机器学习在药物分子结构优化中的研究进展

李星雨[1], 武 潼[1], 徐启兰[1], 王奕文[1], 余河水[1,2,3], 岑世欣[1,2,3*], 李 正[1,2,3,4,5*]

(1. 天津中医药大学中药制药工程学院, 天津 301617; 2. 组分中药国家重点实验室, 天津 301617; 3. 天津市中药智能制药与绿色制药重点实验室, 天津 301617; 4. 现代中医药海河实验室, 天津 301617; 5. 现代中药创制全国重点实验室, 天津 301617)

摘要: 药物分子结构优化是提高药物有效性与安全性的关键环节, 其主要目标是通过合理地调整分子结构, 来满足研发过程中对药物综合性能的严格需求。然而, 庞大的药物化学空间和复杂的多目标优化任务使得这一过程面临显著挑战。近年来, 以深度学习和强化学习为代表的机器学习技术由于其强大的信息挖掘、特征提取以及非线性拟合能力, 已被广泛应用于药物分子结构优化中, 为药物设计提供了新的思路和方法。本文系统梳理了机器学习在药物分子结构优化中的最新研究进展, 深入分析并比较其在不同结构优化场景中的应用优势及特色, 总结了当前药物分子结构优化面临的主要挑战, 并就结构优化未来可能发展的研究方向提供了技术视角。

关键词: 药物设计; 分子结构优化; 机器学习; 深度学习; 强化学习

中图分类号: R914 **文献标识码**: A **文章编号**: 0513-4870(2025)10-2963-20

Advances in machine learning for drug molecule structure optimization

LI Xing-yu[1], WU Tong[1], XU Qi-lan[1], WANG Yi-wen[1], YU He-shui[1,2,3], CEN Shi-xin[1,2,3*], LI Zheng[1,2,3,4,5*]

(*1. College of Pharmaceutical Engineering of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China; 2. State Key Laboratory of Component-based Chinese Medicine, Tianjin 301617, China; 3. Tianjin Key Laboratory of Intelligent and Green Pharmaceuticals for Traditional Chinese Medicine, Tianjin 301617, China; 4. Haihe Laboratory of Modern Chinese Medicine, Tianjin 301617, China; 5. State Key Laboratory of Chinese Medicine Modernization, Tianjin 301617, China*)

Abstract: The optimization of drug molecular structure is a crucial step in enhancing the efficacy and safety of drugs. Its primary objective is to rationally adjust the molecular structure to meet the strict requirements for the comprehensive performance of drugs during the research and development process. However, the vast chemical space of drugs and the complex multi-objective optimization tasks pose significant challenges to this process. In recent years, machine learning techniques, represented by deep learning and reinforcement learning, have been widely applied in the optimization of drug molecular structures due to their powerful capabilities in information mining, feature extraction, and nonlinear fitting, providing new ideas and methods for drug design. This paper systematically reviews the latest research progress of machine learning in the optimization of drug molecular structures, deeply analyzes and compares their application advantages and characteristics in different structural optimization scenarios, summarizes the main challenges currently faced in the optimization of drug molecular structures, and provides a technical perspective on the possible research directions for future structural optimization.

Key words: drug structure design; molecular structure optimization; machine learn-ing; deep learning; reinforcement learning

收稿日期: 2025-04-14; 修回日期: 2025-06-05.

基金项目: 现代中药新质生产力科技创新工程专项 (24ZXZKSY00010); 天津市科技计划项目 (24ZXZSSS00460).

*通讯作者 Tel: 15522200298, E-mail: censhixin@tjutcm.edu.cn;

Tel: 15868496961, E-mail: lizheng@tjutcm.edu.cn

DOI: 10.16438/j.0513-4870.2025-0412

药物研发是一项高风险、高成本的复杂系统工程，通常需要消耗数十亿美元的资金并经历长达10余年的周期，才能将一种新药推向市场[1]。经研究数据显示，从新药发现至最终获批的全流程中，仅约10%的候选药物能通过临床试验，其余的候选药物则会因其临床疗效不足、毒性难以控制或药物不良反应等原因而淘汰[2]。在此过程中，药物分子结构优化作为新药研发的关键环节，其目的是通过合理调整药物分子结构，在早期研发阶段提高候选药物的活性、选择性及药代动力学性能（如吸收、分布、代谢、排泄和毒性，ADMET），提高研发效率。索拉菲尼和奥美拉唑的成功开发是药物分子优化策略的典型案例。3-噻吩基脲是索拉菲尼的先导化合物，由于其选择性和活性较差，限制了其作为抗肿瘤药物的应用潜力。通过基于结构的药物设计，优化其分子骨架，增强了其对酪氨酸激酶靶点的结合能力，并在苯环对位引入甲基，成功提高了10倍活性。最终，索拉菲尼成为第一个多靶点酪氨酸激酶抑制剂，广泛用于治疗晚期肾细胞癌[3]。类似地，奥美拉唑的先导化合物具有强效的抗胃酸作用，但对肝脏和甲状腺有较强的毒性。经过多次分子结构优化和修饰，稳定活性代谢中间体，在保持抗胃酸分泌效力的同时显著降低肝毒性。奥美拉唑后来作为首个质子泵抑制剂成功上市，用于胃食管反流病和胃溃疡的治疗[4]。这两个案例不仅验证了药物分子优化在提高新药研发成功率和降低研发风险方面的关键作用，也为未来药物设计提供了重要的理论支持与技术借鉴。

然而，传统药物分子结构优化在很大程度上依赖于药物化学家的经验，这一过程通常需要大量人力和物力的投入[5]。由于药物化学空间极度庞大（约存在10^{60}种潜在分子结构），在有限时间内设计出具有低毒性和高效性的理想分子结构是一项极具挑战性的任务[6]。同时，药物分子优化还面临多目标权衡的难题，例如，在优化药物的疏水性、溶解度、选择性和活性的同时，如何协同降低其毒性，这进一步增加了优化过程的复杂性和技术难度[7]。此外，实验数据的质量和稀缺性也对研究进展形成了限制。因此，开发一种能够兼顾计算速度、预测准确性和药物性能的先导化合物优化方法，已成为药物分子优化领域的迫切需求[8]。

为了推动药物分子结构优化领域的发展，机器学习（machine learning, ML）算法的引入为药物结构优化提供了强有力的技术支持。机器学习已成为推动药物设计与药物分析领域发展的重要技术手段。基于Web of Science数据库，本文以“机器学习”，“深度学习”，“强化学习”和“分子优化”等关键词检索了全球范围内的相关文献，并使用CiteSpace 6.4软件对检索数据进行可视化分析，生成了研究热点的网络图和关键词突现图（图1A、B）。研究表明，机器学习在药物分子结构优化领域展现出巨大潜力，其核心在于通过高效建模和模拟复杂分子结构，实现对化学和生物数据的精准挖掘与利用。这些数据涵盖了公开数据库（如ChEMBL、PubChem、ZINC）中的药物化学数据、药物生物数据、药物实验数据等，构成了机器学习技术应用的关键数据支撑。此外，包括深度学习（deep learning, DL）和强化学习（reinforcement learning, RL）等在内的机器学习技术，能够高效提取分子图结构、解析简化分子线性输入规范（simplified molecular input line entry system, SMILES）、处理自参照嵌入字符串（self-referencing embedded strings, SELFIES）及剖析分子化学特征。借此，可显著提升分子属性预测、多目标优化与分子生成的有效性和准确性，为药物研发提供坚实有力的技术支撑[9]。

本文系统综述了深度学习和强化学习算法在药物分子优化领域的实际应用，重点分析了其在不同场景中的具体成功案例与研究进展。具体而言，本文详细阐述了大语言模型、图神经网络以及强化学习算法的基本原理，梳理了算法的演进脉络并讨论它们在药物分子生成、属性预测和多目标优化任务中的典型应用。这些算法通过突破传统分子优化在化学空间探索效率、多目标权衡精度及属性预测能力上的瓶颈，实现了从单一属性优化到复杂系统建模的跨越；大语言模型通过化学语言与自然语言结合，推动分子表征从静态符号体系向动态空间进化；图神经网络借助图结构数据建模，实现了从简单拓扑特征提取到复杂分子构象预测、从单任务学习到多目标优化的技术跃升；强化学习则通过智能体与化学空间的交互，将分子生成从随机搜索转变为目标导向的策略优化。此外，本文还深入分析了当前算法在实际应用中所面临的主要挑战，如数据稀缺性、立体结构建模不足、计算成本高昂等。在此基础上，进一步展望了这些算法的未来发展方向，并提出应对现有问题的可能解决方案。通过上述分析，本文旨在为读者提供清晰的学术脉络和技术启示，以期为提升新药研发效率和推动创新药物开发贡献新的视角与思路。

1 深度学习算法在药物分子优化中的应用

深度学习是机器学习的重要分支，近年来在图像识别、无人驾驶、自然语言处理及医药领域取得了突破性进展，并在各类复杂任务中展现出卓越的性能[10]。深度学习这一概念最早可以追溯至1986年，并在2000年逐步应用于人工神经网络[11]。2012年，随着AlexNet模型在ImageNet大规模视觉识别竞赛中取得的成功，

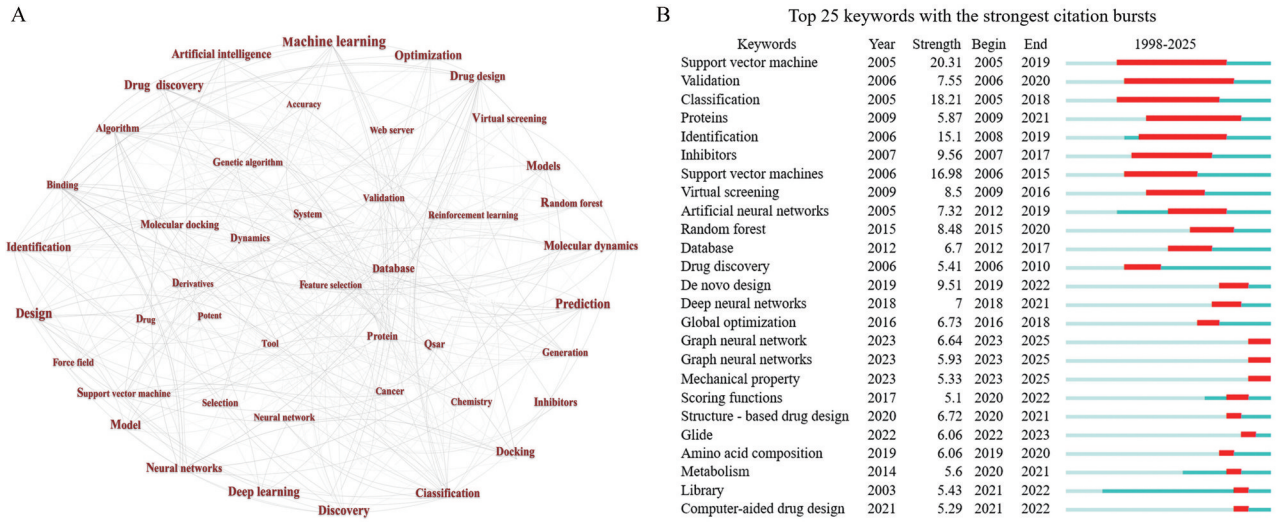


Figure 1 Visualization of keyword analysis graph. A: Network graph of research hotspots; B: Graph of keyword prominence analysis

深度学习正式进入主流研究领域, 掀起了计算机视觉领域的革命[12,13]。这一发展为深度学习在其他领域的广泛应用奠定了坚实基础。

深度学习通过多层非线性信息处理, 对高维数据的特征表示和复杂关系进行建模与预测[14,15]。相较于传统机器学习, 深度学习在多个方面具有显著优势。首先, 深度学习在处理大规模数据集时表现尤为突出, 而传统方法则在小样本数据集上相对具有优势[16]; 其次, 深度学习能够自动提取数据的深层次特征, 无需人工特征工程, 大幅提升了模型的表征学习能力[17]。此外, 通过复杂的深层网络结构, 深度学习能够有效捕捉非线性关系, 使其在应对复杂任务时的性能显著优于传统机器学习方法 (如线性回归、支持向量机、决策树等)[18]。尽管深度学习对计算资源需求更高, 但其在计算机视觉、自然语言处理等高复杂场景中的应用价值远超过传统方法, 已成为推动科技进步的关键技术手段[13-15,19-21]。

根据算法特性及其在药物分子优化中的应用范式, 深度学习算法模型可以归纳为生成式模型 (generative models)、判别式模型 (discriminative models) 和图神经网络模型 (graph neural networks, GNN) 3大类。其中, 生成式算法模型如生成对抗网络 (generative adversarial networks, GAN)、变分自编码器 (variational autoencoder, VAE) 和变换器 (Transformer) 通过捕捉化学空间的隐式分布实现分子设计与结构优化; 判别式算法模型如卷积神经网络 (convolutional neural networks, CNN)、循环神经网络 (recurrent neural network, RNN) 侧重于分子性质的精准预测与分类; 图神经网络如图卷积网络 (graph convolutional network, GCN)、图注意力网络 (graph attention network, GAT) 等则基于图结构数据表征, 解析分子拓扑与化学键之间的复杂关系[13,14,22-25]。近年来, 基于Transformer结构的大语言模型与图神经网络模型的新型技术显著推动了药物分子优化的范式革新—前者通过模拟化学语言 (SMILES或SELFIES等) 的生成逻辑来实现多目标分子优化设计, 后者则通过学习原子与键之间的关系来预测分子的药效及毒性, 二者的飞速发展为化学空间的高效探索与多目标优化提供了新的方法论[26-28]。

1.1 大语言模型在药物分子优化中的应用

大语言模型 (large language model, LLM) 是基于深度学习和自然语言处理技术开发的系统, 通过大规模训练使其能够理解和生成自然语言文本。Transformer结构是大语言模型的核心架构, 由自注意力机制、多头注意力机制、前馈网络、残差连接和层归一化及位置编码等模块组成, Transformer架构如图2所示。

1.1.1 大语言模型驱动分子优化演进过程

大语言模型在自然语言处理及其他领域展现出的强大文本生成与理解能力, 为药物分子优化领域提供了新的技术范式。基于Transformer结构的大语言模型, 经过不断改进, 逐渐进入药物研发领域, 推动了分子数据处理和药物优化的创新。这种技术的进步不仅限于文本生成, 还涉及到药物分子与自然语言的深度融合, 使分子优化过程更加精准高效[29]。

在分子优化范式演进过程中, 药物分子优化的研究重点始终围绕化学空间的探索与属性约束的满足两大核心展开。基于大语言模型的分子优化技术演进过程如图3所示, 经历了从启发式规则到数据驱动、再到知识蒸馏的三阶段跃迁, 其本质是分子表征学习与生成策略的协同进化。在这一背景下, 化学语言的同构性建模、在多模态约束的联合嵌入, 以及零样本知识蒸

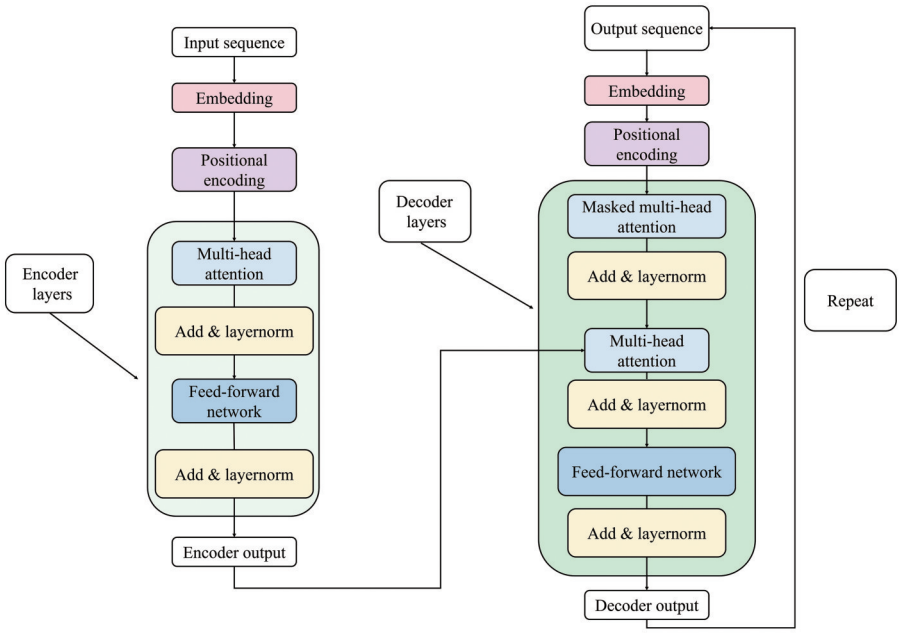


Figure 2 Transformer architecture diagram

馏框架等技术逐步成为主流, 极大地推动了药物分子设计与优化的高效化和精准化[30]。

1.1.2 化学语言的同构性建模技术

Bran 等[31]首次提出化学语言理论框架, 证明了 SMILES 或 SELFIES 等分子线性表征与自然语言在离散符号集、语法规则、序列性和语义性等方面具有一定共性。基于这一理论, 后续研究开始利用 Transfomer 架构来处理分子表示, 并在分子优化中采用自回归生成和注意力机制来建模化学空间的特征分布。然而, 将 Transformer 直接应用于这些分子线性表示进行生成任务, 以期发现新的药物分子, 面临两大核心挑战, 这些挑战直接影响了模型学习分子结构的能力以及生成分子的质量和优化效率。

其一, SMILES 等线性表征存在固有的顺序依赖性, 即同一个分子结构可能对应多个不同的 SMILES 字符串, 例如,“CCO”和“OCC”都代表乙醇分子[32]。这

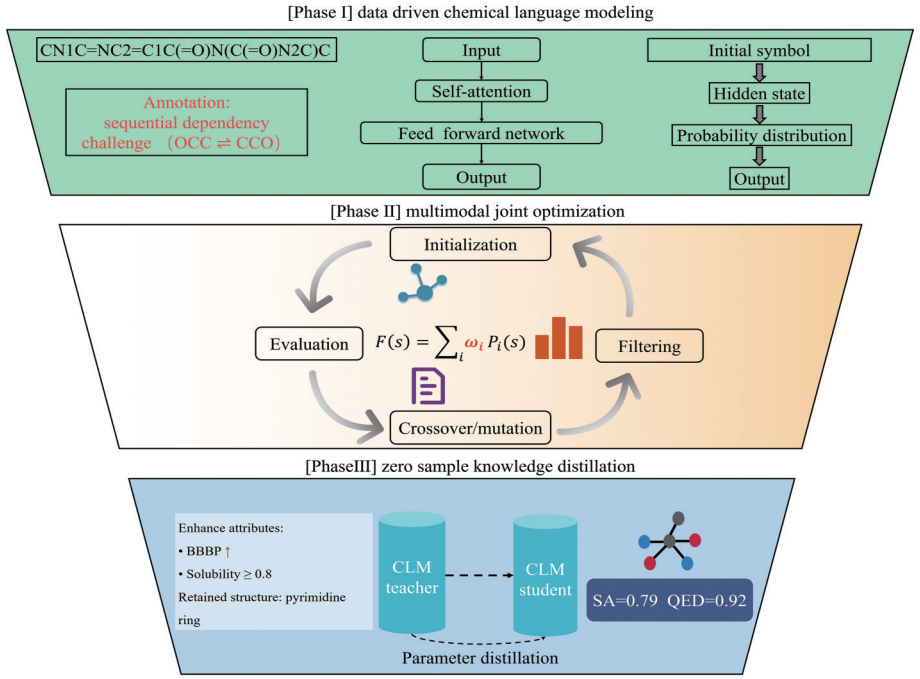


Figure 3 Schematic diagram of the evolution of the molecular optimization paradigm driven by large language models. BBBP: Blood-brain barrier penetration; CLM: Chemical language model; SA: Synthetic accessibility; QED: Quantitative estimate of druglikeness

种现象使得模型在训练过程中需要学习同一个分子的不同表示方式, 导致学习难度增加、模型训练收敛不稳定。甚至在分子生成的过程中, 由于模型难以掌握所有分子的等价表示, 使得生成的SMILES表达式出现语法错误, 从而生成大量无效分子。在早期SMILES序列生成模型的研究中, 研究人员发现即便用的训练数据集全部由有效分子构成, 模型输出的分子仍难以实现100%有效性。事实上, 部分模型生成有效分子的比例甚至低于50%, 这一现象导致了计算资源的严重损耗与生成效率的显著下降[33]。这种低效率不仅延长了药物研发等相关领域的探索周期, 也使得大量计算资源未能得到充分有效的利用, 阻碍了基于SMILES序列生成技术的快速发展与实际应用。

其二, 分子空间具有离散性, 主要表现在原子类型固定、化学键类型限定和结构变化非连续等方面。然而, Transformer的连续向量嵌入机制与分子结构的离散特性存在本质性矛盾[34], 这种表示鸿沟使得模型在连续的潜在空间进行探索时, 难以准确的映射到有效的、期望的离散化学结构[35]。虽然精调奖励函数和采样策略可缓解此问题, 但基于SMILES的Transformer模型在生成任务中仍面临严峻考验。

1.1.3 多模态约束的联合嵌入技术

多模态学习通过整合不同数据模态 (如文本、图像、图结构) 的特征表示, 能够更全面地捕捉复杂系统的内在规律。在药物分子优化中, 这一方法尤为关键, 它结合了化学语言、分子图拓扑结构、实验数据等信息, 提供了更为全面的优化视角。此类方法可以在多个目标之间进行平衡, 如分子的药效、溶解度和毒性等, 为药物发现提供更加精确的优化策略[36]。

1.1.3.1 基于大语言模型与进化策略的联合优化 在多模态学习的背景下, Guevorguian等[37]利用大语言模型开发了Chemlactica与Chemma模型, 其核心在于将大语言模型的分子表示学习和生成能力与群体进化策略相结合, 从而高效地进行分子优化和化学空间探索。与遗传算法原理相似, 该方法维护了一个包含P个高性能分子的池 (Pool)。在每一次迭代进化中, 从随机Pool中选择S个分子作为“父本”, 利用训练好的大语言模型, 根据群体中的分子信息, 生成N个新的“子代”分子。后续通过一个外部评价函数 (black-box oracle) 来评估新生分子的特定化学性质并给出评分。根据新的子代分子的评分, 更新Pool, 并只保留评分最高的P个分子。这种迭代的过程使得分子群体逐渐向目标属性最优化的区域靠近[38]。

以优化分子对多巴胺受体Drd2的结合能力为例, 这是一个典型的多目标属性分子优化问题。其优化目标是最大化分子与Drd2靶点的对接评分, 同时要满足分子质量小于500 Da和高药物相似性评分。该研究团队使用Chemlactica与Chemma模型进行此任务。为了量化评估模型的优化效率和生成分子的质量, 实验采用生成产量和Oracle调用负载等指标。研究结果显示, 在要求分子评分达到0.7及以上时, Chemma-2B模型以3 848的生成量显著优于基线方法 (Beam Structure15: 3 474; Reinvent: 1 879)。甚至在更严格的0.8评分阈值下, Chemma-2B模型的生成产量依然最高, 达到2 985。并且, Chemlactica-125M和Chemlactica-1.3B也显著优于基线方法。在寻找高潜力分子的Oracle调用计算负载指标评估中, Chemlactica-125M和Chemlactica-1.3B模型均表现出色。在找到1个评分高于0.8的分子的目标下, Chemlactica-125M平均仅用20次Oracle调用, 而基线方法Beam Structure15需要126次, Reinvent需要168次。

这些定量结果表明, Chemlactica与Chemma模型结合进化策略在Drd2分子优化中能够更有效地生成高评分分子并显著降低所需的评价成本, 展现出其在探索复杂化学空间方面的强大潜力, 为分子优化和生成提供了新的技术手段。

1.1.3.2 基于自然语言指令与分子图的联合嵌入 为突破传统多模态优化中静态权重分配的瓶颈, 研究者引入指令调优技术, 将自然语言指令与化学语言进行深度融合, 开辟了多属性约束下分子优化的新范式[39]。基于自然语言的指令描述可将定量属性约束 (如溶解度、靶标亲和力) 嵌入到分子生成过程中, 实现跨模态的知识迁移。

传统分子优化模型在处理多目标约束时需依赖手工设计的权重函数 (如ω_i系数), 难以动态适应复杂的药性平衡需求。对此, Ye等[40]提出的DrugAssist框架, 首次采用自然语言指令作为属性约束的隐式嵌入空间。通过构建分子优化专用指令数据集, 将溶解度、血脑屏障穿透性等6大属性的定量约束转化为可微调的语言指令。在药物研发相关的场景下, 研究人员可在DrugAssist模型的交互界面输入特定指令, 如“实现分子溶解度的显著增加, 并有效降低其hERG毒性”。DrugAssist模型基于其独特的算法架构与自然语言理解模块, 能够精准解析这类以自然语言形式呈现的复杂研究需求。

DrugAssist在多属性分子优化任务中展现出显著优势, 并在多个评估指标上取得了优异的性能。例如, 在同时优化溶解度和血脑屏障穿透性等多目标优化任务中, DrugAssist取得了62%的成功率, 而对标基线方法, Mol-Transformer为59%, Mol-Seq2Seq仅为35%。根据

实验结果显示, DrugAssist生成分子的拓扑有效性高达98%, 表明其在生成具有实际合成潜力的分子方面表现优异。这种基于自然语言交互的优化方式, 为复杂约束下的分子设计提供了更灵活和高效的解决方案。

1.1.4 零样本优化的知识蒸馏架构

1.1.4.1 Prompt-MolOpt: 零样本多属性优化模式 如今, 在多属性分子优化中, 传统的生成模型常常受到多目标平衡和数据稀缺的限制[41]。为克服这些挑战, Wu等[42]开发的Prompt-MolOpt框架, 通过融合化学语言模型与提示工程, 开创了基于知识蒸馏的零样本多属性优化范式。其方法核心在于构建跨属性知识共享架构: 模型通过单属性数据集的训练, 学习原子级属性嵌入, 结合提示符动态组合多属性优化目标, 实现了无需额外多属性标签的零样本优化能力。

Prompt-MolOpt的核心技术包括: 用于构建分子数据集的子结构屏蔽解释方法、生成属性特定的原子嵌入的多任务图注意力网络, 以及基于Transformer架构的分子生成模型。通过引入属性提示, 该模型在优化方向上具有高度的灵活性, 能够针对单一或多目标进行优化, 甚至支持药效团的保留。这种灵活性在新药开发中具有重要意义, 尤其是在同时优化溶解度、毒性、血脑屏障通透性等多个目标时, 该方法能够有效生成优化后的分子。

经实验结果表明, 在多巴胺受体亲和力与药物相似性的多目标优化任务中, Prompt-MolOpt在零样本场景下实现了17.87%的成功率 (δ = 0.6), 较Modof模型提升15%, 证明其跨属性泛化能力。此外, 模型在血脑屏障通透性与脂溶性联合优化任务中, 通过提示符“BBBP lipop”指导生成分子Slog*P*值显著提升, 同时保留关键药效团 (如嘧啶环), 验证了模型在多目标权衡与结构保留上的协同优势。

1.1.4.2 FluID: 基于数据的联邦知识蒸馏框架 在药物结构优化的实际应用中, 数据隐私与异构性成为关键技术的挑战, 为了突破这一限制, 联邦学习 (federated learning, FL) 于2016年提出, 作为一种分布式机器学习范式, 旨在解决医疗及药物研发领域数据孤岛问题与保护数据隐私[43]。目前, FL主要基于模型驱动 (MD-FL) 方式, 在多个分布式站点上训练模型, 通过参数更新来训练中心模型。这种方法可以减少对数据共享的依赖, 同时通过差分隐私等技术增强数据保护。传统的MD-FL模型仍存在通信开销大、计算成本高、泛化能力低和模型架构受限等一系列瓶颈[44]。针对MD-FL的局限性, Hanser等[45]创新性地将知识蒸馏 (knowledge distillation, KD) 技术引入MD-FL, 就此开发了FLuID (federated learning using information distillation) 框架, 用于药物设计和毒性预测[46]。其核心流程分为3个步骤: ① 知识提取: 各参与方使用本地数据训练私有教师模型, 并用该模型为公共数据集生成预测标签; ② 知识整合: 将所有教师模型的预测结果合并, 形成一个统一的联邦数据集; ③ 知识融合: 使用联邦数据训练学生模型, 并可进一步与私有数据结合, 优化性能。FLuID在涉及8家制药公司的合作实验中得到了验证, 成功从企业数据中提取并融合hERG活性知识。hERG结合分类实验证明, 该框架显著提升了模型的性能和适用范围。

在药物结构优化领域, 数据孤岛与隐私保护问题制约着众多研究机构的优化效率, 传统联邦学习的高通信开销与数据泛化能力, 难以满足药物结构优化的需求。随着计算算力的提升与多模态数据融合技术的发展, FLuID框架通过隐私保护型知识共享架构与联邦蒸馏范式的深度融合, 未来有望为药物结构的智能化优化开辟跨机构协同创新的新路径, 助力药物结构优化效率的提升。

1.2 图神经网络在药物分子优化中的应用

图神经网络算法通过特征提取药物分子的拓扑结构, 从根本上已经超越了人工筛选实验和传统的基于SMILES的方法, 现已成为药物结构优化领域的变革型范式[47]。在图神经网络框架中, 其中分子被表示为图, 节点表示原子, 边则表示为化学键[48]。图神经网络是图形结构化数据上构建分子结构预测模型的先进方法, 其通过消息传递机制实现分子图的数学建模。这一特性使GNN成为推动药物结构优化领域发展的生力军, 从预测出治疗效果佳、毒性低和合成可行性高的分子结构再到发现新颖的药物分子[49]。

1.2.1 分子图特性与图神经网络

在传统的欧几里得空间中, 数据常常以向量的形式表示, 每个节点的距离可以通过几何方法来测量。但是, 无论是进行药物分子从头发现, 还是先导化合物优化任务中, 药物分子并不是由简单的点和线构成的, 而是由原子和化学键构成的复杂非欧几里得拓扑系统, 其官能团之间的关联性、化合物环结构的三维构象无法被欧式空间向量充分表明[50-54]。GNN直接操作图结构的能力提升了对分子特征进行建模的效率, 例如, Zeng等[55]提出的UAlign模型, 该模型是一个新颖的基于图到序列的无模板逆合成预测方法, 通过将图神经网络与Transformer相结合, 使其更有效地利用分子的图结构信息来预测反应物。

UAlign模型的核心是编码器−解码器结构。编码器是一个图神经网络, 负责提取分子图中包括原子和化学键的特征。解码器是基于Transformer解码器构建

而成, 负责将编码器输出的结构信息转换为反应物的SMILES字符串。通过大量的实验证明, UAlign在逆合成预测任务上取得了显著的性能提升。在USPTO-50K数据集上, UAlign在Top-5和Top-10的准确度上相较于无模板的基线方法提高了5%和5.4%[56]。

以治疗骨髓纤维化JAK-2抑制剂Pacritinib为例, UAlign模型成功预测了文献记载的8步合成路径, 从5-硝基水杨醛和2,6-二氯嘧啶出发, 精准预测到最终产物。其中关键合成步骤包括: 烯烃复分解反应、芳香族取代反应、连续烯丙基取代和铃木偶联反应、硝基还原和另一个烯丙基取代, 以及最终的醛基还原和亲核取代。与传统方法Retroformer等模型相比, UAlign通过直接处理分子图结构, 更好的捕捉了分子的拓扑结构信息, 使其在处理多反应中心复杂药物分子时表现出强大的预测能力[57]。而传统的基于模板的方法虽然预测准确, 但是受限于模板库的覆盖范围, 难以处理模板库中不存在的反应类型。UAlign的无模板特性以及对分子图结构的有效利用, 为药物分子优化和合成路线提供了强大而灵活的工具。

1.2.2 图神经网络核心流程和训练步骤

图神经网络通过迭代式消息传递, 学习分子图的拓扑特征, 其中节点 (原子) 聚合邻域信息 (如化学键邻接关系), 并通过堆叠多层网络扩大节点的感受野, 使得节点可以学习更广泛的图信息。在训练过程中, 梯度反向传播结合任务驱动的损失函数优化模型权重[24,48,58-62]。典型的图神经网络药物性质预测框架如图4A所示, 展示了如何通过多层图卷积层和激活函数处理输入的分子并进行性质预测。

图神经网络善于处理具有非欧几里得拓扑结构的图数据, 这类数据与规则的网状数据结构存在本质差异。其核心机制在于通过层级式的消息传递实现特征学习, 在每个图卷积层中, 节点通过聚合邻域信息并融合自身特征实现状态更新[48]。从数学模型的角度上来说, 许多图神经网络模型都可以被归纳到一个通用的消息传递网络框架下。Gilmer等[24]在2017年的工作中形式化了这一框架, 他们将图神经网络向前传播的过程分为两个阶段, 分别为消息传递阶段和读出阶段。其中消息传递阶段又分为消息计算与聚合以及节点状态更新。首先, 在消息计算与聚合阶段, 模型通过聚合邻域节点信息来构建消息向量, 其数学表达根据公式(1)计算:

$$m_v^{t+1} = \sum_{\omega \in N(v)} M_t\left(h_v^t, h_\omega^t, e_{v\omega}\right) \tag{1}$$

其中, m_v^{t+1}表示图神经网络中节点v在第t+1时间步生成的聚合消息向量, h_{tv}和h_ω^t分别是节点v和邻居ω在时间步长t的隐藏状态, $e_{v\omega}$是连接节点v和邻居ω的边的特征。在这一过程中, 每个节点能够有效地收集到邻居节点的信息, 使得节点的表示能够反映其局部图结构和邻居的特征。随着消息传递阶段的迭代,

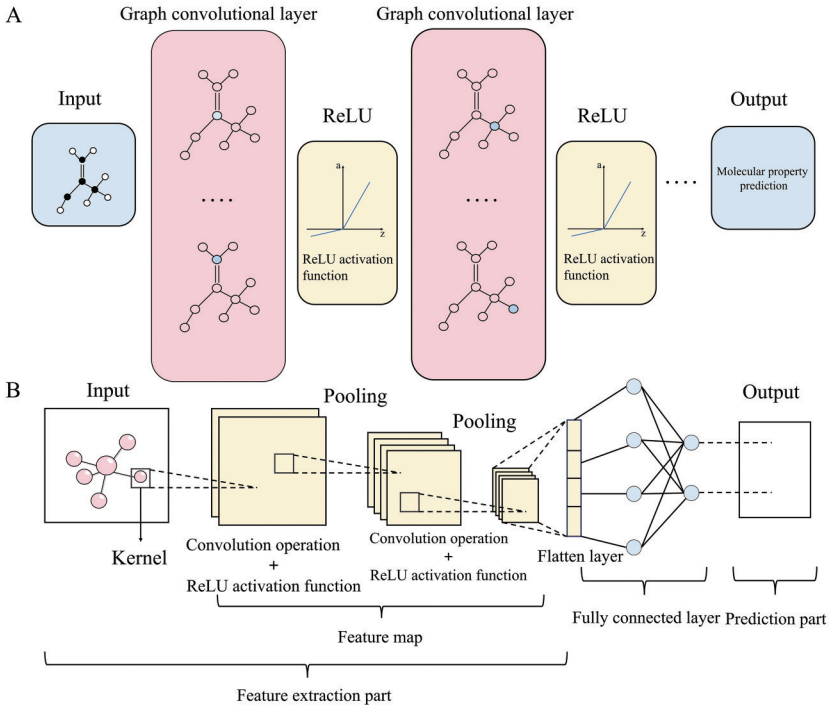


Figure 4 Framework diagram of graph neural networks and convolutional networks for drug design. A: Framework diagram of graph neural networks for drug design; B: The core convolution rule of convolutional neural networks. ReLU: Rectified linear unit

信息可以传播到更远的节点, 让节点能够感知更多的图结构信息。

然后, 在节点状态更新过程中, 模型利用上一阶段计算得到的聚合消息向量 m_v^{t+1} 和节点自身信息的前一时刻隐藏状态 h_v^t, 通过一个可学习的更新函数 U_t 来更新得到节点 v 在当前 $t+1$ 时刻的新隐藏状态 h_v^{t+1}, 其计算过程如公式 (2) 所示:

$$h_v^{t+1} = U_t\left(h_v^t, m_v^{t+1}\right) \tag{2}$$

最后, 在读出阶段, 当所有节点均经过 T 轮消息传递获得最终的隐藏状态 h_v^T 后, 模型采用一个可学习函数 R 对图中所有节点的最终隐藏状态集合 $\{h_v^T | v \in G\}$ 进行整合, 以生成整个图的表示, 其表达式根据公式 (3) 计算:

$$\hat{y} = R\left(\left\{h_v^T | v \in G\right\}\right) \tag{3}$$

1.2.3 图卷积网络与卷积神经网络的比较

图卷积神经网络作为图神经网络的一种典型实现方法, 其核心思想与卷积神经网络相似, 即通过局部邻域聚合来进行特征提取, 其本质都是滤波操作[63]。传统卷积神经网络依赖于卷积运算、激活函数及池化等核心操作进行特征提取, 其典型流程图如图4B所示。然而, 图神经网络算法和卷积神经网络在任务适应面上是不同的。CNN通常依赖于规则的网格状结构数据, 例如在处理分子数据时, 需要将SMILES字符串强制映射为2D图像等欧氏空间表征, 这种预处理方式可能导致关键的原子间空间关系与化学键信息丢失。具体而言, 在药物−药物相互作用等预测任务中, 若采用CNN, 通常需将分子结构转换为图像或分子指纹, 这使得模型难以学习到更深层次的原子间动态关系; 相比之下, GCN更加适合处理不规则结构的应用场景, GCN直接以分子图为输入, 通过邻接矩阵和图传播规则聚合周围节点的特征信息。例如, Zeng等[64]提出了一种基于三方跨网络嵌入图卷积网络的药物重新定位 (tripartite graph convolutional network for drug repositioning, TGCNDR) 法。在预测药物与疾病之间关联的任务中, TGCNDR表现出理想的性能, 其AUROC为0.922、AUPR为0.916[48,65,66]。

1.2.4 图卷积神经网络在药物结构优化中的范式演进

药物分子优化从传统经验驱动走向人工智能驱动的过程中, 图神经网络通过“拓扑建模−动态生成−多维融合”的技术跃迁, 逐步突破了化学空间探索、多目标权衡、数据稀缺性等挑战, 推动了药物设计与结构优化领域的革命性发展。其中, 关键技术方法与算法模型的演进过程如图5所示。

1.2.4.1 静态表征阶段: 从图卷积基础到注意力机制的分子拓扑特征提取 图神经网络这一概念最早由Scarselli等[47]于2008年提出, 是一种专门用于处理图结构数据的深度学习模型, 其核心目标是将传统神经网络扩展到非欧几里得空间, 从而有效分析具有复杂关系的数据。此时GNN模型仍处于理论阶段, 主要应用于验证图结构数据的建模能力, 尚未涉及实际应用, 且早期GNN的训练效率较低, 缺乏大规模计算资源和数据集支持, 难以应用于复杂药物研发任务。

随着算法模型的不断进步和药物结构优化任务复杂性不断增加, 早期传统分子属性预测模型 (如指纹结构算法) 由于仅依赖原子连接关系的浅层遍历, 无法挖掘原子间的高阶相互作用, 导致难以捕捉分子拓

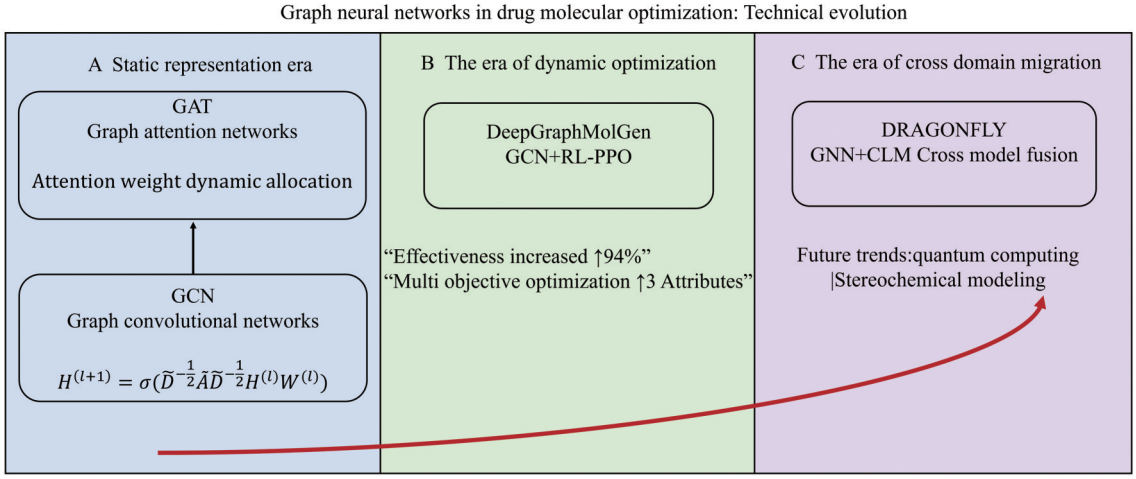


Figure 5 The evolution of model technology in drug molecule optimization using graph neural networks (GNN). A: Static representation era; B: The era of dynamic optimization; C: The era of cross domain migration. GAT: Graph attention network; GCN: Graph convolutional network; RL-PPO: Reinforcement learning - proximal policy optimization; DRAGONFLY: Drug-target interactome-based generation of novel biologically active molecules

扑的深层关联性。并且, 指纹算法通常将不同类型的化学键简化为统一连接符号, 忽视了分子的化学性质差异[53]。

为了针对上述问题, Kipf等[67]于2017年首次提出使用图卷积网络进行半监督任务, 为药物分子拓扑特征建模提供了新范式。在图卷积网络中, 节点特征的层传播根据公式 (4) 计算:

$$H^{(l+1)}=\sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{4}$$

其中, $\tilde{A}=A+I_N$为引入自连接的邻接矩阵, $\tilde{D}$是$\tilde{A}$度矩阵, $H^{(l)}$为第l层图卷积网络的节点特征矩阵, $W^{(l)}$为第l层的可训练权重矩阵。通过对邻接矩阵进行归一化处理, 该方法能够有效聚合邻居节点信息, 实现多层特征传播。该团队在PubChem数据集上验证了图卷积网络在药物肝毒性预测任务中的有效性, 其曲线下面积 (area under curve, AUC) 相比支持向量机提升了0.07。这一工作为基于图神经网络的分子特征学习奠定了重要基础, 推动了GNN在药物设计领域的广泛应用。

初期的GCN在其均质化邻域聚合策略中,对所有邻域节点赋予相同权重, 因此难以区分关键官能团的差异化贡献。为了改进GCN这一局限性, Veličković等[68]将动态权重分配机制引入GCN, 从而提出图注意力网络。该网络中动态权重分配, 即计算节点对其邻居节点的注意力系数过程, 如公式 (5) 所示:

$$\alpha_{ij}=\frac{exp\left(LeakyReLU\left(\vec{\mathrm{a}}^{T}\left[W\vec{h}_i\|W\vec{h}_j\right]\right)\right)}{\sum_{k\in N_i}exp\left(LeakyReLU\left(\vec{\mathrm{a}}^{T}\left[W\vec{h}_i\|W\vec{h}_k\right]\right)\right)} \tag{5}$$

其中, α_{ij}表示节点i对邻居节点j的注意力系数, $\vec{h}_i$和$\vec{h}_j$分别是节点特征, W是共享权重矩阵, $\vec{\mathrm{a}}$是可学习的注意力向量, N_i表示节点i的邻域。该机制使模型能够自适应地评估不同邻居节点的重要性, 有效区分关键官能团的差异化贡献, 显著提升了模型在节点分类、预测等任务上的性能, 同时保持了计算效率和模型可解释性。

为了验证GAT的有效性, 该研究团队在多个数据集上进行了对比实验。在Cora和Citeseer数据集上的实验结果显示, GAT的分类准确率相比GCN均提高了1.6%, 证明了注意力机制在图结构数据建模中的优势。基于GAT在节点分类任务上的优异性能, Shao等[69]进一步将其扩展应用于药物–靶点相互作用预测任务, 成功构建了DTI-HETA模型, 为后续的药物设计研究提供了新的技术路径。

图卷积神经网络和图注意力网络的发展为后续中期发展阶段 (多目标强化学习、几何散射生成) 奠定了重要的理论基础和技术框架, 标志着图神经网络从静态表征迈向动态表征的阶段式演进[70]。

1.2.4.2 动态优化阶段: 从单一属性预测到多目标强化学习的生成范式突破 尽管GCN和GAT的提出解决了药物分子拓扑特征提取的困难, 但是早期阶段的图神经网络模型存在一些局限性。首先, GCN和GAT只能对单一目标任务进行训练和预测, 缺乏多目标优化能力; 其次, 模型只能捕捉二维拓扑特征, 无法考虑立体化学结构特征; 最后, 当前算法只能对于属性进行预测, 无法进行结构优化, 未实现在庞大化学空间进行充分探索。

Khemchandani 等[71]所提出的DeepGraphMolGen多目标强化学习框架突破了GCN仅能进行单一目标优化的不足。其方法核心步骤分为两步: 分子属性预测与分子生成模块。其中, 分子属性预测模块扮演着“评估者”的角色, 负责预测分子结构的属性得分。此模块以分子图为输入, 利用GCN提取分子的核心拓扑结构信息, 然后通过前馈神经网络输出预测的属性值。为了减少实际的实验策略数据中的误差, 该框架在训练的部分采用鲁棒损失函数来处理实验数据中的噪声和异常值。这种损失函数能有效减弱数据中的异常值, 从而提升属性预测的稳定性与可靠性。该鲁棒损失函数的数学表达式根据公式 (6) 计算:

$$f\left(x,\alpha,c\right)=\frac{|\alpha-2|}{\alpha}\left(\left(\frac{(x/c)^2}{|\alpha-2|}+1\right)^{\alpha/2}-1\right) \tag{6}$$

其中, x代表预测值与真实值之间的误差, α和c作为可学习的参数, 它们控制着损失函数的形状和对误差的敏感度, 通过优化这些参数模型能够更好地应对数据中的噪声, 提升模型预测的鲁棒性。

DeepGraphMolGen框架的第二个核心组件为分子生成模块, 该模块基于强化学习机制驱动分子结构的逐步构建。此生成过程依赖于一个图卷积策略网络, 该网络将分子生成视为序列决策问题, 并依据当前已构建的部分分子结构来预测后续可行动作的概率 (例如, 在特定原子上添加官能团)。在此图卷积策略网络中, 节点 (原子) 表示的更新遵循特定的图卷积更新规则, 其数学表达式根据公式 (7) 计算:

$$H^{(l+1)}=AGG\left(ReLU\left(\left\{\tilde{D}_i^{-\frac{1}{2}}\tilde{E}_i\tilde{D}_i^{-\frac{1}{2}}H^{(l)}W_i^{(l)}\right\},\forall_i\in(1\cdots b)\right)\right) \tag{7}$$

其中, $H^{(l+1)}$是下一层的节点嵌入矩阵, $\tilde{E}_i$为邻接矩阵, $\tilde{D}_i$为度矩阵, $H^{(l)}$为第l层的节点嵌入, $W_i^{(l)}$是可训练的权重矩阵。图卷积更新公式在本质上描述了信息如

何在分子图中传递。每个节点收集来自其邻居节点的信息, 然后通过权重矩阵和ReLU激活函数进行变换, 最后聚合得到新的节点表示。

除图卷积更新规则以外, DeepGraphMolGen框架的另一核心数学表达在于其强化学习过程。为指导分子生成分子生成策略网络的学习, 使其能够生成具有期望属性的分子, 该框架采用了特定的强化学习优化目标函数, 其具体表达根据公式 (8) 计算:

$$maxL^{\mathrm{CLIP}}(\theta) = \mathrm{IE}_t\left[\min\left(r_t(\theta)\hat{A}_t, clip\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t\right)\right] \quad (8)$$

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta old}(a_t|s_t)}$$

其中, L^{CLIP}是需要最大化的目标函数, $r_t(\theta)$为策略比率, $\hat{A}_t$为广义优势估计, *clip*( ) 函数将比率限制在[1-∈,1+∈] 范围内。此目标函数通过限制策略更新的步长来实现稳定训练。目标函数保证新策略不会与旧策略偏离太远, 从而避免过大的策略变换而导致训练不稳定。当优势为正时, 鼓励增加相应行动的概率; 当优势为负时, 则减少该行动的概率。强化学习的优化目标函数可以指导模型构建所需属性的分子。

该研究团队使用DeepGraphMolGen框架在生成与多巴胺转运蛋白具有高结合力分子的单目标优化任务中取得了显著成果。其方法成功生成了大量有效分子, 并且这些分子在药物相似性和合成可行性等关键指标上表现优异。在生成的分子中, QED值大于0.8和SA值大于0.8的比例较高, 充分展现了该方法在高质量候选药物生成方面的潜力。具体而言, 该研究团队基于不同实验室设置, 如初始分子和专家数据集的选择, 以及原子数量的限制, 对生成的分子进行了分析。以实验一 (初始分子为碳原子, 专家数据集为ZINC) 和实验二 (初始分子为碳原子, 初始数据集为多巴胺数据集) 为例: 实验一生成的分子集合QED平均值为0.843 (标准差0.027), 合成可行性平均值为0.842 (标准差为0.023); 实验二在初始分子不变的情况下, 改变数据集, 生成的分子集合QED平均值为0.847 (标准差0.029), SA平均值为0.850 (标准差0.029)。这些具体的定量统计数据表明, DeepGraphMolGen框架在不同的生成任务和约束条件下, 都能够稳定地生成具有高药物相似性和合成可行性的分子, 有效地展现了其在药物发现等领域的应用潜力。

在后续多目标优化实验, 例如生成对多巴胺转运体高亲和力但对去甲肾上腺素转运体低亲和力的分子上, 通过调整奖励函数, 成功生成具有差异化亲和力的分子。在消融实验中, 结合GCN特征提取和RL策略优化的DeepGraphMolGen能显著提升生成分子的有效性和多样性。

在药物分子优化的研究中, 尽管基于图神经网络 (如DeepGraphMolGen框架) 通过结合强化学习与多目标奖励机制显著提升了生成分子的多样性与目标属性, 但是传统图神经网络的固有局限性在复杂分子结构生成中仍未被彻底解决。如, 传统消息传递图神经网络因局部邻域聚合受限, 难以生成全局连贯的分子拓扑, 导致无效分子比例高达50%以上且方法相对耗时[72]。为此, Bhaskar等[73]将几何散射理论与GNN相结合, 提出GRASSY (graph scattering synthesis network) 框架用于药物分子生成与优化。该框架的创新性体现为3个核心模块: ① 几何散射变换通过多尺度扩散小波捕获分子图的全局和局部特征, 构建置换不变性散射矩; ② 正则化自动编码器联合优化散射矩重建损失与物理化学性质回归损失; ③ 对抗分子生成网络利用逐步插值策略直接生成完整分子图, 避免逐节点生成导致的化学计量错误。GRASSY框架在ZINC分子数据集的分子生成任务中表现出优势, 其实验基于从该数据集不同分子量区间的子集各采样2 500个分子的数据。具体而言, GRASSY在FBAB子集上生成的分子有效性达到了0.94。该方法通过对几何散射矩进行正则化自编码学习, 使潜空间表征受到关键分子物理化学性质如QED和拓扑极性表面积等回归损失的约束。例如, 在对ZINC数据库BBAB子集的实验中, GRASSY对QED性质预测的平均绝对误差为0.167 7, 对拓扑极性表面积的平均绝对误差为13.20。这些结果共同显示了GRASSY框架在学习分子结构与性质之间关联, 并利用这种关联生成具有期望特性新分子方面的潜力。

DeepGraphMolGen框架的创新之处在于其成功实现了从传统的单一属性预测向多目标强化学习的生成范式转变。这一突破性进展打破了早期图神经网络仅能进行单一目标优化的局限, 实现了多种药物属性的协同优化。尽管这一范式转变代表了分子生成领域的重要里程碑, 但GRASSY等后续框架的出现也表明, 解决复杂分子拓扑生成的挑战依然存在, 这促使药物分子优化和生成技术向更高维度、更全局化的方向不断演进。

1.2.4.3 跨域知识迁移阶段: 从数据稀缺到多任务泛化的零样本优化范式 通过剖析图神经网络演进脉络, 可以发现随着图神经网络算法助力药物结构优化不断发展, 用于训练和预测的数据的质量在这一过程中占据关键地位。但是, 高质量数据 (如生物分子−配

体相互作用) 的获取成本高昂, 且数据集通常规模有限, 例如PDBBind数据库[74]作为一个专注于蛋白质–配体结合数据的在线数据库和服务器, 提供了许多资源用于研究蛋白质结构与功能之间的关系。截止至2025年3月, PDBBind数据库也仅仅包含不到3万条蛋白质–配体复合物结构[75,76]。数据的质量和稀疏性限制了现有图神经网络模型的泛化能力。

针对数据稀缺性问题, Buterez等[77]提出了一种基于迁移学习的图神经网络方法。该方法通过两步策略实现迁移, 首先使用低保真数据训练监督变分图自编码器, 学习分子结构的跨保真潜在特征; 随后通过自适应读出与动态特征增强策略, 将低保真知识加入高保真任务的模型优化中。

在蛋白–配体的相互作用预测实验验证环节中, 该研究团队使用包含2 800多万条实验性蛋白–配体相互作用的数据集和12种量子属性的数据集进行验证, 通过其提出的方法, 利用低保真数据, 有效提高了筛选阶段高保真数据的预测性能, 并且减少了一定的高保真数据的需求量。与传统的多保真状态嵌入等算法相比, 在平均绝对误差上降低了20%～60%, 甚至在数据量极少的情况下, 依然保证了稳定的预测精确度。此外, 在实际的药物筛选和量子力学模拟中, 该方法在跨域迁移等场景中也有着优异的成绩。尽管如此, 该方法对低保真数据质量和数量依赖较大, 在极端不平衡或高噪声数据环境下仍需进一步优化。基于图神经网络的迁移学习方法, 相较于多保真状态嵌入等现有方法, 其创新在于通过动态权重分配, 为数据稀缺下的药物分子优化提供了新范式。

在前期的迁移学习方法中, 尽管通过低保真数据协同训练策略显著缓解了生物分子相互作用数据稀缺问题, 但这类方法依赖目标域的微调和复杂的强化学习步骤。针对这些不足, Atz团队[78]提出了DRAGONFLY框架。该框架创新性地结合了图神经网络和化学语言模型, 利用药物–靶点互作图对药物分子与靶点之间的复杂关系进行建模, 开创了无需特定任务微调的“零样本”分子生成的新范式, 兼顾生物活性、合成可行性和结构新颖性。

该研究团队从ChEMBL数据库和PDBbind数据库中获取数据, 通过互作图将分子视为节点, 靶点结合关系视为边, 构建了一个包含约36万个配体、2 989个靶点和约50万条生物活性数据的药物–靶点互作网络, 以及约17 824个蛋白质–配体三维结构, 为模型训练奠定了数据基础。

在后续实验中, 该研究团队将DRAGONFLY框架用于过氧化物酶体增殖物激活受体 (PPARγ) 靶点的分子优化与设计。他们以PPARγ的三维结合位点为输入, 在考虑分子质量、新颖性和合成可行性的条件下, 生成了超过30万种候选分子。然后利用定量构效关系模型对生成分子的活性进行预测。实验结果表示, 在PPARγ靶点的验证中, 相较于传统RNN, DRAGONFLY框架生成分子合成可行性通过率 (RA ≥ 0.5) 达86.0%。相比于决策树的基线方法, 半数抑制浓度的负对数 (pIC_{50}) 预测的平均绝对误差降低小于或等于0.6。

DRAGONFLY框架在分子生成的多样性和新颖性方面展现出显著优势。和传统算法不同点在于DRAGONFLY无需依赖单一分子的微调和强化学习, 这一特点减少了额外机器学习的步骤。在多项关键性能指标上, DRAGONFLY表现优异, 尤其是在分子新颖性和骨架多样性等方面, 远超传统化学语言模型。与其他的GNN模型相比, DRAGONFLY结合图神经网络和化学语言模型, 在捕捉分子图局部信息和高维几何特性上实现了突破。

1.2.4.4 图神经网络模型挑战与未来方向 图神经网络通过“静态表征–动态优化–跨域迁移”的技术推进, 驱动了药物分子优化的范式革新。在静态表征时代, GCN和GAT突破了传统指纹方法对分子结构特征提取的限制; 动态优化时代的强化学习与多目标奖励机制解决了化学空间探索和分子生成的难题; 迁移技术与多模态知识蒸馏, 降低了数据稀缺对模型泛化能力的制约。图神经网络在药物分子优化方面目前还存在一些挑战, 如对数据依赖性强、3D构象建模不足、计算成本高昂等。就现存GNN变体模型来说, 往往在可拓展性、精确性和计算效率方面存在不足。未来, 图神经网络有望与量子计算结合, 利用量子线路嵌入GNN的消息传递过程, 提升预测与分子生成的效率[79]。此外, 代表性及补充的GNN模型的性能与特点如表1所示[80-82]。

2 强化学习算法在药物分子优化中的应用

2.1 强化学习的核心框架与分子优化中的映射

强化学习是一种通过智能体在环境中进行探索和学习, 以最大化累计奖励的机器学习方法。其核心概念就是智能体与环境的交互作用 (图6A)。与监督学习和无监督学习不同, 强化学习不依赖于标注数据, 而是通过与环境的互动获得即时反馈, 从而改变决策策略。强化学习的目标是学习采取何种行动, 以获得最大化长期回报奖励[83]。强化学习核心概念应用于分子优化场景下 (图6B), Agent一般为药物设计模型; Environment为药物性质评估环境, 用于评估药物的各项理化和药理学性质; State为药物分子的表现形式 (如SMILES或图结构); Action为对分子结构做出的调整, 如添加、删除、替换的官能团等操作; Reward为药物分

Table 1 Performance and characteristics of the GNN model. PPO: Proximal policy optimization; GRASSY: Graph scattering synthesis network; MOLER: Molecule-level rewards; GraphMVP: Graph multi-view pre-training; HierG2G: Hierarchical graph encoder-decoder

Model	Core technology	Strength	Limitation	Applicable scenario
GCN	Neighborhood feature aggregation	Efficient extraction of molecular local features	Ignore node heterogeneity, weak global feature modeling capability	Basic ADMET prediction
GAT	Graph attention mechanism	Multi-attribute joint optimization	High computational complexity, difficult to scale to large molecules	Target selective optimization
DeepGraphMolGen	GCN + multi-objective reinforcement learning (PPO)	Multi-objective attribute prediction/generation	Long training time consuming	Multi-attribute molecular generation
GRASSY	Geometric scattering transform + regularized autoencoder	Molecular validity, high synthetic feasibility	Dependent on scattering moment precomputation	Complex topology molecular design
DRAGONFLY	GNN-CLM fusion + zero sample generation	Rich in molecular backbone diversity	Dependent on protein structure data	Targeted drug synthesizability design
MOLER	Diffusion modeling + subgraph template generation	Excellent stability of generated molecules	Limited coverage of template library	Modification of derivatives of known structures
GraphMVP	Multi-view pre-training (2D/3D)	Enhanced generalization of molecular characterization	High 3D data requirements and pre-training costs	Small sample molecular property prediction
HierG2G	Hierarchical diagram generation	Support for macromolecule generation	Slow generation speed	Antibody/peptide drug design

子优化后的奖励值, 如活性、毒性、溶解度等。

2.2 马尔可夫决策过程: 强化学习的数学基础与应用

马尔可夫决策 (markov decision process, MDP) 是强化学习的基础框架 (图6C)。马尔可夫决策过程为强化学习提供了一个数学模型, 描述了在部分可控和部分随机的决策过程中如何选择最佳行动。MDP主要由5个部分组成: ① 状态空间 S, 描述系统内所有可能发生的状态, 用状态 S_t 表示; ② 动作空间 A, 描述智能体可以采取的所有动作, 用 a_t 表示; ③ 转移概率 $P(s'|s, a)$, 表示智能体在状态 S_t 执行动作后, 转移到下一个状态的概率; ④ 奖励函数 $R(s, a)$, 其表示在状态 s 执行动作 a 后, 智能体获得的奖励; ⑤ 策略 $\pi(a|s)$, 其表示智能体在状态 s 下选择动作 a 的规则或者是概率分布。MDP映射到分子优化中, S 表示分子状态空间 (SMILES/图/3D构象); A 代表结构编辑工作 (原子或基团替换、修饰); R 则代表多属性奖励函数[84-86]。MDP的应用例子, 如Fu等[87]提出了一种进化马尔可夫决策过程 (evolutionary markov decision process, EMDP) 用于药物设计。传统的MDP通常以单个分子作为状态, 而EMDP的创新在于将种群 (population) 作为状态, 将遗传算法进化步骤 (突变和交叉) 作为MDP的动作。新改造的EMDP则可以通过智能探索化学空间、抑制随机游走行为和知识迁移来提高遗传算法的效率和稳定性。实验结果证明, 基于EMDP的强化遗传算法 (reinforced genetic algorithm, RGA) 在药物设计中能更稳定和高效地搜索分子与疾病相关的蛋白质结合, 并且RGA在分子对接方面的表现也优于Autogrow 4.0、Graph-GA等基线方法, 为加速药物发现过程提供了新方法。

2.3 强化学习的关键优化目标: 价值函数和动作价值函数的理论

强化学习旨在通过优化策略以实现累积奖励的最大化。为评估策略性能, 强化学习引入了价值函数 (value function) 和动作价值函数 (action-value function) 来判断策略的质量高低[84]。价值函数 $V^{\pi}(x)$ 用于评估在给定初始状态 x 下, 遵循策略 π 所能获得的期望总折扣奖励[85]。其数学表达式根据公式 (9) 计算:

$$V^{\pi}(x) = \mathrm{IE}\left[\sum_{t=0}^{\infty}\gamma^{t}R_{t+1}|X_0 = x\right], x \in \chi \tag{9}$$

其中, IE代表期望, γ 是折扣因子, R_{t+1} 表示在时间步 t 到 t+1发生的转移获得的奖励。价值函数 $V^{\pi}(x)$ 衡量了从状态 x 开始, 执行策略 π 的长期价值。

与价值函数相似, 动作价值函数 $Q^{\pi}(x,a)$ 用于评估在给定状态下, 首先采取动作 a, 然后遵循策略 π 所能获得的期望累积奖励, 其数学表达式根据公式 (10) 计算:

$$Q^{\pi}(x,a) = \mathrm{IE}\left[\sum_{t=0}^{\infty}\gamma^{t}R_{t+1}|X_0 = x, A_0 = a\right], \quad x \in \chi, a \in \mathrm{A} \tag{10}$$

将药物分子优化过程建模为MDP, 其中状态是分子, 行动是分子修改, 奖励是性质评估。在这一过程中, 动作价值函数尤为重要, 通过函数逼近方法来接近最优函数, 从而以迭代的方式引导分子朝着高奖励 (理想药物性质) 的方向演化。同时, 价值函数可以作为补充, 用于评估分子的整体潜力。

Zhou等[86]结合化学领域知识提出了“分子深度Q网络” (molecule deep q-networks, MolDQN) 框架, 该框架运用双Q学习与随机化价值函数, 借助神经网络

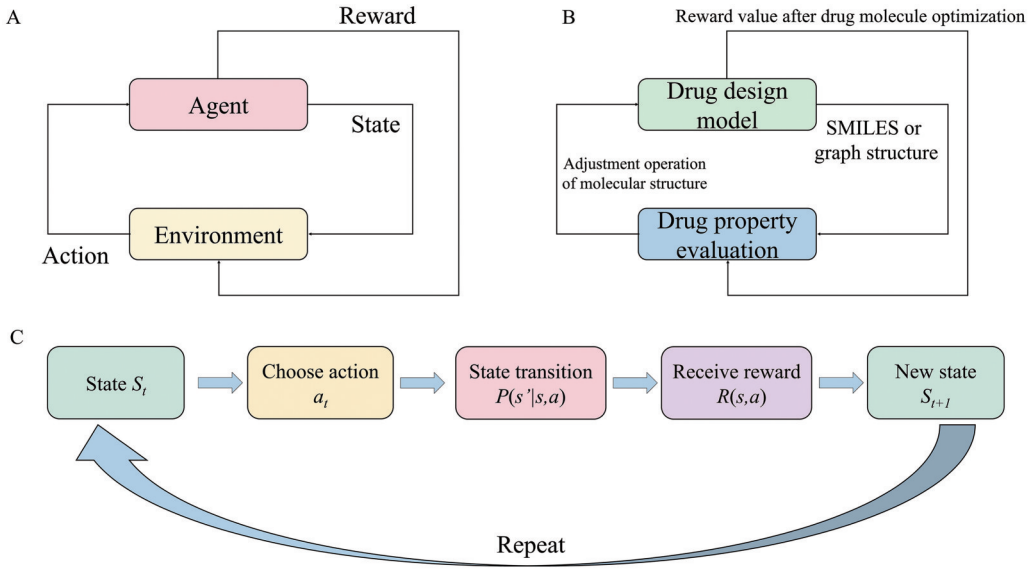


Figure 6 Core concepts and application scenarios of reinforcement learning. A: Core concepts of reinforcement learning; B: Application of reinforcement learning core framework in molecular optimization; C: Markov decision process (MDP)

来估算动作价值函数, 并以动作价值函数中的最大值计算状态价值函数, 使得预测收益最大化。在优化分子的脂水分配系数对数 (log*P*) 的任务中, 研究团队使用 MolDQN 与 GCPN、JT-VAE、ORGAN 等基线模型进行对比, 并以惩罚化的脂水分配系数对数 (penalized log*P*) 作为奖励函数的关键衡量指标, 模型每执行一次分子修改操作, 便会依据修改后分子的 penalized log*P* 值获取相应奖励, 最终通过不断调整策略以最大化累积奖励。经过多轮训练和评估, MolDQN 取得了 11.84 的 penalized log*P* 分数。在相同的评判标准下, GCPN、JT-VAE 和 ORGAN 所取得的 penalized log*P* 分数均低于 11.84。此外, 经 MolDQN 优化的分子, 其分子有效性达到了 100%, 而 ORGAN 仅为 0.4%。在药物相似性优化任务中, MolDQN 达到了 0.948 的最优结果, 优于 JT-VAE 的 0.925 和 ORGAN 的 0.896。MolDQN 框架显著提高了分子优化过程的稳定性和样本利用率, 高效地实现了目标分子优化, 在分子设计领域展现出独特的优势。

2.4 强化学习在药物分子优化中的技术演进与应用突破

强化学习与深度学习作为机器学习的两大重要分支体系, 在方法论框架上与任务执行上存在本质性差异。深度学习主要通过从大规模数据中提取特征, 经过多层非线性特征变换实现高维数据的识别, 其在进行分类或者回归任务中表现出强大的学习能力。与之不同的是, 强化学习以马尔可夫决策过程为理论基础, 其核心原理在于构建智能体与环境之间的动态交互范式, 通过策略梯度优化和奖励机制来实现决策目标[84,88]。

在药物分子优化场景中, 强化学习被用于在离散或连续的化学空间中进行分子生成、优化与探索, 旨在发现具有目标属性的候选分子。在药物分子优化领域中, 强化学习经历了从早期的启发式搜索方法逐步发展到现代深度学习驱动的生成技术。Q-learning 作为强化学习的早期算法, 由 Watkins 等[89]于 1992 年正式提出, 其主要任务就是学习动作价值函数, 用于指导智能体在特定状态下选择能够获得最大累积奖励期望值的动作, 但是此阶段的 Q-learning 并未用于药物分子优化。后来, Zhou 等[86]首次将 Q-learning (特别是 double Q-learning) 应用于分子优化框架, 通过将深度强化学习与化学领域知识相结合, 开发了用于分子优化的 MolDQN 模型。中期, You 等[90]提出了图卷积策略网络, 实现了将图神经网络与强化学习相结合, 并首次用于分子优化, 是 GNN 与 RL 在药物发现中的奠基性工作。当前, 强化学习在药物分子优化的核心应用包括: 生成模型设计、图结构优化、多参数权衡和可合成性验证。

2.4.1 分子生成范式: 从序列到图结构的强化学习演进

2.4.1.1 序列驱动的分子生成: SMILES 与 Transformer 的结合 分子生成与优化领域存在着诸多挑战。首先, 在庞大而复杂的高维度化学空间进行高效的探索是一项艰巨的任务, 而高质量标记数据资源的稀缺进一步制约了算法的训练准确度[91]; 其次, 传统生成模型与变分自编码器等方法虽然能够处理较长序

列, 但是在优化分子的药物特性上仍有不足。强化学习框架虽可通过奖励机制强化目标属性, 但在新型结构多样化创造方面表现薄弱, 容易陷入局部最优解的困境。值得注意的是基于图的生成方法在生成有效分子方面取得了一些进展, 可是如何生成多样分子与避免过拟合问题上仍然存在挑战[92,93]。

Mazuz等[94]提出了一种基于Transformer和策略梯度强化学习相结合的新方法—Taiga。模型采用两个阶段进行分子生成和优化。第一阶段, 模型基于大量SMILES字符串学习化学语言的语法和规则, 使得模型可以生成化学上有效的分子结构; 第二阶段为强化学习阶段, 该阶段将分子生成视为马尔可夫决策过程, 与上述MDP原理相同, 将SMILES字符串视为“状态”, 添加SMILES字符串为“动作”“奖励函数”用于指导生成目标属性分子。若生成的SIMLES字符串是无效分子, 则奖励为0。例如, 在针对于BACE靶点的优化任务中, Taiga在GDB13、Moses和ZINC数据集上将平均pIC_{50}值基线的4.9～5.2提升至5.6～10.5。此外, 在优化pIC_{50}的同时, Taiga能够维持与基线相似的QED (例如在ZINC数据集上, 优化后QED为0.671) 和可合成性分数 (例如在ZINC数据集上, 优化后可合成性分数为2.99), 这表明生成的分子在提升目标活性的同时, 其药物相似性和合成可行性未显著恶化。

与传统的生成方法如LSTM-PG相比, Taiga在不同数据集上实现了85.5%～97.2%的分子有效性, 显著优于部分利用SMILES的RNN模型。在新颖性和多样性方面, Taiga生成的分子也表现优异, 新颖性达到了95.6%～100%, 多样性可达98.9%～100%。Taiga通过结合Transformer架构与强化学习的目标导向优化, 成功克服了传统方法在生成有效、新颖且易于合成的分子方面的瓶颈, 为药物分子优化提供了高效的工具。

2.4.1.2 多肽分子生成: 基于图注意力机制的优化 肽类药物分子是基于天然肽或合成肽设计的一类药物, 近年来在药物研发中具有愈发重要的地位, 肽类药物分子介于小分子药物与大分子生物药之间[95-97]。现有的肽类药物在稳定性、生物利用度和生产等方面面临巨大挑战。为了解决这些问题, Wang等[98]提供了基于图注意力机制和强化学习驱动的多肽生成模型。该模型中, 强化学习充当了进行决策优化的核心角色, 以生物活性和ADMET作为奖励机制来调整肽分子的结构, 使其生成新型药物样肽。

该模型的核心框架包含两个结构: 图注意力网络和Actor-Critic (A2C) 算法的强化学习生成框架。图注意力网络由图注意力机制和门控循环单元 (gated recurrent unit, GRU) 组成。GAT负责构建以氨基酸残基为节点, 以残基间关系为边的图结构, 而GRU则擅长提取时间序列中的依赖关系, 生成多维特征。Actor网络通过策略梯度优化技术不断更新策略参数, 生成具有更高价值的氨基酸残基片段。该价值通过综合考虑ADMET药物性质评分与生物活性对接评分所构建的奖励函数来计算, 并结合时间差异误差来获取长期预期回报, 旨在同时优化肽分子的药物学特性及其与靶蛋白之间的结合亲和力。

该模型为肽类药物的生成和优化提供了一种全新的方法, 显著提高了药物设计的效率和准确度, 不仅减少了肽类药物开发和优化过程中的人力、财力和物力的消耗, 还为新药研发提供了更多可能性。

2.4.2 基于图生成模型在分子优化中的应用

2.4.2.1 MolGAN框架: GAN与强化学习的融合 由于离散化学空间的庞大性, 在如此广阔的空间中寻找具有特定目标性质的分子是一项极为复杂且具有挑战性的任务。传统的基于SMILES字符串的生成模型, 虽然可以生成分子, 但是在处理复杂化学结构及生成的分子有效性方面常常面临困难。为了缓解这些问题, De Cao等[99]提出了一种小尺寸分子图的隐式生成模型, 名为MolGAN。MolGAN创新性地将对抗生成网络与强化学习奖励机制相耦合, 构建了基于分子图的联合优化框架。其模型由3个主要组件构成: ① 生成器 (generator) 从一个先验分布中采样向量, 生成一个表述分子的图; ② 鉴别器 (discriminator) 负责区分生成的分子图和真实数据集中的分子图, 通过对抗训练使生成器逐步学会生成与真实数据分布相似的分子; ③ 奖励网络 (reward network) 用于评估生成分子的特定化学性质 (可溶性、药效等), 并通过强化学习目标优化生成器, 使其生成具有理想性质的分子。

MolGAN的创新点在于直接在分子图空间进行生成, 有别于传统方法生成SMILES字符串的模式。在QM9数据集开展的实验中, MolGAN的这种方法充分彰显了其优势。经实验验证显示, 其生成的分子有效性接近100%。MolGAN的多样性数值为2.0%～2.3%, 在生成的有效分子中, MolGAN的新颖性数值达到了94.2%。与传统的基于SMILES表示的生成模型相比, 例如CharacterVAE和GrammarVAE, 二者在QM9数据集上生成的分子有效性分别为10.3%和60.2%, 远低于MolGAN接近100%的有效性。对比的结果证明, MolGAN直接在图空间种生成的分子在化学有效性方面具有突出优势。

然而, MolGAN存在一定的局限性。虽然生成的分子有效性极高, 但在多样性方面, 相比于CharacterVAE的67.5%和GrammarVAE的9.3%, MolGAN并没

有体现出优势, 甚至在某些情况下低于部分基于SMILES的方法。这主要由于MolGAN容易受到模式坍缩的影响, 使得其模型容易偏执于生成有限且集中结构。尽管模型存在上述缺陷, MolGAN通过结合GAN和RL, 为直接在图空间进行目标靶向的分子生成提供了新的解决方案, 在生成有效小分子方面发挥了重要作用。

2.4.2.2 图神经网络的强化学习优化: BAR策略加速收敛 Atance等[100]提出了一种基于图结构的深度生成模型 (deep generative model, DGM) 与强化学习相结合的方法, 用于从头分子设计。该模型通过强化学习的策略梯度方法优化分子的目标属性, 重点优化药物相似性和多巴胺受体Drd2的活性。

其核心创新点在于提出最佳智能体提醒 (best agent reminder, BAR) 机制的强化学习框架。该机制通过在损失函数中保留最高分智能体的行动序列, 有效解决了传统强化学习方法在分子生成过程中容易遗忘高质量分子构建路径的难题。BAR机制通过引入平衡参数α来调节当前智能体与最佳智能体的贡献权重。实验结果表明, BAR机制相较于传统REINVENT方法具有显著优势。在无BAR机制条件下 (α=0.0), 模型学习过程缓慢且容易产生震荡。而采用BAR机制后 (α = 0.5), 模型的收敛速度明显增加, 生成分子质量大幅提升。定量分析表明, BAR模型在活性分子生成方面表现卓越: 活性分子生成率高达95%, 较预训练模型的1%提升了94个百分点; 平均QED从0.59提升至0.76; 平均Drd2活性预测值从0.03提升至0.94。BAR机制模型的显著优势, 不仅标志着分子生成技术的重要进展, 更凸显了该机制在优化活性分子产出效率上的巨大潜力, 为药物研发等领域提供了更高效的分子设计解决方案。

尽管DGM在活性分子生成领域展现出显著优势, 但该模型仍面临一些局限性。首先, 模型训练过程需要消耗大量计算资源, 且预训练周期较长; 其次, 在优化较大分子结构时, 模型可能会生成含有不稳定环状化学结构的分子, 影响实际应用效果。

2.4.3 多参数优化: 多目标驱动的分子优化与生成策略

2.4.3.1 DeepFMPO: 片段级多目标优化 Ståhl等[101]提出了一种名为 (deep fragment-based multiparameter optimization, DeepFMPO) 的强化学习分子优化方法, 通过引入Actor-Critic强化学习框架, 利用片段级优化策略逐步替换分子片段, 实现多目标分子属性优化。DeepFMPO的创新性在于其独特的分子表示和生成策略。首先, 将分子分解为有效片段; 其次, 基于片段的结构相似性和平衡二叉树来构建片段的二进制编码; 最后, 强化学习智能体选择学习某个片段后, 将其替换为结构相似、性质更优的新片段。

为了模拟真实药物优化场景, 该研究团队将一组多巴胺D2/D4受体活性化合物优化至预设的“理想区间”, 即分子质量、脂水分配系数和极性表面积达到特定范围内。研究采用了动态调整型奖励函数, 根据各属性目标的难易程度对已生成分子进行评估和更新, 使智能体能够有效平衡多重优化目标。实验表明, DeepFMPO最终生成了93%的化学有效分子, 其中33%的分子成功进入了预先定义的“理想区间”。具体而言, 在满足脂水分配系数和极性表面积目标的情况下, 分子比例分别从最初的10.1%和20.4%显著上升至52.1%和66.1%。但该研究并未与其他先进的生成和优化模型进行定量比较, 其优化范围主要局限于初始分子库及其片段组合生成的邻近化学空间, 对于发现全新骨架或远离初始结构的分子能力有限。尽管如此, DeepFMPO证明了基于片段的强化学习方法是处理多参数目标药物分子优化的一个具有潜力的方向。

2.4.3.2 多样性驱动的分子生成: 双生成器策略 在药物分子优化中, 如何高效生成既满足生物活性要求又兼顾结构多样性的候选分子, 是一项亟待解决的核心技术难题。当前主流的分子生成方法普遍面临多样性与目标特征难以协同优化的重大挑战。

针对于分子生成和优化领域的痛点, Pereira等[102]开发了一种基于SMILES表示法和强化学习策略梯度算法的分子生成框架。该框架采用双生成器策略分别进行分子的探索和开发, 通过动态调整两者权重来平衡多样性与目标特征。一个生成器负责探索广泛的化学空间, 另一个生成器则通过强化学习不断更新, 根据所获得的经验去生成具有更高目标属性得分的分子。该分子生成框架由生成器和预测器组成。生成器基于循环神经网络, 用于学习SMILES字符串的语法规则生成有效的分子序列。预测器为一个定量构效模型, 用于评估分子与特定靶点结合能力, 并提供奖励信号给强化学习过程。该框架创新性地融入基于Tanimoto相似度的动态奖惩机制。具体而言, 当新生成分子与分子库中最佳生成分子的平均Tanimoto相似度超过预设阈值时, 模型将自动对该分子的奖励值进行衰减处理 (乘以0.85系数)。这一机制通过量化分子结构间的相似程度, 形成负反馈调节回路, 有效引导生成器规避重复生成相似分子结构, 从而推动模型探索更具结构多样性的化学空间, 确保生成分子的新颖性与独特性。

在针对腺苷A2A和阿片受体抑制活性的优化任务中, 该双生成器框架模型展现了良好的性能。以

A2A 为例, 在优化 pIC_{50} 的任务中, 模型生成了 90.4% 的化学有效性分子, 其中 95.1% 的分子满足目标属性要求, 即 p 值大于 6.5。并且, 经优化器生成的分子在多样性上也保持了较高水平, 其多样性指标平均可达到 0.879, 百分比表示的独特性高达 96.4%。以上实验结果表明, 双生成器策略和奖惩机制可以有效平衡目标属性优化和生成分子的新颖性和多样性, 为药物分子优化领域提供了新的多目标权衡的强化学习方法。

2.4.4 可合成性验证: 实验驱动的强化学习闭环

在分子优化的流程中, 生成分子的可合成性和实验验证是决定化合物优化价值的关键因素。传统分子生成模型多依赖片段组合或物理性质预测合成可行性, 但未深度整合模块化化学反应规则, 导致生成分子常因反应条件严苛、不良反应多等问题难以实际合成。针对这一挑战, Wang 等[103]提出了一种名为 ClickGen 的新型分子生成模型, 将强化学习与模块化化学反应规则 (如点击化学和酰胺化反应) 相结合, 显著提高了生成分子的可合成性和生物活性。ClickGen 的核心组件包括 3 个模块: ① 化学反应组合器, 该模块基于点击化学和酰胺化反应规则, 组合可合成的分子片段; ② 修复生成模型, 其模型借鉴图像修复技术, 通过填补缺失的分子片段生成结构新颖的分子; ③ 基于蒙特卡洛树搜索的强化学习模块, 此模块利用蛋白质口袋的对接分数引导分子生成, 平衡可合成性与靶点结合力。

在 PARP1 抑制剂开发任务的实验验证中, ClickGen 展现出卓越的性能。研究人员仅用 20 天便成功合成了多个候选化合物, 并验证了它们的生物活性。其中, 两个先导化合物在癌细胞系中表现出显著的抗增殖效果, 且对 PARP1 展现了纳摩尔级的抑制活性。与同类基于反应的生成模型 (如 BBAR 和 Synnet) 相比, ClickGen 生成的分子在结构新颖性和合成可行性方面具有明显优势。未来, ClickGen 若与自动化合成平台相结合, 有望实现从分子设计到实验验证的闭环流程, 这将极大提升药物研发的效率和自动化水平。

3 总结和展望

虽然近年来深度学习和强化学习算法在药物分子优化中取得了显著进展, 但仍面临诸多挑战。首先, 化学空间的覆盖与探索依然是关键难题。已知的化学空间包括超过 10^{60} 种潜在分子, 现有算法对庞大化学空间的探索能力有限, 生成的分子通常局限于训练数据的分布范围, 难以生成显著偏离已知化学骨架的分子[104,105]。例如, 基于图的生成模型 (如 MolGAN) 在探索新颖化学区域方面取得了一定成果, 但是在生物合成可行且具有生物活性的分子方面仍表现欠佳。未来的研究应集中于整合拓扑和几何描述符, 以增强对未知化学空间的探索能力。其次, 多目标优化的权衡问题实际上是平衡多个竞争性属性的过程[106]。药物分子优化需要在多种性质 (如活性、毒性和药代动力学性能) 之间找到最佳平衡, 传统的强化学习框架通常使用加权奖励函数, 但这类方法可能难以准确捕捉这些属性之间的复杂关系[86]。近期兴起的多智能体强化学习 (multi-agent reinforcement learning, MARL) 作为强化学习的子领域, 侧重于研究共享环境种共存的多个学习代理的行为, 每个代理都收到自身奖励的激励, 并采取行动。MARL 有望引入分子优化领域, 用于解决多目标优化的复杂难题。此外, 数据库的稀缺性和模型的鲁棒性也是主要限制因素[107]。药物研发中高质量数据的稀缺, 直接影响了模型的训练效果。尤其在数据有限的场景下, 深度学习和强化学习模型需要具备更强的外推能力和鲁棒性。尽管 ChEMBL 和 PubChem 等公共数据库具有重要价值, 但其数据一致性和注释准确性往往存在不足[108]。值得注意的是, 生成分子的化学有效性与可合成性仍然是药物设计的重要考量[104], 部分生成分子在实际合成中面临较大困难, 因此结合化学反应规则开发生成式模型, 并通过实验验证生成分子的实际可行性, 将是未来研究的重点。最后, 探索高维化学空间的计算需求对强化学习和深度学习框架提出了重大挑战。如何通过量子计算和硬件加速技术提高计算效率, 特别是在强化学习和深度学习框架下, 将是下一阶段的重要研究方向[107]。

药物分子优化领域的创新突破将依赖于算力的提升与多学科技术融合。一方面, 深度强化学习算法通过解析分子化学语言特征空间, 已成功构建起基于数据驱动的化合物生成方法。然而, 随着分子数据集向海量、高维、复杂方向发展, 传统计算架构在处理这类数据时逐渐显示出效率瓶颈。基于量子计算技术的量子态并行处理特性可突破经典算法的复杂度限制, 这种计算能力的提升, 有望在分子设计的关键环节实现指数级加速, 从而打开更广阔的药物创新空间[109]。另一方面, 跨领域协同已成为驱动药物研发的核心动力。传统的由单一学科主导的研究模式正在逐步让位于多学科交叉体系, 在新药研发与药物结构优化的关键环节, 需要研究人员掌握化学、生物学、计算机科学等多领域的知识[110-112]。未来, 人工智能技术的深度融合将引领药物研发方式从经验驱动转向智能计算驱动的重大变革, 为新药研发提供了新的技术支撑[113]。

作者贡献: 李星雨撰写论文初稿; 武潼对文章进行了修改和定稿; 徐启兰和王奕文进行文献收集; 余河水提供了修改意见; 岑世欣和李正负责论文的框架设计、核心观点提炼

以及学术指导。

利益冲突: 本文所有作者声明不存在利益冲突关系。

References

[1] Wang MY, Li D, Hou TJ, et al. Research progress on novel drug design based on deep learning [J]. Sci Sin Chim (中国科学: 化学), 2023, 53: 95-106.

[2] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs [J]. J Health Econ, 2016, 47: 20-33.

[3] Wilhelm S, Carter C, Lynch M, et al. Discovery and development of sorafenib: a multikinase inhibitor for treating cancer [J]. Nat Rev Drug Discov, 2006, 5: 835-844.

[4] Lindberg P, Brändström A, Wallmark B, et al. Omeprazole: the first proton pump inhibitor [J]. Med Res Rev, 1990, 10: 1-54.

[5] Patne AY, Dhulipala SM, Lawless W, et al. Drug discovery in the age of artificial intelligence: transformative target-based approaches [J]. Int J Mol Sci, 2024, 25: 12233.

[6] Polishchuk PG, Madzhidov TI, Varnek A. Estimation of the size of drug-like chemical space based on GDB-17 data [J]. J Comput Aided Mol Des, 2013, 27: 675-679.

[7] Andricopulo AD, Guido RV, Oliva G. Virtual screening and its integration with modern drug design technologies [J]. Curr Med Chem, 2008, 15: 37-46.

[8] Yu J, Li Z, Chen G, et al. Computing the relative binding affinity of ligands based on a pairwise binding comparison network [J]. Nat Comput Sci, 2023, 3: 860-872.

[9] Li W, Yang JC, Huang N. Deep learning in drug design and discovery [J]. Acta Pharm Sin (药学学报), 2019, 54: 761-767.

[10] Dai QQ, Yu JL, Li GB. Research progress of deep learning assisted drug discovery [J]. Prog Pharm Sci (药学进展), 2022, 46: 60-70.

[11] Wang XZ, Zhao YX, Pourpanah F. Recent advances in deep learning [J]. Int J Mach Learn Cybern, 2020, 11: 747-750.

[12] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge [J]. Int J Comput Vis, 2015, 115: 211-252.

[13] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks [J]. Commun ACM, 2017, 60: 84-90.

[14] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521: 436-444.

[15] Bengio Y, Goodfellow I, Courville A. Deep learning [M]. Cambridge: MIT press, 2016: 351-354.

[16] Wang P, Fan E, Wang P. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning [J]. Pattern Recognit Lett, 2021, 141: 61-67.

[17] Çayir A, Yenidoğan I, Dağ H. Feature extraction based on deep learning for some traditional machine learning methods [C]// 2018 3rd International Conference on Computer Science and Engineering (UBMK). Sarajevo, Bosnia and Herzegovina: IEEE, 2018: 494-497.

[18] Chauhan NK, Singh K. A review on conventional machine learning vs deep learning [C]// 2018 International Conference on Computing, Power and Communication Technologies (GUCON). Greater Noida: IEEE, 2018: 347-352.

[19] Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning [J]. Electron Mark, 2021, 31: 685-695.

[20] Rusk N. Deep learning [J]. Nat Methods, 2016, 13: 35.

[21] Gupta R, Srivastava D, Sahu M, et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery [J]. Mol Divers, 2021, 25: 1315-1360.

[22] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks [J]. Commun ACM, 2020, 63: 139-144.

[23] Liang SS, Pan Z, Liu W, et al. A survey on variational autoencoders in recommender systems [J]. ACM Comput Surv, 2024, 56: 1-40.

[24] Gilmer J, Schoenholz SS, Riley PF, et al. Neural message passing for quantum chemistry [C]// Proceedings of the 34th International Conference on Machine Learning. Sydney: PMLR, 2017: 1263-1272.

[25] Schmidhuber J. Deep learning in neural networks: an overview [J]. Neural Netw, 2015, 61: 85-117.

[26] Nigam A, Pollice R, Krenn M, et al. Beyond generative models: superfast traversal, optimization, novelty, exploration and discovery (STONED) algorithm for molecules using SELFIES [J]. Chem Sci, 2021, 12: 7079-7090.

[27] Cai H, Zhang H, Zhao D, et al. FP-GNN: a versatile deep learning architecture for enhanced molecular property prediction [J]. Brief Bioinform, 2022, 23: bbac408.

[28] Zhang N, Wang XY, Han B, et al. Research progress of artificial intelligence technology in pharmacology of traditional Chinese medicine [J]. Acta Pharm Sin (药学学报), 2025, 60: 550-558.

[29] Chang YP, Wang X, Wang JD, et al. A survey on evaluation of large language models [J]. ACM Trans Intell Syst Technol, 2024, 15: 1-45.

[30] Zhang Q, Ding KY, Lv TW, et al. Scientific large language models: a survey on biological & chemical domains [J]. ACM Comput Surv, 2025, 57: 1-38.

[31] Bran AM. Transformers and large language models for chemistry and drug discovery [M]// Schwaller P. Drug Development Supported by Informatics. Singapore: Springer, 2024: 143-163.

[32] Leon M, Perezhohin Y, Peres F, et al. Comparing SMILES and SELFIES tokenization for enhanced chemical language modeling [J]. Sci Rep, 2024, 14: 25016.

[33] Kusner MJ, Paige B, Hernández-Lobato JM. Grammar variational autoencoder [C]// Proceedings of the 34th International Conference on Machine Learning. Sydney: PMLR, 2017: 1945-1954.

[34] Vogt M. Using deep neural networks to explore chemical space

[J]. Expert Opin Drug Discov, 2022, 17: 297-304.

[35] Gómez-Bombarelli R, Wei JN, Duvenaud D, et al. Automatic chemical design using a data-driven continuous representation of molecules [J]. ACS Cent Sci, 2018, 4: 268-276.

[36] Zhu Y, Wu Y, Sebe N, et al. Vision+ x: a survey on multimodal learning in the light of data [J]. IEEE Trans Pattern Anal Mach Intell, 2024, 46: 9102-9122.

[37] Guevorguian P, Bedrosian M, Fahradyan T, et al. Small molecule optimization with large language models [J]. arXiv preprint arXiv, 2024. DOI:10.48550/arXiv.2407.18897.

[38] Lambora A, Gupta K, Chopra K. Genetic algorithm-a literature review [C]// 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). Faridabad: IEEE, 2019: 380-384.

[39] Bran AM, Cox S, Schilter Oliver, et al. Augmenting large language models with chemistry tools [J]. Nature, 2024, 6: 525-535.

[40] Ye GY, Cai XB, Lai HT, et al. Drugassist: a large language model for molecule optimization [J]. Brief Bioinform, 2025, 26: bbae693.

[41] Yoshizawa T, Ishida S, Sato T, et al. A data-driven generative strategy to avoid reward hacking in multi-objective molecular design [J]. Nat Commun, 2025, 16: 2409.

[42] Wu Z, Zhang O, Wang X, et al. Leveraging language model for advanced multiproperty molecular optimization *via* prompt engineering [J]. Nat Mach Intell, 2024, 6: 1359-1369.

[43] Kairouz P, McMahan HB, Avent B, et al. Advances and open problems in federated learning [J]. Found Trends Mach Learn, 2021, 14: 1-210.

[44] Li L, Fan Y, Tse M, et al. A review of applications in federated learning [J]. Comput Ind Eng, 2020, 149: 106854.

[45] Hanser T, Ahlberg E, Amberg A, et al. Data-driven federated learning in drug discovery with knowledge distillation [J]. Nat Mach Intell, 2025, 7: 423-436.

[46] Gou J, Yu B, Maybank SJ, et al. Knowledge distillation: a survey [J]. Int J Comput Vis, 2021, 129: 1789-1819.

[47] Scarselli F, Gori M, Tsoi AC, et al. The graph neural network model [J]. IEEE Trans Neural Netw, 2008, 20: 61-80.

[48] Zhou J, Cui G, Hu S, et al. Graph neural networks: a review of methods and applications [J]. AI open, 2020, 1: 57-81.

[49] Corso G, Stark H, Jegelka S, et al. Graph neural networks [J]. Nat Rev Methods Primers, 2024, 4: 17.

[50] Bronstein MM, Bruna J, LeCun Y, et al. Geometric deep learning: going beyond euclidean data [J]. IEEE Signal Process Mag, 2017, 34: 18-42.

[51] Wang Y. Graph neural networks for molecules [M]// Li Z, Farimani A. Machine learning in molecular sciences. Cham: Springer, 2023: 21-66.

[52] Kearnes S, McCloskey K, Berndl M, et al. Molecular graph convolutions: moving beyond fingerprints [J]. J Comput Aided Mol Des, 2016, 30: 595-608.

[53] Duvenaud DK, Maclaurin D, Iparraguirre J, et al. Convolutional networks on graphs for learning molecular fingerprints [C]// NIPS'15: Proceedings of the 29th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015, 2224-2232.

[54] Jiang Y, Jin S, Jin X, et al. Pharmacophoric-constrained heterogeneous graph transformer model for molecular property prediction [J]. Commun Chem, 2023, 6: 60.

[55] Zeng K, Yang B, Zhao X, et al. Ualign: pushing the limit of template-free retrosynthesis prediction with unsupervised SMILES alignment [J]. J Cheminform, 2024, 16: 80.

[56] Asif NA, Sarker Y, Chakrabortty RK, et al. Graph neural network: a comprehensive review on non-euclidean space [J]. IEEE Access, 2021, 9: 60588-60606.

[57] Wan Y, Hsieh CY, Liao B, et al. Retroformer: pushing the limits of end-to-end retrosynthesis transformer [C]// Proceedings of the 39th International Conference on Machine Learning. Baltimore: PMLR, 2022: 22475-22490.

[58] Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks [J]. IEEE Trans Neural Netw Learn Syst, 2020, 32: 4-24.

[59] Tiezzi M, Marra G, Melacci S, et al. Deep constraint-based propagation in graph neural networks [J]. IEEE Trans Pattern Anal Mach Intell, 2021, 44: 727-739.

[60] Reiser P, Neubert M, Eberhard A, et al. Graph neural networks for materials science and chemistry [J]. Commun Mater, 2022, 3: 93.

[61] Zhang R, Wu C, Yang Q, et al. MolFeSCue: enhancing molecular property prediction in data-limited and imbalanced contexts using few-shot and contrastive learning [J]. Bioinformatics, 2024, 40: btae118.

[62] Yu L, Qiu W, Lin W, et al. HGDTI: predicting drug-target interaction by using information aggregation based on heterogeneous graph neural network [J]. BMC Bioinformatics, 2022, 23: 126.

[63] Stanković L, Mandic D. Convolutional neural networks demystified: a matched filtering perspective-based tutorial [J]. IEEE Trans Syst Man Cybern Syst, 2023, 53: 3614-3628.

[64] Zeng P, Zhang B, Liu A, et al. Drug repositioning based on tripartite cross-network embedding and graph convolutional network [J]. Expert Syst Appl, 2024, 252: 124152.

[65] Gilmer J, Schoenholz SS, Riley PF, et al. Message passing neural networks [M]// Machine Learning Meets Quantum Physics. Cham: Springer, 2020: 199-214.

[66] Jin HF, Tan JW, Liu M. Drug-target affinity prediction based on graph convolutional network with residual structure [J]. Adv Biomed Eng (生物医学工程学进展), 2023, (4): 371-380.

[67] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks [J]. arXiv preprint arXiv, 2016. DOI:

10.48550/arXiv.1609.02907.

[68] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks [J]. arXiv preprint arXiv, 2017. DOI: 10.48550/arXiv.1710.10903.

[69] Shao K, Zhang Y, Wen Y, et al. DTI-HETA: prediction of drug-target interactions based on GCN and GAT on heterogeneous graph [J]. Brief Bioinform, 2022, 23: bbac109.

[70] Wang Y, Hou Y, Che W, et al. From static to dynamic word representations: a survey [J]. Int J Mach Learn Cybern, 2020, 11: 1611-1630.

[71] Khemchandani Y, O′Hagan S, Samanta S, et al. DeepGraphMolGen, a multi-objective, computational strategy for generating molecules with desirable properties: a graph convolution and reinforcement learning approach [J]. J Cheminform, 2020, 12: 1-17.

[72] Tang M, Li B, Chen H. Application of message passing neural networks for molecular property prediction [J]. Curr Opin Struct Biol, 2023, 81: 102616.

[73] Bhaskar D, Grady J, Castro E, et al. Molecular graph generation *via* geometric scattering [C]// 2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP). Xi'an: IEEE, 2022: 1-6.

[74] Wang R, Fang X, Lu Y, et al. The PDBbind database: methodologies and updates [J]. J Med Chem, 2005, 48: 4111-4119.

[75] Durant G, Boyles F, Birchall K, et al. The future of machine learning for small-molecule drug discovery will be driven by data [J]. Nat Comput Sci, 2024, 4: 735-743.

[76] Siebenmorgen T, Menezes F, Benassou S, et al. MISATO: machine learning dataset of protein-ligand complexes for structure-based drug discovery [J]. Nat Comput Sci, 2024, 4: 367-378.

[77] Buterez D, Janet JP, Kiddle SJ, et al. Transfer learning with graph neural networks for improved molecular property prediction in the multi-fidelity setting [J]. Nat Commun, 2024, 15: 1517.

[78] Atz K, Cotos L, Isert C, et al. Prospective *de novo* drug design with deep interactome learning [J]. Nat Commun, 2024, 15: 3408.

[79] Ryu JY, Elala E, Rhee JK. Quantum graph neural network models for materials search [J]. Materials (Basel), 2023, 16: 4300.

[80] Fu TF, Xiao C, Glass LM, et al. MOLER: incorporate molecule-level reward to enhance deep generative model for molecule optimization [J]. IEEE Trans Knowl Data Eng, 2022, 34: 5459-5471.

[81] Liu S, Wang H, Liu W, et al. Pre-training molecular graph representation with 3d geometry [J]. arXiv preprint arXiv, 2021. DOI:10.48550/arXiv.2110.07728.

[82] Chen Z, Min MR, Parthasarathy S, et al. A deep generative model for molecule optimization *via* one fragment modification [J]. Nat Mach Intell, 2021, 3: 1040-1049.

[83] Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: a survey [J]. J Artif Intell Res, 1996, 4: 237-285.

[84] Shakya AK, Pillai G, Chakrabarty S. Reinforcement learning algorithms: a brief survey [J]. Expert Syst Appl, 2023, 231: 120495.

[85] Szepesvári C. Algorithms for reinforcement learning [M]// Synthesis Lectures on Artificial Intelligence and Machine Learning. Cham: Springer, 2010: 1-103.

[86] Zhou ZP, Kearnes S, Li L, et al. Optimization of molecules *via* deep reinforcement learning [J]. Sci Rep, 2019, 9: 10752.

[87] Fu T, Gao W, Coley C, et al. Reinforced genetic algorithm for structure-based drug design [J]. Adv Neural Inform Process Syst, 2022, 35: 12325-12338.

[88] Matsuo Y, LeCun Y, Sahani M, et al. Deep learning, reinforcement learning, and world models [J]. Neural Netw, 2022, 152: 267-275.

[89] Watkins CJ, Dayan P. Q-learning [J]. Mach Learn, 1992, 8: 279-292.

[90] You J, Liu B, Ying R, et al. Graph convolutional policy network for goal-directed molecular graph generation [C]// NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2018: 6412-6422.

[91] Jin W, Barzilay R, Jaakkola T. Junction tree variational autoencoder for molecular graph generation [C]// Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018: 2323-2332.

[92] Chang P. Improving Transformer-Based Molecular Optimization Using Reinforcement Learning [D]. Uppsala: Uppsala University, 2021.

[93] Abeer AN, Urban NM, Weil MR, et al. Multi-objective latent space optimization of generative molecular design models [J]. Patterns (N Y), 2024, 5: 101042.

[94] Mazuz E, Shtar G, Shapira B, et al. Molecule generation using transformers and policy gradient reinforcement learning [J]. Sci Rep, 2023, 13: 8799.

[95] Muttenthaler M, King GF, Adams DJ, et al. Trends in peptide drug discovery [J]. Nat Rev Drug Discov, 2021, 20: 309-325.

[96] Craik DJ, Fairlie DP, Liras S, et al. The future of peptide-based drugs [J]. Chem Biol Drug Des, 2013, 81: 136-147.

[97] Zhu YS, Tang K, Lv J. Peptide-drug conjugate-based novel molecular drug delivery system in cancer [J]. Trends Pharmacol Sci, 2021, 42: 857-869.

[98] Wang Q, Hu XT, Wei ZQ, et al. Reinforcement learning-driven exploration of peptide space: accelerating generation of drug-like peptides [J]. Brief Bioinform, 2024, 25: bbae444.

[99] De Cao N, Kipf T. MolGAN: an implicit generative model for small molecular graphs. arXiv 2018 [J]. arXiv preprint arXiv, 2018. DOI: 10.48550/arXiv.1805.11973.

[100] Atance SR, Diez JV, Engkvist O, et al. *De novo* drug design using reinforcement learning with graph-based deep generative models [J]. J Chem Inf Model, 2022, 62: 4863-4872.

[101] Ståhl N, Falkman G, Karlsson A, et al. Deep reinforcement learning for multiparameter optimization in de novo drug design [J]. J Chem Inf Model, 2019, 59: 3166-3176.

[102] Pereira T, Abbasi M, Ribeiro B, et al. Diversity oriented deep reinforcement learning for targeted molecule generation [J]. J Cheminform, 2021, 13: 21.

[103] Wang MY, Li S, Wang JK, et al. ClickGen: directed exploration of synthesizable chemical space *via* modular reactions and reinforcement learning [J]. Nat Commun, 2024, 15: 10127.

[104] Horwood J, Noutahi E. Molecular design in synthetically accessible chemical space *via* deep reinforcement learning [J]. ACS Omega, 2020, 5: 32984-32994.

[105] Sadybekov AA, Sadybekov AV, Liu YF, et al. Synthon-based ligand discovery in virtual libraries of over 11 billion compounds [J]. Nature, 2022, 601: 452-459.

[106] Gunantara N. A review of multi-objective optimization: methods and its applications [J]. Cogent Eng, 2018, 5: 1502242.

[107] Liang L, Deng CL, Zhang YM, et al. Applications and challenges of artificial intelligence in drug discovery [J]. Prog Pharm Sci (药学进展), 2020, 44: 18-27.

[108] Liu RZ, Song JK, Liu AL, et al. Progress on the application of artificial intelligence technology in ligand-based and receptor structure-based drug screening [J]. Acta Pharm Sin (药学学报), 2021, 56: 2136-2145.

[109] Ajagekar A, You F. Molecular design with automated quantum computing-based deep learning and optimization [J]. NPJ Comput Mater, 2023, 9: 143.

[110] Ye Q, Xu H, Ye J, et al. Mplug-owl2: revolutionizing multi-modal large language model with modality collaboration [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2024: 13040-13051.

[111] Santagati R, Aspuru-Guzik A, Babbush R, et al. Drug design on quantum computers [J]. Nat Phys, 2024, 20: 549-557.

[112] Goshisht MK. Machine learning and deep learning in synthetic biology: key architectures, applications, and challenges [J]. ACS Omega, 2024, 9: 9921-9945.

[113] Li GJ. AI for research (AI4R): the fifth research paradigm [J]. Bull Chin Acad Sci (中国科学院院刊), 2024, 39: 1-9.