, 2025, 41(3): 1197-1218.

摘要: 生物过程是利用活性生物细胞或酶实现底物生物转化的一种绿色可持续、环境友好的加工过程，在生物制造中起着关键作用。具有双重属性的生物过程多层次多尺度间的复杂关联导致

资助项目：中国科学院战略性科技先导专项(XDC0120402)；国家自然科学基金(22478397, 22421003)

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDC0120402) and the National Natural Science Foundation of China (22478397, 22421003).

*Corresponding authors. E-mail: ZHANG Qinghua, qhzhang@ipe.ac.cn; YANG Chao, chaoyang@ipe.ac.cn

Received: 2024-07-23; Accepted: 2024-10-25; Published online: 2024-10-25

生物过程的优化十分困难, 深入认识介尺度机理是了解生物过程动态变化和梳理多尺度复杂关系的关键之一。介尺度的数值模拟为介尺度现象的认识提供了一种新途径, 人工智能(artificial intelligence, AI)优化与介尺度模拟的结合为生物过程的优化注入了新的活力。本文综述了生物过程中介尺度模拟和 AI 优化的研究进展, 探讨了可能的发展方向, 以期促进介尺度模拟和 AI 优化在生物过程中的应用与发展。

关键词: 生物过程; 介尺度模拟; 计算流体力学; 人工智能优化; 机器学习

Mesoscale simulation and AI optimization of bioprocesses

WANG Zhihui¹, WANG Cong², ZHANG Qinghua^{1,3*}, XIA Jianye⁴, CONG Wei¹, YANG Chao^{1,3*}

1 State Key Laboratory of Petroleum Molecular & Process Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China

2 School of Chemical & Environmental Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China

3 School of Chemical Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

4 Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

Abstract: As green, sustainable, and environmentally friendly material processing processes using biological cells or enzymes to achieve substance conversion, bioprocesses play an increasingly important role in biomanufacturing. It is difficult to optimize bioprocesses because of the complex relationship at multiple levels and multiple scales. The knowledge of mesoscale behaviors is the key to understanding the dynamics of bioprocesses and to sort out the complex relationships of parameter variations in the spatial-temporal domain. Mesoscale numerical simulation paves a way for understanding these phenomena, and the integration of artificial intelligence (AI) and mesoscale simulation offers new vitality into the optimization of bioprocesses. This article reviews the progress in mesoscale simulation and AI optimization of bioprocesses and discusses the possible development directions, aiming to promote the development of this field.

Keywords: bioprocess; mesoscale simulation; computational fluid dynamics; artificial intelligence (AI) optimization; machine learning

生物过程是利用生物催化剂生成所需的化合物或破坏不需要的物质或危险物质^[1]实现物质的转化。由于催化剂和降解产物不产生附加污染, 生物过程是一种绿色可持续、环境友好的生物加工过程, 在生物医药、能源、环境修复和食品等领域得到了越来越广泛的应用^[2]。

生物过程兼具物质加工科学和生命科学的属性。例如, 微生物细胞的发酵过程既包含微

生物细胞本身的增殖, 又包含基质消耗、产物合成等化学成分的转化, 呈现两大科学领域所包含的多层次研究系统的多尺度特征^[3]。目前, 对物质加工体系和生命体系的各个层次(水平)涉及的边界尺度(单元和系统尺度)的静态特征, 传统理论研究已较深入。但是, 生物过程的双重属性及各层次、各尺度之间存在复杂的关联^[4], 呈现复杂的动态特征, 目前对生物过程动态特

征的认识还十分有限。

生物过程中, 介于单元和系统之间的尺度(介尺度)往往表现出动态非均匀(非平衡、非线性)的特征, 采用基于实验测试或理论分析的传统方法对该区间现象及作用机理进行解析十分困难。在反应器层次、介观尺度的动态非平衡性多是由介尺度的结构和行为变化导致。近年来, 随着计算机技术和计算流体力学(computational fluid dynamics, CFD)的迅速发展, 介尺度的数值模拟为认识生物过程的介尺度行为提供了一种途径。Hulburt 等^[5]在颗粒技术中提出的群体平衡方程(population balance equation, PBE), 可以描述离散相的分布特性以及引起分布变化的离散相微观行为^[6-7]。采用 PBE 方程模拟多相体系中分散相大小与分布随时空变化的数学模型被称为群体平衡模型(population balance model, PBM), 与传统计算流体力学(CFD)结合, 可以揭示流场与颗粒间的相互影响, 耦合 PBM 的 CFD 模拟已被越来越多地应用在生物过程的模拟中。例如, 用于揭示活性污泥絮体等颗粒及颗粒群、生物过程中气泡群等的时空变化以及其他生物过程介尺度结构及行为的动态变化^[8-17]。不同于从连续介质角度考虑流体流动的传统流体力学, 格子玻尔兹曼方法(Lattice Boltzmann method, LBM)从微观动力学角度出发, 将流体的宏观运动看作大量微观粒子运动的统计平均, 既克服了宏观方法需要处理复杂边界的劣势, 又克服了微观方法计算尺度受限的局限性, 是一种介于宏观和微观这 2 种方法之间的介观数值模拟方法^[18], 已被用于生物膜、多孔介质等的模拟^[19-33], 直观显示介尺度的流场, 为介尺度现象的认识和结构的优化提供了有力依据。

生物过程双重属性下多层次多尺度的特征, 使得生物过程关键参数的识别及各参数间

的相关性分析十分困难, 限制了基于机理分析的过程优化方法的应用及发展。近年来, 计算机技术和互联网的发展, 推动了大数据和人工智能(artificial intelligence, AI)技术的发展, 新型 AI 技术——数据驱动的机器学习技术的发展, 将人工智能推向新的高度。机器学习(machine learning, ML)面向数据分析与处理, 可以识别复杂数据中的关联性, 建立数学模型, 预测大数据涵盖范围下的生物过程行为, 为生物过程的优化提供策略。基于机器学习的 AI 过程优化正成为生物过程高效优化的重要手段, 在生物医药、生物燃料、生物废水处理等生物过程优化方面表现出良好的应用性能^[34]。然而, 由于生物反应器内部环境的动态变化、微生物间的相互作用以及微生物与环境间的相互作用等诸多关键因素尚且未知, 限制了 AI 对复杂生物过程的优化性能的进一步提升。分子动力学模拟(molecular dynamics simulation, MDS)和计算流体力学(CFD)技术的进步, 使认识生物过程的动态特征成为可能, 也为 AI 优化提供了新的思路。MDS/CFD 与新一代 AI 技术(机器学习)的结合, 为生物过程的发展注入新的活力, 正在打破传统的生物过程的研究及发展模式。

本文主要概述了近年来生物过程中基于 PBM/LBM 的介尺度模拟仿真、机器学习方法的 AI 优化、AI 与 MDS/CFD 结合优化方面的研究进展, 探讨了可能的发展方向, 以期促进介尺度模拟和 AI 优化在生物过程中的应用与发展。

1 介尺度模拟

生物过程涉及生命体系和物质转化体系, 生命体系中的介尺度主要有生物大分子层次的二级结构、细胞层次的多分子(包括生物大分子和其他分子)形成的超大分子机器; 物质转化体

系中涉及的介尺度主要有底物层次的分子聚集体和反应器层次的颗粒聚集体。任瑛等^[35]综述了生命大分子和细胞层次的分子动力学模拟的理论模型和算法, 内容包括了生命体系的二级结构、超大分子机器及物质转化体系中分子聚集体的介尺度模拟。本文主要概述了反应器层次的介尺度计算流体力学模拟——PBM和LBM在生物过程中的应用。

1.1 群体平衡模型(PBM)

非均相颗粒体系中群体平衡模型是研究分散相分布特性的有效工具。采用基于粒数密度函数的群体平衡方程描述多相体系中颗粒状态随时间及空间的连续变化, 将颗粒的离散行为(聚并、破碎等)与统计平均属性(粒径等)联系起来。将 PBM 与传统计算流体力学(CFD)的 Euler-Euler 多相流模型耦合^[17,36] (图 1), 可以呈现反应器内流体与分散相以及分散相的颗粒间相互作用下的局部非平衡态的介尺度现象。PBM 与 CFD 双向耦合过程中, 通过包含相间作用力的 Navier-Stokes (N-S)方程计算流场, 获得反应器内的局部速度和相含率等参数; 求解包含颗粒聚并和破碎模型(表 1)的 PBE 方程, 通过 PBM 获得分散相(颗粒群)的颗粒尺寸和数量的变化, 修正 CFD 模型中的相间作用力; 通过 CFD 与 PBM 的双向耦合获得反应器内相间

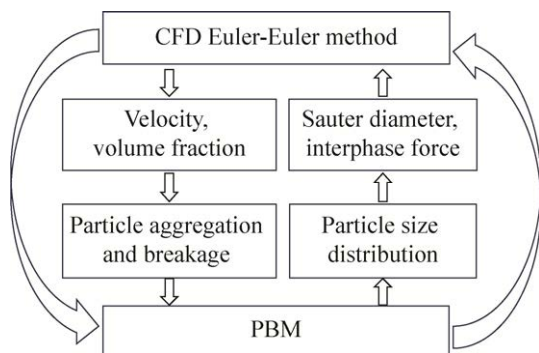


图 1 CFD-PBM 耦合方法^[17,36]

Figure 1 Schematic diagram of the CFD-PBM coupling model principle^[17,36].

作用影响下的连续相流动行为、颗粒群组成及分布特性的时空变化, 为生物过程的工艺优化以及生物反应器的设计优化及放大提供依据。

PBM 用于生物反应器中细胞培养过程的模拟, 可以描述由细胞生长、死亡和分裂引起的细胞分布随时间的变化, 能够捕获细胞培养过程中的局部非平衡态变化, 帮助研究人员加深对细胞周期动力学和细胞周期相位异质性的认识, 为更精准地调控细胞培养工艺提供依据。Györgya 等^[9]采用 PBM 模型模拟胚胎干细胞的分化过程, 捕获到传统的细胞增殖常微分方程 (ordinary differential equation, ODE) 难以提供的分化初期的细胞总量和未分化细胞数量细微波动的振荡期。Alhuthali 和 Kontoravdi^[14]采用单阶段基于体积的 PBM 方法, 捕捉到了仓鼠细胞培养过程中单克隆抗体和宿主细胞蛋白的积累动态, 计算了单克隆抗体和宿主细胞蛋白积累速率和最大积累量, 分析了各参数之间的相互关系, 为优化抗体滴度、降低杂质含量以及降低生产成本提供了依据。细胞培养过程的 PBM 模拟主要考察了细胞分化对细胞分布和代谢对物质含量分布的影响, 尚未考虑培养环境的流体力学因素与细胞生长及代谢的相互作用。一方面, 流体剪切或颗粒碰撞对剪切敏感细胞的活性有显著影响; 另一方面, 细胞代谢产物中的蛋白或多糖等物质也会影响流体环境的物理性质。对具有上述特征的细胞培养过程的模拟, 还需考虑培养环境的流体力学因素与细胞生长代谢的相互影响。

在好氧发酵等涉及气液两相的生物过程中, 气液传质性能显著影响生物过程效率, 气泡尺寸分布和气含率是影响气液传质性能的关键因素。采用考虑气泡聚并与破碎的 PBM 模型, 可以获取气泡群介尺度结构(气泡群的气泡数量和尺寸分布)随时空的变化, 为气液传质性

表 1 气泡聚并模型与破碎模型^[15]Table 1 Bubble coalescence models and breakage models^[15]

Model	Equation	Characteristics
Luo coalescence model ^[37]	$c(V_i, V_j) = w_{ij}^T P_c(d_i, d_j)$ $P_c(d_i, d_j) = \exp \left\{ - \frac{[0.75(1+x_{ij}^2)(1+x_{ij}^3)]^{0.5}}{\left(\frac{\rho_k}{\rho_l} + 0.5\right)^{0.5} (1+x_{ij})^3} \right\} We_{ij}^{0.5}$ $We_{ij}^{0.5} = \frac{\rho_l d_i \dot{A}_{<ij}^2}{\sigma}, \dot{A}_{<ij} = 1.43 \varepsilon^{\frac{1}{3}} \left(d_i^{\frac{2}{3}} + d_j^{\frac{2}{3}} \right)^{\frac{1}{2}}$	The most used kernel for bubble coalescence due to low computational requirements and optimal precision, needs parameters tuning to reduce the coalescence rate; over-predicts the collision frequency parameter ^[38]
Mc-Luo coalescence model ^[39-40]	$c(V_i, V_j) = w_{ij}^T P_c(d_i, d_j) M_c$ $M_c = \begin{cases} 0.13491 + 1.13403e^{-\frac{U_g}{0.01148}} & U_g \leq 0.101 \text{ m/s} \\ 0.10164 + 1.53737e^{-\frac{U_g}{0.04961}} & U_g > 0.101 \text{ m/s} \end{cases}$	Based on the energy-minimization multi-scale concept, the meso-scale energy dissipation was considered
Luo breakage model ^[41]	$\omega_1(\lambda, d_0) = 0.923(1 - \alpha_g) \varepsilon^{\frac{1}{3}} (1+x)^2 x^{\frac{11}{3}} d_0^{\frac{5}{3}},$ $x = \lambda / d_0, \lambda \leq d_0$	Energy spectrum combined with vortex energy
Han breakage model ^[42-44]	$\omega_2(\lambda, d_0) = 0.923(1 - \alpha_g) \varepsilon^{\frac{1}{3}} (1 + C_0 + x)^2 x^{\frac{11}{3}} d_0^{\frac{5}{3}}, \lambda \leq d_0$ $\omega_3(\lambda, d_0) = \frac{0.43(1 - \alpha_g) \varepsilon^{\frac{1}{3}}}{x^{\frac{14}{3}} d_0^{\frac{5}{3}} \min \left[1, 1 / \left(2\sqrt{2} x \sin^2 \left(\frac{\pi}{4x} \right) \right) \right]}, \lambda > d_0$	The collision frequencies of bubbles with large vortices and small vortices were modeled separately
Solsvik breakage model ^[45]	$\omega_4(\lambda, d_0) = \frac{\pi}{4} (\lambda + d_0)^2 \left[\frac{0.5}{\pi} (d_0 + \lambda) \gamma_l \right] \frac{0.822(1 - \alpha_g)}{\lambda^4}, \lambda \leq d_0$	Collision frequency caused by viscous shear was considered
Shi breakage model ^[46]	$\omega_5(\lambda, d_0) = C_4 (1 - \alpha) n_i \frac{(H\xi)^2}{d_0 \varepsilon^{\frac{2}{3}}} \sqrt{C_b \frac{\gamma g u_{slip}}{V}}, \lambda \leq d_0$ $C_4 = \frac{\pi}{4} C_3 C_\lambda^{\frac{1}{2}}, C_3 = \frac{12}{\pi C_\lambda (2\pi)^2}$	The mean turbulent velocity of eddy under the influence of turbulence and the characteristic corresponding to BIT characteristic wave number/length was considered

能优化提供依据, 指导反应器的优化及放大。Mishra 等^[12]将 PBM 与 CFD 以及传质模型耦合, 模拟得到的气液传质系数与实验值基本符合, 相对误差约为 3.8%–8.7%, 通过数值模拟考察了反应器操作条件对液相流动行为、气泡尺寸时空变化以及气液传质的影响, 确定了用于生产单克隆抗体药物的 300 L 的通气搅拌式反应器的转速和通气速率。在对通气搅拌式工业生物反应器的放大设计中, 万景等^[15]对比了 2 种聚并模型(Luo 聚并模型^[37-38]和基于能量最小多

尺度(energy minimization multi-scale, EMMS)理论的 Mc-Luo 修正聚并模型^[39-40]与 4 种气泡破碎模型(Luo 等^[41]提出的湍流涡诱导碰撞气泡破碎的模型、Han 等^[42-44]基于表面能量密度增加提出的考虑表面振荡和大尺度涡影响的气泡破碎模型、Solsvik 等^[45]考虑黏流剪切影响的气泡破碎模型和 Shi 等^[46]基于气泡诱导湍流的修正气泡破碎模型)相组合形成的 8 种组合模型(表 1)模拟得到的气泡尺寸及传质系数, 发现基于介尺度理论的 Mc-Luo 聚并模型与考虑黏

流剪切影响的气泡破碎模型组合模拟得到的气泡尺寸和体积传质系数与实验值更接近；基于上述的最优聚并-破碎组合模型，考察了桨型对反应器内气泡分散效果的影响，依据平均体积传质系数和混合死区体积，确定了 400 m³ 生物反应器的最优桨型组合。PBM 模型在气液两相生物反应器模拟中呈现的介尺度现象，帮助解释了部分放大效应产生的原因，为反应器放大提供了更可靠的依据。但是，聚并和破碎模型存在可调参数，这些参数主要源于实验或通过实验的对比确定，实验的依赖性使模型的拓展性受到限制。

在涉及活性污泥絮体等固体颗粒的通气式生物反应器中，分散相包括气泡和絮体 2 种分散相，气泡群和絮体聚集体 2 种介尺度结构的时空变化均对生物过程效率有重要影响。Hussain 和 Kumar^[16] 基于 Euler-Euler 方法，建立了半理论分析-半数值模拟方法，引入 Elzaki 积分变换处理域分解，用于曝气浆态搅拌槽反应器中的活性污泥絮体变化，提高了模拟的准确性和收敛性，获得反应器内颗粒在较大尺寸跨度下的聚集与破碎行为。Wodołażski^[10] 分别建立了气泡和絮体的群体平衡方程，采用标准矩方法求解气泡的群体平衡方程，与 CFD 耦合，考察了搅拌桨转速和气体流速对絮体尺寸分布、气泡尺寸分布及生物质生长的影响，为反应器结构和操作条件优化提供了依据。PBM 虽已拓展应用于气-液-固三相生物反应器的模拟，但是仍局限于分散相与连续相间的相互作用，缺乏两种分散相间的相互作用的研究。分散相间的相互作用往往对剪切、界面反应、相间传质等有显著影响，气-液-固三相生物反应体系中更多时空变化的介尺度现象尚待探索。

PBM 方法可以捕获培养细胞的非平衡态变化，直观显示生物过程中微生物絮体、气泡

群/固体颗粒群的变化，有助于揭示生物反应器内部介尺度的行为机制，为生物过程的工艺优化及反应器的结构优化及放大提供有力的依据。但是，目前 PBM 的聚并、破碎模型存在可调参数，可调参数仍主要来源于实验，在没有实验数据的情况下，难以保证模拟的准确性，限制了 PBM 的应用。基于 EMMS 的介尺度模拟性能的提升提示是否可以从热力学角度出发，探索介尺度的共性机制，基于统计学或变分原理等数学方法，建立简单易用的数学模型，代替可调参数，突破介尺度模拟的实验依赖性。

1.2 格子玻尔兹曼方法(LBM)

LBM 是一种介于宏观和微观这 2 种方法之间的介观数值模拟方法，从微观动力学角度出发，将流体的宏观运动看作大量微观粒子运动的统计平均，通过分布函数描述粒子分布状态和宏观物理量之间的关系，来获得密度、速度等宏观流动信息。LBM 通过分子运动论和统计力学，将微观和宏观连接，具有边界条件处理简单、易于并行计算等优点，在许多传统模拟方法难以胜任的微/纳尺度流动、多孔介质流等领域取得了诸多进展，在生物过程模拟中也逐渐得到应用。

生物体器官多具有复杂的多孔隙结构或多微通道结构，通过 LBM 仿真器官中介尺度通道内的流体动力学行为及细胞群体变化的介尺度行为，有助于理解器官功能和疾病机制。崔建强^[24] 采用 LBM 仿真不可压缩的流场，结合浸没边界方法(immersed boundary method, IBM) 处理固-液边界的相互作用，模拟得到了红细胞所处流体的速度剖面图、红细胞的坦克履带式运动、单细胞在微血管中变形和运动以及多细胞的分解图。赫轶男^[25] 采用 LBM 模拟了淋巴管的自发收缩-舒张过程，并呈现了多段淋巴

管内的淋巴液流动情况和一氧化氮 (nitric oxide, NO) 的产生及分布。在器官模拟方面, LBM 与 IBM 结合可以模拟变化的界面, 但是由于计算方法的限制, 目前只研究了较简单的多细胞或多段微通道中的介尺度行为, 具有三维各向异性多孔结构器官中的介尺度行为尚待揭示。

在含多孔介质或微通道结构的生物过滤或酶催化等生物反应器中, 多孔介质内部及表面的流动和传质对生物过程效率起着至关重要的作用。实验测试难以获得微生物生长富集及其对孔隙结构变化的影响, IBM 浸没边界方法可以描述复杂的边界, 四参数随机生成法 (quartet structure generation set, QSGS) 可以构建复杂的几何结构, 与上述方法之一结合, LBM 数值模拟使多孔介质中的流动及微生物生长情况得以呈现 (表 2)。Benioug 等^[27]采用 LBM-IBM 耦合方法模拟了多孔介质中的生物膜生长, 发现复杂多孔介质中生物膜的生长模式与 Damköhler

数、Péclet 数和无量纲剪切应力相关, 除孔隙率外, 微生物(膜)分布对多孔介质的渗透性有显著影响。马苏宁等^[29]将 LBM 与 IBM 耦合模拟多孔介质中的流场, 采用细胞自动机模型模拟多孔介质表面微生物的生长衰亡过程, 模拟得到的微生物生长及衰亡过程造成的多孔介质相对渗透率衰减与实验的变化趋势一致, 揭示了生物堵塞的动态发展过程及其造成多孔介质渗透性能改变的本质, 为避免生物堵塞提供了依据。杨勇等^[30]采用 LBM-IBM 耦合模型模拟多孔介质中流场, 采用细胞模型描述微生物生长, 建立了孔隙尺度微生物生长与表征体元尺度 (representative elementary volume, REV) 多孔介质渗透率衰减的耦合算法 (图 2), 通过 REV 尺度的渗流模型获得孔隙尺度的边界条件, 根据孔隙尺度的模拟计算平均孔隙率, 更新 REV 尺度的渗透率, 该方法不仅详细描述微生物生长特性, 还能准确反映孔隙尺度微生物生长与衰亡过程对多孔介质宏观渗流特性的影响, 发现

表 2 用于生物过程的 LBM 模型

Table 2 LBM models used in bioprocess

Models	Characteristics
IBM ^[27-30,32]	Using interpolation functions on discrete grids to describe the position and properties of solid boundaries, it can handle boundary problems with complex geometric shapes. The accuracy in simulating high Reynolds number flow needs to be improved
QSGS ^[31,33]	The morphological characteristics of porous medium constructed through the adjustment of four parameters [growth probability of the initial growth point, probability of growth direction, the volume (or area) ratio occupied by the growing phase, the growth probability between two contacting phases]. It can describe well the microscopic pore distribution and shape of real materials. It requires significant computational resources and time due to the complexity

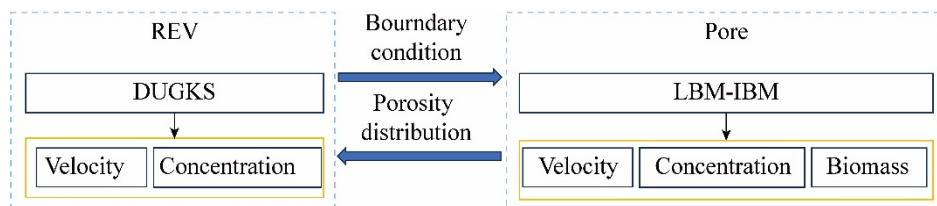


图 2 孔隙-表征体元尺度 LBM 耦合方法^[30]

Figure 2 Schematic diagram of the Pore-REV coupling LBM model principle^[30].

微生物生长对多孔介质渗透率衰减的影响规律,为有效控制地下水回灌过程中的微生物堵塞提供了理论依据。为了获得生物膜多孔结构内较为详细的信息和节约计算时间,杨艳霞等^[31]采用多块网格(生物膜采用细网格,主流区采用粗网格)离散方法,通过四参数随机生成法重构具有不同多孔结构的生物膜,用LBM模拟了平板式膜生物反应器内的流场及传质,揭示了反应器内介尺度的生物膜结构(孔隙率和孔隙结构)对物质传输及生化反应的影响规律,为生物膜的结构优化指明了方向,即通过提高生物膜孔隙率或改善生物膜的孔隙分布,可促进物质传输,强化生化反应,提高底物降解效率。目前,LBM用于生物过程多孔结构的数值模拟主要采用二维模型,能够呈现某些关键因素变化引起的介尺度现象的主要变化规律,但是多孔介质往往是各向异性,三维空间效应的影响难以体现;另一方面多孔介质内部参数难以检测,一般通过宏观参数的实验测试来验证模型,微观尺度上定量的准确性尚难判断。

采用浸没边界法或者多块网格方法处理曲面边界,LBM可以模拟具有复杂边界结构的器官或反应器中多孔介质孔隙的介观尺度的单相流流动和传质。同时,LBM方法易于实现孔隙尺度与表征体元尺度的耦合,可以为器官或生物膜反应器中介观尺度的行为调控提供支撑。但是,由于LBM模型在模拟湍流流动的数值稳定性、气液相界面变化的模拟准确度等方面尚待提高,在涉及气液固三相体系的生物过程的应用研究尚待开展。另外,目前LBM方法在生物过程的应用主要采用的是二维模型,三维模型报道较少,而三维模型能呈现更多物理空间效应方面的信息,比较三维模型与二维模型结果的异同,有助于更全面地了解介尺度的行为,分析介尺度的作用机制。

2 AI 优化

生物过程双重属性多层次多尺度间的复杂关联导致基于过程机理的优化十分困难。随着大数据时代的到来以及检测技术的发展,越来越多生物过程的实验数据得以积累,现代人工智能技术(AI)——机器学习(machine learning)的发展,为生物过程的优化注入了新的活力,生物过程的AI优化成为近期的研究热点。

2.1 机器学习方法简介

机器学习是人工智能的子领域,研究智能体(目前一般指计算机)基于经验提升表现的能力,机器学习利用数据构建模型,并利用构建的模型来求解问题^[47]。

传统的机器学习,根据数据的标注情况,一般分为监督学习、半监督学习和无监督学习^[48]。监督学习使用包含标签的数据对模型进行训练,再利用模型对未知数据进行预测,常见的监督学习算法包括决策树(decision tree)、支持向量机(support vector machine, SVM)、线性回归(linear regression)、逻辑回归(logistic regression)、神经网络(neural network)等。无监督学习采用不包含人为标签的数据作为训练集,通过k-means聚类(k-means clustering)、层次聚类(hierarchical clustering)、主成分分析(principal component analysis, PCA)等算法对数据进行分类,尝试挖掘输入数据背后的规律性。半监督学习针对仅部分数据含有标签的情况,基于有监督学习训练模型使用含标签的数据学习输入到输出的映射关系,然后使用大量的未标注数据去调整模型,常用的算法包括转导支持向量机(transductive support vector machine)、生成模型算法(generative model algorithm)、自训练算法(self-training algorithm)等。还有一种弱监督学习(weakly supervised learning)。针对的问题分为2种情况,

一种是标签不够精细，一种是标签不够准确。传统的机器学习通常基于数学和统计的原理，通过人工设计的特征进行数据的分类和预测，对数据量的需求相对较小，能够在小数据集上表现出色；但是，由于基于特征识别的机器学习方法，通常需要相关领域知识和大量的实验来选择和提取输入数据的特征，具有较强的人工依赖性，容易引入主观性和误差。

深度学习(deep learning)是一种基于多层神经网络的机器学习的新方法^[49]。不同于传统的机器学习一般需要人工构造特征，深度学习一般是通过梯度下降算法学习多层迭代的网络架构，核心是多层神经网络，每一层通过非线性变换来提取和学习数据的抽象特征。深度学习的主要优势在于能够从大规模数据中自动学习高级特征。根据学习策略，深度学习包括强化学习(reinforcement learning, RL)和迁移学习(transfer learning, TL)等多种模式。强化学习从一系列奖励信号中进行学习，这些信号提供了其行为质量的一些信息，包括有模型强化学习和无模型强化学习。在有模型强化学习中，智能体使用环境的转移模型来帮助解释奖励信号并决定如何行动。在无模型强化学习中，智能体不知道环境的转移模型，并且也不会学习它，而是通过动作效用函数或策略搜索方法直接学习如何采取行为方式。迁移学习可以将一个领域或任务中学到的知识或模式应用在另一个相关领域以解决目标问题的方法，允许从非常大的未标注语料库中开发模型，并应用于一系列任务，在目标领域进行微调后，通过预测缺失词预训练的模型处理该领域的任务。基于多层神经网络的深度学习能够从大规模数据中自动学习数据的特征，学习过程无需人工干预，用于处理复杂数据时，具有较高的效率和准确率。

在大数据的驱动下，随着多学科的协同发

展，例如数学、神经科学、心理学、控制学等多学科的共同发展，不断探索高级智能的构建逻辑，AI正在向着对人依赖越来越少的方向发展。近年来，具有不同学习模式的机器学习方法在各学科中的应用也不断拓展，模型架构日益丰富。

2.2 AI 在生物过程优化中的应用

近年来，AI优化在生物过程不同领域得到了广泛的应用，研究成果不断涌现。从生命科学角度，李丹丹等^[50]综述了近年来人工智能应用于生命体系的研究进展，包括了在生物分子层次到生命体方面的应用。本文主要综述以通过酶或活性细胞实现高的物质转化效率为优化目标的AI优化的应用进展。

2.2.1 酶催化过程

酶催化反应具有反应条件温和以及环境友好等优点，在医药、食品、能源等领域广泛应用。酶的选择性及其在体外反应环境中的活性和稳定性是影响酶催化过程工业化应用的主要问题，是酶催化反应过程优化关注的关键指标。机器学习具有高效处理数据的优势，可以覆盖广阔的搜索空间，快速发现具有映射关系的数据对或组合。目前，机器学习已用于酶的定向设计优化^[51-55]、酶功能的预测^[56-60]以及酶反应预测和优化^[61-65]等方面(表3)。

酶的特性取决于酶的结构，酶的定向设计优化是酶催化反应优化关注的主要方向之一。传统的突变扫描实验耗时长且成功率低，机器学习方法在预测具有某种特性的酶的序列方面已具有较高的准确性；而且数据驱动的机器学习方法可以广泛搜索性能-序列关系，显著提高酶的定向设计优化效率，扩充酶的序列空间。在蛋白质结构预测方面，基于深度学习的AlphaFold 2的全局距离检测(global distance test, GDT)得分中位数达到了92.4，预测的均方根误差约为1.6 Å，已具有很高的准确性^[52]。

表 3 AI 在酶催化过程中的应用

Table 3 AI models used in enzyme bioprocess

Subject	Task	AI algorithm	References
Enzyme design	<i>De novo</i> design malate dehydrogenase and fructose-1,6-bisphosphate aldolase	Generative adversarial networks (GAN)	[52]
	Predict halohydrin dehalogenase sequences	Leak generative adversarial networks (LeakGAN)	[53]
	Enhance the diastereoselectivity of ketoreductase for dihydrotetrabenazine synthesis	Gated recurrent unit (GRU)	[54]
Enzyme function	EC coding of multifunctional enzyme	Graph convolutional network and convolutional neural network (GCN-CNN)	[57]
	Function prediction of understudied enzymes, promiscuous enzymes with two or more EC numbers	Feed-forward neural networks Transformer (FFNN-Transformer)	[58]
	Function prediction of multi-level enzyme	Multi-view deep interactive network (MVDINET)	[59]
	Prediction of metabolic pathway	Graph convolutional network and random forest (GCN-RF)	[63]
Enzymatic reactions	Prediction of polysaccharide yield	Random forest (RF), eXtreme gradient boosting (XGB), deep neural networks (DNN)	[64]
	Prediction of enzymatic reactions with a molecular transformer	Multi-task transfer learning	[65]

很多已发现的酶由于功能未知,限制了这些酶的应用。通过机器学习方法可以预测已知序列或序列标注不足的酶的功能。钟灵茜^[57]建立了一种融合序列和多标签嵌入信息的多视角深度学习多功能酶功能预测方法(multi-label deep learning GCN-CNN net, mlDGCnet),将多视角学习、多标签分类机制和图卷积深度学习网络结构相结合,通过提取酶序列的序列相关性特征和序列无关特征构建多视角特征集;使用图卷积网络对酶分类标签信息进行深度特征提取,在此基础上,又使用了带多头注意力机制的卷积神经网络(convolutional neural networks, CNN)-双向门控循环单元(bi-directional gated recurrent unit, BiGRUs)混合网络来提取了多功能酶序列的深度局部特征和深度全局特征,预测了多功能酶的完整 EC 编码, EC 编码 1-4 层各子集的准确率分别是 97.84%、93.14%、90.27%和 72.49%。Tang 等^[59]提出了一种多视图深度学习方法

(multi-view deep interactive network, MVDINET)用于预测酶的多级功能,基准数据集的评估结果表明 MVDINET 对酶多级功能的预测性能显著优于传统方法。目前,单独采用机器学习方法即使深度学习尚不能成功预测未知序列酶的功能,仍需要融入其他类型的证据,并且由相关领域的专家审查^[60]。

机器学习在酶反应预测及过程优化中也体现了优良性能。Lincoln^[62]通过监督式机器学习方法,确定了与葡萄糖氧化酶反应性增强相关的醌氧化还原剂的结构组分,确定了酶电催化反应介质设计时需要优化的关键化学参数。Baranwal 等^[63]开发了一种基于图卷积神经网络(graph convolutional neural network, GCNN)和随机森林(random forest, RF)的机器学习方法,通过图卷积神经网络提取分子形状特征,作为随机森林分类器的输入,正确预测了 95.16%的测试生物化合物的相应代谢通路类别,对多标

签任务的预测准确率达到 95.62%。与传统的化学反应相比,酶反应的数据还较少,传统的依赖于大数据的机器学习方法如图神经网络(graph neural network, GNN)和循环神经网络(recurrent neural network, RNN)的应用受到了一定的限制。Kreutter 等^[65]采取了深度学习中的多任务迁移学习(multi-task transfer learning, MTL)的策略,利用美国专利商标局(United States Patent and Trademark Office, USPTO)的有机合成反应对模型进行预训练,再利用酶反应继续训练,通过有机反应的预训练和酶信息的加入提高了模型预测的准确性,模型性能受制于训练数据集的数据数量和质量。当使用完整的酶信息时,结合 USPTO 和 ENZR 数据集的 MTL 的酶反应结果预测准确率约为 62.2% (仅使用 ENZR 模型的准确率仅约为 34.3%),在数据集中数量最多的酶(脂肪酶和脱氢酶)的酶反应结果预测准确率最优,脂肪酶酶反应结果预测的准确率达到约 70%^[65]。可见,虽然采用深度学习中的迁移学习策略可以提高酶反应结果预测的准确率,但其准确率仍尚待提高。

数据驱动的 AI 优化可以绕过实验的突变筛选,覆盖广阔的搜索空间,挖掘潜在的序列和活性空间,识别功能与结构的映射,可以加速酶的定向设计优化,预测已知序列酶的功能,可以预测常见酶的反应路径及效率,显著提高酶反应过程的优化效率。机器学习不会由于先验知识而造成人为偏差,可以实现酶序列及反应基本性能的高效准确预测。但是,目前机器学习还不能成功预测无标注酶的功能,原因可能在于酶的结构复杂(酶的基础结构包含四级结构),机器学习对未知酶的构型构象及其变化的识别和数据化还不够精准,高质量的可信数据集还不足,而机器学习是基于数据的映射,预测的成功与否取决于输入到系统中的数据数

量与质量,描述酶的构型构象与功能的数据标准及数据质量的评判方法尚需完善。另一方面,迁移学习提高酶反应预测准确率的案例提示,不同学科之间具有潜在的相同机制,随着机器学习方法等 AI 技术的发展,有望挖掘到不同学科间存在的共同机制,建立可跨领域推广的更具普适性的数学模型或控制方程。

2.2.2 活性细胞生物转化过程

生物过程利用微生物或动植物细胞的生命活动来制备细胞本身、初级代谢产物或次级代谢产物。生物反应体系大多是多相和多组分的复杂体系,生物体内代谢及体外反应均极易受到反应环境的影响,反应路径易变,而且反应过程中众多影响参数间的关系复杂、隐蔽且随时空多变。传统的生物过程优化高度依赖于经验,而多因素多关联的复杂性使基于经验的数据分析和处理十分困难。AI 技术具有强大的数据读取与映射挖掘能力,在细胞识别和改造^[66-70]、反应路径及过程优化^[71-83]、生物过程的智能化检测与控制^[84-91]等方面展现出良好的性能(表 4)。

筛选高效稳定的细胞是提高生物过程效率的主要途径之一。传统的实验筛选一般需要多批次多轮次,各轮次细胞的识别还要依靠专业经验,费时费力,过程控制难,准确率也不高。具有大数据处理甚至图像分析能力的机器学习方法可以高效分析大量的光谱等数据提高细胞识别效率,提高细胞筛选通量;另一方面,机器学习方法可以搜索大量的基因组数据,挖掘功能和基因组间的关系,为细胞的基因改造提供依据。Shang 等^[66]分别使用支持向量机(support vector machine, SVM)和 k 近邻(k-nearest neighbor, KNN)机器学习算法分析拉曼光谱数据识别菌落, SVM 的菌落识别率超过 95%, KNN 的菌落识别率超过 90%。Aliev 等^[67]通过多层感知模

型(multilayer perceptron model, MPM)处理电化学数据, 预测样品中细菌浓度, 浓度识别范围为 10^2 – 10^9 CFU/mL, 平均准确率 97%, 将大肠杆菌检测时间从 24–48 h 缩短至 30 min。张和平和高广琦^[69]综述了人工智能在益生乳酸菌精准筛选方面的应用, 指出基于基因组大数据和人工智能的乳酸菌精准筛选技术, 显著提高了筛选速度和准确率, 也推动了益生乳酸菌的产业化进程。通过机器学习方法也可以高效准确

获得动物细胞的动态演变。Migdadi 等^[70]使用机器学习方法分析核磁共振 H-1-H-1 全相关谱, 获得了脂肪组织来源的人间充质干细胞(AT-derived hMSCs)以及脂肪细胞和骨细胞分化的动态演变, 并建立了多类分类以及代谢物检测方法, 其中分类器 kernel null Foley-Sammon transform (KNFST)和 kernel density estimation (KDE)的分类错误率不高于 3.6%, 假阳性和假阴性率均为 0。

表 4 AI 在涉及活性细胞的生物过程中的应用

Table 4 AI models used in bioprocess with cells

Subject	Task	AI algorithm	References
Cell screening	Ultra-fast identification of lactic acid bacteria colonies	Support vector machine (SVM), k-nearest neighbors (KNN)	[66]
	Detecting <i>Escherichia coli</i> bacteria by a bioelectrochemical platform combined with machine learning model	Multilayer perceptron model (MPM)	[67]
	Gene regulation in <i>Escherichia coli</i>	Deep mySQL group replication (DeepMGR)	[68]
	Capture the differentiation of adipose-tissue-derived human mesenchymal stem cells	Kernel null Foley-Sammon transform (KNFST), kernel density estimation (KDE)	[70]
Bioprocess optimization	High cell density cultivation of <i>Corynebacterium glutamicum</i>	Deep neural network, Bayesian optimization and genetic algorithm (DNN-BO-GA)	[78]
	Enhance performance in plant-based microbial fuel cells	Principal component analysis (PCA), classification tree (CT)	[79]
	Improve lipid production by <i>Rhodotorula glutinis</i> for renewable fuel production	Artificial neural network (ANN), support vector machine (SVM), multiple linear regression (MLR)	[80]
	Optimization continuous biohydrogen production in dynamic membrane bioreactor	Particle swarm optimization and functional link artificial neural network (PSO-FLN)	[81]
	Optimize of CHO-K1 fed-batch process	Artificial neural network and response surface methodology (ANN-RSM)	[82]
	Optimize biogas production from palm oil mill effluent	Feed-forward neural networks (FFNN)	[83]
Monitoring and control	Temperature control of a fermentation bioreactor	Twin delayed deep deterministic (TD3), deep deterministic policy gradient (DDPG)	[87]
	Predict cell concentration	Attention social scene long short-term memory (Attn-SSLSTM)	[88]
	Prediction of oxygen concentration	Direct memory access and recurrent neural network (DMA-RNN)	[87]
	Monitoring of biopharmaceutical processes	Principal component analysis (PCA)	[89]
	Rapid determination of starch and alcohol contents in fermented grains	Support vector machine (SVM)-AdaBoost/XGBoost	[90]
	Temperature control strategy	Random forest (RF)	[91]

生物过程细胞的增殖、代谢与环境参数往往呈现多耦合、非线性以及随时空多变的特征。为提高优化性能,对生物过程的反应路径或效率优化,一般需要将多种机器学习方法结合或将机器学习方法与机理模型结合^[71-77]。为提高谷氨酸棒杆菌的细胞生长速率, Konishi^[78]通过深度神经网络(deep neural network, DNN)辅助设计,结合贝叶斯优化(Bayesian optimization, BO)和遗传算法(genetic algorithm, GA)优化培养基组成,模型预测测试数据的 R^2 达到 0.98,采用最优培养基组分,与优化前相比,谷氨酸棒杆菌的生长速率提高了约 46%。Zhang 等^[80]使用多种机器学习算法对脂质发酵过程进行建模和优化,基于人工神经网络和支持向量机(support vector machine, SVM)的机器学习方法建立了生物油脂发酵过程的预测模型,模型的 R^2 高于 0.93,使用有监督学习方法中的多元线性回归(multiple linear regression, MLR)评估了受培养基成分影响的各个目标参数,采用遗传算法(GA)对生物油脂发酵进行了单目标和多目标优化,最大生物量为 50.3 g/L,油脂浓度最高达到 14.1 g/L,预测值与实验值之间的误差小于 5%。Pandey 等^[81]将粒子群优化(particle swarm optimization, PSO)和功能链接人工神经网络(functional link artificial neural network, FLANN)结合,开发了 PSO-FLN 混合模型,使用 PSO-FLN 混合模型预测动态膜反应器中的氢气产率和氢气产量,氢气产率和产量预测的 R^2 分别为 0.97 和 0.80,平均相对误差分别为 0.014%和 0.023%,PSO-FLN 模型能够在较短的计算时间内(氢气产率预测时间为 9.8 s,氢气产量预测时间为 10.0 s)处理复杂的数据集,而且具有较高的精度;采用 SHAP (shapely additive explanations)分析发现有机负荷率和丁酸对膜反应器性能影响显著。Pinto 等^[82]比较了基于不

同前馈神经网络(feed-forward neural networks FFNN)深度(3-5 层)的混合模型对中国仓鼠卵巢细胞培养过程中的参数的预测性能,发现深度模型具有更好的泛化能力,与浅层模型相比,训练和测试误差分别降低了 14.0%和 23.6%,CPU 时间增加 31.6%,可以预测培养过程中关键代谢产物(如乳酸、铵、谷氨酰胺和谷氨酸)的代谢转变。Tan 等^[83]将响应面法(response surface methodology, RSM)与人工神经网络(artificial neural network, ANN)相结合,优化后化学需氧量(chemical oxygen demand, COD)去除率、悬浮固体(suspended solid, SS)去除率和沼气产量分别提高了 7.38%、8.37%、16.18%。通过机器学习方法可以揭示 MSC 细胞进化过程中伴随的代谢变化,为细胞分化代谢途径的研究提供支持。

受限于体系的复杂性,很多过程参数难以通过传感器或光学设备直接测试,利用可检测/监测参数,通过机器学习获取其他过程参数,实时调控生物过程正成为生物过程优化的新态势^[84-86]。Rajasekhar 等^[87]对比了双延迟深度确定性策略梯度(twin delayed deep deterministic, TD3)和深度确定性策略梯度(deep deterministic policy gradient, DDPG)这 2 种强化学习方法对反应器温度的调控性能,采用无约束控制时,基于积分绝对误差奖励的 TD3 控制器的均方差为 0.22,低于 DDPG 的 0.29;在受限控制的情况下,基于积分绝对误差奖励的 TD3 控制器的均方差为 0.38,低于 DDPG 的 0.48。Sammaknejad 等^[89]开发了一种基于无监督学习中主成分分析(PCA)的实时多元统计监测系统,可以监测无历史数据的批式生物制药过程的稳态运行状态,及早预测过程中的故障或异常情况,为最终用户提供足够的时间来作出反应并采取适当的行动,保证工艺过程的稳健和高效。由此可见,

在生物过程中,针对不同参数可能需要建立不同的数据处理方法或 AI 模型。通过实时监测,基于多种 AI 方法结合其他机理方法获得关键参数的动态调控,实现对生物过程的实时监控和自动调节,有望减少人为错误和干预,提高生产效率,降低成本,提高产品质量。

数据驱动的 AI 技术在生物过程的各个领域已得到广泛应用,与传统的基于实验的优化相比,在某些方面的优化效率可以提高 1 到 2 个数量级,在生物过程优化方面体现出了显著的优势。从公开资料看, AI 优化使用的数据和知识主要来源于实验测试和基于实验测试的经验公式,如在线传感器测试或离线取样测试所获得的数据或基于该类型数据推导获得的描述状态或反应动力学的公式,一般为局域的宏观状态参数或描述状态或反应动力学的方程。虽然机器学习方法可以处理复杂非线性问题,但是生物过程中过程参数的相互影响使生命体及其环境所涉及的关键参数复杂多变,在实时和多变量系统中,机器学习的预测精度还有待提高。

3 AI 与数值模拟的结合

生物过程中生命体和其所在环境间的复杂关系的动态变化制约着 AI 优化的可解释性、性能提升以及应用拓展。近年来,数值模拟方法的发展及其在生物过程的应用,为生物过程的 AI 优化提供了新思路,如分子动力学模拟(MDS)或计算流体力学(CFD)模拟获得、经过验证的数据,为生物过程的 AI 优化提供新数据源。将 MDS/CFD 与 AI 技术结合,不仅可以为 AI 优化提供实验难以获取的动态环境数据,还可以提高 AI 的可解释性;另一方面,还可以通过 AI 技术提高 MDS/CFD 模拟效率^[92-93]。数值模拟与机器学习结合的新方法开始应用于生物过程优化。近几年,面向生物过程优化,研究人

员开始开展数值模拟与 AI 技术(主要是机器学习)结合的新方法的研究。接下来将介绍 AI 技术与 MDS/CFD 结合的研究进展。

3.1 AI 与 MDS 的结合

酶的性质和功能显著依赖于酶分子的构象变化, MDS 可以揭示分子构象变化, AI 与 MDS 的结合一方面有助于提高 AI 的可解释性,还可以补充 AI 优化的数据,提高酶功能预测和结构设计的精度;另一方面,对分子动力学模拟而言,提高精度需要大幅延长计算时间导致计算成本显著增加, AI 优化的加入有望提高效率,降低计算成本。

通过分子动力学模拟揭示酶构象变化机制,可以提高机器学习的可解释性。马睿^[94]分别以甲醇耐受性和酶活性为目标,使用包含不同数据量的数据集(突变文库)在 One-Hot、Georgiev 和 MSA transformer 编码方式下使用 10 倍交叉验证进行训练和测试,筛选到甲醇耐受性表现最强的 G202C/K208G/G266S 突变体,甲醇耐受性为野生型的 174.1%,通过分子动力学模拟轨迹分析,比较了表现最优突变体 G202N/K208G/G266S 与野生型在整体结构、局部结构、溶剂行为和非共价相互作用力的差异,发现突变使得盖子区域 α 螺旋的溶剂暴露降低,并且在突变区域引入了更加致密的氢键网络,提高了催化三联体在甲醇体系下的构象稳定性,揭示了 AI 优化获得的突变体具有较高甲醇耐受性的原因。通过分子动力学模拟显示酶的构象变化,可以解释酶特性或功能提高的原因,为酶的性能优化提供有力依据。

通过机器学习方法学习分子动力学模拟轨迹可以推断酶的构象变化,拓展酶的构象集合。朱镜璇^[95]引入深度学习(deep learning, DL)中的图神经网络(GNN)来学习分子动力学模拟轨迹,将蛋白质(酶)的变构或激活等过程公式化为

残基相互作用的动态网络,成功地推断了引发功能改变的残基/结构域相互作用变化,为调控蛋白质(酶)的构象变化以实现特定功能提供了依据。Zubatyyuk 等^[96]开发了 AIMNet-NSE 模型,使用从分子动力学模拟中采样的构象,构建了与化学反应相关的构象集合,扩展了酶的构象集合。该方法与量子力学模拟相比,效率显著提高。

将 AI 技术与分子动力学模拟双向耦合,可以提高分子动力学模拟的效率。Glielmo 等^[97]介绍了无监督学习技术在分子模拟数据分析中的应用,通过将轨迹数据转换为低维集体变量,将“原始”的笛卡尔坐标转换为保留了模拟轨迹相关信息的紧凑数值表示,通过无监督学习技术转换数据,可以有效地提取并保留分子模拟中的关键信息,提高效率的同时不降低精度。Köchl 等^[98]采用分子动力学模拟评估刺突的结合力,采用新型卷积神经网络架构对力场对进行高效搜索提高模拟效率,基于深度学习引导的分子动力学模拟,高效获得了 SARS-CoV-2 变体的特异性治疗诱饵。

分子动力学模拟与 AI 优化的结合主要是通过分子动力学模拟获得的介尺度数据提高 AI 的可解释性或者将分子动力学模拟的数据作为 AI 优化的数据源,优化蛋白质/酶的构象,或设计具有特定功能的酶,可以显著减小实验负担,节约资源、提高效率。AI 与 MDS 双向耦合的研究较少,下一步可能需要思考,如何将 AI 与 MDS 及其他学科(如统计学)结合,将不同类型的酶反应及反应过程不同阶段的数据分析与归纳,建立具有一定普适性的数学模型,为相似过程提供依据,使具有大数据处理能力的 AI 技术发挥更大的作用。

3.2 AI 与 CFD 结合

CFD 方法的进步及其在生物过程中的应

用,使研究人员对生物反应器内混合、传质及剪切等生物环境参数时空变化的认识日益深入,流体力学因素对生物的影响也日益受到重视。CFD 模拟产生大量的数据,与数据驱动的 AI 结合,将解释 AI 优化结果产生的原因,提高 AI 优化的精度;高效率的 AI 优化嵌入 CFD 程序中,通过时间或网格数据的优化来降低 CFD 模拟时间,有望在不降低精度的前提下提高效率。

CFD 可以为 AI 优化提供数据源或提高 AI 的可解释性,为工艺优化提供指导。Wodolazski^[11]采用元启发式优化对 CFD 得到的数据优化,比较了元启发式算法中的遗传算法(GA)、蚁群优化(ant colony optimization, ACO)和粒子群优化(PSO-bird)的优化效果,PSO-bird 算法得到的最优单位体积产率最高,比 ACO 算法得到的最优产率高约 6%,但 ACO 算法得到的最优产率所对应的操作范围最宽。Kaya 等^[99]根据 CFD 模拟获得的曝气搅拌槽反应器中不同条件(工作体积、搅拌浆转速和气体流速)下气相和液相的三维流场数据,计算不同条件下特定区域的水力学特性(流体剪切力和 Kolmogorov 尺度)的体积平均分布,之后分别采用量化软最大网络(quantized softmax networks, QSN)、混合密度网络(mixture density network, MDN)和核均值嵌入(kernel mean embeddings, KME)这 3 种机器学习方法建立水力学特性分布和过程输入参数间的关系。结果表明,这 3 种机器学习方法都可以实时预测相关的过程参数,其中, KME 模型在准确性、预测的平滑性和对微小变化的建模能力方面更优。Wodolazski^[11]在优化过程中发现,即使采用几何形状相似的放大准则,放大依然会影响反应器的功率,通过 PBM 与 CFD 耦合的数值模拟解释了上述问题产生的原因——放大导致的流体波动使功率增加。

AI 与 CFD 的双向耦合,可以提高 CFD 的

模拟效率和 AI 优化的准确性。韦振宇^[100]利用深度学习技术加速了 CFD 模拟的操作和计算：利用卷积神经网络(convolutional neural network, CNN)和长短期记忆(long short-term memory, LSTM)神经网络建立了 CFD 软件界面识别模型、CFD 模拟操作手册的自然语言处理及基于 CFD 时序瞬时流率数据的反应器产量预测三部分联合训练,通过深度学习技术加速 CFD 操作；利用长短期记忆神经网络建立了 CFD 瞬时流率数据驱动的生物质快速热解反应器产量预测模型,对比了 LSTM 模型分别耦合注意力机制(attention mechanism)和卷积神经网络(CNN)模型对反应器各组分瞬时流率及产量的预测效果,发现耦合卷积神经网络的 CLSTM 模型在反应器各组分瞬时流率及产量预测上误差最小,对生物油、裂解气及生物炭产量的预测误差分别为 2.44%、0.33%和 2.73%；采用长短期记忆神经网络耦合卷积神经网络的方法有效降低了 30%的 CFD 模拟计算时间,实现了 CFD 模拟的加速。AI 技术使 CFD 模拟降阶,提高 CFD 模拟效率；CFD 提高 AI 的优化性能。

目前来看, AI 优化与 CFD 的结合主要有以下几种作用：将 CFD 模拟的数据作为 AI 优化的数据源,获得实现目标指标最优化的操作条件；通过 CFD 解析 AI 优化产生的原因；通过 AI 优化提高 CFD 的模拟效率。前两种是数据的单向流通,最后一种 AI 优化提高 CFD 效率需要双向耦合。AI 优化提高 CFD 效率主要通过空间和时间优化这 2 方面,通过 CFD 获得粗网格的数据,利用 AI 对粗网格的数据扩展到细网格,更新原 CFD 粗网格的数据,多次循环,在达到与细网格 CFD 模拟相似的计算精度时,能大幅节约计算时间；对于瞬态问题,将 CFD 一定时间步的数据作为 AI 预测的输入,利用 AI 预测代替多时间步 CFD 的计算,然后 AI 的

数据再作为 CFD 新时间步的初始数据,如此循环,获得最终结果,从而提高 CFD 的效率,还能获得各流体力学参数随时间的进化情况。目前生物过程中介尺度模拟与 AI 结合或组合的研究较少,介尺度的数据非常缺乏,需要关注生物过程的介尺度现象和变化。如图 3 所示,在 AI 与 CFD 耦合方法中,完善生物信息和流场信息中的介尺度数据,从而揭示生物过程的动态演化机制,为精准调控生物过程提供有力依据。

4 结语

生物过程是具有生命体系与物质转化体系双重属性的多层次多尺度过程,具有时变性、滞后性、非线性以及多参数耦合性等复杂特性,生化过程的深入研究亟需多学科的交叉与融合。介尺度模拟和 AI 技术的进步,为生物过程的研究注入了新的活力。

生物过程中的介尺度模拟研究刚刚起步,前期的研究主要集中在不同边界尺度的模拟与尺度间的耦合方面。目前,生物过程中的介尺度模拟可以揭示孔隙中的流动与传质以及颗粒群中颗粒尺寸分布和颗粒数量的变化,但是尚缺乏亚颗粒尺度的物质分布及流场信息。例如菌体界面的受力是影响菌体活性的关键,菌体界面的物质分布是影响物质转化的关键。因此,生物过程中的介尺度模拟需要由颗粒间的作用向亚颗粒尺度(如界面问题)深入,呈现更多时空变化的细节；进而建立介尺度范围内不同物理空间尺度和不同时间尺度的问题的耦合方法,厘清生物过程的动态关联特征,解析多尺度间的耦合机制,为优化生物过程提供有力依据。

目前,基于计算技术与信息学的发展, AI 技术发展迅速,已经在生物过程优化中展现出了强大的应用潜力。虽然从智能的本质看, AI

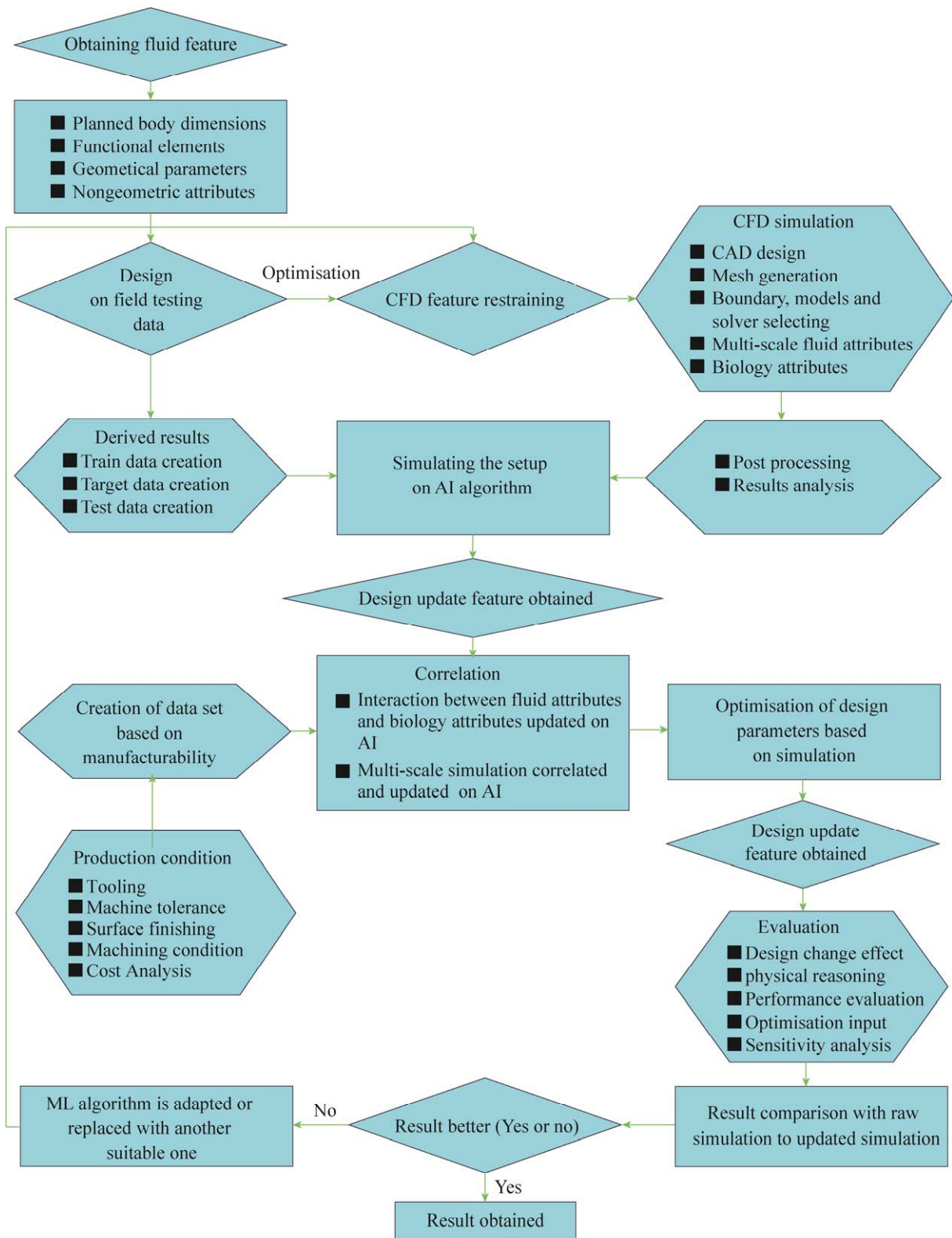


图 3 AI 与 CFD 结合的生物过程优化

Figure 3 Schematic diagram of the integration of AI and CFD in bioprocess.

技术的跨越式发展依赖于脑科学和神经作用机制的解析。但是,从物质基础看, AI 技术不依赖于生物体的组成,可以利用更广泛的物质材料,例如各种生物体不具备或耐受极端环境的传感器、执行器和控制器等,深入更广阔的空间探索,实时传感、计算流体力学与 AI 技术的发展与融合,有望推动基于实时智能优化的生物过程的智能控制的实现,进而带动生物技术的跨越式发展。

作者贡献声明:

王智慧:资料收集,方案设计、图表绘制、初稿写作,稿件修改;王聪:资料收集,图表绘制,初稿写作;张庆华:初稿写作,方案设计、图表修改、稿件修改、经费支持;夏建业:提供材料、稿件润色修改;丛威:提供材料、稿件润色修改;杨超:方案设计、稿件修改、经费支持。

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] 迈克尔 L. 舒勒, 费克莱特·卡基(著); 陈涛, 赵学明(译). 生物过程工程: 基本概念, 第二版[M]. 北京: 化学工业出版社, 2008: 3-9.
SHULER ML, KARGE F(Authors); Chen T, Zhao XM(Translators). *Bioprocess Engineering: Basic Concepts*, 2nd Edition[M]. Beijing: Chemical Industry Press, 2008: 3-9 (in Chinese).
- [2] 谭天伟, 秦培勇. 工业生物过程高效转化与系统集成[J]. 生物产业技术, 2017, (4): 25-30.
TAN TW, QIN PY. High efficient biotransformation and system integration in the process of industrial biotechnology[J]. *Biotechnology & Business*, 2017, 4: 25-30 (in Chinese).
- [3] LI JH. Exploring the logic and landscape of the knowledge system: multilevel structures, each multiscaled with complexity at the mesoscale[J]. *Engineering*, 2016, 2(3): 276-285.
- [4] 张嗣良, 潘杭琳, 黄明志, 谢梅野. 生物过程大数据分析智能化[J]. 生物产业技术, 2018(1): 86-93.
ZHANG SL, PAN HL, HUANG MZ, XIE MY. Big data and intelligentized method for industrial bioprocess[J]. *Biotechnology & Business*, 2018(1): 86-93 (in Chinese).
- [5] HULBURT HM, KATZ S. Some problems in particle technology: a statistical mechanical formulation[J]. *Chemical Engineering Science*, 1964, 19(8): 555-574.
- [6] CHENG JC, YANG C, MAO ZS. CFD-PBE simulation of premixed continuous precipitation incorporating nucleation, growth and aggregation in a stirred tank with multi-class method[J]. *Chemical Engineering Science*, 2012, 68(1): 469-480.
- [7] 李倩, 程景才, 杨超, 毛在砂. 群体平衡方程在搅拌反应器模拟中的应用[J]. 化工学报, 2014, 65(5): 1607-1615.
LI Q, CHENG JC, YANG C, MAO ZS. Application of population balance equation in numerical simulation of multiphase stirred tanks[J]. *CIESC Journal*, 2014, 65(5): 1607-1615 (in Chinese).
- [8] MORCHAIN J, GABELLE JC, COCKX A. A coupled population balance model and CFD approach for the simulation of mixing issues in lab-scale and industrial bioreactors[J]. *AIChE Journal*, 2014, 60(1): 27-40.
- [9] GYÖRGYÁ, R, KLONTZAS ME, KOSTOGLÓU M. A population balance model for stem cell differentiation bioprocesses[J]. *Computer Aided Chemical Engineering*, 2017, 40: 2761-2766.
- [10] WODOŁAŻSKI A. Co-simulation of CFD-multiphase population balance coupled model aeration of sludge flocs in stirrer tank bioreactor[J]. *International Journal of Multiphase Flow*, 2020, 123: 103162.
- [11] WODOŁAŻSKI A. Metaheuristic optimization of CFD-multiphase population balance and biokinetics aeration stirrer tank bioreactor of sludge flocs for scale-up study with bio(de/re)floculation[J]. *Biochemical Engineering Journal*, 2022, 184: 108477.
- [12] MISHRA S, KUMAR V, SARKAR J, RATHORE AS. CFD based mass transfer modeling of a single use bioreactor for production of monoclonal antibody biotherapeutics[J]. *Chemical Engineering Journal*, 2021, 412: 128592.
- [13] CHEN H, ZHANG XB, LUO ZH. A CFD-PBM coupled method to optimize a pilot-scale stirred bioreactor[J]. *Industrial & Engineering Chemistry Research*, 2022, 61(23): 8302-8312.
- [14] ALHUTHALI S, KONTORAVDI C. Population balance modelling captures host cell protein dynamics in CHO cell cultures[J]. *PLoS One*, 2022, 17(3): e0265886.
- [15] 万景, 张霖, 樊亚超, 刘颀民, 骆培成, 张锋, 张志炳. 基于介尺度 PBM 模型生物反应器放大模拟及实验研究[J]. 化工学报, 2022, 73(6): 2698-2707.
WAN J, ZHANG L, FAN YC, LIU XM, LUO PC, ZHANG F, ZHANG ZB. Bioreactor scale-up simulation and experimental study based on mesoscale PBM model[J]. *CIESC Journal*, 2022, 73(6): 2698-2707 (in Chinese).
- [16] HUSSAIN S, KUMAR R. Elzaki projected differential

- transform method for multi-dimensional aggregation and combined aggregation-breakage equations[J]. *Journal of Computational Science*, 2024, 75: 102211.
- [17] NI SZ, ZHAO T, SUN ZH, WANG W, SU KZ. CFD simulation for comparative of hydrodynamic effects in biochemical reactors using population balance model with varied inlet gas distribution profiles[J]. *International Journal of Chemical Reactor Engineering*, 2024, 22(3): 277-297.
- [18] 何雅玲. 格子 Boltzmann 方法的理论及应用[M]. 北京: 高等教育出版社, 2023: 1-9.
HE YL. *Lattice Boltzmann Method: Theory and Applications*[M]. Beijing: Higher Education Press, 2023: 1-9 (in Chinese).
- [19] YONG YM, YANG C, JIANG Y, JOSHI A, SHI YC, YIN XL. Numerical simulation of immiscible liquid-liquid flow in microchannels using lattice Boltzmann method[J]. *Science China Chemistry*, 2011, 54(1): 244-256.
- [20] MO HY, YONG YM, YU K, CHEN WQ, DAI JL, YANG C. An integrated Lattice-Boltzmann model of immiscible two-phase flow and bulk mass transfer with Marangoni effect[J]. *Journal of Computational Physics*, 2023, 481: 112037.
- [21] MO HY, YONG YM, CHEN WQ, DAI JL, YANG C. A regularized lattice Boltzmann model with filter for multiphase flow with diffusion-dominated mass transfer considering two-film theory[J]. *Physics of Fluids*, 2023, 35(11): 112121.
- [22] 张潇丹, 雍玉梅, 李文军, 赵元生, 李媛媛, 杨巧文, 杨超. REV 尺度多孔介质格子 Boltzmann 方法的数学模型及应用的研究进展[J]. *化工进展*, 2016, 35(6): 1698-1712.
ZHANG XD, YONG YM, LI WJ, ZHAO YS, LI YY, YANG QW, YANG C. Models and application of lattice Boltzmann method at REV-scale in porous media[J]. *Chemical Industry and Engineering Progress*, 2016, 35(6): 1698-1712 (in Chinese).
- [23] 张亚男. 基于格子-Boltzmann 方法的流体-颗粒非均匀流动结构及相互作用研究[D]. 哈尔滨: 哈尔滨工业大学, 2016.
ZHANG YN. Study on fluid-particle heterogeneous flow structure and interaction based on lattice-Boltzmann method[D]. Harbin: Harbin Institute of Technology, 2016 (in Chinese).
- [24] 崔建强. 微血管红细胞流动特性的理论与仿真[D]. 济南: 山东大学, 2010.
CUI JQ. Theory and simulation of flow characteristics of microvascular red blood cells[D]. Jinan: Shandong University, 2010 (in Chinese).
- [25] 赫铁男. 生物膜材料的 GPU 并行运算和图像处理[D]. 桂林: 桂林电子科技大学, 2020.
HE YN. GPU parallel operation and image processing of biofilm materials[D]. Guilin: Guilin University of Electronic Technology, 2020 (in Chinese).
- [26] 张乾毅. 生物膜材料的晶格玻尔兹曼方法的研究与 GPU 并行计算[D]. 桂林: 桂林电子科技大学, 2021.
ZHANG QY. Research on lattice Boltzmann method of biofilm materials and GPU parallel computing[D]. Guilin: Guilin University of Electronic Technology, 2021 (in Chinese).
- [27] BENIOUG M, GOLFIER F, OLTÉAN C, BUÈS MA, BAHAR T, CUNY J. An immersed boundary-lattice Boltzmann model for biofilm growth in porous media[J]. *Advances in Water Resources*, 2017, 107: 65-82.
- [28] STRNIŠA F, BAJIĆ M, PANJAN P, PLAZL I, SESAY AM, ŽNIDARŠIČ-PLAZL P. Characterization of an enzymatic packed-bed microreactor: experiments and modeling[J]. *Chemical Engineering Journal*, 2018, 350: 541-550.
- [29] 马苏宁, 陈丽萍, 龚延风, 王世林. 多孔介质中微生物生长与衰亡过程介观分析[J]. *环境工程*, 2022, 40(6): 154-161.
MA SN, CHEN LP, GONG YF, WANG SL. Mesoscopic analysis on microbial growth and decay process in porous media[J]. *Environmental Engineering*, 2022, 40(6): 154-161 (in Chinese).
- [30] 杨勇, 陈丽萍, 龚延风, 金椿明, 王世林. 基于多尺度方法研究微生物生长对多孔介质渗透率的影响[J]. *环境工程*, 2023, 41(4): 49-54, 153.
YANG Y, CHEN LP, GONG YF, JIN CM, WANG SL. Effect of microbial growth on permeability of porous media based on multi-scale method[J]. *Environmental Engineering*, 2023, 41(4): 49-54, 153 (in Chinese).
- [31] 杨艳霞, 李静. 膜生物反应器内生化反应过程的格子 Boltzmann 模拟[J]. *计算物理*, 2018, 35(5): 571-576.
YANG YX, LI J. Lattice Boltzmann simulation of flow and mass transfer in membrane bioreactor[J]. *Chinese Journal of Computational Physics*, 2018, 35(5): 571-576 (in Chinese).
- [32] 张凯. 关于两类流动问题的格子 Boltzmann 方法研究[D]. 银川: 宁夏大学, 2023.
ZHANG K. Study on lattice Boltzmann method for two kinds of flow problems[D]. Yinchuan: Ningxia University, 2023 (in Chinese).
- [33] 韩亚鲁. 不同加热方式土中热湿迁移规律研究[D]. 长春: 吉林大学, 2023.
HAN YL. Study on heat and moisture migration in soil with different heating methods[D]. Changchun: Jilin University, 2023 (in Chinese).
- [34] MONDAL PP, GALODHA A, VERMA VK, SINGH V, SHOW PL, AWASTHI MK, LALL B, ANEES S, POLLMANN K, JAIN R. Review on machine learning-based bioprocess optimization, monitoring, and control systems[J]. *Bioresource Technology*, 2023, 370: 128523.
- [35] 任瑛, 徐骥. 蛋白质体系分子动力学模拟的前沿进展: 从介科学角度重新审视[J]. *过程工程学报*, 2018, 18(6): 1126-1138.
REN Y, XU J. Frontiers of molecular dynamics simulations of protein systems: reexamine from the mesoscience perspective[J]. *The Chinese Journal of Process Engineering*, 2018, 18(6): 1126-1138 (in Chinese).
- [36] YAN WC, LUO ZH, GUO AY. Coupling of CFD with

- PBM for a pilot-plant tubular loop polymerization reactor[J]. *Chemical Engineering Science*, 2011, 66(21): 5148-5163.
- [37] LUO H. Coalescence, breakup and liquid circulation in bubble column reactors[D]. Trondheim: Norwegian Institute of Technology, 1993.
- [38] MISHRA S, KUMAR V, SARKAR J, RATHORE AS. Mixing and mass transfer in production scale mammalian cell culture reactor using coupled CFD-species transport-PBM validation[J]. *Chemical Engineering Science*, 2023, 267: 118323.
- [39] 肖颀, 杨宁. 基于 EMMS 模型的搅拌釜内气液两相流数值模拟[J]. *化工学报*, 2016, 67(7): 2732-2739. XIAO X, YANG N. Numerical simulation of gas-liquid flow in stirred tanks based on EMMS model[J]. *CIESC Journal*, 2016, 67(7): 2732-2739 (in Chinese).
- [40] YANG N, XIAO Q. A mesoscale approach for population balance modeling of bubble size distribution in bubble column reactors[J]. *Chemical Engineering Science*, 2017, 170: 241-250.
- [41] LUO HA, SVENDSEN HF. Theoretical model for drop and bubble breakup in turbulent dispersions[J]. *AIChE Journal*, 1996, 42(5): 1225-1233.
- [42] HAN LC, LUO HA, LIU YJ. A theoretical model for droplet breakup in turbulent dispersions[J]. *Chemical Engineering Science*, 2011, 66(4): 766-776.
- [43] HAN LC, GONG SG, LI YQ, GAO NN, FU J, LUO HA, LIU ZF. Influence of energy spectrum distribution on drop breakage in turbulent flows[J]. *Chemical Engineering Science*, 2014, 117: 55-70.
- [44] HAN LC, GONG SG, DING YW, FU J, GAO NN, LUO HA. Consideration of low viscous droplet breakage in the framework of the wide energy spectrum and the multiple fragments[J]. *AIChE Journal*, 2015, 61(7): 2147-2168.
- [45] SOLSVIK J, TANGEN S, JAKOBSEN HA. On the constitutive equations for fluid particle breakage[J]. *Reviews in Chemical Engineering*, 2013, 29(5): 241-356.
- [46] SHI WB, YANG XG, SOMMERFELD M, YANG J, CAI XY, LI G, ZONG Y. Modelling of mass transfer for gas-liquid two-phase flow in bubble column reactor with a bubble breakage model considering bubble-induced turbulence[J]. *Chemical Engineering Journal*, 2019, 371: 470-485.
- [47] 斯图尔特·罗素, 彼得·诺维格. 人工智能: 现代方法[M]. 张博雅, 等译. 北京: 人民邮电出版社, 2022: 86-934. RUSSELL S, NORVIG P. *Artificial Intelligence: a Modern Approach*[M]. ZHANG BY, et al. Trans. Beijing: Posts & Telecom Press, 2022: 86-934 (in Chinese).
- [48] 宫思吟, 俞晓东, 葛思涵, 陈艳琳, 钟汉斌. 基于 CFD 和机器学习的反应器内瞬时状态预测综述[J]. *信息记录材料*, 2023, 24(11): 13-15. GONG SY, YU XD, GE SH, CHEN YL, ZHONG HB. A review on predicting transient characteristics in reactors based on machine learning and CFD[J]. *Information Recording Materials*, 2023, 24(11): 13-15 (in Chinese).
- [49] 温视中, 韩越夏, 孙剑飞. AI 赋能的生物医学研究[J]. *科学(上海)*, 2024, 76(3): 26-31. WEN YZ, HAN YX, SUN JF. Biomedical research on AI empowerment[J]. *Science (Shanghai)*, 2024, 76(3): 26-31 (in Chinese).
- [50] 李丹丹, 马征远, 张博文, 江源, 冯心宜, 史彤, 毛开云. 人工智能驱动生命科学研究的发展态势[J]. *生命科学*, 2024, 36(1): 122-128. LI DD, MA ZY, ZHANG BW, JIANG Y, FENG XY, SHI T, MAO KY. Development trend of life science research driven by artificial intelligence[J]. *Chinese Bulletin of Life Sciences*, 2024, 36(1): 122-128 (in Chinese).
- [51] 蒋迎迎, 曲戈, 孙周通. 机器学习助力酶定向进化[J]. *生物学杂志*, 2020, 37(4): 1-11. JIANG YY, QU G, SUN ZT. Machine learning-assisted enzyme directed evolution[J]. *Journal of Biology*, 2020, 37(4): 1-11 (in Chinese).
- [52] 任梦媛. 基于生成式对抗网络的酶智能设计[D]. 广州: 华南理工大学, 2023. REN MY. Intelligent design of enzyme based on generative antagonism network[D]. Guangzhou: South China University of Technology, 2023 (in Chinese).
- [53] 潘彦园. 新型生物酶的挖掘算法研究[D]. 成都: 电子科技大学, 2019. PAN YY. Research on mining algorithm of new biological enzyme[D]. Chengdu: University of Electronic Science and Technology of China, 2019 (in Chinese).
- [54] HUANG CM, ZHANG L, TANG T, WANG HJ, JIANG YQ, REN HW, ZHANG YT, FANG JL, ZHANG WH, JIA X, YOU S, QIN B. Application of directed evolution and machine learning to enhance the diastereoselectivity of ketoreductase for dihydrotetrabenazine synthesis[J]. *JACS Au*, 2024, 4(7): 2547-2556.
- [55] ORSI E, SCHADA von BORZYSKOWSKI L, NOACK S, NIKEL PI, LINDNER SN. Automated *in vivo* enzyme engineering accelerates biocatalyst optimization[J]. *Nature Communications*, 2024, 15(1): 3447.
- [56] 池燕飞, 李春, 冯旭东. 机器学习在蛋白质功能预测领域的研究进展[J]. *生物工程学报*, 2023, 39(6): 2141-2157. CHI YF, LI C, FENG XD. Research progress of machine learning in the field of protein function prediction[J]. *Chinese Journal of Biotechnology*, 2023, 39(6): 2141-2157 (in Chinese).
- [57] 钟灵茜. 多功能酶智能预测新方法研究[D]. 无锡: 江南大学, 2023. ZHONG LQ. Study on a new method of intelligent prediction of multifunctional enzymes[D]. Wuxi: Jiangnan University, 2023 (in Chinese).
- [58] 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.

- [59] TANG WL, DENG ZH, ZHOU HW, ZHANG W, HU FP, CHOI KS, WANG ST. MVDINET: a novel multi-level enzyme function predictor with multi-view deep interactive learning[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2024, 21(1): 84-94.
- [60] de CRÉCY-LAGARD V, DIAS R, FRIEDBERG I, YUAN YF, SWAIRJO MA. Limitations of current machine-learning models in predicting enzymatic functions for uncharacterized proteins[J]. bioRxiv, 2024: 2024.07.01.601547.
- [61] 曾涛, 巫瑞波. 数据驱动的酶反应预测与设计[J]. 合成生物学, 2023, 4(3): 535-550.
- ZENG T, WU RB. Data-driven prediction and design for enzymatic reactions[J]. Synthetic Biology Journal, 2023, 4(3): 535-550 (in Chinese).
- [62] LINCOLN M. Rational design and engineering of quinone redox mediators for enzymatic bioelectrocatalysis[D]. Michigan: Michigan State University ProQuest, 2024.
- [63] BARANWAL M, MAGNER A, ELVATI P, SALDINGER J, HERO AO. A deep learning architecture for metabolic pathway prediction[J]. Bioinformatics, 2020, 36: 2547-2553.
- [64] TIAN YS, YANG X, CHEN NH, LI CY, YANG WL. Data-driven interpretable analysis for polysaccharide yield prediction[J]. Environmental Science and Ecotechnology, 2024, 19: 100321.
- [65] KREUTTER D, SCHWALLER P, REYMOND JL. Predicting enzymatic reactions with a molecular transformer[J]. Chemical Science, 2021, 12(25): 8648-8659.
- [66] SHANG LD, WANG Y, CHEN FY, PENG H, BAO XD, TANG XS, LIU KX, XU L, XIAO DY, LIANG P, LI B. Ultra-fast identification of lactic acid bacteria colonies based on droplet microcavity label-free SERS[J]. LWT-Food Science and Technology, 2024, 204: 116435.
- [67] ALIEV TA, LAVRENTEV FV, DYAKONOV AV, DIVEEV DA, SHILOVSKIKH VV, SKORB EV. Electrochemical platform for detecting *Escherichia coli* bacteria using machine learning methods[J]. Biosensors and Bioelectronics, 2024, 259: 116377.
- [68] KWON MS, ADIDJAJA JJ, KIM HU. Predicting the effects of cultivation condition on gene regulation in *Escherichia coli* by using deep learning[J]. Computational and Structural Biotechnology Journal, 2023, 21: 2613-2620.
- [69] 张和平, 高广琦. 基于人工智能的益生乳酸菌精准筛选及产业化关键技术[J]. 中国食品学报, 2023, 23(6): 1-7.
- ZHANG HP, GAO GQ. The key technologies of precise isolation and industrialization of probiotic lactic acid bacteria based on artificial intelligence[J]. Journal of Chinese Institute of Food Science and Technology, 2023, 23(6): 1-7 (in Chinese).
- [70] MIGDADI L, SHARAR N, JAFAR H, TELFAH A, HERGENRÖDER R, WÖHLER C. Machine learning in automated monitoring of metabolic changes accompanying the differentiation of adipose-tissue-derived human mesenchymal stem cells employing 1H-1H TOCSY NMR[J]. Metabolites, 2023, 13(3): 352-370.
- [71] ANDRADE CRUZ I, CHUENHART W, LONG F, SURENDRA KC, RENATA SANTOS ANDRADE L, BILAL M, LIU H, TAVARES FIGUEIREDO R, KHANAL SK, FERNANDO ROMANHOLO FERREIRA L. Application of machine learning in anaerobic digestion: perspectives and challenges[J]. Bioresource Technology, 2022, 345: 126433.
- [72] DU YH, WANG MY, YANG LH, TONG LL, GUO DS, JI XJ. Optimization and scale-up of fermentation processes driven by models[J]. Bioengineering, 2022, 9(9): 473.
- [73] 王冠, 田锡炜, 夏建业, 储炬, 张嗣良, 庄英萍. 大数据-模型混合驱动下生物过程优化与放大的新机遇与挑战[J]. 生物工程学报, 2021, 37(3): 1004-1016.
- WANG G, TIAN XW, XIA JY, CHU J, ZHANG SL, ZHUANG YP. New opportunities and challenges for hybrid data and model driven bioprocess optimization and scale-up[J]. Chinese Journal of Biotechnology, 2021, 37(3): 1004-1016 (in Chinese).
- [74] 夏建业, 刘晶, 庄英萍. 人工智能时代发酵优化与放大技术的机遇与挑战[J]. 生物工程学报, 2022, 38(11): 4180-4199.
- XIA JY, LIU J, ZHUANG YP. Opportunities and challenges for fermentation optimization and scale-up technology in the artificial intelligence era[J]. Chinese Journal of Biotechnology, 2022, 38(11): 4180-4199 (in Chinese).
- [75] HELLECKES LM, HEMMERICH J, WIECHERT W, von LIERES E, GRÜNBERGER A. Machine learning in bioprocess development: from promise to practice[J]. Trends in Biotechnology, 2023, 41(6): 817-835.
- [76] WANG XT, MOHSIN A, SUN YF, LI C, ZHUANG YP, WANG G. From spatial-temporal multiscale modeling to application: bridging the valley of death in industrial biotechnology[J]. Bioengineering, 2023, 10(6): 744-766.
- [77] CHENG Y, BI XY, XU YM, LIU YF, LI JH, DU GC, LV XQ, LIU L. Artificial intelligence technologies in bioprocess: opportunities and challenges[J]. Bioresource Technology, 2023, 369: 128451.
- [78] KONISHI M. High cell density cultivation of *Corynebacterium glutamicum* by deep learning-assisted medium design and the subsequent feeding strategy[J]. Journal of Bioscience and Bioengineering, 2024, 137(5): 396-402.
- [79] GÜRBÜZ T, GÜNAY ME, TAPAN NA. Machine learning solutions for enhanced performance in plant-based microbial fuel cells[J]. International Journal of Hydrogen Energy, 2024, 78: 1060-1069.
- [80] ZHANG LH, ZHANG CW, ZHAO X, HE CL, ZHANG X. Improving lipid production by *Rhodotorula glutinis* for renewable fuel production based on machine learning[J]. Frontiers of Chemical Science and

- Engineering, 2024, 18(5): 51.
- [81] PANDEY AK, NAYAK SC, KIM SH. Functional link hybrid artificial neural network for predicting continuous biohydrogen production in dynamic membrane bioreactor[J]. *Bioresource Technology*, 2024, 397: 130496.
- [82] PINTO J, RAMOS JRC, COSTA RS, ROSSELL S, DUMAS P, OLIVEIRA R. Hybrid deep modeling of a CHO-K1 fed-batch process: combining first-principles with deep neural networks[J]. *Frontiers in Bioengineering and Biotechnology*, 2023, 11: 1237963.
- [83] TAN VWG, CHAN YJ, ARUMUGASAMY SK, LIM JW. Optimizing biogas production from palm oil mill effluent utilizing integrated machine learning and response surface methodology framework[J]. *Journal of Cleaner Production*, 2023, 414: 137575.
- [84] 夏建业, 田锡炜, 刘娟, 庄英萍. 人工智能时代的智能生物制造[J]. *生物加工过程*, 2020, 18(1): 13-20. XIA JY, TIAN XW, LIU J, ZHUANG YP. Intelligent biological manufacturing in the age of artificial intelligence[J]. *Chinese Journal of Bioprocess Engineering*, 2020, 18(1): 13-20 (in Chinese).
- [85] 刘可意, 汪俊卿, 傅凯, 李峰, 王瑞明, 李丕武. 智能生物反应器装备制造进展[J]. *食品与发酵工业*, 2024, 50(15): 409-416. LIU KY, WANG JQ, FU K, LI F, WANG RM, LI PW. Progress in the manufacture of intelligent bioreactor equipment[J]. *Food and Fermentation Industries*, 2024, 50(15): 409-416 (in Chinese).
- [86] 张楠, 程思远, 余永建, 窦帅伟, 刘稼鑫, 唐瑞骏, 朱圆圆, 于振. 封闭式固态发酵体系监测与调控的研究进展[J]. *食品科学*, 2025, 46(1): 274-283. ZHANG N, CHENG SY, YU YJ, DOU SW, LIU JX, TANG RJ, ZHU RR, YU Z. Research progress on monitoring and regulation of closed solid-state fermentation system[J]. *Food Science*, 2025, 46(1): 274-283 (in Chinese).
- [87] RAJASEKHAR N, RADHAKRISHNAN TK, MOHAMED SN. Reinforcement learning based temperature control of a fermentation bioreactor for ethanol production[J]. *Biotechnology and Bioengineering*, 2024, 121(10): 3114-3127.
- [88] 宋利豪. 基于深度学习的溶菌酶发酵工艺优化[D]. 杭州: 杭州电子科技大学, 2023. SONG LH. Optimization of lysozyme fermentation process based on deep learning[D]. Hangzhou: Hangzhou Dianzi University, 2023 (in Chinese).
- [89] SAMMAKNEJAD N, LEE J, AUSTRIA JM, DUENAS N, HEIBA L, REED D, EPHRAIM B, SRIDHARAN G, JUSTICE J, DAVIS J, UNDEY C. Real-time multivariate statistical monitoring of biopharmaceutical processes with no prior product-specific history[J]. *Computers & Chemical Engineering*, 2024, 189: 108788.
- [90] LIANG Y, TIAN JP, HU XJ, HUANG YX, HE KL, XIE LL, YANG HL, HUANG D, ZHOU YF, XIA YY. Rapid determination of starch and alcohol contents in fermented grains by hyperspectral imaging combined with data fusion techniques[J]. *Journal of Food Science*, 2024, 89(6): 3540-3553.
- [91] XU ZX, ZHU XF, MOHSIN A, GUO JF, ZHUANG YP, CHU J, GUO MJ, WANG G. A machine learning-based approach for improving plasmid DNA production in *Escherichia coli* fed-batch fermentations[J]. *Biotechnology Journal*, 2024, 19(6): 2400140.
- [92] PANCHIGAR D, KAR K, SHUKLA S, MATHEW RM, CHADHA U, SELVARAJ SK. Machine learning-based CFD simulations: a review, models, open threats, and future tactics[J]. *Neural Computing and Applications*, 2022, 34(24): 21677-21700.
- [93] ZHOU JH, HUANG ML. Navigating the landscape of enzyme design: from molecular simulations to machine learning[J]. *Chemical Society Reviews*, 2024, 53(16): 8202-8239.
- [94] 马睿. 机器学习辅助定向进化提高奇异变形杆菌脂肪酶的甲醇耐受性[D]. 无锡: 江南大学, 2023. MA R. Machine learning assisted directed evolution improves methanol tolerance of *Proteus mirabilis* lipase[D]. Wuxi: Jiangnan University, 2023 (in Chinese).
- [95] 朱镜璇. 分子动力学模拟结合深度学习探究氨基酸突变或配体结合对酶活性的影响[D]. 长春: 吉林大学, 2021. ZHU JX. Molecular dynamics simulation combined with deep learning to explore the influence of amino acid mutation or ligand binding on enzyme activity[D]. Changchun: Jilin University, 2021 (in Chinese).
- [96] ZUBATYUK R, SMITH JS, NEBGEN BT, TRETIK S, ISAYEV O. Teaching a neural network to attach and detach electrons from molecules[J]. *Nature Communications*, 2021, 12(1): 4870.
- [97] GLIELMO A, HUSIC BE, RODRIGUEZ A, CLEMENTI C, NOÉ F, LAIO A. Unsupervised learning methods for molecular simulation data[J]. *Chemical Reviews*, 2021, 121(16): 9722-9758.
- [98] KÖCHL K, SCHOPPER T, DURMAZ V, PARIGGER L, SINGH A, KRASSNIGG A, CESPUGLI M, WU W, YANG XL, ZHANG YC, WANG WWS, SELLUSKI C, ZHAO TH, ZHANG X, BAI CH, LIN L, HU YX, XIE ZW, ZHANG ZH, YAN J, et al. Optimizing variant-specific therapeutic SARS-CoV-2 decoys using deep-learning-guided molecular dynamics simulations[J]. *Scientific Reports*, 2023, 13(1): 774.
- [99] KAYA U, GOPIREDDY S, URBANETZ N, NOPENS I, VERWAEREN J. Predicting the hydrodynamic properties of a bioreactor: conditional density estimation as a surrogate model for CFD simulations[J]. *Chemical Engineering Research and Design*, 2022, 182: 342-359.
- [100] 韦振宇. 基于深度学习的智能 CFD 模拟研究[D]. 西安: 西安石油大学, 2023. WEI ZY. Research on intelligent CFD simulation based on deep learning[D]. Xi'an: Xi'an Shiyou University, 2023 (in Chinese).