·研究论文·

曲妥珠单抗细胞培养过程模型建立与智能算法对比

王文慧[1], 万宇翔[1,2], 焦静雨[2], 高　丹[2], 高　栋[2], 瞿海斌[1*]

(1. 浙江大学药学院, 药物信息学研究所, 浙江 杭州 310058;　2. 浙江博之锐生物制药有限公司, 浙江 杭州 311404)

摘要: 机理模型通过数学方程描述细胞代谢行为, 是优化和预测生物过程的重要理论工具, 本研究基于Monod生长代谢模型, 构建了曲妥珠单抗生物类似药的细胞培养机理模型。针对微分方程参数拟合复杂度高且容易陷入局部最优的问题, 通过引入粒子群优化算法 (particle swarm optimization, PSO)、灰狼优化算法 (grey wolf optimizer, GWO) 和差分进化算法 (differential evolution, DE) 对模型参数拟合。结果表明, GWO在全局搜索能力和拟合精度方面表现最佳, 在处理多参数拟合问题时具有更高的稳定性和适应性。进一步对GWO算法的位置更新与速度更新方法进行改进, 改进的灰狼算法 (improved grey wolf optimizer, IGWO) 显著提升了模型的预测能力, 能够精准预测细胞密度、代谢物浓度及抗体滴度。本研究为提高生物药物生产效率、降低开发成本及优化细胞培养工艺提供了重要理论支持。

关键词: 细胞培养; 单克隆抗体; 曲妥珠单抗; 机理模型; 智能算法

中图分类号: R283　　**文献标识码**: A　　**文章编号**: 0513-4870(2025)09-2788-10

Mechanistic modeling for cell culture process of trastuzumab and intelligent algorithm comparison

WANG Wen-hui[1], WAN Yu-xiang[1,2], JIAO Jing-yu[2], GAO Dan[2], GAO Dong[2], QU Hai-bin[1*]

(*1. Pharmaceutical Informatics Institute, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China;　2. BioRay Pharmaceutical Co., Ltd., Hangzhou 311404, China*)

Abstract: Mechanistic modeling, which describes cellular metabolic behavior through mathematical equations, is an important theoretical tool for optimizing and predicting biological processes. In this study, a cell culture mechanistic model for biosimilar trastuzumab was constructed based on the Monod growth metabolism model. To address the high complexity of differential equation parameter fitting and the tendency to fall into local optima, particle swarm optimization (PSO), grey wolf optimization (GWO), and differential evolution (DE) algorithms were introduced for model parameter fitting. The results show that GWO outperforms the others in global search capability and fitting accuracy, demonstrating higher stability and adaptability in handling multi-parameter fitting problems. Furthermore, improvements were made to the position and velocity update methods of the GWO algorithm. The improved grey wolf optimization (IGWO) significantly enhanced the model's predictive ability, enabling accurate predictions of cell density, metabolite concentration, and antibody titer. This study provides important theoretical support for improving the production efficiency of biopharmaceuticals, reducing development costs, and optimizing cell culture processes.

Key words: cell culture; monoclonal antibody; trastuzumab; mechanistic modeling; intelligent algorithm

收稿日期: 2025-01-20;　修回日期: 2025-04-30.

基金项目: 浙江省重点研发计划 (2023C03116).

*通讯作者 Tel: 86-571-88208428, E-mail: quhb@zju.edu.cn

DOI: 10.16438/j.0513-4870.2025-0071

哺乳动物细胞是生产多种生物药物 (如重组蛋白药物、病毒疫苗和激素) 的常用表达宿主, 能够进行复杂的翻译和修饰 (如N-糖基化), 并确保蛋白质的正确折叠与组装[1,2]。曲妥珠单抗是一种靶向人表皮生长因子受体-2 (human epidermal growth factor receptor 2, HER2) 的重组DNA衍生人源化单克隆抗体。曲妥珠单抗作为单克隆抗体药物的代表, 广泛应用于HER2阳性乳腺癌等疾病的治疗[3]。曲妥珠单抗注射液由瑞士罗氏制药公司原研, 1998年首次由FDA批准上市, 商品名为赫赛汀®。2019年, 赫赛汀®全球销售额达60亿美元, 赫赛汀®的专利分别于2014和2019年在欧洲与美国到期[4]。此后, 曲妥珠单抗生物类似药的开发迎来了迅速发展期, 生物类似药是指在质量、安全性和药效等方面与已批准的参考药物高度一致的药物[5]。如何进一步提高细胞的生长密度、优化培养策略以提升产品品质, 并通过智能制造技术实现培养过程的自适应调控仍是亟待解决的核心技术难题。

在早期细胞培养工艺开发中, 实验设计 (design of experiments, DoE) 方法被广泛用于探索工艺参数[6,7]。然而, DoE方法存在明显的局限性。首先, DoE方法需要大量实验验证, 投入大量的时间与资源, 成本较高[8]; 其次, DoE方法更关注变量间的统计关联, 而对细胞培养过程的机理缺乏深入理解, 导致其结果在未经验证条件下的适用性有限, 难以实现全局优化[9]。此外, 随着研究对象复杂度的增加, DoE方法在多变量和高维空间中的应用难度进一步加剧, 限制了其在现代生物制药工艺开发中的有效性[10]。随着生物制药行业的快速发展, 基于机理的细胞培养模型逐渐受到研究者的关注。机理模型通过数学方程描述细胞代谢行为, 涵盖细胞生长、凋亡、代谢物消耗和产物合成等关键过程, 为培养过程提供更深层次的理解[11]。利用机理模型, 研究人员可以预测不同条件下细胞的生长过程, 优化工艺参数。与传统的DoE方法相比, 机理模型不仅能够显著减少实验次数, 还能在多变量高维空间中预测培养过程的动态行为, 为工艺优化提供更加精准的指导[12]。通过结合具体的细胞系、培养条件及工艺参数, 机理模型能够模拟多种潜在操作场景, 支持跨条件和跨项目的预测与优化, 帮助开发者高效筛选最佳工艺窗口[13]。随着计算能力的提升和优化算法的不断进步[14], 机理建模已成为细胞培养工艺开发中不可或缺的工具, 为提高单克隆抗体等生物药物的生产效率与质量提供了重要的理论与实践支持[15]。

参数估计在细胞培养的机理建模中至关重要, 它直接决定了模型的精度和预测性能。机理模型通常由大量非线性且强耦合的方程组成, 其参数数量随着代谢网络复杂度的增加而显著增加[16]。这使得参数求解成为一个具有高度复杂性的非凸优化问题, 面临局部最优解、维度复杂性及计算成本高昂等多重难题。传统优化方法 (如梯度下降法) 虽然在处理规模较小的问题时能够取得一定成效, 但在高维和复杂参数空间中表现出明显不足[17]。它们对初始值的依赖性较高, 易陷入局部最优解, 同时对目标函数的可微性要求较强, 而这种要求在许多生物动力学系统中往往难以满足。为应对上述挑战, 智能优化算法通过模仿自然界中的群体行为、进化法则或物理规律, 提供了一种灵活而高效的求解方法[18]。这类算法具备出色的全局搜索能力和强大的适应性, 特别适合于解决高维非线性优化问题, 在复杂系统的参数估计中展现出巨大潜力[19]。传统参数估计方法对高维、非线性问题的适应性不足, 导致模型预测精度受限。本研究采用粒子群优化算法 (particle swarm optimization, PSO)、差分进化算法 (differential evolution, DE)、灰狼优化算法 (grey wolf optimizer, GWO) 进行参数估计, 并将改进的灰狼算法 (improved grey wolf optimizer, IGWO) 引入细胞培养建模, 通过动态权重与莱维飞行等策略提升参数估计效率, 为解决多变量强耦合模型的优化难题提供了新方法。

材料与方法

药品与试剂 曲妥珠细胞株由博之锐生物制药有限公司提供。培养基包括CD1005-N-1代培养基、CD1005-N代培养基, 以及补料培养基CD1005-A和CD1005-B, 由于培养基成分涉及浙江博之锐生物的企业机密, 具体成分未公开。SIGMA台盼蓝溶液、SIGMA胰蛋白酶 (德国Merck KGaA集团, 批号分别为RNBL3669、SLCL2409); 十二水合磷酸氢二钠、氯化钠、二水磷酸二氢钠、氢氧化钠、甘氨酸 (上海国药集团化学试剂有限公司, 批号分别为20220514、20231013、20221220、20220908、20220915); 盐酸 (湖南尔康制药股份有限公司, 批号为000420220701)。

实验仪器 BioCore QC台式细胞生物反应器玻璃 (上海楚怡生物科技有限公司); M-900 Plus细胞培养生化分析仪 (深圳西尔曼科技有限公司); TD5A-WS台式低速大容量离心机 (湖南湘仪实验室仪器开发有限公司); Countstar BioTech自动细胞计数仪 (上海睿钰生物科技有限公司); 色谱柱 (PORPS 20A, 2.1 mm×30 mm, 美国Thermo公司); Agilent 1200/1260液相色谱仪 (美国安捷伦科技有限公司)。

细胞复苏 在细胞复苏及摇瓶培养阶段, 先将曲妥珠细胞株 (博之锐药业开发的CHO-K1细胞系) 从

液氮中解冻, 并以合适的接种密度 (每毫升 0.4×10^6～0.6×10^6 个细胞) 接种至 CD1005-N-1 代培养基中。每代培养 48 ± 2 h 后进行扩增, 扩增后的细胞继续在摇瓶中培养。培养过程中, 将摇瓶置于培养箱摇床内 (37 ℃, 5.0% CO_2, 80% 湿度, 125 r·min^{-1}, 摇床直径 50 cm)。摇瓶培养阶段结束的最终每毫升细胞数在 2.0×10^6～3.0×10^6 个。

细胞培养 曲妥珠细胞在 3.0 L 玻璃生物反应器中培养, 初始工作体积为 1.0 L, 并以 1∶1 的传代比例接种至反应器中, 初始种子密度为每毫升 1.0×10^6～1.5×10^6 个。培养条件设定为: 在整个培养过程中, 温度维持在 35 ℃, 溶解氧 (dissolved oxygen, DO) 维持在 40% 的空气饱和度水平, 搅拌速度为 250 r·min^{-1}。pH 值通过二氧化碳气体和碳酸钠溶液调控, 在培养的前 7 天 (D0～D7), pH 值范围为 6.90 ± 0.30; 从第 8～16 天 (D8～D16), pH 值范围调整为 6.80 ± 0.20。补料包括 CD1005-A 培养基、CD1005-B 培养基以及浓度为 40% 的葡萄糖 (glucose, Glc) 溶液。其中, CD1005-A 和 CD1005-B 培养基在培养的第 2、4、6、8、10、12 和 14 天按 10∶1 的体积比补充; Glc 溶液则根据需求添加以维持目标浓度。

生化测定 在整个细胞生长培养过程中, 每日固定时间通过自动取样器取样一次, 取样前废弃 10 mL 样品以保证取样均匀性, 在取样后 2 h 进行补料操作, 通过自动细胞计数仪测定活细胞密度 (viable cell density, Vcd) 和细胞存活率, 检测方法为台盼蓝染色 (细胞液∶台盼蓝∶胰蛋白酶按 1∶1∶2 的比例混合), 每个样品取样 3 次后取平均值。样品经离心机以 1 000 r·min^{-1} 离心 5 min 后, 用 25 μm 滤芯进行过滤, 过滤后的上清液通过细胞培养生化分析仪测定 Glc、乳酸 (lactate, Lac)、氨 (ammonia, Amm)、谷氨酰胺 (glutamine, Gln)、谷氨酸 (glutamate, Glu) 等生化代谢数据。

滴度测定 抗体滴度使用液相色谱法进行测量, 需要配置流动相 A、流动相 B 与标准曲线溶液。流动相 A: 分别称取二水磷酸二氢钠 15.6 g 和氯化钠 17.5 g, 加入 2 000 mL 纯化水, 搅拌溶解, 用 10 mol·L^{-1} 氢氧化钠调 pH 值到 7.0, 经 0.22 μm 滤膜过滤后超声即得流动相 A。流动相 B: 称取甘氨酸 7.5 g, 加入 1 000 mL 纯化水搅拌溶解, 用 6 mol·L^{-1} 盐酸调整 pH 到 2.5, 经 0.22 μm 滤膜过滤后超声即得流动相 B。将参比品用流动相 A 梯度稀释为 4、2、1、0.5、0.25、0.125 mg·mL^{-1}, 做 6 点标准曲线, 保证被测样品的浓度应该包含在标准曲线的浓度范围内, 其中以 1 mg·mL^{-1} 的标准曲线作为系统适用性溶液。若样品预估浓度大于方法线形范围的最高点, 则用流动相 A 将供试品稀释至线形范围, 记录稀释倍数 N, 若供试品预估浓度在方法的线形范围内, 则无需稀释。供试品使用 0.22 μm 滤膜进行过滤。样品盘温度设定为 8 ℃, 柱温设定为 30 ℃, 流速设定为 3.0 mL·min^{-1}, 进样体积为 20 μL, 波长设定为 280 nm, 运行时间设定为 3 min, 运行时间分别为 0、0.05、0.51、1.50、1.51、3.00 min 时, 流动相 A 的比例分别为 100%、100%、0%、0%、100%、100%。

建模方法 本研究采用 Wang 等[20]发表的机理模型, 以 Monod 方程作为生长代谢模型, 主要描述细胞的生长、死亡及蛋白质表达的动态过程。在模型中, 营养物质的消耗主要包括 Glc、Glu 和 Gln, 用于支持细胞的生长及维持细胞活力, Lac 和 Amm 被认为是抑制细胞生长的因素, 以 Michaelis-Menten 动力学方程形式进行描述。在微分方程组中 [X] 代表活细胞密度; [Glc]、[Glu]、[Gln]、[Lac]、[Amm]、[IgG] 分别代表对应代谢物的浓度; μ_G 与 k_d 分别为生长速率与死亡速率, KI_{Glc}、KI_{Lac}、KI_{Amm}、KI_{Glu} 分别为对应代谢物的 Monod 方程生长常数; KD_{Glc}、KD_{Lac}、KD_{Amm}、KD_{Glu} 分别为对应代谢物的 Monod 方程死亡常数; m_{Glc}、m_{Gln}、m_{Glu} 分别为对应代谢物的维持系数; $YF_{x,Glc}$、$YF_{x,Amm}$、$YF_{x,Gln}$、$YF_{x,Glu}$、$YF_{Lac,Glc}$、$YF_{Gln,Glu}$ 为产率系数; k_{Lac}、k_{Amm} 为消耗速率; K_{Lac}、K_{Amm} 为米氏常数; $Q_{p,max}$ 和 KS_{Gln} 分别为最大比生产速率和饱和常数, 具体模型描述如方程 1～9 所示:

$$\frac{d[X]}{dt}=(\mu_G-k_d)\cdot[X] \tag{1}$$

$$\mu_G=\mu_{G,max}\cdot\frac{[Glc]}{KI_{Glc}+[Glc]}\cdot\frac{[Glu]}{KI_{Glu}+[Glu]}\cdot\frac{KI_{Lac}}{KI_{Lac}+[Lac]}\cdot\frac{KI_{Amm}}{KI_{Amm}+[Amm]} \tag{2}$$

$$k_d=k_{d,max}\cdot\frac{KD_{Glc}}{KD_{Glc}+[Glc]}\cdot\frac{KD_{Glu}}{KD_{Glu}+[Glu]}\cdot\frac{[Lac]}{KD_{Lac}+[Lac]}\cdot\frac{[Amm]}{KD_{Amm}+[Amm]} \tag{3}$$

$$\frac{d[Glc]}{dt}=-\left(m_{Glc}+\frac{\mu_G}{YF_{x,Glc}}\right)\cdot[X] \tag{4}$$

$$\frac{d[Glu]}{dt}=-\left(m_{Glu}+\frac{\mu_G}{YF_{x,Glu}}\right)\cdot[X] \tag{5}$$

$$\frac{d[Gln]}{dt}=\left(\left(m_{Glu}+\frac{\mu_G}{YF_{x,Glu}}\right)\cdot YF_{Gln,Glu}-\left(m_{Gln}+\frac{\mu_G}{YF_{x,Gln}}\right)\right)\cdot[X] \tag{6}$$

$$\frac{d[Lac]}{dt}=\left(\left(m_{Glc}+\frac{\mu_G}{YF_{x,Glc}}\right)\cdot YF_{Lac,Glc}-\frac{k_{Lac}\cdot[Lac]}{K_{Lac}+[Lac]}\right)\cdot[X] \tag{7}$$

$$\frac{\mathrm{d}[\mathrm{Amm}]}{\mathrm{dt}}=\left(\frac{\mu_{\mathrm{G}}}{YF_{\mathrm{x,Amm}}}-\frac{k_{\mathrm{Amm}}\cdot[\mathrm{Amm}]}{K_{\mathrm{Amm}}+[\mathrm{Amm}]}\right)\cdot[\mathrm{X}] \quad (8)$$

$$\frac{\mathrm{d}[\mathrm{IgG}]}{\mathrm{dt}}=Q_{\mathrm{p,max}}\cdot\frac{[\mathrm{Gln}]}{KS_{\mathrm{Gln}}+[\mathrm{Gln}]}\cdot[\mathrm{X}] \quad (9)$$

评价指标 本研究利用ode45函数求解一阶微分方程, 并结合实验数据, 通过智能优化算法拟合求解机理模型参数。模型参数求解与拟合过程基于Matlab_R2022b版本运行, 运行电脑配置为MacBook Pro, 采用2.3 GHz四核Intel Core i5处理器, 内存为8 GB。实验数据包括Vcd、Glc、Lac、Amm、Glu、Gln、IgG及取样时间, 实验数据作模型输入, 并通过最小化测量结果与拟合结果之间的残差拟合25个参数。为消除拟合变量在数量级上的差异对残差的影响, 需要进行归一化处理[7,20]。变量的拟合性能定义为适应度(fitness), fitness值越小拟合越准确, 为确保Vcd和滴度的拟合优先级, 目标函数设置为拟合度与变量权重乘积的最小和。通过计算决定系数(coefficient of determination, R^2)[21,20]和均方根误差(root mean square error, RMSE)[20,22]评估拟合与预测性能, 具体评价指标描述如方程10～13所示。

$$R^2=1-\frac{\sum_i^n\left(y_i-\widehat{y}_i\right)^2}{\sum_i^n\left(y_i-\bar{y}\right)^2} \quad (10)$$

$$\mathrm{Fitness}_j=\frac{\sum_i^n\left(y_i-\widehat{y}_i\right)^2}{\sum_i^n\left(y_i-\bar{y}\right)^2}=1-R^2 \quad (11)$$

$$\mathrm{Fitness}_j=\sum_j^7\left(\omega_j*\mathrm{Fitness}_j\right) \quad (12)$$

$$\mathrm{RMSE}=\sqrt{\frac{\sum_i^n\left(y_i-\widehat{y}_i\right)^2}{n}} \quad (13)$$

算法原理

GWO是一种群体智能的优化算法, 通过模拟灰狼社会中的捕猎行为进行目标优化, 为解决复杂优化问题提供了创新性方法。GWO的主要特点在于, 通过领导者(alpha、beta、delta)与追随者(omega)的分层结构, 动态调整个体位置, 从而在全局搜索与局部开发之间实现有效平衡[23]。近年来, GWO逐步应用于细胞培养等复杂生物过程的机理建模中, 用以解决参数估计这一关键问题。在函数寻优过程中, GWO首先通过数学方法描述狼群的等级结构, 其中, alpha表示适应度排名第一的最优解, beta和delta分别表示排名第二和第三的次优解, 其余可行解由omega表示[24]。与其他群体智能算法类似, GWO以狼群的随机初始化为起点。在多维搜索空间中, 狼群每个成员的位置均以矢量表示。在迭代过程中, 狼群在搜索空间内持续探索、信息交换, 并动态调整其位置。标准GWO流程包括狼群初始化、适应度值计算以及位置矢量的更新等步骤[25], 灰狼算法流程示意图如图1所示。GWO通过以下3个核心步骤来模拟捕猎行为。

包围猎物 灰狼通过围捕猎物来缩小捕猎范围, 位置更新公式如方程14～17所示。

$$\vec{D}=\left|\vec{C}\cdot\vec{X}_{\mathrm{prey}}-\vec{X}\right| \quad (14)$$

$$\vec{X}(\mathrm{t}+1)=\vec{X}_{\mathrm{prey}}-\vec{A}\cdot\vec{D} \quad (15)$$

$$\vec{A}=2\cdot\vec{a}\cdot\vec{r}_1-\vec{a} \quad (16)$$

$$\vec{C}=2\cdot\vec{r}_2 \quad (17)$$

其中: $\vec{X}_{\mathrm{prey}}$是猎物的位置, 视为当前最优解, $\vec{X}$是灰狼的位置, $\vec{A}$和$\vec{C}$是动态系数, 控制灰狼对解空间的探索能力, $\vec{a}$是线性递减参数, $\vec{r}_1$和$\vec{r}_2$为随机数。

猎物位置更新 为了模拟不同层次个体的协作, GWO分别以α、β和δ 3个位置来指导位置更新, 具体公式如方程18所示。

$$\vec{X}(\mathrm{t}+1)=\frac{\vec{X}_1+\vec{X}_2+\vec{X}_3}{3} \quad (18)$$

攻击猎物 随着搜索迭代的进行, 系数$\vec{A}$逐渐减小至0, 使得灰狼从全局搜索转向局部搜索, 从而实现对解的逐步精细化。

改进策略 在初始的3种算法的预实验(PSO、DE、

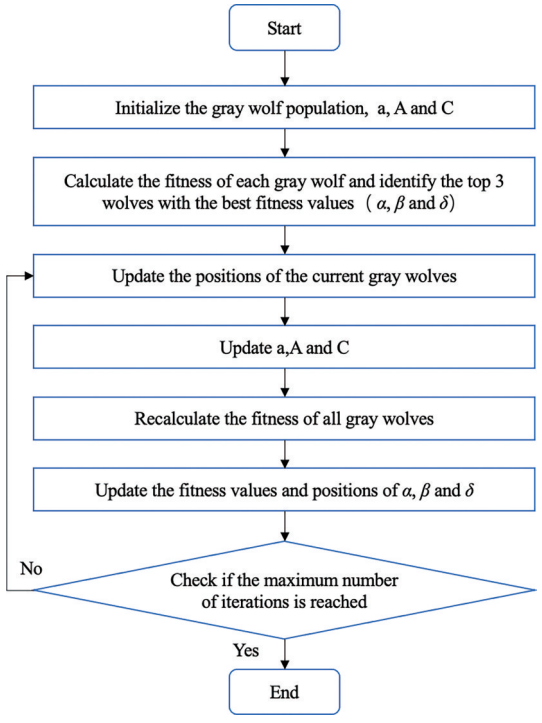


Figure 1 Flowchart of the Grey Wolf algorithm

GWO) 中, GWO 算法在收敛速度、搜索能力指标上表现最佳, 通过对群智能算法的改进 (如优化初始种群、调整迭代公式、增添新策略方法) 可以有效地提升算法的性能, 为进一步提升灰狼算法的优越性, 本研究对其进行了改进和创新。首先, 算法基于线性递减权重策略 W 引入了一个随迭代次数减少的参数 q, 通过动态调整权重和参数的方式[26], 使算法在初期能够保持较大的探索范围以增强全局搜索能力, 而在后期则提高收敛精度, 从而在不同阶段自适应地调整搜索行为, 更好地适应问题特征。其次, 引入了基于莱维飞行的位置更新方法[27], 通过增加随机性和长距离跳跃的特性, 有效提升了搜索过程中的全局探索能力, 使算法能够跳出局部最优解, 从而进一步提高求解效率和优化精度。此外, 改进后的算法在速度更新机制中加入了随机值 r1 和 r2, 这一增强机制使速度更新、更加灵活[28], 能够平衡全局搜索与局部搜索之间的关系, 并且根据当前状态选择是否结合莱维飞行方法, 从而进一步增强搜索的随机性与适应性, 最终显著提升了算法的整体优化性能。

结果

1 参数拟合结果与算法性能对比

在模型建立阶段, 为评估不同优化算法 (PSO、GWO、IGWO、DE) 在细胞培养建模中的性能与算法的差异性, 使用一批 3 L 的细胞培养数据进行参数拟合与模型建立[16]。为了增加算法的可靠性, 各算法均重复运行了 10 次, 每次迭代次数 (iterations) 固定为 100 次。Fitness 随 iterations 的变化被可视化, 收敛曲线的结果如图 2 所示, 当迭代次数达到 100 时, PSO、DE、GWO 和 IGWO 算法的 fitness 数值分别为 20.79、22.86、19.30、15.35。结果表明, IGWO 在优化性能和稳定性方面均优于其他算法; PSO 和 GWO 的优化性能接近 IGWO, 但相比之下稍逊一筹; DE 的性能明显不如 IGWO, 尤其是 DE 的最终适应度值偏高, 表明其对问题的适应能力较弱; 图 2 中阴影区域代表的置信带进一步说明, IGWO 的结果更稳定。

为了更加直观展示 GWO 和 IGWO 在优化过程中粒子 (或候选解) 分布的可视化对比, 图中的每个点表示一个候选解, 其颜色深度代表着迭代次数的增加, 而黑点表示的全局最优解 (global optimum) 的近似解。GWO 粒子的分布范围较广 (图 3A), 集中性较弱, 说明 GWO 的收敛能力有限, 尽管部分粒子较接近全局最优解 (黑点), 但整体分布较分散, 表明算法可能存在搜索精度不足或收敛不充分的问题。而 IGWO 粒子分布范围较小 (图 3B), 明显集中在全局最优解附近, 说明 IGWO 的收敛性能和搜索精度显著提高, 改进后的 IGWO 算法显著增强了在优化问题上的表现, 能够更有效地找到全局最优解, 并提升解的质量。

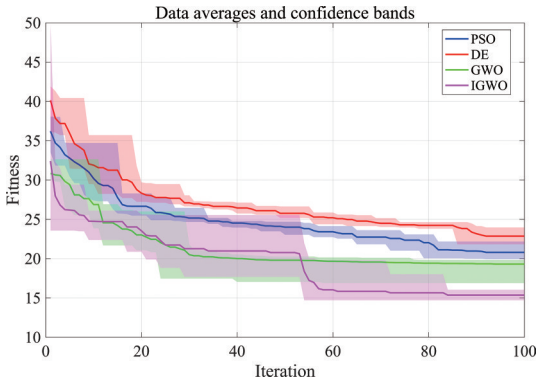


Figure 2 Convergence curves of four different algorithm. PSO: Particle swarm optimization; DE: Differential evolution; GWO: Grey wolf optimization; IGWO: Improved grey wolf optimization

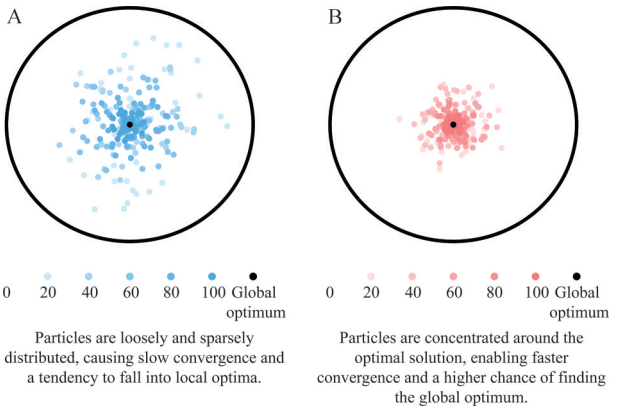


Figure 3 Radiation visualization of particle positions during search process. A: Particle position visualization of the GWO algorithm; B: Particle position visualization of the IGWO algorithm

除了不同算法 (PSO、DE、GWO、IGWO) 会影响算法性能外, 对 IGWO 算法中的关键参数进行设置也可优化算法性能, 如狼群规模、搜索速度、迭代次数、初始化狼群位置、搜索步长等。根据结果可以得出: ① 狼群规模 (图 4A) 在适中范围 (Num = 90) 时, 算法的拟合性能较高且稳定; 在更小规模 (Num = 50) 的情况下, 性能有所下降, 而过大的种群规模 (Num > 110) 虽然仍能保持较高性能, 但计算复杂度也会增加; ② 搜索速度 (图 4B) 对 fitness 值的影响较小, 各值之间差异不显著, 但搜索速度为 0.4 或 0.5 时, fitness 的波动较大, 说明较小的搜索速度可以获得更加稳定的结果; ③ 随着最大迭代次数 (图 4C) 的增加, fitness 值明显下降, 且误差波动极大, 幅度减少。较少的迭代次数 (iter = 50) 会导致算法的结果不理想, 在 iter = 100 附

近, 算法拟合性已达较优程度, 进一步增加迭代次数虽能略微提升性能, 但计算效率显著降低; ④ 不同初始种群分布方式 (图 4D) 对算法性能有显著影响, Random 初始化策略表现出更大的波动性, 而其他策略 (如 Gaussian、Grid、Radial) 则表现得较为稳定, fitness 值接近; ⑤ 步长设置 (图 4E) 的优化效果相对不明显, 采用 Fixed、Decay、Adaptive、Random、Levy 等步长设置所得 fitness 值接近, 说明步长参数的设置对算法结果的影响较弱, 但 Levy 步长下的 fitness 值波动较大, 证明其相比于其他步长设置具备一定跳出局部最优的可能性。

IGWO 算法改进策略通过动态权重调整机制实现搜索行为的自适应优化, 初期大范围探索避免陷入局部最优解, 后期精细收敛确保高维参数空间的精准定位, 特别适合处理多参数耦合情况; 莱维飞行有效避免局部最优提升算法的全局搜索能力和优化效率, 即使在参数间存在强非线性耦合时仍保持良好搜索效率; 在速度更新中引入随机值 r1 和 r2, 结合莱维飞行策略, 进一步优化搜索的随机性和适应性, 有效平衡全局与局部搜索。为了量化各部分对性能提升的贡献, 对改进策略 (动态权重、莱维飞行、随机值) 进行敏感性分析。本研究利用蒙特卡洛模拟这 3 种改进策略的 IGWO 算法的寻优结果, 对比 GWO 的方差和均值变动程度, 判断不同搜索策略对 IGWO 算法的提升效果, 模拟运行 50 次 (每次迭代 100 次)。在稳定性评估中 (图 5A), 3 种策略的权重差异显著, 随机值策略以 58.91% 的权重占比成为主导因素, 表明其对系统稳定性的影响最为关键; 莱维飞行策略的贡献次之; 而动态权重策略呈现异常值为−13.11%, 暗示其参数设置可能对稳定性产生负面影响, 需进一步优化。在平均最优解分析中 (图 5B), 随机值策略仍以 53.08% 的权重保持最高; 莱维飞行与动态权重的权重占比趋于接近, 表明两者对优化结果的影响程度相当。

2 参数拟合结果

使用 IWGO 算法对曲妥珠细胞分批补料培养的过程数据进行参数拟合, 虽然细胞生长的培养过程中温度恒定, 但在实际拟合过程中为保证有更好的拟合效果, 仍以 D11 为界分为两段进行拟合, 25 个关键参数均为拟合求解得到[29]。在实际拟合过程, 第一段参数作为起始拟合参数对整体模型的拟合效果和预测精度有重要影响, 因此本文重点对第一段 (D0～D11) 进行拟合优化以确保模型能够准确捕捉细胞生长和代谢的主要特征。为减少数量级对拟合效果的影响, 计算过程中细胞密度统一以 10^6 为数量级, 代谢物的浓度均以 $mol \cdot L^{-1}$ 作为单位, 模型参数拟合结果如表 1 所示。

图 6 展示了建模批次的实验数据预测结果, 其中

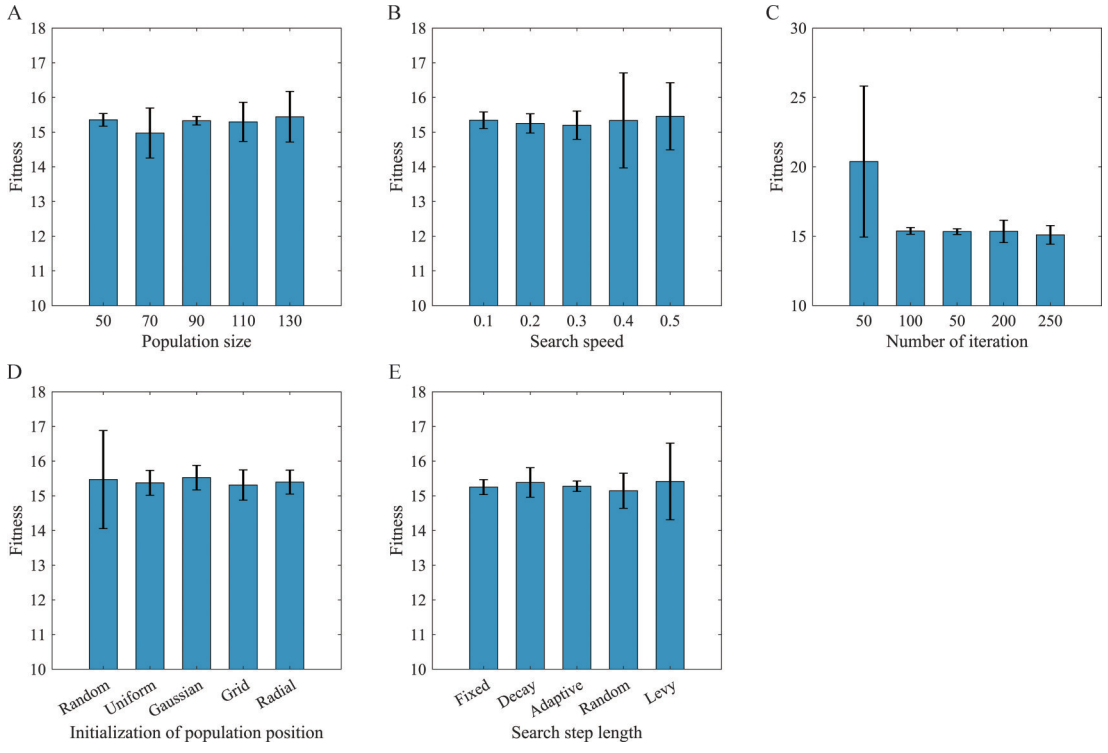


Figure 4 Comparison of fitting performance through different parameter settings. A: Comparison results for different wolf pack size settings; B: Comparison results for different search speed parameter settings; C: Comparison results for different iteration number settings; D: Comparison results for different initial wolf position settings; E: Comparison results for different search step size settings

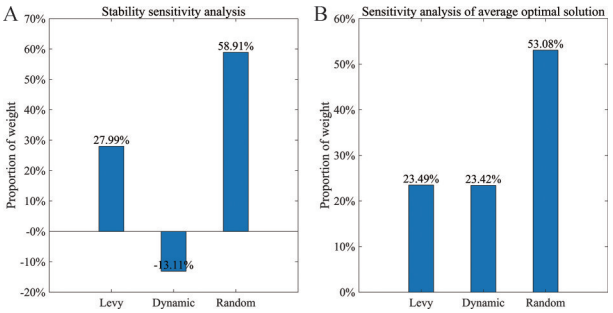


Figure 5 Sensitivity analysis of IGWO improvement strategy. A: Stability sensitivity analysis; B: Sensitivity analysis of average optimal solution

黑圆点表示实际取样数据, 蓝圆点为模型预测值, 绿圆点则对应补料操作后的实验数据。对于 Vcd 和 IgG (图 6A、G), 模型的预测效果优秀, 曲线与实验数据高度吻合。对于 Glc 和 Glu (图 6B、F), 模型较好地反映了补料操作对浓度的影响。然而, 对于 Glc 而言, 模型预测存在一定偏差, 实际浓度并不会为零。这种偏差主要源于模型参数拟合中的权重调整以及数学公式的过度简化, 这些简化限制了机理模型的预测能力[30]。

对于 Lac、Amm、Gln (图 6C～E), 模型能够较好地捕捉到整体趋势, 但局部存在一定误差, 这可能是因为模型未能完全反映细胞行为的变化[31], 需要从生化代谢的角度进一步修正模型[32]。此外, 由于取样方法与生化仪采用酶标法的缘故, 存在一定程度的校准误差, 可能会导致模型的拟合结果误差增加。

3 模型验证

为验证模型的适用性, 选取了不同起始浓度的培养批次进行模型验证, 评估该模型在其他批次上的预测效果, 采用 R^2 和 RMSE 作为主要的评价指标, 对不同批次数据的模型性能进行全面评估。建模批次与模型验证批次的主要工艺差别如表 2 所示, 其中 A 批次的数据用于参数拟合与模型建立, B 批次与 A 批次工艺相同, 均为 1∶1 传代接种, 并分别在实验的第 2、4、6、8、10、12、14 天补料。批次 C、D 为 1∶3 传代接种, 初始密度低于 A、B 批次, 并分别在实验的第 3、5、7、9、11、13 天补料。

Table 1 Model parameter fitting results

Parameter/unit	Parameter meaning	D0−D11	D12−D16
$\mu_{G,max}$/h^{-1}	Maximum specific growth rate	0.27	0.17×10^{-2}
$k_{d,max}$/h^{-1}	Maximum specific death rate	0.024	0.05
KI_{Glc}/mmol·L^{-1}	Cell growth Monod constant for glucose	10.88	29.35
KI_{Lac}/mmol·L^{-1}	Cell growth Monod constant for lactate	2.40	42.18
KI_{Amm}/mmol·L^{-1}	Cell growth Monod constant for ammonium	71.08	95.48
KD_{Glc}/mmol·L^{-1}	Cell death Monod constant for glucose	25.09	326.63
KD_{Lac}/mmol·L^{-1}	Cell death Monod constant for lactate	0.46	6.06
KD_{Amm}/mmol·L^{-1}	Cell death Monod constant for ammonium	6.77	45.71
m_{Glc}/mmol^{-1}·cells·h^{-1}	Glucose maintenance coefficient	0.00	1.60×10^{-4}
$YF_{x,Glc}$/mmol^{-1}·cells	Yield coefficient of cell growth to glucose consumption	0.46×10^{-2}	0.34×10^{-2}
$YF_{Lac,Glc}$/mmol·mmol^{-1}	Yield coefficient of lactate formation to glucose uptake	1.21	0.42×10^{-2}
$YF_{x,Amm}$/mmol^{-1}·cells	Yield coefficient of cell growth to ammonium formation	8.13×10^{-4}	0.02
k_{Lac}/mmol·cells^{-1}·h^{-1}	Lactate consumption rate	7.28×10^{-4}	1.58×10^{-4}
K_{Lac}/mmol·L^{-1}	Michaelis constant for lactate	101.20	82.64
k_{Amm}/mmol·cells^{-1}·h^{-1}	Ammonium consumption rate	1.40×10^{-5}	0.00
K_{Amm}/mmol·L^{-1}	Michaelis constant for ammonium	0.15	0.07
KI_{Glu}/mmol·L^{-1}	Cell growth Monod constant for glutamate	0.03	2.47
KD_{Glu}/mmol·L^{-1}	Cell death Monod constant for glutamate	47.55	154.29
m_{Gln}/mmol·cells^{-1}·h^{-1}	Glutamine maintenance coefficient	0.00	3.18×10^{-7}
$YF_{x,Gln}$/mmol^{-1}·cells^{-1}	Yield coefficient of cell growth to glutamine consumption	2.22×10^{-4}	1.65×10^{-4}
$YF_{Gln,Glu}$/mmol·cells^{-1}	Yield coefficient of glutamine formation to glutamate uptake	0.13×10^{-2}	1.04
m_{Glu}/mmol·cells^{-1}·h^{-1}	Glutamate maintenance coefficient	1.85×10^{-6}	4.13×10^{-6}
$YF_{x,Glu}$/mmol^{-1}·cells	Yield coefficient of cell growth to glutamate consumption	0.00	4.27×10^{-4}
$Q_{p,max}$/mg·cells^{-1}·h^{-1}	Maximum specific productivity	5.58×10^{-4}	0.15×10^{-2}
KS_{Gln}/mmol·L^{-1}	Protein production saturation constant of glutamine	0.00	0.00

Table 2 The main process differences between the modeling batch and the model validation batch

Batch ID	Seed density (10^6 cells mL^{-1})	Max-Vcd (10^6 cells mL^{-1})	Passage ratio of N generations	Days of cell culture	Feeding strategy
A	1.19	27.2	1:1	16	D2/4: 4% CD1005-A + 0.4% CD1005-B; D6/8/10/12/14: 5% CD1005-A + 0.5% CD1005-B
B	1.10	20.5	1:1	16	
C	0.79	20.1	1:3	15	D3/5: 4% CD1005-A + 0.4% CD1005-B; D7/9/11/13: 5% CD1005-A + 0.5% CD1005-B
D	0.61	17.1	1:3	15	

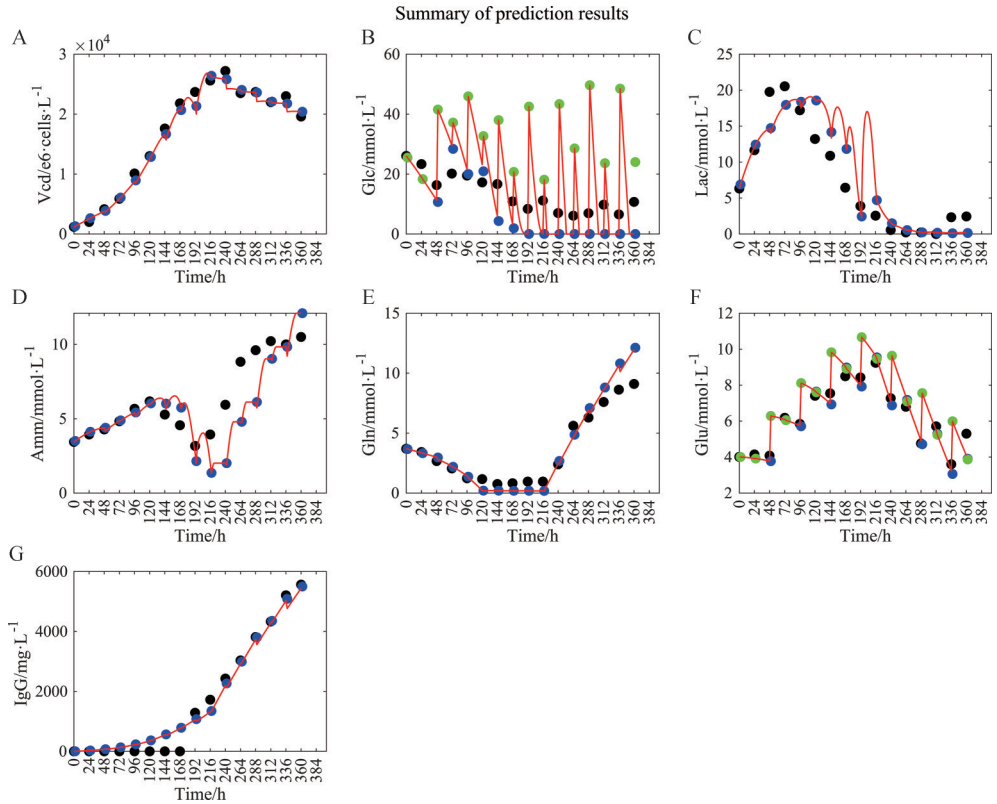


Figure 6 Experimental data and fitting results for fed batch cell culture process. A: Prediction results for Vcd; B: Prediction results for Glc; C: Prediction results for Lac; D: Prediction results for Amm; E: Prediction results for Gln; F: Prediction results for Glu; G: Prediction results for IgG. Vcd: Viable cell density; Glc: Glucose; Lac: Lactate; Amm: Ammonia; Gln: Glutamine; Glu: Glutamate

整体来看, 模型的预测能力稳定, 预测值与测量值之间具有较高的一致性。A 批次与 B 批次 (图 7A、B) 的 R^2 和 RMSE 结果表明模型的拟合效果较好, 预测值与实际值之间有较高的线性相关性。在 C 与 D 批次上 (图 7C、D), 即使初始接种密度与补料工艺有所变化, 模型仍然保持了较强的预测能力, 表明模型在应对工艺变化时具备高度的稳健性和适应性。

讨论

本研究针对曲妥珠单抗细胞培养过程建立了机理模型, 结果表明机理模型可以很好地描述细胞密度、代谢物与产物浓度等随着时间的变化, 并且很好地体现了补料操作带来的浓度变化与影响。在参数估计方面, 创新的 IGWO 算法在收敛速度、适应度优化及解空间搜索能力方面均优于其他算法, 特别是在平衡全局搜索与局部搜索能力方面表现出色, 为微分方程多参数优化问题提供了新思路。此外, IGWO 算法在处理多维、高复杂度的生物过程建模时, 能够高效克服传统算法可能面临的局部最优解和计算效率低下的问题。

但由于在实际培养过程中, 环境变量如 pH 值、溶氧值、转速、温度的改变[29], 使得机理模型的适用性和

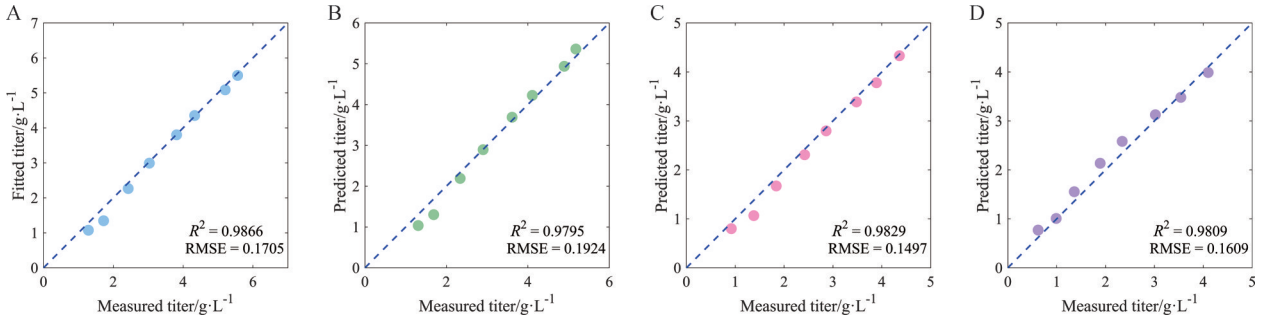


Figure 7 Predicted results of model validation. A: Antibody titer validation for Batch A; B: Antibody titer validation for Batch B; C: Antibody titer validation for Batch C; D: Antibody titer validation for Batch D

泛化能力受到挑战, 纳入环境变量的机理模型[33]或混合模型[34]将是未来更深一步的研究方向。

作者贡献: 王文慧负责文献调研及代码编写; 万宇翔、焦静雨负责培养细胞; 高丹、高栋负责文章修改; 瞿海斌负责论文选题和拟定论文框架指导、把关。

利益冲突: 所有作者声明不存在利益冲突。

References

[1] Kyriakopoulos S, Ang KS, Lakshmanan M, et al. Kinetic modeling of mammalian cell culture bioprocessing: the quest to advance biomanufacturing [J]. Biotechnol J, 2018, 13: 1700229.

[2] Zhao CX, Hu ZW, Cui B. Recent advances in monoclonal antibody-based therapeutics [J]. Acta Pharm Sin (药学学报), 2017, 52: 837-847.

[3] Wang J, Wu SG. Breast cancer: an overview of current therapeutic strategies, challenge, and perspectives [J]. Breast Cancer, 2023, 15: 721-730.

[4] Kirshner G, Makai P, Brouns C, et al. The impact of an 'evergreening' strategy nearing patent expiration on the uptake of biosimilars and public healthcare costs: a case study on the introduction of a second administration form of trastuzumab in The Netherlands [J]. Eur J Health Econ, 2024, 25: 1147-1163.

[5] Cui J, Han DM, Xu LC, et al. The discussion on a proposed quality "similarity assessment criteria" of trastuzumab biosimilar [J]. Acta Pharm Sin (药学学报), 2021, 56: 3190-3197.

[6] Polanco A, Liang G, Park S, et al. Trace metal optimization in CHO cell culture through statistical design of experiments [J]. Biotechnol Prog, 2023, 39: e3368.

[7] Fu Z, Wang Z, Chen G. Enhanced parameter estimation with improved particle swarm optimization algorithm for cell culture process modeling [J]. AIChE J, 2024, 70: e18388.

[8] Bayer B, Duerkop M, Striedner G, et al. Model transferability and reduced experimental burden in cell culture process development facilitated by hybrid modeling and intensified design of experiments [J]. Front Bioeng Biotechnol, 2021, 9: 740215.

[9] Ladiwala P, Dhara VG, Jenkins J, et al. Addressing amino acid-derived inhibitory metabolites and enhancing CHO cell culture performance through DOE - guided media modifications [J]. Biotechnol Bioeng, 2023, 120: 2542-2558.

[10] Zhou T, Reji R, Kairon RS, et al. A review of algorithmic approaches for cell culture media optimization [J]. Front Bioeng Biotechnol, 2023, 11: 1195294.

[11] Ramin E, Cardillo AG, Liebers R, et al. Accelerating vaccine manufacturing development through model-based approaches: current advances and future opportunities [J]. Curr Opin Chem Eng, 2024, 43: 100998.

[12] Sha S, Huang Z, Wang Z, et al. Mechanistic modeling and applications for CHO cell culture development and production [J]. Curr Opin Chem Eng, 2018, 22: 54-61.

[13] Pennington O, Ríos SE, Sebastian MT, et al. Dynamic multiscale hybrid modelling of a CHO cell system for recombinant protein production [J]. IFAC Pap OnLine, 2024, 58: 133-138.

[14] Khuat TT, Bassett R, Otte E, et al. Applications of machine learning in antibody discovery, process development, manufacturing and formulation: current trends, challenges, and opportunities [J]. Comput Chem Eng, 2024, 182: 108585.

[15] Pinto J, Ramos JRC, Costa RS, et al. Hybrid deep modeling of a CHO-K1 fed-batch process: combining first-principles with deep neural networks [J]. Front Bioeng Biotechnol, 2023, 11: 1237963.

[16] Selișteanu D, Șendrescu D, Georgeanu V, et al. Mammalian cell culture process for monoclonal antibody production: nonlinear modelling and parameter estimation [J]. Biomed Res Int, 2015, 2015: 598721.

[17] Elnady SM, El-Beltagy M, Radwan AG, et al. A comprehensive survey of fractional gradient descent methods and their convergence analysis [J]. Chaos Solitons Fractals, 2025, 194: 116154.

[18] Li W, Wang GG, Gandomi AH. A survey of learning-based intelligent optimization algorithms [J]. Arch Comput Methods Eng, 2021, 28: 3781-3799.

[19] Tang J, Liu G, Pan Q. A review on representative swarm intelligence algorithms for solving optimization problems: applications and trends [J]. IEEE/CAA J Autom Sin, 2021, 8: 1627-1643.

[20] Wang Z, Wang C, Chen G. Kinetic modeling: a tool for temperature shift and feeding optimization in cell culture process development [J]. Protein Expr Purif, 2022, 198: 106130.

[21] Fernandes S, Richelle A, Amribt Z, et al. Extended and unscented Kalman filter design for hybridoma cell fed-batch and continuous cultures [J]. IFAC Pap OnLine, 2015, 48: 1108-1113.

[22] Okamura K, Badr S, Ichida Y, et al. Modeling of cell cultivation for monoclonal antibody production processes considering lactate metabolic shifts [J]. Biotechnol Prog, 2024, 40: e3486.

[23] Makhadmeh SN, Al-Betar MA, Doush IA, et al. Recent advances in grey wolf optimizer, its versions and applications: review [J]. IEEE Access, 2024, 12: 22991-23028.

[24] Liu Y, As'arry A, Hassan MK, et al. Review of the grey wolf optimization algorithm: variants and applications [J]. Neural Comput Appl, 2024, 36: 2713-2735.

[25] Dada E, Joseph S, Oyewola D, et al. Application of grey wolf optimization algorithm: recent trends, issues, and possible horizons [J]. Gazi Univ J Sci, 2022, 35: 485-504.

[26] Wang S, Zhu D, Zhou C, et al. Improved grey wolf algorithm based on dynamic weight and logistic mapping for safe path planning of UAV low-altitude penetration [J]. J Supercomput, 2024, 80: 25818-25852.

[27] Chen Y, Xi J, Wang H, et al. Grey wolf optimization algorithm based on dynamically adjusting inertial weight and levy flight strategy [J]. Evol Intell, 2023, 16: 917-927.

[28] Long W, Jiao J, Liang X, et al. A random opposition-based learning grey wolf optimizer [J]. IEEE Access, 2019, 7: 113810-113825.

[29] Xing Z, Nguyen TB, Kanai-Bai G, et al. Construction of a novel kinetic model for the production process of a CVA6 VLP vaccine in CHO cells [J]. Cytotechnology, 2024, 76: 69-83.

[30] Chiu KC, Du D. A neural ordinary differential equation model for predicting the growth of Chinese hamster ovary cell in a bioreactor system [J]. Biotechnol Bioprocess Eng, 2025, 30: 100-115.

[31] Lu Y, McCann MG, Hu W, et al. Multi-cell-line learning for the data-driven construction of mechanistic metabolic models [J]. Biotechnol Bioeng, 2024, 121: 2833-2847.

[32] Mowbray MR, Wu C, Rogers AW, et al. A reinforcement learning-based hybrid modeling framework for bioprocess kinetics identification [J]. Biotechnol Bioeng, 2023, 120: 154-168.

[33] Doyle K, Tsopanoglou A, Fejér A, et al. Automated assembly of hybrid dynamic models for CHO cell culture processes [J]. Biochem Eng J, 2023, 191: 108763.

[34] Yang S, Fahey W, Truccollo B, et al. Hybrid modeling of fed-batch cell culture using physics-informed neural network [J]. Ind Eng Chem Res, 2024, 63: 16833-16846.