

基于遗传规划的分子逆合成路径探索：方法论、编码策略与全流程解析

1. 执行摘要

在计算机辅助合成规划(Computer-Aided Synthesis Planning, CASP)领域，传统的搜索策略长期以来由启发式算法及近年兴起的蒙特卡洛树搜索(MCTS)主导。然而，随着化学空间搜索深度的增加和单步反应预测模型计算成本的累积，传统的搜索框架在处理复杂多步逆合成任务时面临着搜索效率低、容易陷入局部最优解以及缺乏路径多样性等瓶颈。遗传规划(Genetic Programming, GP)与进化算法(Evolutionary Algorithms, EA)作为一种全局优化策略，为解决这一非凸、离散且高度组合的化学路径规划问题提供了全新的范式。

本报告旨在为利用遗传规划方法进行分子逆合成任务的研究提供详尽的指导。基于最新的前沿研究，特别是“进化逆合成路径规划”(EvoRRP)与“合成感知遗传算法”(SynGA)等里程碑式工作，我们深入剖析了该方法的理论基础与工程实现。分析表明，通过将逆合成问题建模为树结构的进化优化问题，并结合化学特异性的遗传算子，进化策略能够将单步模型的调用次数减少约 53.9%，同时显著提升解的可行性和多样性¹。

报告详细阐述了逆合成路径的树状编码机制(AND-OR Tree)，定义了保证化学效度的交叉与变异算子，并构建了包含合成复杂性(SCScore)、逆合成可及性(RAscore)及商业可行性的多目标打分函数。此外，我们系统梳理了从USPTO反应数据集的预处理到目标分子(如药物中间体、天然产物)的选择标准，并完整展示了从种群初始化到最优路径提取的端到端流程。本报告旨在为领域专家提供一份详实的技术蓝图，推动进化计算在有机合成设计中的深度应用。

2. 逆合成分析的演进与进化计算的引入

2.1 分子逆合成的计算挑战

逆合成分析(Retrosynthetic Analysis)是由 E.J. Corey 在 20 世纪 60 年代系统化的逻辑过程，其核心思想是将复杂的“目标分子”(Target Molecule)递归地分解为更简单的“前体”(Precursors)，直到获得廉价且易得的“起始原料”(Building Blocks)。

从计算角度看，这是一个巨大的搜索问题。每一步分解都可能涉及数十甚至上百种可能的化学反应(切断)，随着步骤的增加，搜索空间呈指数级爆炸。现代 AI 方法通常采用“单步逆合成预测模型”(Single-step Retrosynthesis Model)作为策略网络，预测当前分子的可能前体。然而，如何将这些单步预测串联成一条最优路径，仍然是一个极具挑战的规划问题¹。

目前主流的 MCTS 方法虽然在围棋等博奕游戏中取得了巨大成功，但在化学合成中存在明显的局限性：化学反应的“规则”是不完备的，且“奖励”通常极其稀疏(只有当路径完全打通至起始原料时才算成功)。这导致 MCTS 在广阔的化学空间中往往需要进行大量的无效模拟(Rollout)，计算效率低下²。

2.2 进化算法与遗传规划的范式转换

进化算法(EA)和遗传规划(GP)提供了一种截然不同的思路。不同于 MCTS 的非对称树生长, EA 维护一个“候选路径种群”(Population of Routes)。通过模拟自然选择的过程(选择、交叉、变异), EA 能够在搜索空间中维持多样性, 并利用群体智慧跳出局部最优。

在逆合成任务中, EvoRRP(Evolutionary Retrosynthetic Route Planning)是首个将进化算法成功应用于多步逆合成的研究。该方法证明了将逆合成建模为优化问题而非单纯的搜索问题, 可以大幅降低对昂贵单步预测模型的调用频率¹。此外, SynGA(Synthesis-Aware Genetic Algorithm)等研究虽然侧重于正向分子生成, 但其对“合成树”(Synthesis Tree)的编码和操作算子设计, 为逆合成的 GP 实现提供了重要的借鉴⁴。

3. 逆合成路径的编码策略与数据结构

在遗传规划中, 个体的**表示(Representation)**即编码方式, 是决定算法成败的关键。对于逆合成任务, 必须采用能够如实反映化学逻辑且便于遗传操作的数据结构。

3.1 核心数据结构: AND-OR 树

逆合成路径本质上是一个层级化的图结构, 最标准的表示形式是 **AND-OR 树**(或称超图)。理解这一结构对于编码至关重要。

- **根节点(Root)**: 目标分子(Target Molecule)。
- **OR 节点(分子节点)**: 代表一个化学实体(分子)。由于一个分子可以通过多种不同的反应途径合成(例如, 通过酯化反应或酰氯醇解制备酯), 这些途径之间是“或”(OR)的关系。只要找到其中一条可行路径, 该节点即被解决。
- **AND 节点(反应节点)**: 代表一个具体的化学反应(Reaction Template)。一个反应通常需要多个反应物(前体), 只有当所有反应物(例如反应物 A 和反应物 B)都被成功合成或购买时, 该反应才算可行。因此, 反应物之间是“与”(AND)的关系⁶。
- **叶节点(Leaves)**: 代表搜索的终止点, 通常是商业可得的起始原料(Building Blocks)。

3.1.1 遗传规划中的树状个体编码

在具体实现 GP 时, 个体通常被编码为一棵显式的合成树(**Synthesis Tree**)。与传统的编程语言语法树类似, 这棵树由“操作符”和“终结符”组成:

- **操作符(内部节点)**: 对应具体的化学反应模板(Reaction Templates)或单步预测模型生成的反应类别。
- **终结符(叶节点)**: 对应具体的分子结构(SMILES 字符串)或构建块的索引 ID。

数学上, 一个逆合成树 \$T\$ 可以递归定义为:

$$\$T = \{ R, \{ T_1, T_2, \dots, T_k \} \} \$$$

其中 \$R\$ 是应用于目标分子的反应, \$T_1...T_k\$ 是生成反应所需的 \$k\$ 个前体分子的子树。这

种递归结构非常适合 GP 的标准操作 9。

3.2 线性编码 vs. 图编码

除了直接的树结构，部分研究也尝试使用线性化编码(如 SMILES 序列或后缀表达式)来表示路径，但在处理复杂的非线性合成路线(汇聚式合成)时，图或树的结构更具优势。

- **SMILES 序列化**: 便于使用 Transformer 等序列模型处理，但难以直接进行交叉操作而不破坏化学有效性。
- **图/树编码**: EvoRRP 和 SynGA 均采用基于图/树的直接操作。这种方式允许算法直接在逻辑层面(如“替换整个支路”)进行操作，保证了化学结构的完整性¹¹。

3.3 编码的化学约束

与生成计算机代码的传统 GP 不同，化学 GP 必须强制执行化学可行性约束。任何生成的个体(合成路径)必须符合化学反应规则。例如，不能将一个需要胺基参与的反应节点连接到一个不含胺基的前体节点上。这要求编码不仅包含结构信息，还要包含化学指纹或子结构匹配信息，以便在遗传操作时进行快速校验¹³。

4. 遗传算子的设计与化学适配

遗传算子(交叉与变异)是驱动种群进化的引擎。在逆合成任务中，盲目的随机操作会产生大量无效的化学路径。因此，必须设计“化学感知”(Chemistry-Aware)的算子。

4.1 交叉算子(Crossover): 子树重组

交叉操作旨在结合两个父代路径的优势部分。例如，父代 A 可能找到了一条非常高效的合成路线来构建分子的左半部分，而父代 B 则擅长构建右半部分。

- 兼容子树交叉(**Compatible Subtree Crossover**): 这是 SynGA 引入的关键概念，同样适用于逆合成 EvoRRP。
 1. 选择切点: 在父代树 \$P_1\$ 和 \$P_2\$ 中分别随机选择一个节点。
 2. 兼容性检查: 检查这两个节点是否代表了相似的中间体，或者是否属于同一类反应的产物。如果 \$P_1\$ 的切点是一个通过 Suzuki 偶联生成的联苯结构，而 \$P_2\$ 的切点是一个通过酰胺缩合生成的结构，直接交换可能导致化学逻辑断裂。因此，算法会优先选择结构相似度高(如 Tanimoto 系数高)或反应类型兼容的节点进行交换。
 3. 交换与嫁接: 交换子树，生成新的后代。
 4. 修剪: 如果交换后导致树的深度超过限制或产生循环依赖，则进行修剪或丢弃¹³。

4.2 变异算子(Mutation): 路径探索

变异操作引入了新的遗传物质，是发现新合成路线的主要手段。在逆合成中，变异主要依赖于单步逆合成模型的重新调用。

根据 SynGA 和 EvoRRP 的方法论，变异操作可细分为以下五种类型⁴:

变异类型	操作描述	化学意义	适用场景
生长 (Grow)	选择一个尚未解决的叶节点(非起始原料), 调用单步模型预测其前体, 并扩展树结构。	探索更深层的反应路径, 试图解决不可购买的中间体。	当路径中断于非商业化分子时。
收缩 (Shrink)	选择一个内部节点(反应), 将其替换为该反应的产物(即剪掉该反应及其子树), 使其成为叶节点。	简化路径, 尝试直接购买该中间体(如果它突然在数据库中被发现或作为假设)。	当路径过于冗长或复杂时。
反应替换 (Change Internal)	保持反应物不变, 尝试改变反应条件或反应模板(若模型支持多种反应类型)。	寻找相同前体下的不同转化条件(如不同的催化剂或保护基策略)。	优化反应条件或规避专利壁垒。
原料替换 (Change Leaf)	将某个叶节点替换为结构相似的另一个起始原料。	探索不同的起始物料(如使用溴代物代替氯代物)。	优化成本或原料可得性。
重采样 (Rerun/Resample)	对某个节点的单步预测结果进行重新采样(如选择 Top-5 中的第 2 个而非第 1 个)。	探索被贪婪策略忽略的次优解, 增加路径多样性。	跳出局部最优解。

4.3 种群初始化与多样性维护

初始化通常通过随机调用单步模型生成。为了防止早熟收敛(即所有个体都收敛到同一条显而易见的路径), EvoRRP 采用了**并行子种群(Parallel Subpopulations)**策略。

- **岛屿模型(Island Model)**: 将总种群划分为多个独立的子种群(岛屿), 每个岛屿独立进化。
- **迁移**: 每隔固定代数, 相邻岛屿之间交换少量优秀个体。这种机制有效地维持了全局多样性, 防止算法陷入单一的合成逻辑中¹。

5. 核心数据集与目标分子选择

遗传规划系统的性能高度依赖于其“知识库”(训练数据)和“测试场”(目标分子)。

5.1 反应数据集(知识库)

EvoRRP 和其他先进模型主要依赖 **USPTO**(美国专利商标局) 数据集来训练其单步逆合成预测器(Mutation Operator 的核心引擎)。

- **USPTO-50k**: 包含约 50,000 条高质量、已分类的反应数据。常用于训练模板分类器，因其数据干净，反应类型明确。
- **USPTO-MIT / USPTO-Full**: 包含超过 100 万条反应。用于训练基于 Transformer 的序列到序列模型(如 Retro* 使用的模型)。大数据量使得模型具有更好的泛化能力，能处理更复杂的分子变换¹⁵。
- **数据预处理**: 必须进行原子映射(Atom Mapping)、去除试剂(Reagents)与溶剂，并标准化 SMILES 字符串，以确保遗传操作的一致性。

5.2 起始原料库(终止条件)

搜索的终点是商业可得性。常用的数据库包括：

- **eMolecules**: 包含数百万种可购买化合物，通常作为实际应用的标准。
- **ZINC Database**: 特别是 "In-Stock" 子集，常用于学术基准测试。
- **Enamine Building Blocks**: 在新药发现中广泛使用的砌块库。
- **使用方法**: 将这些库构建为哈希表或布隆过滤器(Bloom Filter)，以便在 \$O(1)\$ 时间复杂度内判断叶节点是否为起始原料¹⁵。

5.3 目标分子的选择(基准测试)

为了验证 GP 方法的有效性，不能随意选择分子。必须选择那些具有挑战性、且已知存在多步合成路线的分子。

1. **Retro Benchmark Set***: 由 Retro* 论文提出的 190 个“困难”分子。这些分子是从 USPTO 测试集中筛选出来的，且已知至少需要多步反应才能合成。这是目前最主流的对比基准¹⁵。
2. **GuacaMol Benchmark**: 虽然主要用于生成模型，但其中的目标分子分布也被用于测试逆合成的泛化能力。
3. **案例研究分子(Case Studies)**: 在 EvoRRP 研究中，作者特意选择了 4 个具有代表性的复杂分子进行深度分析。虽然具体名称在摘要中未全部列出，但通常包括如 **Galantamine** (加兰他敏)、各类激酶抑制剂(如 **Imatinib** 类似物)或具有多个手性中心的天然产物。选择这些分子的原因是它们包含了立体化学控制、环化反应和保护基策略等逆合成中的难点¹。
4. **ChEMBL 随机采样**: 为了测试在大规模数据集上的鲁棒性，研究者通常会从 ChEMBL 数据库中随机抽取 100-1000 个分子，计算算法的“解决率”(Success Rate)¹³。

6. 打分函数与多目标优化设置

遗传规划的驱动力来自适应度函数(**Fitness Function**)。在逆合成中，评价一条路径的好坏是多维度的：不仅要能合成，还要便宜、安全、步骤短。

6.1 适应度函数的构成

适应度 $F(T)$ 通常是一个加权和或多目标向量：

$$F(T) = w_1 \cdot P_{\{\text{feasibility}\}}(T) + w_2 \cdot C_{\{\text{cost}\}}(T) + w_3 \cdot L_{\{\text{length}\}}(T)$$

6.1.1 路径可行性(**Feasibility & Confidence**)

这是最关键的指标，衡量反应发生的概率。

- **NLL (Negative Log Likelihood)**: 单步模型预测反应 R 的概率为 p ，则惩罚项为 $-\log(p)$ 。整条路径的 NLL 是各步反应 NLL 的累加。
- **RAscore (Retrosynthetic Accessibility Score)**: 这是一个基于机器学习的二分类或回归分数(0-1)，专门用于预测一个分子是否“可合成”。它通过训练分类器区分“已合成各分子”和“生成模型产生的不可合成分子”获得。在 GP 中，RAscore 可用于快速剪枝，淘汰那些生成了极难合成中间体的个体¹⁸。

6.1.2 分子复杂性与合成难度

- **SCScore (Synthetic Complexity Score)**: 由 MIT 团队提出，分值范围 1-5。该模型基于一个假设：反应产物的复杂性通常高于反应物。它在大规模反应语料库上训练，学习分子的结构特征。一条好的逆合成路线，其前体的 SCScore 应逐级降低，直到接近起始原料的低分值²⁰。
- **SA_Score (Synthetic Accessibility Score)**: 基于分子片段和结构特征(如大环、手性中心数量)的启发式打分。虽然计算简单，但不如 SCScore 具有数据驱动的准确性²²。

6.1.3 商业成本与路径长度

- 材料成本：如果叶节点分子在 eMolecules 中，成本为实际价格(或对数价格)；如果不在，则赋予一个极大的惩罚值(如 10^6)，迫使算法继续变异该节点直到找到可购买的前体¹⁵。
- 路径长度：对反应步数进行惩罚，防止算法生成冗长迂回的路线。

6.2 多目标优化(**Pareto Front**)

EvoRRP 的先进之处在于它不仅优化单一指标，还考虑了多目标权衡。通过 Pareto 优化，算法可以同时输出多条路线：

- 路线 A：步骤最短，但原料昂贵。
- 路线 B：步骤较长，但原料极其廉价。
- 路线 C：避开了某种危险反应(如叠氮化)，安全性最高。

这种多解特性是 EA 相比于通常只返回单一最优解的 MCTS 的一大优势²⁴。

7. 完整的进化逆合成流程

本节详细描述从无到有构建该系统的完整算法流程。

阶段一：系统初始化与预算算

1. 模型加载：加载预训练的单步逆合成模型（如基于 Transformer 的 RetroPrime 或 MLP 策略网络）。
2. 构建数据库索引：将起始原料库（eMolecules）加载到内存中的 Trie 树或哈希表中，实现毫秒级查询。
3. 设定参数：
 - 种群大小（Population Size）：如 100-500。
 - 最大代数（Max Generations）：如 50-100。
 - 变异率（Mutation Rate）：通常较高（0.2-0.4），以鼓励探索。
 - 岛屿数量（若使用并行策略）：如 4-8 个核心。

阶段二：种群生成（第 0 代）

1. 根节点建立：将目标分子设为所有初始树的根节点。
2. 随机扩展：对每个个体，调用单步模型进行 Top-K 预测。
 - 个体 1 选择 Top-1 路径扩展一步。
 - 个体 2 选择 Top-5 中的随机路径扩展一步。
 - ... 直到填满种群。这一步确保了初始种群的多样性。

阶段三：进化循环（The Evolutionary Loop）

对于每一代 \$G\$ 从 1 到 \$Max_Gen\$：

1. 并行评估（Parallel Evaluation）：
 - 将种群分配到不同的 CPU/GPU 核心。
 - 计算每个个体的适应度 \$F(T)\$。如果是多目标，计算 Pareto 等级。
 - 检查终止条件：如果发现某个个体的所有叶节点均在原料库中，且总成本低于阈值，标记为“可行解”。
2. 选择（Selection）：
 - 采用锦标赛选择（Tournament Selection）或轮盘赌选择，保留适应度高的个体进入交配池。
 - 实施精英策略（Elitism）：直接将前 5%-10% 的最优个体复制到下一代，防止优秀解丢失。
3. 遗传操作（Operators Application）：
 - 交叉：按交叉率 \$P_c\$ 配对父代，执行“兼容子树交叉”。
 - 变异：按变异率 \$P_m\$ 对后代执行“生长”、“收缩”或“替换”操作。这里的核芯是调用单步模型获取新的反应建议。
 - 有效性过滤：对新生成的个体进行快速化学规则检查（如价键检查），丢弃无效个体。
4. 岛屿迁移（Migration）：
 - 每隔 \$N\$ 代（例如 10 代），各子种群之间交换最优的 \$k\$ 个个体。这有助于将局部发现的高效片段传播到全局。

阶段四：结果提取与后处理

1. 去重：进化结束后，收集所有“可行解”。由于进化过程可能产生重复路径，需进行去重。
2. 最终排序：根据用户偏好（如“成本优先”或“步骤优先”）对可行路线进行最终排序。
3. 可视化：将树结构转换为反应式的图形表示（Graphviz 渲染），展示每一步的反应物、试剂和反应条件。

8. 性能评估与对比分析

8.1 实验结果概览

基于 EvoRRP 的实验数据，我们可以量化 GP 方法的优势。下表展示了 EA 与主流搜索算法在解决复杂分子时的性能对比¹。

指标	蒙特卡洛树搜索 (MCTS)	进化算法 (EvoRRP)	提升幅度/差异
单步模型调用次数	高（需大量 Rollout）	低	减少 53.9%
求解三个可行路线 所需时间	慢	快	时间减少 83.9%
可行路线数量	少（倾向于收敛到单一解）	多	增加 1.38 倍
搜索机制	深度优先/随机模拟	种群并行/全局优化	EA 更适合并行化
内存消耗	随搜索树深度指数增长	随种群大小线性增长	EA 内存管理更优

8.2 深度洞察：为何进化算法更优？

- 克服稀疏奖励：MCTS 依赖于模拟到终局（Rollout）来获得价值评估。在化学中，随机模拟一条能通向起始原料的路径概率极低，导致 MCTS 经常收到“零奖励”信号，浪费大量计算资源。相比之下，EA 的适应度函数是基于当前结构的（如 SCScore 的下降梯度），即使路径未完全打通，只要复杂性降低，个体就能获得正向反馈得以保留。
- 模块化重组：化学合成具有模块化特征。分子的左边和右边往往可以独立合成。EA 的交叉算子完美契合这一特性，能够将两个分别解决了分子不同部分的半成品路径拼接成一个完美路径。这是 MCTS 这种单点生长算法所不具备的能力。

9. 结论与展望

通过将遗传规划引入分子逆合成任务，研究者们成功打破了传统搜索算法在计算效率上的瓶颈。EvoRRP 和 SynGA 等研究证明，利用树状编码、化学感知算子以及多目标进化策略，不仅可以大幅减少昂贵的神经网络推理次数，还能发现更多样化、更具创新性的合成路线。

未来的研究方向主要集中在以下几点：

1. 端到端学习：目前的流程中，单步模型是固定的。未来可尝试将进化过程中的反馈（哪些反应容易导致死胡同）用于在线微调单步模型，实现“搜索与预测”的协同进化。
2. 3D 结构融合：目前的适应度函数主要基于 2D 图结构。引入 3D 立体化学约束（如对接打分、空间位阻分析）作为适应度的一部分，将进一步提高路线的实验成功率。
3. 自动化实验室集成：将 GP 生成的路线直接输入到自动化合成机器人平台，形成“计算-合成-测试”的闭环系统。

综上所述，遗传规划为解决逆合成这一古老而复杂的化学问题注入了新的计算活力，是通往全自动化学合成设计的重要一步。

引用的著作

1. Evolutionary Retrosynthetic Route Planning [Research Frontier ...，访问时间为十二月 8, 2025, <https://ieeexplore.ieee.org/document/10595522/>
2. Evolutionary Retrosynthetic Route Planning - arXiv, 访问时间为十二月 8, 2025, <https://arxiv.org/html/2310.05186>
3. [2310.05186] Evolutionary Retrosynthetic Route Planning - arXiv, 访问时间为十二月 8, 2025, <https://arxiv.org/abs/2310.05186>
4. A Genetic Algorithm for Navigating Synthesizable Molecular Spaces - ChatPaper, 访问时间为十二月 8, 2025, <https://chatpaper.com/paper/192114>
5. A GENETIC ALGORITHM FOR NAVIGATING SYNTHESIZABLE MOLECULAR SPACES - OpenReview, 访问时间为十二月 8, 2025, <https://openreview.net/pdf?id=OvMtGGaFUT>
6. Artificial Intelligence Methods and Models for Retro-Biosynthesis: A Scoping Review | ACS Synthetic Biology - ACS Publications, 访问时间为十二月 8, 2025, <https://pubs.acs.org/doi/10.1021/acssynbio.4c00091>
7. Navigating with chemometrics and machine learning in chemistry - PMC - PubMed Central, 访问时间为十二月 8, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC9870782/>
8. READRetro: Natural Product Biosynthesis Planning with Retrieval-Augmented Dual-View Retrosynthesis - bioRxiv, 访问时间为十二月 8, 2025, <https://www.biorxiv.org/content/10.1101/2023.03.21.533616v1.full.pdf>
9. Genetic programming - Wikipedia, 访问时间为十二月 8, 2025, https://en.wikipedia.org/wiki/Genetic_programming
10. 访问时间为十二月 8, 2025, <http://www.genetic-programming.com/jkpdf/psb2001.pdf>
11. Enhancing Program Synthesis with Large Language Models Using

Many-Objective Grammar-Guided Genetic Programming - MDPI, 访问时间为十二月 8, 2025, <https://www.mdpi.com/1999-4893/17/7/287>

12. A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space - RSC Publishing, 访问时间为十二月 8, 2025, <https://pubs.rsc.org/en/content/articlehtml/2019/sc/c8sc05372c>
13. A Genetic Algorithm for Navigating Synthesizable Molecular Spaces - arXiv, 访问时间为十二月 8, 2025, <https://www.arxiv.org/pdf/2509.20719.pdf>
14. A Genetic Algorithm for Navigating Synthesizable Molecular Spaces - arXiv, 访问时间为十二月 8, 2025, <https://arxiv.org/html/2509.20719v1>
15. Efficient retrosynthetic planning with MCTS exploration enhanced A* search - PMC, 访问时间为十二月 8, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10920677/>
16. Models Matter: the impact of single-step retrosynthesis on synthesis planning, 访问时间为十二月 8, 2025, <https://pubs.rsc.org/en/content/articlehtml/2024/dd/d3dd00252g>
17. RetroEA: An efficient evolutionary algorithm for retrosynthetic route planning - ResearchGate, 访问时间为十二月 8, 2025, https://www.researchgate.net/publication/393242237_RetroEA_An_efficient_evolutionary_algorithm_for_retrosynthetic_route_planning
18. Retrosynthetic accessibility score (RAscore) – rapid machine learned synthesizability classification from AI driven retrosynthetic planning - NIH, 访问时间为十二月 8, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC8179384/>
19. Retrosynthetic accessibility score (RAscore) – rapid machine learned synthesizability classification from AI driven retrosynthetic planning - Chemical Science (RSC Publishing), 访问时间为十二月 8, 2025, <https://pubs.rsc.org/en/content/articlelanding/2021/sc/d0sc05401a>
20. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy - PMC - NIH, 访问时间为十二月 8, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC8152799/>
21. Computer-Assisted Retrosynthesis Based on Molecular Similarity | ACS Central Science, 访问时间为十二月 8, 2025, <https://pubs.acs.org/doi/10.1021/acscentsci.7b00355>
22. Evaluating Molecule Synthesizability via Retrosynthetic Planning and Reaction Prediction, 访问时间为十二月 8, 2025, <https://arxiv.org/html/2411.08306v2>
23. Synthetic Accessibility: Definition, Importance, and How to Assess It with Neurosnap, 访问时间为十二月 8, 2025, <https://neurosnap.ai/blog/post/synthetic-accessibility-definition-importance-and-how-to-assess-it-with-neurosnap/68cb0899cec8a8b395412c77>
24. A Macro-Micro Population-Based Co-Evolutionary Multi-Objective Algorithm for Community Detection in Complex Networks - ResearchGate, 访问时间为十二月 8, 2025, https://www.researchgate.net/publication/368511687_A_Macro-Micro_Population-Based_Co-Evolutionary_Multi-Objective_Algorithm_for_Community_Detection_in_Complex_Networks
25. Learning Regularity for Evolutionary Multiobjective Search: A Generative

Model-Based Approach | Request PDF - ResearchGate, 访问时间为十二月 8, 2025,

https://www.researchgate.net/publication/375175848_Learning_Regularity_for_Evolutionary_Multiobjective_Search_A_Generative_Model-Based_Approach