

· 底层技术开发 ·

夏建业 研究员、博士生导师。任中国科学院天津工业生物技术研究所智能生物制造平台实验室主任，兼任中国微生物学会生化过程模型化与控制专委会副主任委员。重点研究方向为发酵过程优化与放大、系统生物技术、过程工程。提出流场和生理结合的发酵过程放大方法，在国内多家企业获得成功应用，取得显著经济效益。先后获得 2011 年国家科技进步二等奖，2019 年中国轻工联合会科技进步二等奖，以第一或通讯作者身份发表 SCI 论文 30 余篇。



人工智能时代发酵优化与放大技术的机遇与挑战

夏建业^{1,2}, 刘晶³, 庄英萍⁴

1 中国科学院天津工业生物技术研究所 智能生物制造平台实验室, 天津 300308

2 国家合成生物技术创新中心, 天津 300308

3 河北工业大学 人工智能与数据科学学院, 天津 300401

4 华东理工大学 生物反应器工程国家重点实验室, 上海 200237

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. 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. Chin J Biotech, 2022, 38(11): 4180-4199.

摘要: 人工智能 (artificial intelligence, AI) 技术正引发一场新的产业革命, 其成功应用正从信息产业迅速渗透到各行各业。传统的发酵工程技术受到巨大挑战的同时更多地迎来了发展变革的机遇。首先, 合成生物技术飞速发展使高性能菌株的可获得性及获取效率显著提升, 对传统低效的发酵优化放大技术提出很大挑战, 亟需对发酵优化放大技术进行升级, 以满足高通量菌种性能验证及工艺开发能力的需求; 其次, 发酵装备技术的持续发展为高效发酵优化技术的进步奠定了良好基础, 加之人工智能技术特别是数字孪生与知识图谱等技术的应用, 将为传统发酵技术的颠覆性发展带来巨大推动力。本文分别从合成生物时代对发酵优化技术的挑战、发酵优化与放大的核心技术、高通量发酵装备技术、数据可视化技术、数字孪生及知识图谱等智能技术在发酵优化放大中的应用等几个方面进行综述, 并对未来工业发酵优化技术的场景以及未来发酵技术对人才培养等提出的新要求进行了展望。

Received: August 2, 2022; **Accepted:** September 30, 2022

Supported by: National Key Research and Development Program of China (2021YFC2101100)

Corresponding author: XIA Jianye. E-mail: xiajy@tib.cas.cn

基金项目: 国家重点研发计划 (2021YFC2101100)

关键词：发酵工程；合成生物学；过程优化与放大；数字孪生；知识图谱

Opportunities and challenges for fermentation optimization and scale-up technology in the artificial intelligence era

XIA Jianye^{1,2}, LIU Jing³, ZHUANG Yingping⁴

1 Smart Biomanufacturing Platform Lab, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

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

3 Artificial Intelligence and Data Science College, Hebei University of Technology, Tianjin 300401, China

4 State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China

Abstract: Artificial intelligence (AI) technology is booming up a new industrial revolution, and its successful application is rapidly spreading from the information industry to many other fields. The traditional fermentation industry also faces more opportunities and great challenges for reforming. First of all, the rapid development of synthetic biotechnology has greatly enhanced the availability and efficiency for obtaining high-performance strains, which poses great opportunities to the traditional fermentation optimization and scale-up technology. It is urgent to upgrade fermentation optimization technology to cope with the requirement for high-throughput verification of strain performance. Secondly, the development of fermentation equipment technology has laid a good foundation for advancing fermentation optimization technology. The application of AI technology, especially the digital twin and knowledge graph technology, will further boost the upgrade of the traditional fermentation technology. This review summarizes the challenges of fermentation optimization technology in the era of synthetic biology, the core technology of fermentation optimization and scale-up, the equipment technology of high-throughput fermentation, data visualization technology, as well as the application of digital twin and knowledge graph in fermentation optimization and scale-up. This review also prospects future industrial fermentation technology, and the associated new requirements for personnel training.

Keywords: fermentation engineering; synthetic biology; process optimization and scale-up; digital twin; knowledge graph

生物制造是利用生物体机能进行物质加工与合成的绿色生产方式，有望在能源、化工和医药等领域改变世界工业制造格局。而发酵优化与放大技术是生物制造技术中决定实验室创新成果向产业化转化能否成功的关键一环。2015年国务院印发《中国制造2025》^[1]以制造业向智能化方向升级为重要任务，2021年底

工业和信息化部等八部委联合发布《“十四五”智能制造发展规划》^[2]明确提出“两步走”：即到2025年，规模以上制造业企业大部分实现数字化网络化，重点行业骨干企业初步应用智能化；到2035年，规模以上制造业企业全面普及数字化网络化，重点行业骨干企业基本实现智能化。2022年5月份国家发展改革委发布

《“十四五”生物经济发展规划》^[3]，明确将生物制造作为生物经济战略性新兴产业发展方向。因此，借助智能制造技术，大力提升生物制造过程中关键核心技术，尤其是严重依赖专家经验、效率低的发酵过程优化与放大技术升级，对我国加快构建绿色低碳循环经济体系，推动生物经济实现高质量发展具有十分重要的意义。

近年来，随着合成生物学、多组学分析技术、高通量筛选技术等快速发展，工业发酵过程高性能菌种的获取效率显著提升。以合成生物学为技术源头的创业公司受到风险资本的广泛青睐，以美国 Amyris、Zymergen 等公司为代表。然而高性能菌种的开发和菌种筛选技术的突破也同时对下游发酵过程优化及放大的效率和通量等提出极大挑战。长期以来，发酵过程优化利用化学工程技术手段，结合菌种生理代谢特性等开展研究，更多依赖发酵工程师的经验或采用试错法进行工艺开发，缺乏理性高效的技术手段。实验室开发的小试发酵优化工艺在进行工业化生产放大时也面临诸多挑战，以至于发酵过程放大一度被称为一门艺术而非技术^[4]。近些年来，我国发酵优化与控制技术取得了长足进步和发展，基于多参数相关分析的发酵优化放大方法、过程控制优化软件应用等，对支撑我国成为国际发酵大国起到了非常重要的作用。即使如此，发酵过程优化放大中还有很多未解决的关键技术难题，如何解决这些技术难题将是实现合成生物技术推动下生物经济得以蓬勃发展的重要课题。

2016 年 AlphaGo 击败世界顶级棋手李世石的消息传遍世界，以深度学习为代表的人工智能技术展现出非凡的“智能”水平。如今已经实现利用氨基酸序列预测蛋白结构的人工智能^[5]，人工智能药物分子设计等也成为现实^[6]。

人工智能技术正在颠覆传统科学研究的范式，其借助数据科学、大数据、人工智能算法等实现了一个个新的技术途径。发酵过程优化与放大过程研究的对象为发酵过程，包括生物反应器内的流场环境以及其中培养的微生物细胞的各种代谢特性，这些环境因素和细胞代谢之间具有复杂的相互作用关系^[7]，利用统计学方法结合反复试错及发酵工程师经验，是传统发酵优化放大中常用的开发方式。如何借助人工智能技术，结合发酵过程自身特点，发展与之适配的人工智能技术，大幅提升发酵优化放大效率，加速实验室成果向产业化转化的进程，是当今发酵工程技术与与时俱进、实现技术升级的全新路径。

本文将从合成生物学飞速发展对发酵技术的挑战出发，以高效发酵优化的装备技术、过程数字化及可视化技术、智能化技术的应用等几个方面综述人工智能时代的发酵优化与放大技术所面临的机遇和挑战，最后对未来的发酵优化放大技术在新发展模式、人才需求等方面进行了展望。

1 合成生物学大背景下的发酵技术

1.1 传统发酵优化与放大技术

工业发酵优化与放大技术最早可以追溯到 20 世纪 40 年代，在利用搅拌通气反应器实现青霉素大规模生产过程中逐渐形成。发酵工程还非独立学科，青霉素的工业化放大生产在生物工程与化学工程专家的共同努力下得以实现，主要借鉴化学工程中“三传一反”理论指导。同时，以抗生素发酵为代表，利用工程化思想进行工艺开发工作的相关学科也逐渐得以发展，我国华东理工大学 1955 年建立的抗生素制造工程专业正是一个例证。著名抗生素专家马誉澂教授编著的《抗生素》^[8]于 1955 年底由人

民卫生出版社出版,此书是国内最早一部系统介绍抗生素研究、制备、检测等知识的专著,成为当时抗生素制造工学专业的必学教学用书。自此,发展了一系列针对次级代谢产物发酵过程优化与放大的方法和理论体系。

另一方面,传统轻工业如醋、酒、酱等的酿造过程,以及大宗发酵产品如氨基酸、有机酸、核酸等产品的发酵产业在我国历史悠久。基于轻工业发展起来的发酵学科则更早些(1952年前后),包括江南大学、华南理工大学、北京工商大学等。在此基础上形成了传统的发酵优化方法,主要包括菌种选育、培养基及培养条件的优化、反应器结构优化等。传统菌种选育包括自然选育、人工诱变选育等在高性能菌种筛选中发挥了重要作用,也为基因工程菌的驯化以适应工业生产环境提供了重要技术基础。早期培养基优化基于单因素试验展开,后期形成了基于统计的优化方法包括单因素设计、均匀设计、实验设计(design of experiment, DoE)等,已经形成相对成熟的培养基优化方法体系。在发酵过程优化方法的研究中,主要集中在最适发酵温度、接种浓度、最适 pH、最适 C/N 比等静态条件的优化^[9-11],这些方法在氨基酸、有机酸、核酸发酵优化、抗生素及次级代谢产物发酵优化等方面获得了广泛应用,并在推动我国发酵产业技术进步方面发挥了重要作用。此外在反应器结构与形式方面也进行了诸多研究,包括不同搅拌桨及通气形式的改进、适用于剪切敏感的丝状菌发酵的气升式反应器开发与应用等,对提升发酵产率和发酵效率起到重要作用。

1.2 合成生物学时代发酵优化技术的机遇与挑战

基因工程、代谢工程以及合成生物技术的飞速发展,给发酵优化提供了前所未有的机遇,

通过插入外源基因实现外源蛋白的表达^[12],构建全新的外源途径表达异源产物^[13],表达 mRNA 疫苗^[14]等大幅拓展了发酵过程研究的对象,但由于外源基因引入会导致宿主细胞代谢特性改变,同一宿主表达不同产物,采用不同策略构建的菌株在发酵过程中表现各异。因此这些新技术的发展为传统发酵行业带来机遇的同时也对其提出了诸多挑战。

近 10 年来合成生物学发展推动菌株构建及高通量自动化筛选技术取得显著发展,使得高性能菌株得以更快获取。然而传统基于实验室规模反应器的发酵工艺开发过程,需要大量人力,开发过程费时耗力,显然无法满足如此大量菌株在性能验证和工艺开发上的需求,随之发展起来的微流控技术在解决高通量、自动化培养方面取得了一定进展,尤其是在高性能菌种的高通量筛选方面发挥了重要作用。然而,微通道反应器内形成的流场环境与工业环境具有较大差别,虽然提高了筛选通量但在发酵过程的放大中还存在一定局限,因此开发高通量、自动化的微型平行反应器,特别是能够准确反映工业生产环境的反应器,成为当前合成生物技术发展对发酵优化装备提出的一个全新挑战^[15]。除此之外,发酵过程是一个复杂的动态过程,过程中需要进行大量在线参数检测,因此高通量发酵工艺优化设备同时带来海量过程数据在存储、可视化、分析等方面的挑战。需要将数据科学引入发酵过程优化研究,利用数据科学理论和工具,对高通量工艺开发过程中形成的海量数据进行处理,这方面可以借鉴高通量筛选数据分析的相关技术。已有数据科学软件包括基于 Python 语言的 scikit-learn 包^[16]、pandas 包^[17]及 Numpy 包^[18],开源 KNIME 软件包^[19]等。

另一方面,代谢工程与合成生物学在构建

高性能菌种过程中,会在底盘宿主菌基础上整合一系列外源基因,或对宿主菌自身基因进行改造而形成大量菌株。了解宿主细胞野生菌在发酵罐中的生理代谢特性,如最佳 pH、最佳温度、生长速率、营养物需求、过量底物和氧气限制下的代谢反应等,对高通量筛选模型选择是非常重要的信息,需要详细研究^[20]。此外,不同宿主菌或不同改造策略形成的大量菌株在反应器中所表现的特有代谢特性,对于更好地改造菌株具有非常重要的参考价值。菌种改造过程中,科研人员可以参考大量的基因组、转录组、蛋白组数据库等,然而对于菌种在发酵过程中代谢特性方面的数据库却严重缺失。这就提出了一个新的挑战,即不同宿主菌以及同一宿主菌不同改造靶点或不同改造策略所形成的菌株,在生物反应器中所表现出的代谢特性数据库的构建,类似研究工作目前还未见报道。这一方面依赖于高通量的发酵设备,另一方面依赖高效的数据科学处理工具,随着以上两方面的发展此类数据库将为后续菌种改造提供更丰富的数据支撑。

此外,如何将发酵过程放大中碰到的问题在构造菌种、选择表达体系之前进行考虑,是目前菌种构造中所欠缺的。比如生产规模反应器中存在的底物浓度分布不均、最大传氧能力限制、诱导剂大规模使用成本、大吨位发酵罐内溶解 CO₂ 浓度过高对细胞活性的负面影响等,都应该在菌种构建和高通量筛选中进行考虑。然而,进行菌种改造等上游生物技术开发,科研人员往往缺乏这方面的认识或重视不足,在此基础上构建的菌株对后续发酵过程优化形成难以避免的瓶颈,因此加强菌种构建科研人员与发酵工程科研人员之间的沟通非常重要。

1.3 发酵过程优化与放大中的核心技术

合成生物技术的飞速发展给发酵工程研究

带来诸多挑战,我们需要重新梳理、认识发酵过程优化与放大中的核心技术。发酵过程的优化与放大技术是相辅相成的。如果不考虑反应器规模对发酵的影响而进行发酵工艺的开发与优化,很有可能导致优化的工艺在放大过程中因放大效应而难以应用。早在 20 世纪 80 年代就提出了基于工业规模反应器来缩小并再现发酵放大的限制条件,并在此基础上实现工艺优化,从而避免工艺开发过程与实际生产环境不一致导致的放大失败。

发酵过程优化的核心是确定操作条件与菌体代谢特性之间的关系,并利用这个关系调控操作条件使菌体的生理代谢特性朝向有利于产物合成的方向,从而达到优化目标。如图 1 所示,反应器中进行的发酵过程涉及人为操作的操作参数 (operation variables)、反应器层面的状态参数 (state variables) 以及细胞层面的生理特性参数 (physiological variables)。然而由于发酵过程中存在复杂的物理化学过程,很难准确地获取这些参数之间的定量关系。目前可以通过不同的方法和策略来达到此目的,以便进行微生物生理代谢特性的调控实现目标产物高产,包括发酵过程多参数相关分析方法^[22],以及基于模型定量描述的方法^[23]。

发酵过程多参数相关分析的理论和方法由华东理工大学张嗣良教授团队提出^[24-25],指出发酵过程中各参数之间存在耦合相关性,并可归结为理化相关与生物相关。发酵过程理化相关主要涉及纯粹的物质理化性质变化所引起的参数相关。这些理化过程包括物质或能量传递、混合、离子解离平衡等导致的相关。例如搅拌转速增加会提高传氧速率,相应的溶解氧 (dissolved oxygen, DO) 水平上升;罐压增加导致溶解 CO₂ 浓度增加,使碳酸解离朝向 H⁺ 浓度增加方向,从而使 pH 值下降。发酵过程的生

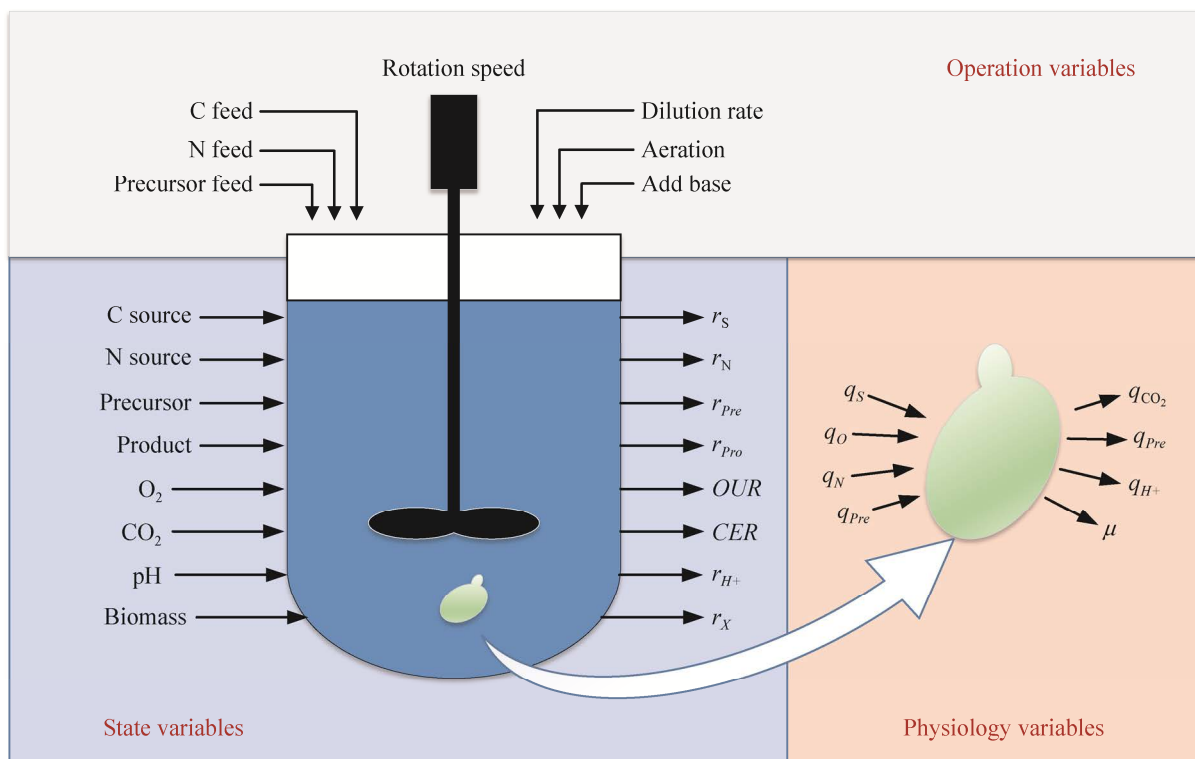


图1 发酵过程中操作参数、状态参数及菌体生理特性参数之间的关系示意图^[21]

Figure 1 Diagram showing the relationship among operating variables, state variables and microorganism physiological variables during a fermentation process^[21].

物相关是指由细胞生命活动所引起的参数之间的耦合相关。比如发酵前期细胞快速生长，耗氧速率指数上升导致 DO 水平快速降低，后期发酵液粘度降低导致传氧系数增加而引起 DO 水平上升。最典型的就是细胞氧气消耗速率 (oxygen uptake rate, OUR) 与溶氧的基本相关特征：一般情况下，OUR 与 DO 曲线的变化呈相反关系，但当 DO 降低到临界值以下时，增加供氧 OUR 将与 DO 呈现相同变化趋势。通过 OUR 与 DO 曲线变化关系可以一定程度上确定微生物代谢的临界氧水平，从而在工艺开发与优化过程中避免临界氧以下状态。发酵过程优化中这些物化及生化过程相互交织、错综复杂，基于发酵过程参数相关分析的方法在多个工业化发酵产品的优化过程中发挥重要作用^[26-28]。

工业发酵优化过程通常采用限制性营养物

质供给速率来调节细胞生长与产物合成间的平衡，尤其对次级代谢产物合成过程尤其重要。细胞在底物限制性 (多为碳源限制) 条件下，胞内代谢网络中各流进流出比速率之间存在相互依赖关系，即通量耦合。此条件下，吸收底物主要用于合成菌体、合成产物以及菌体维持。要想确定这些流量流入、流出间的定量关系，需要建立过程模型进行描述。例如，Tang 等^[29]利用多个不同稀释率和稀释率线性降低下代谢物浓度数据，建立了产黄青霉 9-pool 代谢模型，以便在动态环境中理解、预测和优化细胞工厂的代谢特性与行为。Bayer 等^[23]开发了可用于描述发酵过程的混合模型，将神经网络模型与基于第一性原理的动力学模型整合，用于表征发酵过程动态特性。

发酵过程放大得以成功的关键技术点与发

酵过程优化是一致的,即如何确定反应器内流场环境与细胞代谢特性之间的关系。传统的基于单一放大准则的发酵过程放大在实际应用中碰到各种问题,其中主要体现在供氧能力与混合不均匀导致的放大效应^[30]。理论分析证明,放大过程中无法同时保证3个及以上参数一致^[31]。因此,更多的研究关注于反应器流场和细胞生理代谢特性之间的整合研究。最具代表性的为Enfors等^[32]研究的22 m³工业规模反应器中进行的大肠杆菌补料批培养过程,主要分析了大肠杆菌对非均匀底物浓度场的代谢响应。

通过借助计算流体力学(computational fluid dynamics, CFD)模拟可以实现大型生物反应器内复杂流场特性的描述与统计,这成为联系宏观生物反应器放大设计与局部微观细胞生长环境变化和细胞生理代谢变化的重要工具。在工业实践中,由于各类营养物浓度波动会引起工业过程生产效率的改变,跟踪大型生物反应器中细胞随流体运动路径是评价生产罐发酵性能的非常有价值的工具。工业规模生物反应器中细胞的“生命线”或“运动轨迹”可通过Euler-Lagrange CFD模拟获得,这种方法可以进一步与代谢模型相结合,不仅可以提供由环境波动引发的细胞随运动轨迹上的统计分析,还可以提供对工业生物反应器内异质性代谢反应的全局评估。例如,Haringa等^[33-35]研究了产黄青霉和酿酒酵母的工业规模发酵,旨在量化底物浓度梯度及其影响。此外,将反应器的流场特性与发酵过程的生理特性参数分析相结合,也有助于及时发现反应器几何结构或操作层面的问题,进而调整设备结构或操作工艺。例如,在头孢菌素C的工业发酵过程中,通过多参数相关分析发现,180 m³发酵罐终产物浓度低,且发酵过程的呼吸商(respiratory quotient, RQ)维持在0.9-1.0,而在50 L反应器

中终产物浓度高,RQ维持在理论值0.7左右。结合反应器的计算流体力学模拟结果发现,在180 m³反应器中,原搅拌系统仅配备径向流平叶桨,导致发酵后期从罐顶补加的豆油无法实现全釜范围内混匀,所以豆油利用效率低(RQ偏高)。Yang等^[36]通过替换两层轴流桨最终实现豆油有效利用,过程RQ维持在0.7,头孢菌素C产量提高10%以上,同时能耗降低25%。

2 面向高效发酵过程优化的装备技术

2.1 发酵过程先进传感技术的开发及应用

传感技术是发酵过程优化中必不可少的一项关键技术。常规反应器中配置的pH、DO、温度、压力等传感器是实现生物反应器中稳定pH、供氧水平及恒定温度必不可少的传感器,也是发酵工程师了解反应器中生化反应情况的重要手段。传感器就像为反应器装上了“鼻子”“眼睛”“耳朵”,使得发酵工程师可以实时“闻到”“看到”“听到”反应器中的实际状况,据此判断生物反应过程是否正常,而无需通过取样、离线检测后再进行判断和调整控制策略。

经过多年的发展和实践,发酵优化领域中多种先进传感技术得以应用,其中影响最大、效果最明显的是发酵尾气分析技术。包括早期的发酵尾气分析仪,和近些年发展的在线尾气质谱技术。发酵尾气分析仪与尾气质谱最主要差别在于检测精度,对常规发酵(OUR>10 mmol/(L·h)以上),尤其是高密度发酵,尾气分析仪的检测精度足以满足工艺开发需求;然而对于呼吸较弱的发酵过程(OUR<10 mmol/(L·h)),如兼性厌氧菌的微好氧过程、微好氧的动物细胞培养过程等则需要应用检测精度更高的尾气质谱仪。尾气分析主要分析尾气中CO₂和O₂的百分含量,并通过与进气中这两种气体成分含量对比分析来计算反

反应器中微生物的呼吸代谢状况,从而判断代谢是否正常。如安琪酵母利用尾气分析计算呼吸商 RQ 值来判断补料是否过量,控制 RQ 值在 1 附近^[37],从而避免乙醇积累导致的菌体得率下降,基于此将酵母发酵水平提升 30%以上,为企业赢得大量利润;Zou 等^[26]结合发酵尾气分析计算的 OUR 对比 50 L 与 372 m³ 发酵罐,从而成功将 50 L 发酵罐工艺放大到 372 m³ 罐,为企业创造显著经济效益,国际市场占有率从 3%提升到 30%;嘉必优生物技术(武汉)股份有限公司^[27]在进行花生四烯酸(arachidonic acid, ARA)发酵过程放大中,应用发酵尾气分析获取过程 RQ 值,并通过不同氮源补加策略控制 RQ 值,将发酵工艺直接放大到 200 m³ 发酵罐,发酵效价从 11.93 g/L 提升到 16.82 g/L,成本降低 11.2%。

近年发展起来的以红外为代表的在线光谱检测技术在发酵优化中也获得足够重视。通过光谱获得的海量光谱数据,结合离线测定指标,如产物浓度、底物浓度、菌体浓度等,采用偏最小二乘法建模技术建立谱图与这些离线指标的多元线性模型,从而利用在线实时光谱计算得到各种离线指标的值,从而避免离线检测导致的控制滞后,提升发酵优化效率。Lopez 等^[38]采用红外光谱检测以木质纤维素为碳源的乙醇

发酵过程的残糖浓度,并据此构建发酵过程动力学模型,建立了过程的数字影子系统,实现智能化补料策略。这种利用光谱数据进行在线检测数据的建模思路,同样适用于在线拉曼光谱系统。这类系统的缺点是:进行实际模型使用之前需要收集大量的过程数据,并且建立的模型其外推性很差。即这类检测技术比较适合于工艺相对稳定,用于进行质量控制的场合,不太适用于发酵培养基配方优化过程、成分变化大的情况。

除此之外,近年来发展起来的胞内代谢物实时荧光检测技术也日趋发展成熟,由华东理工大学杨弋教授团队开发^[39]的分子探针可实现 NADH、NADPH 及各种胞内氨基酸分子含量的实时定量测定(表 1 列出了目前已发表的能用于检测的胞内代谢物及相应探针)。将此技术与发酵过程在线传感技术结合,可大幅提升发酵过程在线检测参数的广度和深度,从而更快、更直接地获取微生物胞内代谢状态,使发酵过程更快、更精准地得到预测和控制。

2.2 高通量平行反应器装备开发与应用

高效发酵优化对装备方面的另一个瓶颈是如何提高发酵实验的通量。能再现工业规模反应器内环境的微型化平行反应器装备是实现发

表 1 胞内代谢物实时荧光检测技术及可测定的胞内代谢物列表




Table 1 Real time intracellular metabolite fluorescence detecting techniques and list of available sensors

No.	Intracellular metabolites	Fluorescence probes	References
1	NADH	Frex	[40]
2	NAD ⁺ /NADH	SoNar	[41-42]
3	NADPH	iNap	[43]
4	Redox landscapes	SoNar, iNaps, HyPerRed, rxRFP	[44]
5	Histidine	FHisJ	[45]
6	Inorganic mercury	G-GEIM/R-GEIM	[46]
7	Glucose	FGBP	[47]
8	Redox state	roUnaG	[48]
9	NAD ⁺	FiNad	[49]

酵实验通量提高的关键。近年来围绕这一目标国内外科研工作者、相关装备企业在此方面进行了诸多尝试，也开发了不同类型的微型平行反应器（详细列于表2中）。

表2 不同研发机构或装备公司所开发的微型平行反应器系统对比

Table 2 Comparison of parallel micro-bioreactor systems developed by different research institute or instrument companies

Parallel micro-bioreactor/Creator	Properties	Equipment figures	References
DASGIP/Eppendorf	<p>Pros</p> <ul style="list-style-type: none"> -Modularized easy to extent -Support ORP/OD/Off-gas measurement -Announce real parallel -Up to 8 feed lines per reactor -Support gas mixing -Support DoE <p>Cons</p> <ul style="list-style-type: none"> -Too compacted, hard to mounting -Complicated for assembly & operation -Expensive investment 		[50]
Amber 15/ Sartorius	<p>Pros</p> <ul style="list-style-type: none"> -Single-use bioreactor -Pre sterilized -Up to 48 parallel bioreactors -Possess online OD, pH, DO -Support DoE -Volume as small as 15 mL <p>Cons</p> <ul style="list-style-type: none"> -Very expensive -Large space is needed 		[51]
Amber 250/ Sartorius	<p>Pros</p> <ul style="list-style-type: none"> -Single-use bioreactor -Pre sterilized -Up to 24 parallel bioreactors -Possess online OD, pH, DO -Support DoE -Support PAT -Highly automatic system <p>Cons</p> <ul style="list-style-type: none"> -Very expensive -High cost of consumables -Need more space 		[52]

(待续)

(续表 1)



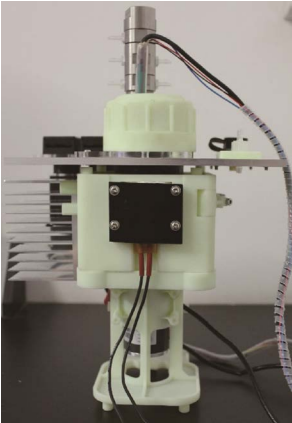

Parallel micro-bioreactor/Creator	Properties	Equipment figures	References
Cloud bioreactor/ Culture bioscience	<p>Pros</p> <ul style="list-style-type: none"> -Computer engineering idea based technology -Remote control -Easy data visualization and analysis -Culture is carried out by service supplier -Up to over 100 parallel bioreactors <p>Cons</p> <ul style="list-style-type: none"> -Strain transfer to the service supplier -Data saved in cloud -Data safety issue 		No publication
Parallel miniature bioreactor/ECUST	<p>Pros</p> <ul style="list-style-type: none"> -Volume limited 80 mL -Pressure controlled -Non-invasive DO and pH -Well parallel performance <p>Cons</p> <ul style="list-style-type: none"> -Need radiation sterilize -Lack data visualization software -Need large mounting space 		[53]
Mini parallel/TIB	<p>Pros</p> <ul style="list-style-type: none"> -High quality design with CFD optimization -More monitoring parameters -Easy to be combined to form different operation mode -Versatile software with several control models -Self designed high quality DO probes <p>Cons</p> <ul style="list-style-type: none"> -Under development status -Need more serious test 		No publication
CloudReady/TJX	<p>Pros</p> <ul style="list-style-type: none"> -Easy to mount ability -Can be easily extended -Video camera is mounted -Announced to be clouded -Powerful software <p>Cons</p> <ul style="list-style-type: none"> -Not stable enough -Parallel quality need to be improved -Temperature control system need to be improved 		No publication

表 2 对国内外不同研发机构及公司所开发的微型平行反应器系统进行了对比, 列举了各

系统的优缺点。从近年来平行生物反应器技术的发展来看, 为解决合成生物学时代在反应器

应用方面的瓶颈, 生物反应器特别是用于工艺开发与工艺优化阶段的生物反应器, 正在朝着微型化、自动化、集群化的方向发展。除了在反应器制造工艺方面提出罐间平行性的要求之外, 对反应器在线检测数据的全面性方面都提出更高的要求。借助人工智能技术, 利用微型反应器集群形成的海量发酵过程数据, 结合实验设计理论, 可更高效地实现发酵菌种验证、发酵工艺开发等工作, 大幅缩短合成生物学创新菌种从实验室走向工业化的时间。

2.3 发酵过程装备及数据处理自动化技术

合成生物学的发展催生了一系列高通量筛选装备及其自动化的相关技术, 目前从平板制备、单克隆挑选、孔板自动化培养及孔板的自动检测等已形成一系列装备和技术, 使得菌种构建后筛选效率大幅提升, 据称利用高通量自动化装备 Biofoundry 平台, 可实现每天千万株筛选水平, 全世界已形成该自动化装备系统的 Biofoundry 联盟已经达到 32 家, 我国共有 5 家。

通过 Biofoundry 高效筛选获得的高性能菌种往往在孔板中培养获得, 由于孔板培养条件受到氧传递速率、混匀速率等方面的限制, 经上述筛选得到的高产菌株在工业生产环境中是否仍能保持其高产性能需要进一步验证。然而与之配套的自动化发酵培养装备的开发却相对落后。近些年围绕这一问题学术界和工业界都在进行积极尝试, 取得一定成果, 将菌种性能验证及发酵工艺优化效率提升一个等级。

其中, 德国相关技术发展处于前列。包括原 m2p lab 公司开发的 Biolector Pro 可实现孔板发酵的补料培养, 既突破了批培养的限制, 还通过配合在线 pH 检测实现恒定 pH 的控制。进一步与液体工作站配合, 建立的 RobLector 已经有通过高通量筛选指导菌种构建案例的报道^[54]。该装备目前已被丹纳赫集团收购, 并应用于高效发酵

工艺优化与菌种性能自动化验证。目前售价在 400 万人民币以上, 高昂的售价使得该装备在发酵工艺开发中的推广受到很大限制, 随着该装备的宣传和应用, 可能会有所改变。

此方面 Sartorius 公司技术也处于前沿, 其开发的 Amber 15 及 Amber 250 系统使用一次性反应器, 搭配非接触式 pH 和 DO 传感器, 利用传统搅拌通气操作方式, 使实验结果更贴近实际工业环境。Amber 250 系统通过使用一次性反应器减轻了发酵工艺人员的工作强度, 软件系统可以实现发酵液的无菌注入, 并采用 DoE 技术提升了培养基配制与优化效率; 此外配置的在线尾气分析系统及对接在线拉曼技术的 PAT 模块, 将质量源于设计 (quality by design, QbD) 理念做到了极致。该系统最多可整合 24 个反应器, 实现发酵过程在线检测及自动化取样分析等, 极大提高了发酵工艺优化的效率。与 RobLector 系统同样的问题, 该系统由于售价超过 1 200 万人民币, 在发酵行业中的推广举步维艰, 在动物细胞培养企业得到一定应用。此外高昂的一次性反应器耗材价格也是该设备在发酵行业获得广泛应用的一个阻碍。因此, 如何实现成本较低的自动化发酵装备将是发酵工业领域亟需解决的一个难题。

在此方面, 美国 Culture Bioscience 公司提出的云端反应器的概念及自动化的装备可能是一条变通的路径, 该公司也因其概念的先进性及实际可操作性获得多家合成生物学初创公司的青睐。2021 年 12 月获得 B 轮融资 8 000 万美元, 这可能是未来发酵过程自动化发展的一个重要方向, 然而对于传统发酵行业长期形成的恶性竞争、菌种知识产权保护困境等都对该方式中菌种安全性提出很大挑战。在国内未形成健全的菌种知识产权保护体制情况下, 该方式的技术推广也存在难度。

除了在反应器培养过程自动化装备技术方面,发酵过程自动化还包括自动化配料装置、自动取样分析装置等的开发,这方面国内天木生物科技有限公司走在前列。其开发的在线自动取样装置在国内多家科研院所、生产企业获得成功应用。此外,随着高通量平行发酵罐系统的不断发展与 DoE 技术配套的发醇培养基自动配制装备也是非常重要的一项装备技术,以避免由于反应器数量大引起的人为操作错误而导致错误实验结论。

同时也要认识到,传感技术、高通量平行反应器及发醇过程自动化装备技术的飞速发展,正推动发醇过程朝向生物制造的大数据时代发展,数据量的激增在提升人们对发醇过程认识的水平,也为高效数据处理提出了新的挑战,仅依靠发醇工程师人工分析远不能满足要求,因此配套的数据可视化以及基于人工智能技术的发醇过程自动化数据分析方法的发展,将是大势所趋。

3 发醇过程网络化数据采集及数据可视化技术

3.1 物联网技术及其在发醇过程监测中的应用

物联网是实现万物互联,远程监测与控制的范式。物联网主要依托无线射频识别技术 (radio frequency identification, RFID),底层依托传感器技术,提供度量、理解物理世界的能力。在万物互联的背景下,物联网快速发展,从简单的传感器、RFID 标签的感知与读取,到传输层与应用层全面发展。感知层是物联网的基础和核心,包括基础传感器、RFID、二维码、摄像头等。传输层负责传递感知层采集的数据,包括 MODBUS、CAN 总线技术、LoRa、NB-IoT、Wifi、5G 等无线通讯技术。应用层指基于物联网技术提供丰富

的物联网应用,在智能家居、智能电网、环境监测、无人驾驶等领域已经取得长足发展。

在微生物发醇过程中,发醇培养环境的监控和动力学参数寻优是常用的优化和控制方法^[55],因此发醇过程参数的实时感知具有重要意义。随着物联网技术中感知层的不断扩大与传输层能力不断增强,发醇过程海量数据的监测成为可能。Cañete-Carmona 等^[56]使用 Arduino 单片机与 CO₂ 传感器,为实时监测酿酒发醇提供了一种低成本解决方案。Angelkov 等^[57]在使用传感器监测温度、pH 值、酒精浓度、CO₂ 浓度基础上,加入以太网控制器实现数据无线传输。物联网技术在这些传统发醇行业中的作用正日益彰显,随着相关技术进步,更多发醇工艺开发及优化过程中将用到智能传感、智能手机应用等物联网技术,从而减少批次间误差、提高产品生产质量等^[58]。

3.2 多源异质传感数据的对接与整合

现场总线技术是工业自控现场 (比如发醇企业中的中控室) 底层互联的通信网络,如工业以太网控制自动化技术 (ether control automatic technology, EtherCAT)、控制器局域网总线 (controller area network, CAN)、过程现场总线 (process field bus, Profibus)、Modbus 等,其中 Modbus 协议较为常见,该协议基于主/从应答式架构,主机发送数据请求与数据更改命令,从机收到请求并响应,实现设备双向通信。发醇企业的控制网络中存在大量不同厂商、不同驱动、不同特性的硬件设备 (比如发醇罐系统、各种在线传感器、过程检测设备),每类设备相互隔离,无法交互且难以集成,缺乏统一的接口标准。用于过程控制的 OLE (OLE for process control, OPC) 标准正是针对该问题设计的工业通讯标准,基于组建对象模型 (component object model, COM)/分布式组建

对象模型 (distributed component object model, DCOM) 技术开发, 将设备驱动与通信程序封装为 OPC 服务器, 屏蔽底层物理设备的差异, 整合了 Modbus 等通讯协议, 统一对应用客户端提供数据交互。

近年来计算机通信技术迅速发展, 各项协议标准不断更新迭代, 工业控制网络的软硬件也随之不断完善与丰富, 同时也暴露出一些问题: (1) 早期现场设备所遵循的协议和标准过于陈旧; (2) 设备供应商提供软件包而不对外开放接口。这就为多协议、多传感器数据融合带来困难, 因此数据融合应从以下 3 点着手实施:

(1) 对于早期设备 (如 Modbus 与 OPC 为例), 需要基于 COM 技术等, 点对点式对接开放标准; (2) 对于封闭式软件包, 需要在数据库层级进行对接整合; (3) 对于暴露接口型设备, 可以对接其通信接口进行整合。如图 2 所示需要构造一个中间件以进行多协议数据的整合, 并提供上行接口, 且所提供的上行接口满足统一标准规范, 从而保证无论数据来自于哪种设备或

协议, 数据都可以通过统一采集指令进行上传, 同时采用统一指令标准下发给设备。

3.3 发酵过程数据清洗及前处理方法

发酵过程数据可能存在缺失值, 而大部分基于数据的智能化模型构建要求数据完整性, 因此在进行模型构建前, 需要针对过程缺失值进行相应处理。发酵过程可能存在以下几类缺失值: (1) 在线采集数据过程中, 经常会遇到一个或多个数据值丢失的情况。这主要由于传感器故障、存储介质故障、冗余等原因所导致。(2) 离线数据采集过程中也可能有缺失。这主要是由于取样人员遗忘, 或因设备故障无法进行测试导致。(3) 相对于在线数据的数据长度来说, 离线数据稀疏很多, 在进行两类数据融合建模时, 需要对离线数据的相应缺失值进行处理。

处理数据缺失值一般有删除和填充两种方法。(1) 删除方法, 将缺失值变量直接删除, 得到一条完整的数据。这种方法仅适用于缺失数据量较少情况下, 当缺失值占比较大时, 可能会使数据发生偏离, 导致模型性能下降。(2) 填

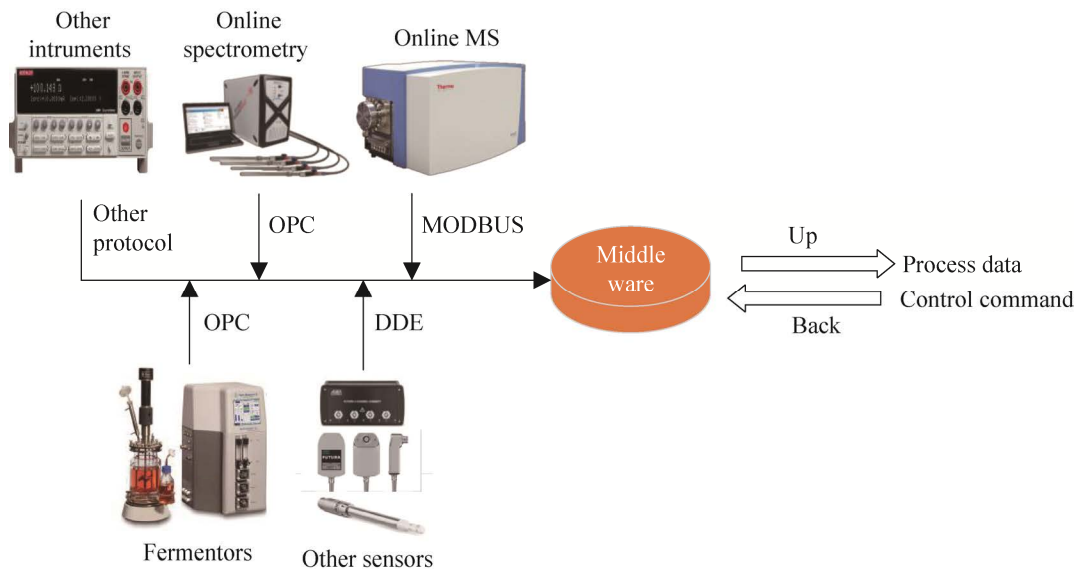


图 2 多源异质数据融合对接技术框架

Figure 2 Framework of heterogeneous multi-source data fusion and docking technology.

充方法,利用一定的值填充缺失值变量,从而使数据完整化。常用的填充补齐方法有均值填充、热卡填充和期望值最大化。

发酵过程数据为多维矢量数据,对多维矢量数据进行特征分析、建模处理时涉及到数据降维。降维主要是降低矢量数据的维数,使数据集中的指标数据量变少,并且保留原数据中指标的大部分有效信息。主要有两种方法,分别为线性判别分析^[59]和主成分分析^[60]。线性判别分析是一种有监督学习的线性降维技术,通过确定一个投影向量,以最大化类间散射矩阵,同时最小化特征空间中的类内散射矩阵。通过放大类间距离,缩小类内距离,就能够获得最优的投影方向。主成分分析是一种无监督学习的线性降维技术,本质是将高维矩阵经过线性变换转为低维度矩阵。首先,通过寻求变量的线性组合求得协方差矩阵,并计算协方差矩阵的特征向量和特征值;其次,依据特征值对特征向量进行降序排列,最后选取最突出的特征值对应的特征向量,将实际数据投影到排序后的特征向量,从而实现降维。

3.4 发酵过程数据特征及可视化技术

发酵过程中获得的数据是菌体代谢信息的载体,好的数据可视化技术可以让工程师更好地对发酵过程信息进行解读,从而做出正确的控制或调控策略,使发酵过程按照正常轨迹运行,亦或通过数据可视化发现发酵优化的关键点,从而对发酵工艺优化有较大的突破。

要更好地对发酵过程数据进行可视化,需要对发酵过程形成的数据有清晰的认识,根据数据所属类别及所包含信息的类型进行确定。发酵过程中形成的数据按照其获取方式不同,大致可以分成在线数据和离线数据两类,这些数据经过数学运算处理可以获得更多具有确切意义的导出型变量。按照导出型变量代表的对

象主体差异又可以将其分成过程变量、反应器变量及细胞生理变量。以上变量都具有明确物理或生理学意义,是通过原始传感数据或离线检测数据经过数学运算获得的,因此可以对其变化进行合理的解释。

另一类变量是通过诸如主成分分析、偏最小二乘法等降维技术处理所获得的数据,是从数据中抽取信息含量最大的隐含特征,这样的特征变量称为隐变量或潜变量。类似于一个具有明确概念的变量,比如要表征一个人是否健康,我们会为其进行各种科目的体检,体检的各项指标正如前面所述发酵过程的各种直接或导出数据,不能直接根据上述任何单一指标判断一个人是否健康,需通过医生的综合分析判断人的健康状况。这就类似于特征提取方法中按照数据方差最大化方向进行数据分解从而得到的一个综合变量,根据该变量的值再对过程特征进行判断。发酵过程中包括至少 20 个以上变量,如何进行数据的抽象和降维,并将降维数据可视化,从而对影响因素最大的原始变量进行反向推导具有重要意义。

4 智能化技术在发酵过程优化与放大中的应用

4.1 发酵过程知识图谱技术

知识图谱是结构化的知识库,知识图谱由定义为“实体-关系-实体”形式的三元组及其属性构成。通过三元组的形式,知识图谱可以描述现实世界中的各种概念及其相互间关系。知识图谱的概念最早由谷歌公司于 2012 年提出^[61],用于辅助搜索引擎在知识库中实现快速查询。对于不同应用场景,知识图谱的定义具有一些差别,但广义上的知识图谱具有 3 个重要特征,即对应于实际对象的节点、实际对象间关系抽象得到的边以及一定数量的节点和边组成的网络结构。

知识图谱通常根据其应用对象分为通用领域知识图谱和垂直领域知识图谱^[62]。通用领域知识图谱包含大量常识性知识,并不针对某一特定领域,具有代表性的如 Wikidata、FreeBase、Dbpedia 等。垂直领域知识图谱则强调针对特定领域,对知识库中的知识深度有较高要求,知识通常来源于专家经验、专业书籍等,目前在医疗、生物、电力、电商等多个领域都有较丰富的实践。发酵过程知识图谱属于垂直领域知识图谱,其构建过程更多需要发酵专家介入,通过专家知识与经验建立发酵过程各种参数(或状态)之间的专家规则,这些规则严格按照质量守恒定则、微生物反应本体动力学以及相关学科技术定则、经验证的相关性特征等确认,并以知识图谱形式保存(图 3)。同时建立生物过程的问题求解集合作为专家规则库,通过知识图谱的实体——关系链接,采用问题归约方法,在问题集合的状态空间通过实体间的有向

图链接形式把问题的求解过程描述出来,最终识别出发酵过程的状态^[63]。

4.2 发酵过程数字孪生模型技术

数字孪生的概念最早出现于 2011 年,在航空航天领域中被应用于飞机结构的寿命预测。2016 年数字孪生的理论首次拓展到制造系统的研发和管理中,并迅速受到学术界和工业界的关注。2017 年中国科协智能制造学术联合体在世界智能制造大会上将数字孪生列为世界智能制造十大科技进展之一。

数字孪生是物理对象的数字模型,该模型可以通过接收来自物理对象的数据而实时演化,从而与物理对象在全生命周期保持一致。基于数字孪生可进行仿真,实现分析、预测、诊断、训练等,并将结果反馈给物理对象,从而辅助针对物理对象的优化和决策。面向数字孪生全生命周期(构建、演化、评估、管理、使用)的技术称为数字孪生技术。近年来,数

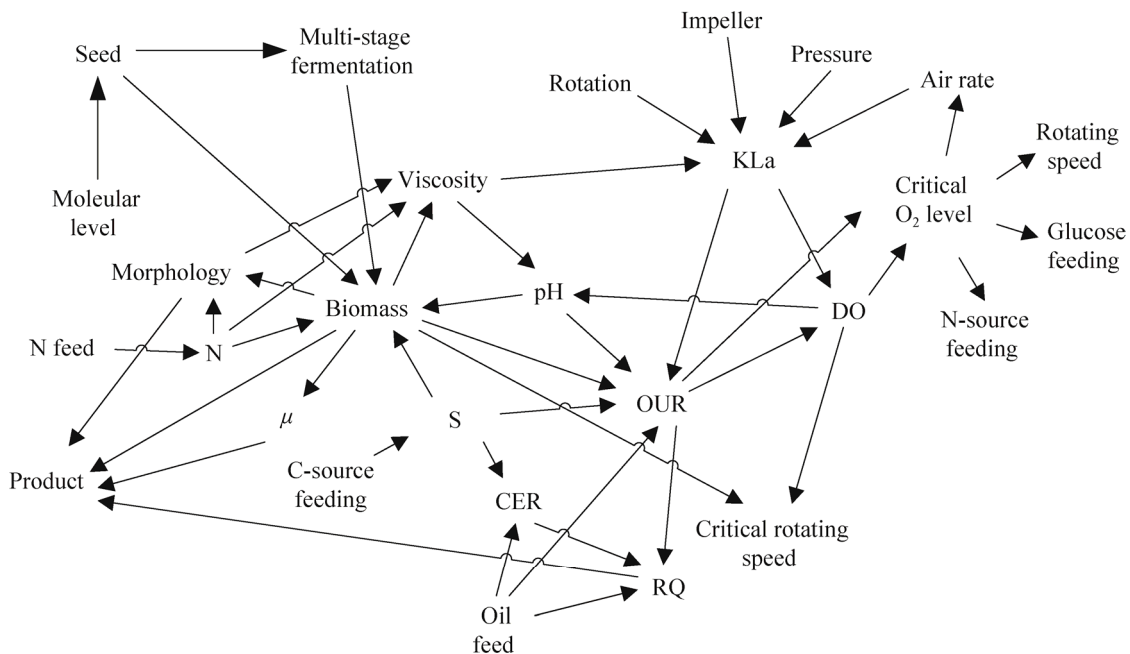


图 3 描述发酵过程参数之间关系的知识图谱^[63]

Figure 3 Knowledge graphs that describe interactions among fermentation parameters^[63].

字孪生技术在智能制造领域显现出巨大潜力,成为航空航天、汽车制造和机器人制造等行业关注和探索的关键技术。

数字孪生有望成为推动发酵过程实现智能化的核心技术。欧美国家在发酵过程的数字孪生技术方面正投入大量研发力量,其中较为活跃的包括通用公司、DSM公司、西门子公司以及代尔夫特理工大学、丹麦理工大学、英国帝国理工大学等。据报道英国著名医药公司GSK与西门子公司合作^[64],在中试规模验证数字孪生技术并获得成功,即将在生产规模实施数字孪生技术。奥地利初创公司Novasign开发了基于混合模型的数字孪生系统,运用该系统对大肠杆菌表达超氧化物歧化酶的过程进行优化^[23],加速发酵过程优化的进程。西门子公司2019年收购PSE公司过程数字化模型软件平台gPROMs,用于布局生物过程数字孪生系统,该系统具有大规模微分方程组求解算法优势,是进行发酵过程动力学模拟的理想平台。我国学者通过与国外科研机构合作在发酵过程数字孪生技术方面做了尝试。例如:华东理工大学与代尔夫特理工大学及DSM公司的国际合作项目,合作发表了将反应器流场和产黄青霉产青霉素的动力学模型整合的数字模型^[35],用于工业规模发酵过程的缩小设计及过程优化。江南大学与丹麦理工大学合作发表综述^[65],介绍建立生物制造过程的数字孪生系统的方法,并分析了其面临的机遇与挑战。

发酵过程的数字化、模型化正成为生物过程研究一个新的研究热点,各国生物过程研究的科研机构,包括科研院所、创新型公司、过程控制巨头等都在发酵过程的数字孪生系统建设中激烈竞争。随着发酵过程数字化基础设施的逐渐完善,以及整个制造业向数字化、智能化升级的进程推进,基于数字孪生系统的过程

模拟、预测、优化策略、甚至自动优化调控等一定会推动发酵过程优化与控制技术产生巨大进步。

4.3 整合数字孪生与知识图谱的发酵过程智能决策辅助

基于第一性原理的过程模型结合基于数据的无参数模型构成的混合模型,及在此基础上建立的数字孪生系统,为建立各种发酵过程控制条件与所关心的状态质量属性(critical quality attribute, CQA)之间的定量关系提供了路径。此方法借助人工智能中数据驱动模型建立控制参数与特定指标之间的模型,并引入第一性原理的结构模型加强模型的预测能力。然而,该方法忽略了发酵专家在长期发酵实践中形成的各种经验知识,而这些知识往往在实际发酵过程优化中发挥重要作用。前述知识图谱技术可以将专家经验以有向图的方式进行存储,并建立一套基于知识图谱的逻辑推理算法,具有数字孪生所不具备的优点。因此,若能将两者结合,即充分利用数字孪生对发酵过程的定量描述又引入包含专家经验的知识图谱,从而增强系统对发酵过程优化辅助决策的准确性。

通过知识图谱,可以将各种概念形成语义连接,使隐含知识显性化;而数字孪生技术,则可以在传统静态模型中,增加实时运行数据的反馈与动态记录,并以可视化的方式进行展现。Banerjee等^[66]即在工业生产线上实现了从知识图谱到生产线数字孪生系统的构建。Sahlab等^[67]则提出基于知识图谱实现数字孪生模型之间的交互融合,从而增强数字孪生系统的性能。这些将数字孪生与知识图谱技术融合的方法正在节能型数字社会构建、商品供应链整合以及制造业中探索性应用。目前,还未见将数字孪生技术与知识图谱结合实现发酵过程

优化及智能决策辅助的报道。

天津工业生物所依托国家合成生物技术创新中心建设,正在建设智能生物制造平台实验室,该实验室将重点围绕设计-构建-测试-学习(design-build-test-learn, DBTL) 范式中 Test 环节,拟在实验室及中试规模实现发酵过程的数字化及智能化。该平台实验室将分为高通量菌种性能验证平台、微生物代谢特性研究平台、高通量发酵过程优化平台以及对接数字孪生的中试自动化工艺开发平台等多个硬件平台,同时将建立基于数据科学与人工智能的生物过程大数据分析的软件平台,致力于将发酵过程数字孪生系统与发酵过程知识图谱结合的发酵过程智能决策辅助技术的开发,加速国家合成生物技术创新中心推动实验室创新成果向产业化转化的进程。

5 未来工业发酵技术展望

伴随着合成生物技术的飞速发展及人工智能技术的迅速崛起,传统工业发酵技术正迎来全新的机遇和挑战。随着这些新兴技术的不断融合渗透,未来工业发酵技术将发生颠覆性转变,以适应未来生物技术进步带来的产业化方面的新需求。发酵工程是一门多学科交叉的工程性学科,涉及微生物学、过程工程、反应工程等。近年来随着合成生物技术发展迎来的高通量、自动化实验室基础设施建设,以及对海量数据进行分析时引入的数据科学、人工智能技术等,都在推动着发酵工程技术的不断进步。相应的创新型企业也逐渐进入到生物技术产业行列,如 TeselaGen Biotechnology、Culture Biosciences、Novasign 公司等。

Culture Biosciences 是美国加利福尼亚州南旧金山的一家技术公司,为生物技术公司提供一种更简单、更快捷、更便宜的生物制造研发

方式。通过利用 Culture Biosciences 开发的云生物反应器平台,需要进行发酵的企业或科研单位可以通过远程提交任务,将发酵过程在 Culture Biosciences 的云发酵罐中进行,过程科学家可以远程控制及监测实际的发酵过程,因此可以投入更多时间设计和分析实验。同在加利福尼亚州的 TeselaGen 公司则利用 AI 技术,围绕合成生物学的设计-构建-测试-学习等几个环节,搭建了一套软件系统 TeselaGen,使生物学家、实验室技术人员以及生物信息学家协同工作,从而加速生物技术的产业化进程。奥地利的 Novasign 公司则依托 AI 技术,利用混合建模方法搭建了描述发酵过程的数字孪生系统,对发酵过程进行预测与控制,加速发酵过程优化的速度。由此可见,AI 技术正在向传统发酵行业技术逐渐渗透,衍生出各种新型的技术型企业,从而使得现代发酵技术的分支更加细化,正在推动发酵技术的整体进步,从而加速推进合成生物学实验室创新成果的产业化进程。

未来随着 AI 技术更加成熟,其与发酵技术的深度融合将颠覆传统发酵技术手段。数字孪生将成为发酵企业的标准配置,借助知识图谱技术辅助,发酵工程师进行发酵过程的智能分析和优化调控将变得更加便利高效,AI 赋能的高效发酵技术为生物经济发展起到高效支撑作用。基于此对未来发酵工程人才也提出了更高的要求,一方面对其知识构成又有了新的要求,除了原生物学、生物工程、系统生物学等方面的专业基础知识外,还应具备各类传感器应用、大数据处理等信息学基本知识;另一方面,对发酵工程相关的学校和专业也应调整原有的培养方案,增加交叉学科的相关课程(或选修课),以拓展学生的知识面应对新技术融合下的新发展;最后更加重要的是要培养学生的实践能力,提供更多的开放式平台,供学生开展实践研究

与创新, 培养出真正拥有交叉学科知识、拥有解决未来发酵工程复杂问题能力的人才。

REFERENCES

- [1] 国务院关于印发《中国制造 2025》的通知[EB/OL]. [2022-09-15]. http://www.gov.cn/zhengce/content/2015-05/19/content_9784.htm
- [2] 八部门关于印发《“十四五”智能制造发展规划》的通知[EB/OL]. [2022-09-15]. http://www.gov.cn/zhengce/zhengceku/2021-12/28/content_5664996.htm
- [3] 国家发展改革委关于印发《“十四五”生物经济发展规划》的通知[EB/OL]. [2022-09-15]. https://www.ndrc.gov.cn/xxgk/zcfb/ghwb/202205/t20220510_1324436.html?code=&state=123
- [4] Humphrey A. Shake flask to fermentor: what have we learned? *Biotechnol Prog*, 1998, 14(1): 3-7.
- [5] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021, 596(7873): 583-589.
- [6] Jayatunga MKP, Xie W, Ruder L, et al. AI in small-molecule drug discovery: a coming wave? *Nat Rev Drug Discov*, 2022, 21(3): 175-176.
- [7] Baliga NS. Systems biology. The scale of prediction. *Science*, 2008, 320(5881): 1297-1298.
- [8] 马誉激. 抗生素. 北京: 人民卫生出版社, 1955. Ma YC, Antibiotics. Beijing: People Health Press, 1955 (in Chinese).
- [9] Nagodawithana TW, Castellano C, Steinkraus KH. Effect of dissolved oxygen, temperature, initial cell count, and sugar concentration on the viability of *Saccharomyces cerevisiae* in rapid fermentations. *Appl Microbiol*, 1974, 28(3): 383-391.
- [10] Sidoruk KV, Davydova LI, Kozlov DG, et al. Fermentation optimization of a *Saccharomyces cerevisiae* strain producing 1F9 recombinant spidroin. *Appl Biochem Microbiol*, 2015, 51(7): 766-773.
- [11] Ming S, Zhang Y, Guan YL, et al. Fermentation optimization of *Bacillus licheniformis*. *IOP Conf Ser Mater Sci Eng*, 2019, 612(2): 022112.
- [12] Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol*, 2014, 5: 172.
- [13] Chen ZY, Shen XL, Wang J, et al. Establishing an artificial pathway for *de novo* biosynthesis of vanillyl alcohol in *Escherichia coli*. *ACS Synth Biol*, 2017, 6(9): 1784-1792.
- [14] Rosa SS, Prazeres DMF, Azevedo AM, et al. mRNA vaccines manufacturing: challenges and bottlenecks. *Vaccine*, 2021, 39(16): 2190-2200.
- [15] Fink M, Cserjan-Puschmann M, Reinisch D, et al. High-throughput microbioreactor provides a capable tool for early stage bioprocess development. *Sci Rep*, 2021, 11(1): 2056.
- [16] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*, 2011, 12: 2825-30.
- [17] The pandas development team, pandas-dev/pandas: pandas. Zenodo. 2020.10.5281/zenodo.3509134
- [18] Harris CR, Millman KJ, Van Der Walt SJ, et al. Array programming with NumPy. *Nature*, 2020, 585(7825): 357-362.
- [19] Berthold M R, Cebron N, Dill F, et al. KNIME-the konstanz information miner: version 2.0 and beyond. *ACM SIGKDD Explorations Newsletter*, 2009, 11(1): 26-31.
- [20] Wehrs M, De Beaumont-Felt A, Goranov A, et al. You get what You screen for: on the value of fermentation characterization in high-throughput strain improvements in industrial settings. *J Ind Microbiol Biotechnol*, 2020, 47(11): 913-927.
- [21] 夏建业, 田锡炜, 刘娟, 等. 人工智能时代的智能生物制造. *生物加工过程*, 2020, 18(1): 13-20. Xia JY, Tian XW, Liu J, et al. Intelligent biological manufacturing in the age of artificial intelligence. *Chin J Bioprocess Eng*, 2020, 18(1): 13-20 (in Chinese).
- [22] Zhang SL, Chu J, Zhuang YP. A multi-scale study of industrial fermentation processes and their optimization. *Adv Biochem Eng Biotechnol*, 2004, 87: 97-150.
- [23] Bayer B, Striedner G, Duerkop M. Hybrid modeling and intensified DoE: an approach to accelerate upstream process characterization. *Biotechnol J*, 2020, 15(9): e2000121.
- [24] 张嗣良, 储炬. 多尺度微生物过程优化. 北京: 化学工业出版社, 2003. Zhang SL, Chu J. Multi-scale Microbiology Process Optimization. Beijing: Chemical Industry Press, 2003 (in Chinese).
- [25] 张嗣良. 发酵过程多水平问题及其生物反应器装置技术研究——基于过程参数相关的发酵过程优化与放大技术. *中国工程科学*, 2001, 3(8): 37-45. Zhang SL. Study on the fermentation processes at multi-levels in bioreactor and its application for special purposes—optimization and scaling up of the fermentation process based on the parameter

- correlation method. *Eng Sci*, 2001, 3(8): 37-45 (in Chinese).
- [26] Zou X, Xia JY, Chu J, et al. Real-time fluid dynamics investigation and physiological response for erythromycin fermentation scale-up from 50 L to 132 m³ fermenter. *Bioprocess Biosyst Eng*, 2012, 35(5): 789-800.
- [27] Li XY, Yu C, Yao JM, et al. An online respiratory quotient-feedback strategy of feeding yeast extract for efficient arachidonic acid production by *Mortierella alpina*. *Front Bioeng Biotechnol*, 2018, 5: 83.
- [28] Wang G, Haringa C, Noorman H, et al. Developing a computational framework to advance bioprocess scale-up. *Trends Biotechnol*, 2020, 38(8): 846-856.
- [29] Tang WJ, Deshmukh AT, Haringa C, et al. A 9-pool metabolic structured kinetic model describing days to seconds dynamics of growth and product formation by *Penicillium chrysogenum*. *Biotechnol Bioeng*, 2017, 114(8): 1733-1743.
- [30] Xia JY, Wang G, Lin JH, et al. Advances and practices of bioprocess scale-up. *Adv Biochem Eng Biotechnol*, 2016, 152: 137-151.
- [31] Lara AR, Galindo E, Ramirez OT, et al. Living with heterogeneities in bioreactors. *Mol Biotechnol*, 2006, 34(3): 355-381.
- [32] Enfors SO, Jahic M, Rozkov A, et al. Physiological responses to mixing in large scale bioreactors. *J Biotechnol*, 2001, 85(2): 175-185.
- [33] Haringa C, Tang WJ, Deshmukh AT, et al. Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: an analysis of organism lifelines. *Eng Life Sci*, 2016, 16(7): 652-663.
- [34] Haringa C, Noorman HJ, Mudde RF. Lagrangian modeling of hydrodynamic-kinetic interactions in (bio)chemical reactors: practical implementation and setup guidelines. *Chem Eng Sci*, 2017, 157: 159-168.
- [35] Haringa C, Tang WJ, Wang G, et al. Computational fluid dynamics simulation of an industrial *P. chrysogenum* fermentation with a coupled 9-pool metabolic model: towards rational scale-down and design optimization. *Chem Eng Sci*, 2018, 175: 12-24.
- [36] Yang YM, Xia JY, Li JH, et al. A novel impeller configuration to improve fungal physiology performance and energy conservation for cephalosporin C production. *J Biotechnol*, 2012, 161(3): 250-256.
- [37] Li X, Huang C, Xu CQ, et al. High cell density culture of baker's yeast FX-2 based on pH-stat coupling with respiratory quotient. *Biotechnol Appl Biochem*, 2019, 66(3): 389-97.
- [38] Lopez PC, Udugama IA, Thomsen ST, et al. Towards a digital twin: a hybrid data-driven and mechanistic digital shadow to forecast the evolution of lignocellulosic fermentation. *Biofuels Bioprod Biorefining*, 2020, 14(5): 1046-1060.
- [39] Zhao YZ, Yang Y, Loscalzo J. Real-time assessment of the metabolic profile of living cells with genetically encoded NADH sensors. *Methods Enzymol*, 2014, 542: 349-367.
- [40] Zhao YZ, Jin J, Hu QX, et al. Genetically encoded fluorescent sensors for intracellular NADH detection. *Cell Metab*, 2011, 14(4): 555-566.
- [41] Zhao YZ, Hu QX, Cheng FX, et al. SoNar, a highly responsive NAD⁺/NADH sensor, allows high-throughput metabolic screening of anti-tumor agents. *Cell Metab*, 2015, 21(5): 777-789.
- [42] Zhao YZ, Wang AX, Zou YJ, et al. *In vivo* monitoring of cellular energy metabolism using SoNar, a highly responsive sensor for NAD⁽⁺⁾/NADH redox state. *Nat Protoc*, 2016, 11(8): 1345-1359.
- [43] Tao R, Zhao Y, Chu H, et al. Genetically encoded fluorescent sensors reveal dynamic regulation of NADPH metabolism. *Nat Methods*, 2017, 14(7): 720-728.
- [44] Zou Y, Wang A, Shi M, et al. Analysis of redox landscapes and dynamics in living cells and *in vivo* using genetically encoded fluorescent sensors. *Nat Protoc*, 2018, 13(10): 2362-2386.
- [45] Hu H, Gu Y, Xu L, et al. A genetically encoded toolkit for tracking live-cell histidine dynamics in space and time. *Sci Reports*, 2017, 7: 43479.
- [46] Tao RK, Shi M, Zou YJ, et al. Multicoloured fluorescent indicators for live-cell and *in vivo* imaging of inorganic mercury dynamics. *Free Radic Biol Med*, 2018, 121: 26-37.
- [47] Hu HY, Wei YF, Wang DC, et al. Glucose monitoring in living cells with single fluorescent protein-based sensors. *RSC Adv*, 2018, 8(5): 2485-2489.
- [48] Hu HY, Wang AX, Huang L, et al. Monitoring cellular redox state under hypoxia using a fluorescent sensor based on eel fluorescent protein. *Free Radic Biol Med*, 2018, 120: 255-265.
- [49] Zou YJ, Wang AX, Huang L, et al. Illuminating NAD⁺ metabolism in live cells and *in vivo* using a genetically encoded fluorescent sensor. *Dev Cell*, 2020, 53(2): 240-252.e7.

- [50] Samaras JJ, Micheletti M, Ducci A. Suspension and mixing characterization of intermittent agitation modes in DASGIP bioreactors. *Chem Eng Technol*, 2019, 42(8): 1587-1593.
- [51] Velez-Suberbie ML, Betts JPJ, Walker KL, et al. High throughput automated microbial bioreactor system used for clone selection and rapid scale-down process optimization. *Biotechnol Prog*, 2018, 34(1): 58-68.
- [52] Xu P, Clark C, Ryder T, et al. Characterization of TAP Ambr 250 disposable bioreactors, as a reliable scale-down model for biologics process development. *Biotechnol Prog*, 2017, 33(2): 478-489.
- [53] Li C, Tian JT, Wang WF, et al. Numerical and experimental assessment of a miniature bioreactor equipped with a mechanical agitator and non-invasive biosensors. *J Chem Technol Biotechnol*, 2019, 94(8): 2671-2683.
- [54] Tenhaef N, Stella R, Frunzke J, et al. Automated rational strain construction based on high-throughput conjugation. *ACS Synth Biol*, 2021, 10(3): 589-599.
- [55] 张嗣良, 潘杭琳, 黄明志, 等. 生物过程大数据分析与智能化. *生物产业技术*, 2018(1): 86-93.
Zhang SL, Pan HL, Huang MZ, et al. Big data and intelligentized method for industrial bioprocess. *Biotechnol Bus*, 2018(1): 86-93 (in Chinese).
- [56] Cañete-Carmona E, Gallego-Martínez JJ, Martín C, et al. A low-cost IoT device to monitor in real-time wine alcoholic fermentation evolution through CO₂ emissions. *IEEE Sens J*, 2020, 20(12): 6692-6700.
- [57] Angelkov D, Martinovska Bande C. Sensor Module for Monitoring Wine Fermentation Process; Proceedings of the Applied Physics, System Science and Computers, Cham, F 2018[C]. Springer International Publishing.
- [58] Saha H, Auddy S. Internet of Things (IoT) on bio-technology [M]. 2017.
- [59] 武天鸿, 翁小清, 单中南. 基于 LDA 符号表示的时间序列分类算法. *计算机应用与软件*, 2020, 37(2): 259-265, 307.
Wu TH, Weng XQ, Shan ZN. Time series classification based on LDA symbolic representation. *Comput Appl Softw*, 2020, 37(2): 259-265, 307 (in Chinese).
- [60] 张玉霖. 基于主成分分析的网络时延特征数据提取仿真. *计算机仿真*, 2020, 37(3): 301-304.
Zhang YL. Simulation of network delay feature extraction based on principal component analysis. *Comput Simul*, 2020, 37(3): 301-304 (in Chinese).
- [61] Singhal A. Introducing the knowledge graph: things, not strings. Official google blog, 2012, 5: 16.
- [62] 中国电子技术标准化研究院. 信息技术与标准化, 2021(S1): 2.
China Electronics Standardization Institute. *Inf Technol & Stand*, 2021(S1): 2 (in Chinese).
- [63] 张嗣良, 张旭蓓. 基于知识图谱有向图的生物过程控制方法: CN109243528A, 2022-02-08.
Zhang SL, Zhang XB. Biological process control method based on knowledge map directed graph: CN109243528A, 2022-02-08.
- [64] Markarian J. Opportunities for digital twins. *BioPharm International*, 2021, 34(9): 30-33.
- [65] Udugama IA, Lopez PC, Gargalo CL, et al. Digital Twin in biomanufacturing: challenges and opportunities towards its implementation. *Syst Microbiol Biomanufacturing*, 2021, 1(3): 257-274.
- [66] Banerjee A, Dalal R, Mittal S, et al. Generating digital twin models using knowledge graphs for industrial production lines. *WebSci'17: Proceedings of the 2017 ACM on Web Science Conference*, 2017: 425-430.
- [67] Sahlab N, Kamm S, Müller T, et al. Knowledge graphs as enhancers of intelligent digital twins; proceedings of the 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), 2021.

(本文责编 郝丽芳)