arXiv:2601.10110v1 [cs.NE] 15 Jan 2026

多重约束进化分子设计框架：结合基于规则的进化与分子交叉的可解释药物设计方法

Shanxian Lin[1 [0009−0001−4855−8252]]、WeiXia[2], Yuichi Nagata[1 [0000−0001−7582−2699]]*和Haichuan Yang[1 [0000−0001−7100−7945]]**

[1] 德岛大学技术、工业与社会科学研究生院，日本德岛 770-8506 utigooselsx@gmail.com、nagata@is.tokushima-u.ac.jp、you.kaisen@tokushima-u.ac.jp[2] 中国石油大学（华东）化学与化学工程学院，中国山东青岛 266580 xiawei@upc.edu.cn

摘要。 本研究提出 MCEMOL（多重约束进化分子设计框架），这是一种将基于规则的进化与分子交叉相结合的分子优化方法。MCEMOL 采用双层进化：在规则层面优化变换规则，同时对分子结构进行交叉和变异。与需要大量数据集和广泛训练的深度学习方法不同，我们的算法能够从少量起始分子出发，以较低的计算开销高效进化。该框架结合了消息传递神经网络和全面的化学约束，确保分子设计高效且可解释。实验结果表明，MCEMOL 通过其进化机制提供透明的设计路径，同时生成有效、多样且符合目标要求的分子。该框架实现了 100% 的分子有效性，具有较高的结构多样性和出色的类药性合规性，并在对称性约束、药效团优化和立体化学完整性方面表现优异。不同于黑箱方法，MCEMOL 带来双重价值：一方面提供研究人员能够理解并信赖的可解释变换规则，另一方面提供可用于实际应用的高质量分子库。这确立了一种新的范式，使可解释的 AI 驱动药物设计与有效的分子生成得以同时实现，从而弥合计算创新与实际药物发现需求之间的差距。

关键词： 进化算法 · 药物设计 · 分子优化 · 可解释人工智能 · 分子交叉。

⋆ *通讯作者

1 引言

传统的分子设计依赖专家直觉和反复试错，使发现过程缓慢且成本高昂。药物开发可能耗时 10-15 年，成本超过 10 亿美元，成功率低于 10% [33]。计算机辅助分子设计通过在计算机上优先筛选候选分子、降低实验成本，并提高靶标特异性和安全性，改善了这一过程。

近期机器学习（ML）方法推动了分子生成的发展[39, 11], ，但仍面临三大主要挑战：（i）确保化学有效性和合成可行性；（ii）避免结构多样性丧失；（iii）提供可解释设计 [15, 30]。

变分自编码器（VAEs） [17] 和生成对抗网络（GANs） [7] 等传统方法展现出良好前景，但往往会生成无效或不可合成的分子。强化学习（RL）方法能够优化特定属性，但缺乏可解释性，且可能发生过拟合 [52, 51]。基于规则的系统能够保持有效性，但受限于固定的变换集合 [35, 40]。

我们提出了 MCEMOL，即多约束进化分子设计框架，将基于规则的进化与分子级交叉相结合。MCEMOL 采用双层进化方法：同时进化分子结构和变换规则，在确保化学有效性的同时提高设计柔性。尤为重要的是，与需要大量计算资源和大型数据集的深度学习方法不同，MCEMOL 仅需极少的计算开销即可高效运行，并且只需少量起始分子便能开展进化。通过整合消息传递神经网络（ATLAS-DMPNN）和多维化学约束系统，MCEMOL 能够优化具有多种属性的分子，包括合成可行性、对称性和药效团完整性。与“黑箱”方法不同，MCEMOL 提供双重输出：通过发现的变换规则给出可解释的设计路径，并生成一个有效性达 100%、类药性合规性优异的高质量分子库。这一双重能力确保研究人员既能获得机理见解，也能得到可用于进一步开发的实用候选分子。我们的贡献包括：（1）提出一种受约束的进化框架，在确保化学有效性的同时生成多属性靶向分子；（2）将规则级学习与分子级进化相结合，以保持多样性和可合成性；（3）发现经实证验证、适用于特定药物分子反应的规则；（4）在提供可直接使用的分子库的同时，给出发现的变换规则，证明可解释性与生成质量并非相互排斥；（5）实现高计算效率，无需大型数据集或长时间训练即可实际部署。

2 文献综述

分子设计的机器学习方法经历了多个发展范式。深度生成模型，尤其是变分自编码器和生成对抗网络，开创了

利用连续潜在空间进行分子生成的方法。Gómez-Bombarelli 等人 [17] 将 SMILES 字符串编码为潜在表示，以进行基于梯度的优化；而 Jin 等人 [24] 则将这种方法扩展到了分子图。基于流的模型 [41] 引入了数据空间与潜在空间之间的可逆变换。然而，这些方法经常生成无效或不可合成的分子。

强化学习将分子生成视为由奖励引导策略驱动的序贯决策过程。Zhou 等人 [52]采用策略梯度方法进行性质优化，而 You 等人 [51] 开发了用于多约束满足的基于图的策略。尽管强化学习能够有效控制性质，但强化学习方法存在可解释性差和策略崩溃的问题，导致结构多样性降低。图神经网络 [16, 10] 提供了关键的性质预测能力，可用于引导这些优化过程。

基于规则的系统利用化学知识来确保有效性和可合成性。CReM [35] 和合成规划框架 [40]采用预定义的变换规则，但静态规则集限制了新化学型的发现。进化算法通过选择、交叉和变异操作 [22, 6], 克服这一局限，但要保持化学有效性仍然具有挑战性。

进化规则学习（ERL）系统将进化搜索与基于规则的可解释性相结合。XCS [45] 和遗传模糊系统[5] 等系统能够在管理不确定性的同时调整规则集。包括片段重组 [14]和 BRICS 分解 [8],在内的化学感知型交叉算子，能够在进化过程中保留关键子结构。

分子设计需要平衡结构有效性（键合、化合价、立体化学） [25, 3], 、可合成性（合成可及性评分） [13], 和药理性质。类药性筛选器 [28, 46] 和药效团模式 [50] 可指导优化。有效的设计需要权衡效力、ADMET 性质和复杂性 [31] ，同时保持可解释性和化学有效性。

3 框架组件

为应对上述分子设计挑战，我们提出了多维约束进化分子设计框架（MCE-MOL）。该框架采用双层进化方法，同时优化变换规则并直接编辑分子结构，以实现化学有效性、性质控制和可解释性。与依赖大量数据的深度学习模型不同，MCEMOL 所需的计算资源极少，无需大量训练，即可从数量有限的起始分子中有效进化。该方法通过可解释的规则表示确保可追溯性，同时生成多样化的有效分子库。借助多策略分子交叉引擎，MCEMOL 提升了结构多样性与创新性，有效平衡了可解释性和生成性能。本节将介绍该框架的基础

组件，随后阐述这些组件如何整合并扩展为我们提出的系统。

3.1 规则表示系统。

遵循既有的进化规则学习原则 [1]„ 我们采用三元组规则表示，为我们的框架提供可解释的基础，并使化学家能够理解并信任设计过程中的分子转化。

$$R = (C, T, w_r) \tag{1}$$

其中，条件集 C 定义规则的适用标准，突变集 T 规定要执行的操作，而规则权重 w_r 则作为适应度评估中的乘法因子，影响进化操作期间规则的选择概率。

条件集 C 通过六个类别确定规则的适用性：结构类（例如芳香环）、理化性质类（例如分子量、LogP）、药理学类、稳定性类（不包括不稳定基团）、类药性类（Lipinski/Veber规则）和高级类（合成/代谢评估）。每个条件均返回布尔值，以确保精确的转化控制。

转化集T 包括七种突变类型和36个函数，例如基团添加（如甲基化）、原子取代（如将碳替换为氮）、基团修饰和环操作（如添加苯环）。例如，苯酚可以突变为环己醇、4-羟基吡啶、茴香醚或苯并[环]己醇。

规则的应用遵循明确的流程：当分子 $m \in M$满足 C中的所有条件时，将从左到右依次应用 T 中的变异；否则，该函数返回 null：

$$\text{apply}(R, m) = \begin{cases} t_1 \circ t_2 \circ \cdots \circ t_k(m), & \text{if } \forall c_j \in C,\ c_j(m) = \text{true}, \\ \text{null}, & \text{otherwise}, \end{cases} \tag{2}$$

其中，M 表示当前分子集（种群），m 表示具有 $m \in M$的单个分子；$C = \{c_j\}$ 是条件集合；$T = \langle t_1, \ldots, t_k \rangle$ 是按顺序排列的变异列表，其中 k 是变异的数量。复合操作 $t_1 \circ t_2 \circ \cdots \circ t_k$被定义为从左到右进行求值，这意味着分子$m$ 依次经历 k 个变换 $t_1, \ldots, t_k$；如果任何变换产生 null，则整个复合操作返回 null。

这种表示既可机器处理，又直观易懂，便于化学家使用，是 MCEMOL 框架[44]可解释性的关键基础。

3.2 多维化学约束系统与性质预测。

我们采用包含十二项互补约束的综合约束系统 [13, 46], , 确保生成的分子具备所需特性，并

满足实际药物开发要求。这些约束会筛除理论上有趣但不切实际的候选分子，同时能够准确预测无法通过简单结构规则推导出的复杂药理性质。

约束分为基本标准和高级标准。基本标准用于确保有效性和核心类药性行为：（i）化学有效分子的比例；（ii）点群对称性 [25],；（iii）通过消息传递神经网络进行性质预测；（iv）基于指纹的结构多样性；以及（v）关键理化性质，如 LogP 和 TPSA [28]。

高级标准侧重于可开发性和功能：（i）合成可行性 [13],；（ii）反应性／稳定性筛查；（iii）药效团符合性 [50],；（iv）构象灵活性；（v）骨架保留；（vi）立体化学完整性 [3],；以及（vii）根据药物化学指南评估的扩展类药性 [46]。

在性质预测方面，我们仅将 ATLAS-DMPNN 用于复杂的药理学终点，而基础理化性质（分子量、LogP、TPSA）则通过成熟算法直接计算。鉴于ADMET终点需要借助机器学习来捕捉非线性的构效关系，ATLAS-DMPNN 通过定向消息传递、注意力引导聚合、拓扑感知编码和结构基序挖掘 [27]来实现这一目标。该模型处理包含原子特征（类型、电荷、杂化状态、芳香性）和键特征（键级、环成员关系、立体化学）的分子图，并输出预测的性质值。我们采用原始架构（隐藏维度为 300、三层消息传递层、dropout 为 0.2），并在任务特定的数据集上从头开始训练。凭借这些特征，模型能够识别代谢软点、毒性基团和蛋白质结合位点——这些对于临床成功至关重要的关键模式，往往由于上下文建模能力有限而被标准消息传递神经网络遗漏。

根据这些约束和预测属性对每个候选分子进行评分，其中每个属性 $S_u(m)$ 都经过归一化并汇总：

$$F(m) = \sum_{i=1}^{n} w_u \cdot S_u(m) \tag{3}$$

其中，w_u 是属性 u 的权重，$S_u(m) \in [0,1]$ 是其归一化得分，综合了直接计算的数值和 ATLAS-DMPNN 的预测结果。权重因任务而异，优先考虑设计目标 [31]。

对于中枢神经系统（CNS）药物，血脑屏障（BBB）穿透需要满足严格的约束：LogP 值为 2.8（范围为 2–3），分子量约为 305 Da，拓扑极性表面积为 44.8 Å（最佳范围为 <90 Å），≤ 2 个氢键供体，以及适中的碱性（酸度系数为 8.4） [47]。中枢神经系统多参数优化评分综合了这些性质，评分为 ≥ 4.0 表示血脑屏障穿透能力较好[48]。这种综合方法确保将计算资源优先用于需要机器学习的性质，而非那些可通过确定性计算获得的性质，同时在整个进化过程中保持全面的分子评估。

3.3 分子交叉策略与连接点选择。

MCEMOL 采用七种互补的交叉策略，并通过智能连接点选择确保产物在化学上有效且结构新颖。这些策略能够生成多样化的分子子代，同时保持化学可行性；这对于探索广阔的化学空间而不产生无效结构至关重要。

七种交叉策略如下：(i) 基于片段的交叉，保留一个亲代分子的中心骨架，同时替换为另一个亲代分子的外围官能团，从而保留核心药效团结构 [26]。(ii) 基于骨架的交叉，替换中心骨架，同时保留外围侧链，从而探索能够维持相似结合相互作用的新化学型 [2]。(iii) BRICS交叉，在符合合成要求的断裂位点进行重组，以提高可及性 [8]。(iv) 子结构交换，在母体分子之间交换环或官能团 [12]。(v) 药效团引导的重组，在交叉过程中保留关键药效团特征 [50]。(vi) 多亲本重组，根据片段适应度评分，通过锦标赛选择从三个或更多亲本中选取互补片段，然后将其组装成一个子代[32]。(vii) 基于反应模板的交叉，使用反应模板提高合成可行性 [4]。

对于连接点选择，我们采用四种启发式策略进行加权概率选择，其中策略权重根据历史成功率进行更新：（i）最佳匹配，利用 Morgan指纹 [37]识别具有相似化学环境的原子。（ii）反应性基团，利用预定义的反应性基团对实现可行的成键 [19]。（iii）最小化张力，根据原子的度数和氢原子数 [36]，优先选择空间位阻可及原子。（iv）保留特征，避免破坏已识别的药效团要素 [9]。

4 方法

如图1所示，MCEMOL包含两个集成组件：分子进化引擎（顶部，标记为蓝色）和规则进化引擎（底部，标记为绿色）。以下小节将详细介绍各个组件。

4.1 规则进化引擎

基于第3.1节中的规则表示，我们开发了一种自适应遗传算法，以持续进化规则集 R。为降低高维组合空间中由局部最优导致的早熟收敛风险，MCEMOL会动态优化规则集。这确保了在有限范围内进行有效利用，从而提升分子设计性能。

进化过程由四个关键步骤组成：
初始化：生成具有多样化条件和转化的初始规则种群，以确保广泛覆盖化学空间。

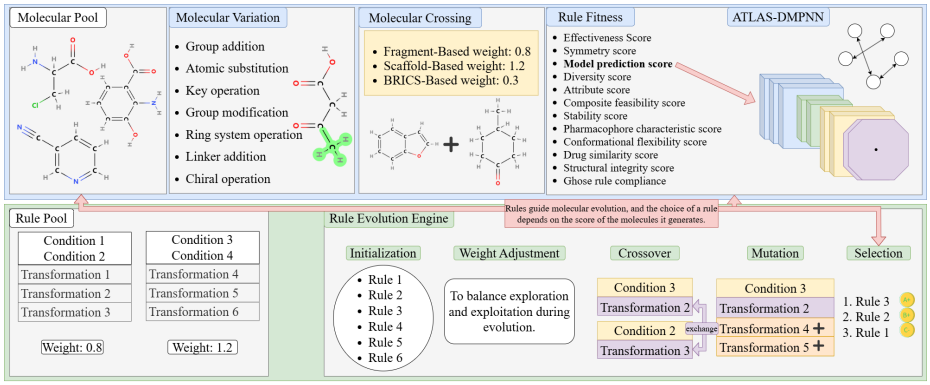


图1。 MCEMOL 的整体架构。

权重调整：为平衡进化过程中的探索与利用，我们采用了一种受强化学习策略梯度方法启发的权重更新机制。规则权重根据性能反馈进行更新，具体如下：

$$w_r = w_r + \alpha \cdot (r - b) + \eta \tag{4}$$

其中，w_r 是规则权重（不同于式（3）中的属性权重 w_u ）， α是学习率， r 是基于规则性能的奖励， b 是用于降低方差的基线，η 是用于防止早熟收敛的探索噪声。

进化操作：规则集通过选择、交叉和变异进行迭代优化。

首先，交叉通过组合亲代规则生成子代规则。由于规则 $R = (C,\ T,\ w_r)$ 由三个组成部分构成，因此条件和转换操作分别进行交叉：

$$R'^i = \begin{cases} (C', T', \bar{w}_r), & \text{if } \rho < p_c \\ R^i_{\text{p1}}, & \text{otherwise} \end{cases} \tag{5}$$

其中

$$C' = C_{\text{p1}}[:\xi_C] \oplus C_{\text{p2}}[\xi_C:], \quad T' = T_{\text{p1}}[:\xi_T] \oplus T_{\text{p2}}[\xi_T:], \quad \bar{w}_r = \frac{w_{r,\text{p1}} + w_{r,\text{p2}}}{2} \tag{6}$$

其中，p1 和 p2 分别表示父代1和父代2， $C[:\xi]$ 表示列表 C中的前 ξ 个元素，$C[\xi:]$ 表示从位置 ξ 开始的后续元素， $\oplus$表示列表拼接， ξ_C和 ξ_T 分别是各自列表长度范围内的随机交叉点。当 $\rho \sim \mathcal{U}(0,1)$满足 $\rho < p_c$时，以 p_c 的概率发生交叉。

其次，变异会引入变化，以保持多样性：

$$R''^i = \begin{cases} \Delta_\sigma(R'^i), & \text{if } \rho < p_m \\ R'^i, & \text{otherwise} \end{cases} \tag{7}$$

其中， $\rho \sim \mathcal{U}(0,\ 1)$ 中的 p_m 是突变概率， Δ_σ 表示随机扰动算子。变异会分别独立应用于条件 C和变换 T。对于条件，会按概率选择策略 $\sigma \in \{$添加、删除、替换$\}$：(i) 添加：从可用池中随机添加一个条件。(ii) 删除：移除一个现有条件（至少保留一个）。(iii) 替换： Δ_σ 用新元素替换一个现有元素。(iv) 重排： Δ_σ仅对变换序列 T应用随机排列。

最后，进化更新过程通过基于适应度的选择确定下一世代的规则。种群层面的适应度计算如下：

$$\overline{F}(M) = \frac{1}{|M|} \sum_{m \in M} F(m) \tag{8}$$

其中， $F(m)$ 是第 3.2 节（公式 (3)）中的分子适应度函数， M 表示分子种群。随后，选择机制更新规则种群：

$$R^{i+1} = \begin{cases} R''^{i}, & \text{if } \overline{F}(M^{i+1}) > \overline{F}(M^{i}) \\ R^{i}, & \text{otherwise} \end{cases} \tag{9}$$

其中， R''^{i} 表示经过交叉和变异操作后的候选规则种群， M^{i+1} 是使用 R''^{i} 生成的分子种群， M^{i} 是当前分子种群。该选择机制确保只有在规则更新能够提升分子种群整体质量时，才会保留这些更新。

4.2 分子进化引擎

分子进化引擎通过集成变异和交叉操作来管理结构进化，并与规则系统协同工作，在优化分子的同时拓展化学空间探索。

分子变异。对于变异操作，我们采用第 3.1 节定义的基于规则的变换系统。对于每个分子 $m \in M$ （其中 M 表示分子种群），当满足 C 中的条件时，按照式（2）的定义依次应用相应的变换 T。如果任何变换产生空值或不满足条件，则变异返回空值，从而确保仅保留化学有效变换。图 2 展示了分子变异的示例，说明不同的变换操作如何改变分子结构。

分子交叉。基于规则的变异确保了可解释性，而分子交叉操作对于多样性和创新至关重要。我们通过自适应引擎实现了第 3.3 节定义的七种交叉策略，该引擎根据历史表现动态选择最有效的策略。如果选定的策略未能生成有效分子，引擎便会依次尝试其他策略。如果所有策略均失败，系统将返回未经修改的亲代分子，确保交叉操作始终产生有效输出。

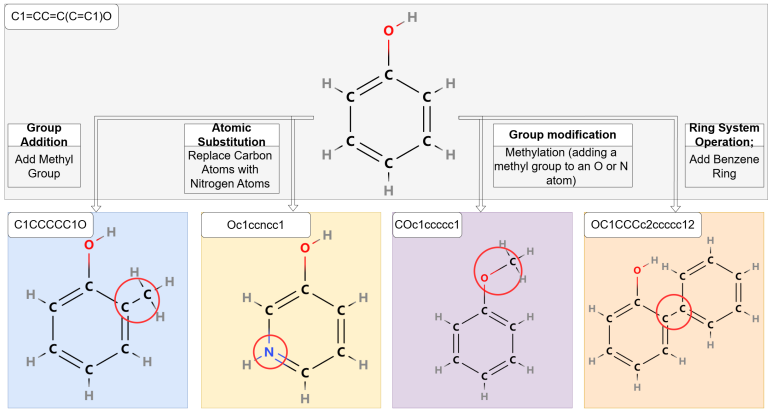


图 2. 分子变异示例。

如图 3 所示，以苯并呋喃和 4-甲基环己酮为亲代分子 A 和 B，我们的框架展示了三种交叉结果：基于片段的交叉生成 2-(4-甲基环己基)苯并呋喃，基于骨架的交叉生成 7-甲基-2,3-二氢苯并呋喃，而 BRICS交叉则生成 2-(4-氧代环己基)苯酚。

为了最大限度地提高交叉成功率，我们采用自适应策略选择器，根据交叉选项的历史表现更新其应用概率，使搜索倾向于近期效果良好的策略[38]:

$$P_y = \frac{S_y}{\sum_j S_j} \tag{10}$$

其中， P_y 是选择策略 y的概率， S_y 是根据近期代计算得出的策略 y 成功率。这种自适应机制使系统能够学习哪些交叉策略最适合当前优化任务和分子种群。

种群管理与更新。 执行变异和交叉操作后，种群中的每个分子都可能产生多个子代。该引擎采用一种兼顾多样性和质量的选择策略:

$$M_{\text{new}} = \text{top}(\text{unique}(M_{\text{current}} \cup M_{\text{offspring}}), N), \tag{11}$$

其中， M_{new} 是更新后的分子种群， M_{current} 是当前种群， $M_{\text{offspring}}$ 是通过进化操作生成的子代分子， N 是最大种群规模。函数 unique(·) 会移除重复分子以保持多样性，而 top(·， N) 则根据使用第 3.2 节评分函数计算的适应度，通过保留排名靠前的 N 个分子来执行选择。

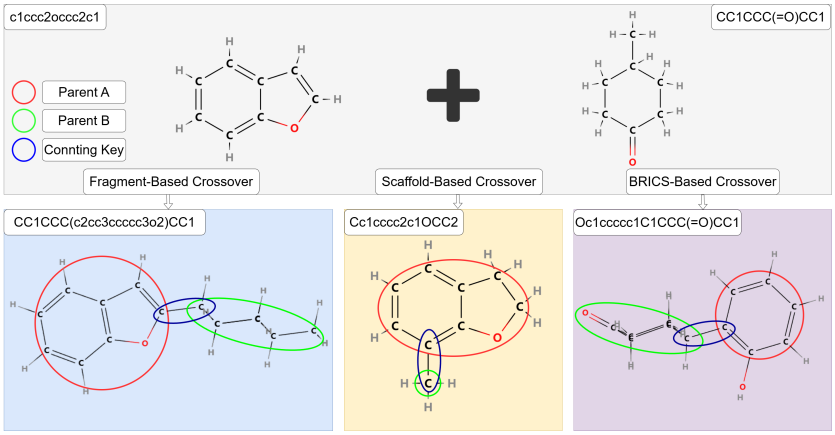


图 3. 分子交叉引擎。

动态参数调优。 分子进化引擎将种群多样性作为自适应控制参数，以平衡探索与利用。该指标通过采样的 Morgan指纹之间的平均成对 Tanimoto 相似度的补数计算得出，并被纳入规则适应度评估，以保持结构多样性，从而确保高效搜索并避免早熟收敛。该自适应机制与第 4.1 节所述的规则权重更新协同运作，形成一个协同进化系统，其中变换规则及其应用概率共同进化，以优化分子设计结果。

5 实验设置

我们使用标准进化算法配置，评估 MCEMOL 在分子生成任务中的表现。目标是优化转换规则，以生成结构多样性高、有效性完整且类药性优异的类药分子。

5.1 计算设置

实验在一台配备 Intel Core i5-12500H CPU 和 16GB 内存的笔记本电脑上进行。MCEMOL 在未使用 GPU 加速的情况下，约 45 小时完成了 100 代，将 20 个种子分子扩展为 829 个独特分子。

5.2 数据集

我们使用包含约 250,000 个类药分子的 ZINC250K 数据集[21], ，这是一个标准基准，可用于在药物发现场景中直接评估分子生成方法。

5.3 评估框架

我们的评估框架结合基线比较与全面的类药性评估，以表征 MCEMOL 的性能。

基线与比较数据。我们将 JT-VAE、GCPN[51, 24], GraphVAE 和 ORGAN 作为比较对象，它们代表了多种分子生成方法。为避免重新实现偏差，我们直接采用其原始论文中的结果。表格中以“–”标记缺失的指标。对于 MCEMOL，所有指标在各项实验中均采用一致的定义。

评估指标与合规标准。分子结构质量通过有效性、内部多样性（Tanimoto距离）、与参考分子的相似性、新颖性和骨架多样性进行衡量。相似性指标采用 Morgan指纹（ECFP4）和 Tanimoto系数；JT-VAE 和 GCPN 在优化过程中跟踪配对分子，而 MCEMOL 基于种群进行进化，因此计算初始种群与最终种群之间的跨世代相似性。性质评估包括 LogP、HOMO-LUMO能隙（由 ATLAS-DMPNN 预测）、QED、分子量、合成可及性、拓扑极性表面积、氢键受体数和可旋转键数。我们依据 Lipinski 五规则、Ghose筛选规则和 Veber规则评估类药性合规情况。

5.4 算法配置

MCEMOL 的配置在探索与利用之间取得平衡，同时保持化学有效性。

进化算法参数。 我们使用的种群规模为 100，世代数为 100，规则级变异率为 0.4，规则级交叉率为 0.2，精英保留率为 0.1。

分子适应度评估。 我们使用针对优化目标定制的归一化权重评估 12 项性质。权重根据影响程度分层分配：药理学相关性（主要，0.11–0.18）优先考虑性质覆盖度（0.18）和对称性（0.13）；标准开发准则（次要，0.08）包括有效性、多样性、合成可行性、灵活性和类药性质；精化准则（次要，0.05–0.06）涵盖稳定性、药效团和类药性；结构完整性（0.02）则作为基线检查。

规则评估标准。根据转换成功率和生成分子的适应度分数来评估规则。复杂的条件和变异会获得更高的权重，以体现其对探索化学空间的贡献。为确保一致性，这些标准在所有实验中保持不变。

6 结果

我们将实验分析分为四个部分：分子结构、分子属性、是否符合类药规则，以及TOP分子分析。

6.1 分子结构

我们使用 GCPN 和 JT-VAE 研究中的真实世界数据 [51, 24]，基于三项关键指标评估 MCEMOL 的性能。

为避免重新实现偏差，基线结果直接引用自原始出版物，未报告的指标标记为"–"。如表 1 所示，MCEMOL 的有效性达到 100%，且多样性具有竞争力。与侧重生成的基线方法不同，MCEMOL 通过双层进化机制，优先发掘可解释的变换规则，并将分子库视为副产品而非首要目标。具体而言，MCEMOL 实现了完美的有效性、1.0 的新颖性（训练集中不存在）和 0.51 的骨架多样性（独特的 Murcko骨架）——这些指标在基线研究中均未报告。其内部多样性（0.834）与现有方法相当，而相似度得分最低（0.26），表明它在探索新颖化学空间方面表现更佳。此外，MCEMOL 还揭示了可迁移的变换原理，提供了实用分子和可解释的设计知识，而传统生成式方法无法做到这一点。

表 1.分子结构

方法	有效性（%）	多样性	相似性	新颖性	骨架多样性
JT-VAE	100	–	0.280	–	–
GCPN	100	0.855	0.320	–	–
GraphVAE	14	–	–	–	–
ORGAN	2.2	0.759	–	–	–
MCEMOL	**100**	**0.834**	**0.262**	**1**	**0.514**

6.2 分子性质

表2总结了平均性质，并利用 ATLAS-DMPNN 高效预测 HOMO-LUMO 能隙，图4则追踪了这些性质的进化。LogP 在理想亲脂性范围内下降，而持续较高的 QED 值证实了良好的类药性。HOMO-LUMO 能隙表明分子具有稳定性且合成难度较低，分子量也符合口服生物利用度的要求。合成可及性（SA）稳定在适中水平，体现了对复杂性的有效控制。良好的拓扑极性表面积、氢键受体数和可旋转键数量确保了渗透性与柔性的平衡。这些性质共同表明，生成的分子具有优异的类药特征，适合用于药物开发。

表2.分子性质

LogP 差值（eV）	QED 分子量（Da）	合成可及性评分	拓扑极性表面积（Å2）	氢键受体数	可旋转键数
2.41	4.4854	0.75	384.5	3.57	94.94	5.98	4.66

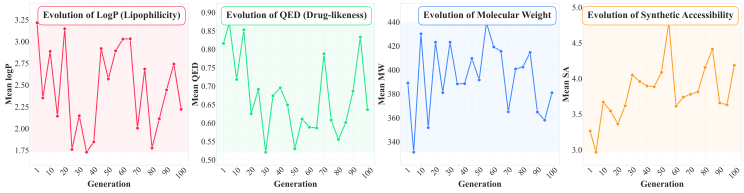


图4. 基本化学性质趋势图。

6.3 药物相似性合规性

该框架生成的分子表现出对类药性规则的出色遵循：

100%符合Lipinski规则：所有分子均满足口服药物的基本标准。98.0%符合Ghose规则：绝大多数分子都符合药物化学所需的理化性质。99.6%符合Veber规则：几乎所有分子都具有良好的口服生物利用度潜力。这些高合规率表明，该框架能够高效生成类药分子，显著减少后续筛选和优化的工作量。表现最佳分子的详细化学分析见第6.4节。

6.4 顶级分子

表 3。顶级分子

SMILES	LogP	QED	结构特征
Cc1ccc(F)c(NC(=O)c2ccc(S(=O)(=O)CO)cc2)c1	2.11	0.902	氟化苯胺结构，磺酰基侧链。
Cc1cc(C(=O)NC2CC2)c(CO)n1-c1ccc2c(c1)OC=CO2	2.41	0.905	含氮杂环的缩合结构。
COc1cc(Cl)c(C)cc1NC(=O)c1ccc(=O)n(CO)c1	2.02	0.902	氯苯胺吡啶酮衍生物。
COCn1cc(C(=O)Nc2nc(C)c(Cl)cc2OC)ccc1=O	2.07	0.903	吡啶酮与氯代吡啶的缩合物。
COC(C)c1c(C(=O)NC2CC2)cc(C)n1-c1cc2c(cn1)OCCO2	2.55	0.89	含有二噁烷的复杂杂环体系。

对表现最佳的分子（表3）的分析表明，MCEMOL 能够生成具有化学意义的结构，且具有多样的药理潜力。以排名前两位的分子为例：

第一个分子（Cc1ccc(F)c(NC(=O)c2ccc(S(=O)(=O)CO)cc2)c1）具有鲜明的类药特征：氟原子可增强代谢稳定性和脂溶性，而酰胺基和磺酰基团可作为氢键受体（后者还能提高水溶性）。再加上羟基作为氢键供体，这些特征表明该分子有望用作酶抑制剂或抗炎/镇痛剂。

第二个分子（Cc1cc(C(=O)NC2CC2)c(CO)n1-c1ccc2c(c1)OC=CO2）展现了精巧的结构设计。环丙烷基团引入了构象约束，并提高了代谢稳定性；二氧六环可能增强靶标结合亲和力。凭借多重氢键相互作用和用途广泛的羟甲基，该结构有望成为激酶抑制剂或中枢神经系统药物。

这些示例说明，MCEMOL 的双层进化机制如何有效地将化学知识与优化算法相结合，从而生成理论上合理且对药物开发具有实际意义的分子。

7 结论

MCEMOL 是一个多约束进化分子设计框架，将规则进化与直接分子操作相结合 [42]。通过结合可解释的基于规则的转化与进化优化，该框架在保持可解释性的同时，解决了人工智能驱动的分子设计中的关键挑战。

我们的结果展示了五项主要成果。首先，规则层面和分子层面的双层进化通过协同进化探索化学空间，优于现有方法 [42, 43]。其次，多维约束系统确保分子在合成可行性、对称性和药效团表示方面的质量 [34, 20]。第三，自适应规则加权和动态种群管理提升了学习效率，并增强了对多样化设计目标的适应能力 [29, 18]。第四，该框架在不牺牲性能的前提下实现了高度可解释性 [49]。第五，MCEMOL 的计算效率使其能够从少量起始分子出发进行实际部署，因此适用于资源有限的研究环境。

实验结果表明，该方法具有很强的新颖性、多样性和类药性，有效性达到 100%，药物规则合规性近乎完美，同时保留了可解释性 [23, 49]。第 6.4 节的定性分析证实了其化学合理性和药理学潜力。

当前的局限性包括：依赖预定义规则库，这可能会限制对非常规化学空间的探索；难以处理灵活的三维构象 [53], ；以及针对不同目标进行超参数调优的需求。

未来的工作将探索帕累托和多目标贝叶斯策略，整合合成规划，纳入蛋白质结构和生物活性数据以提高靶标特异性，研究基于片段的设计，并添加药物化学专家规则。

总的来说，MCEMOL 提供了一种可解释且高效的分子设计方法，通过变换规则揭示机制层面的洞见，并生成种类丰富的高质量类药分子库，有助于弥合计算设计与实际药物发现之间的差距。

参考文献

1. Bacardit, J., Burke, E.K., Krasnogor, N.: 提升基于规则的进化学习的可扩展性。Memetic Computing **1**(1), 55–67 (2009). https://doi.org/10.1007/s12293-008-0005-4, https://doi.org/10.1007/s12293-008- 0005-4 2. Bemis, G.W., Murcko, M.A.: 已知药物的性质。1. 分子骨架。Journal of Medicinal Chemistry **39**(15), 2887–2893 (1996) 3. Brooks, W.H., Guida, W.C., Daniel, K.G.: 手性在药物设计与开发中的重要性。Current Topics in Medicinal Chemistry **11**(7), 760–770 (2011). https://doi.org/http://dx.doi.org/10.2174/156802611795165098, http://www.eurekaselect.com/article/31944 4. Coley, C.W., Green, W.H., Jensen, K.F.: 基于分子相似性的计算机辅助逆合成。ACS Central Science **3**(12), 1237–1245 (2017) 5. Cordon, O., Herrera, F., Gomide, F., Hoffmann, F., Magdalena, L.: 遗传模糊系统十年：当前框架与新趋势。载于：第九届 IFSA 世界大会暨第二十届 NAFIPS 国际会议论文集（目录号 01TH8569）。第3卷，第1241–1246页，第3卷（2001）。https://doi.org/10 943725 6 1109 2001 . /NAFIPS. . . Daeyaert, F., Deem, M.W.: 一种用于高效从头设计多功能分子的帕累托算法。Molecular Informatics **36**(1-2), 1600044 (2017). https://doi.org/https://doi.org/10.1002/minf.201600044, https://onlinelibrary.wiley.com/doi/abs/10 201600044 7 1002. /minf. . De Cao, N., Kipf, T.: MolGAN：一种用于小分子图的隐式生成模型。arXiv 电子预印本 arXiv:1805.11973（2018年5月）。https://doi.org/10 11973 8 48550 1805 . /arXiv. . . Degen, J., Wegscheid-Gerlach, C., Zaliani, A., Rarey, M.: 编纂和运用“类药”化学片段空间的艺术。ChemMedChem **3**(10), 1503–1507 (2008). https://doi.org/https://doi.org/10.1002/cmdc.200800178, https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cmdc.200800178 9. Dixon, S.L., Smondyrev, A.M., Knoll, E.H., Rao, S.N., Shaw, D.E., Friesner, R.A.: Phase：用于药效团识别、三维定量构效关系模型开发和三维数据库筛选的新引擎：1. 方法学与初步结果。Journal of Computer-Aided Molecular Design **20**(10), 647–671 (2006) 10. Duvenaud, D.K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., Adams, R.P.: 用于学习分子指纹的图卷积网络。载于：Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R.（编）《神经信息处理系统进展》。第28卷。Curran Associates, Inc. (2015), https://proceedings.neurips.cc/paper_ files/paper/2015/file/f9be311e65d81a9ad815

0a60844bb94c-

Paper.pdf 11. Elton, D.C., Boukouvalas, Z., Fuge, M.D., Chung, P.W.: 分子设计中的深度学习——前沿进展综述。Mol. Syst. Des. Eng. **4**, 828–849 (2019年)。https://doi.org/10.1039/C9ME00039A, http://dx.doi.org/10.1039/C9ME00039A 12. Ertl, P.: 有机取代基的化学信息学分析：识别最常见的取代基、计算取代基性质，并自动识别类药性生物电子等排基团。Journal of Chemical Information and Computer Sciences**43**(2), 374–380 (2003)

13. Ertl, P., Schuffenhauer, A.: 基于分子复杂度和片段贡献估算类药性分子的合成可及性评分。《化学信息学杂志》 **1**(1), 8 (2009年)。https://doi.org/10.1186/1758-2946-1-8, https://doi.org/10 8 14 1186 1758 2946 1 1 . / - - - . Fechner, U., Schneider, G.: Flux ( ): 一种用于基于片段的从头设计的虚拟合成方案。《化学信息与建模杂志》 **46**(2), 699–707 (2006 0503560 15 10 1021 0503560, 10 1021)。https://doi.org/ . /ci https://doi.org/ . /ci . Gao, W., Coley, C.W.: 生成模型提出分子的可合成性。《化学信息与建模杂志》 **60**(12), 5714–5723 (2020年)。https://doi.org/10.1021/acs.jcim.0c00174, https://doi.org/10 00174 16 1021 0 . /acs.jcim. c . Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E.: 用于量子化学的神经消息传递。载于：Precup, D., Teh, Y.W. (编)《第34届国际机器学习会议论文集》。《机器学习研究论文集》，第70卷，第1263–1272页。PMLR (2017年8月6–11日)，https://proceedings.mlr.press/v70/gilmer17a.html 17. Gómez-Bombarelli, R., Wei, J.N., Duvenaud, D., Hernández-Lobato, J.M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T.D., Adams, R.P., Aspuru-Guzik, A.: 利用分子的数据驱动连续表示实现自动化学设计。ACS Central Science **4**(2), 268–276 (2018年)。https://doi.org/10.1021/acscentsci.7b00572, https://doi.org/10 00572 18 1021 7 . /acscentsci. b . Hansen, N.: 协方差矩阵自适应进化策略：比较性综述，第75–102页。Springer Berlin Heidelberg, Berlin, Heidelberg (2006年)。https://doi.org/10.1007/3- 540 32494 1_4 10 1007 3 540 32494 1_4 - - , https://doi.org/ . / - - - . Hartenfeller, M., Zettl, H., Walter, M., Rupp, M., Reisen, F., Proschak, E., Weggen, S., Stark, H., Schneider, G.: 用于计算机模拟分子设计的一组稳健的有机合成反应。《化学信息与建模杂志》 **51**(12), 3093–3098 (2011年) 20. Hoksza, D., Škoda, P., Voršilák, M., Svozil, D.: Molpher：用于系统化学空间探索的软件框架。《化学信息学杂志》 **6**(1), 7 (2014年)。https://doi.org/10 7 21 1186 1758 2946 6 7, 10 1186 1758 2946 6 . / - - - https://doi.org/ . / - - - . Irwin, J.J., Shoichet, B.K.: Zinc——一个用于虚拟筛选的市售化合物免费数据库。《化学信息与建模杂志》 **45**(1), 177–182 (2005年)。https://doi.org/10.1021/ci049714+, https://doi.org/ 10 1021 049714+ . /ci . Jensen, J.H.: 一种基于图的遗传算法及生成模型7蒙特卡洛树搜索，用于探索化学空间。Chem. Sci. **10**, 3567–3572 (2019年)。https://doi.org/10.1039/C8SC05372C, http://dx.doi.org/10.1039/C8SC05372C 23. Jiménez-Luna, J., Grisoni, F., Schneider, G.: 借助可解释人工智能进行药物发现。《自然·机器智能》 **2**(10), 573–584 (2020年10月)。https://doi.org/10.1038/s42256-020-00236-4, https://doi.org/10.1038/s42256-020-00236-4 24. Jin, W., Barzilay, R., Jaakkola, T.: 用于分子图生成的连接树变分自编码器。载于：Dy, J., Krause, A. (编)《第35届国际机器学习会议论文集》。《机器学习研究论文集》，第80卷，第2323–2332页。PMLR (2018年7月10–15日)，https://proceedings.mlr.press/v80/jin18a.html 25. Largent, R.J., Polik, W.F., Schmidt, J.: Symmetrizer：近似对称分子点群的算法判定。《计算化学杂志》**33**(19), 1637–1642 (2012年)

26. Lewell, X.Q., Judd, D.B., Watson, S.P., Hann, M.M.: RECAP——逆合成组合分析程序：一种强大的新技术，可用于识别在组合化学中具有实用价值的优势分子片段。Journal of Chemical Information and Computer Sciences **38**(3), 511–522 (1998) 27. Lin, S., Cui, G., Tang, C., Zhang, C., Nagata, Y., Yang, H.: ATLAS-DMPNN：一种用于提升ADMET性质预测的注意力引导的拓扑框架。载于：2025年国际计算智能新趋势会议（NTCI）。第142页 11308181 28 147 2025 10 1109 67886 2025 – ( )。https://doi.org/ . /NTCI . . . Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J.: 在药物发现与开发中估算溶解度和渗透性的实验与计算方法。《先进药物递送评论》**23**(1), 3–25 (1997)。https://doi.org/https://doi.org/10.1016/S0169-409X(96)00423-1, https://www.sciencedirect.com/science/article/pii/S0169409X96004231，体外模型在开发候选药物筛选中的应用 29. Liu, T., Xiong, G., Wagdy Mohamed, A., Nagaratnam Suganthan, P.: 具有阀点效应和多燃料选项的大规模经济负荷调度问题的反向互学差分进化混合变异策略。《信息科学》**609**, 1721–1745 (2022)。https://doi.org/https://doi.org/10.1016/j.ins.2022.07.148, https://www.sciencedirect.com/science/article/pii/S0020025522008386 30. Maziarka, Ł., Pocha, A., Kaczmarczyk, J., Rataj, K., Danel, T., Warchoł, M.: Mol-CycleGAN：一种用于分子优化的生成模型。《化学信息学杂志》**12**(1), 2 (2020)。https://doi.org/10.1186/s13321-019-0404-1，https://doi.org/10 1 31 1186 13321 019 0404 . /s - - - . Nicolaou, C.A., Brown, N.: 药物设计中的多目标优化方法。《今日药物发现：技术》**10**(3), e427–e435 (2013)。https://doi.org/https://doi.org/10.1016/j.ddtec.2013.02.001, https://www.sciencedirect.com/science/article/pii/S1740674913000085 32. Nicolaou, C.A., Brown, N., Pattichis, C.S.: 使用计算多目标方法进行分子优化。《药物发现与开发最新观点》**10**(3), 316–326 (2007) 33. Paul, S.M., Mytelka, D.S., Dunwiddie, C.T., Persinger, C.C., Munos, B.H., Lind-borg, S.R., Schacht, A.L.: 如何提高研发生产力：制药行业面临的重大挑战。《自然综述：药物发现》**9**(3), 203–214 (2010)。https://doi.org/10 3078 34 1038 3078，10 1038 . /nrd https://doi.org/ . /nrd . Polishchuk, P.G., Madzhidov, T.I., Varnek, A.: 基于GDB-17数据估算类药化学空间的规模。Journal of Computer-Aided Molecular Design **27**(8), 675–679 (2013)。https://doi.org/10.1007/s10822-013-9672-4，https://doi.org/10 4 35 1007 10822 013 9672 . /s - - - . Polishchuk, P.: CREM：用于结构生成的化学合理突变框架。《化学信息学杂志》**12**(1), 28 (2020)。https://doi.org/10.1186/s13321-020-00431-w, https://doi.org/10.1186/s13321- 020-00431-w 36. Riniker, S., Landrum, G.A.: 更可靠的距离几何：利用已知信息改进构象生成。《化学信息与建模杂志》**55**(12), 2562–2574 (2015) 37. Rogers, D., Hahn, M.: 扩展连接指纹。《化学信息与建模杂志》**50**(5), 742–754 (2010) 38. Salgotra, R., Mirjalili, S.: 用于约束工程的多算法进化策略及自适应变异机制

设计问题。《专家系统与应用》**258**, 125055 (2024)。 https://doi.org/https://doi.org/10.1016/j.eswa.2024.125055, https://www.sciencedirect.com/science/article/pii/S0957417424019225 39. Sanchez-Lengeling, B., Aspuru-Guzik, A.: 利用机器学习进行逆向分子设计：用于物质工程的生成模型。《科-学》**361**(6400), 360–365 (2018)。 https://doi.org/10.1126/science.aat2663，https://www.science.org/doi/abs/10 2663 40 1126. /science.aat . Segler, M.H.S., Preuss, M., Waller, M.P.: 利用深度神经网络和符号人工智能规划化学合成。《自然》**555**(7698), 604–610 (2018)。 https://doi.org/10 25978 41 1038 25978，10 1038 . /nature https://doi.org/ . / nature . Shi, C., Xu, M., Zhu, Z., Zhang, W., Zhang, M., Tang, J.: GraphAF：一种用于分子图生成的基于流的自回归模型。arXiv预印本 arXiv:2001 09382 42 09382 2020 10 48550 2001 4 . (1月)。https://doi.org/ . /arXiv. . . Spiegel, J.O., Durrant, J.D.: Autogrow：一种用于从头药物设计和先导化合物优化的开源遗传算法。《化学信息学杂志》**12**(1), 25（2020年4月）。https://doi.org/10.1186/s13321-020-00429-4, https://doi.org/10 4 43 1186 13321 020 00429 4 . /s - - - . Spiegel, J.O., Durrant, J.D.: Autogrow：一种用于从头药物设计和先导化合物优化的开源遗传算法。《化学信息学杂志》**12**(1), 25（2020年4月）。https://doi.org/10.1186/s13321-020-00429-4, https://doi.org/10 4 44 1186 13321 020 00429 . /s - - - . Swan, J., Adriaensen, S., Brownlee, A.E., Hammond, K., Johnson, C.G., Kheiri, A., Krawiec, F., Merelo, J., Minku, L.L., Özcan, E., Pappa, G.L., García-Sánchez, P., Sörensen, K., Voß, S., Wagner, M., White, D.R.: 大规模元启发式算法。《欧洲运筹学杂志》**297**(2), 393–406 (2022)。https://doi.org/https://doi.org/10.1016/j.ejor.2021.05.042, https://www.sciencedirect.com/science/article/pii/S0377221721004707 45. Urbanowicz, R.J., Moore, J.H.: 学习分类器系统：完整介绍、综述与路线图。《人工进化与应用杂志》**2009**(1), 736398 (2009)。 https://doi.org/https://doi.org/10.1155/2009/736398, https://onlinelibrary.wiley.com/doi/abs/10 736398 46 1155 2009. / / . Veber, D.F., Johnson, S.R., Cheng, H.Y., Smith, B.R., Ward, K.W., Kop-ple, K.D.: 影响候选药物口服生物利用度的分子性质。Journal of Medicinal Chemistry **45**(12), 2615–2623 (2002)。 https://doi.org/10.1021/jm020017n，https://doi.org/10.1021/jm020017n 47. Wager, T.T., Hou, X., Verhoest, P.R., Villalobos, A.: 超越规则：中枢神经系统多参数优化（CNS MPO）方法的开发，以实现类药性质的协调。《ACS化学神经科学》**1**(6), 435–449 (2010)。https://doi.org/10.1021/cn100008c 48. Wager, T.T., Hou, X., Verhoest, P.R., Villalobos, A.: 中枢神经系统多参数优化期望度：在药物发现中的应用。《ACS化学神经科学》**7**(6), 767–775 (2016)。https://doi.org/10 00029 49 1021 6 . / acschemneuro. b . Wellawatte, G.P., Seshadri, A., White, A.D.: 分子反事实解释的模型无关生成。《化学科学》**13**, 3697–3705 (2022)。 https://doi.org/10.1039/D1SC05259D，http://dx.doi.org/10.1039/D1SC05259D 50. Wolber, G., Langer, T.: LigandScout：源自蛋白质结合配体的三维药效团及其作为虚拟筛选过滤器的应用。《化学信息与建模杂志》**45**(1), 160–169 (2005)。 https://doi.org/10.1021/ci049885e，https://doi.org/10.1021/ci049885e

51. You, J., Liu, B., Ying, Z., Pande, V., Leskovec, J.：用于目标导向的分子图生成的图卷积策略网络。载于：Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (编)《神经信息处理系统进展》，第31卷。Curran Associates, Inc. (2018), https://proceedings.neurips.cc/paper_ files/paper/2018/file/d60678e8f2ba9c540798ebbde31177e8- Paper.pdf 52. Zhou, Z., Kearnes, S., Li, L., Zare, R.N., Riley, P.：通过深度强化学习优化分子。《科学报告》**9**(1), 10752 (2019)。https://doi.org/10.1038/s41598-019-47148-x, https://doi.org/10.1038/s41598-019- 47148-x 53. Zhung, W., Kim, H., Kim, W.Y.：用于相互作用引导的药物设计的三维分子生成框架。《自然·通讯》**15**(1), 2688 (2024年3月)。https://doi.org/10.1038/s41467-024-47011-2, https://doi.org/10.1038/s41467-024- 47011-2