

# Multiretro多步逆合成 Agent 项目

## 1. 项目概述与核心功能

**Multiretro** 是一个基于TF模型与模板结合方法，并使用大语言模型 (LLM) 与 ReAct 推理框架增强的智能逆合成规划系统。它旨在解决复杂有机分子的多步合成路线设计问题，通过模拟化学家的“直觉 + 验证”思维模式，提供高可行性的合成方案。

### 核心功能亮点

- 自动化逆合成规划**: 输入目标分子 SMILES，自动生成多步逆合成树。
- 交互式人机协同 (HITL)**: 支持化学家在规划过程中实时干预（如 `switch` 切换分支、`reopen` 回溯重开、`expert` 专家策略注入）。
- 多维度 5D 评分体系**: 摒弃单一概率评分，引入 **复杂度、反应活性、选择性、经济性、保护基成本** 五维评价。
- 全透明推理 (Glass-Box)**: 每一决策步骤均提供完整的化学推理链条 (Chain of Thought)，而非黑盒输出。
- 自愈与纠错**: 具备 SMILES 自动修复及“幽灵节点”自动清理机制，确保数据与逻辑的强一致性。

## 2. 深度技术原理审计 (Deep Dive)

目前项目不是单一简单的 LLM 调用，构建了具备**认知架构 (Cognitive Architecture)** 的 AI 化学家。以下是核心技术实现的深度审计：

### 2.1 双层候选评价机制 (Two-Stage Evaluation)

传统的逆合成模型通常直接从 LLM 获取 Top-k，而 MoleReact 在 `agent_react.py` 中实现了更为严谨的“**先分析，后决策**”双层agent架构：

- 第一层：原子工具分析 (Chemical Analysis)**
  - 代码定位**: `_action_expand_node` (Lines 311-391)
  - 实现逻辑**: agent-1通过 `MoleculeAnalysisTool` 和 `StockCheck` 对 15+ 候选前体进行物理性质扫描、骨架提取和库存匹配。
  - 输出**: 生成一份包含“库存状态 + 结构风险”的元数据事实块 (`candidate_blocks`)，而非直接筛选。
- 第二层：全局一致性评价 (Consistency Selection)**
  - 代码定位**: `_action_expand_node` (Lines 396-410) -> 调用 `v3.4 Selector Agent`
  - 实现逻辑**: 将第一层生成的元数据注入 Context，启用基于 JSON 的 **Selector Agent**。要求 LLM 在看到“全量候选数据的化学事实”后，严格基于 5D 标准进行一致性排序。
  - 价值**: 彻底规避了 LLM 因“缺乏化学常识”而产生的幻觉，确保决策是基于事实 (Fact-Based) 而非单纯的语言概率。

## 2.2 长短期记忆融合 (Long-Short Term Memory)

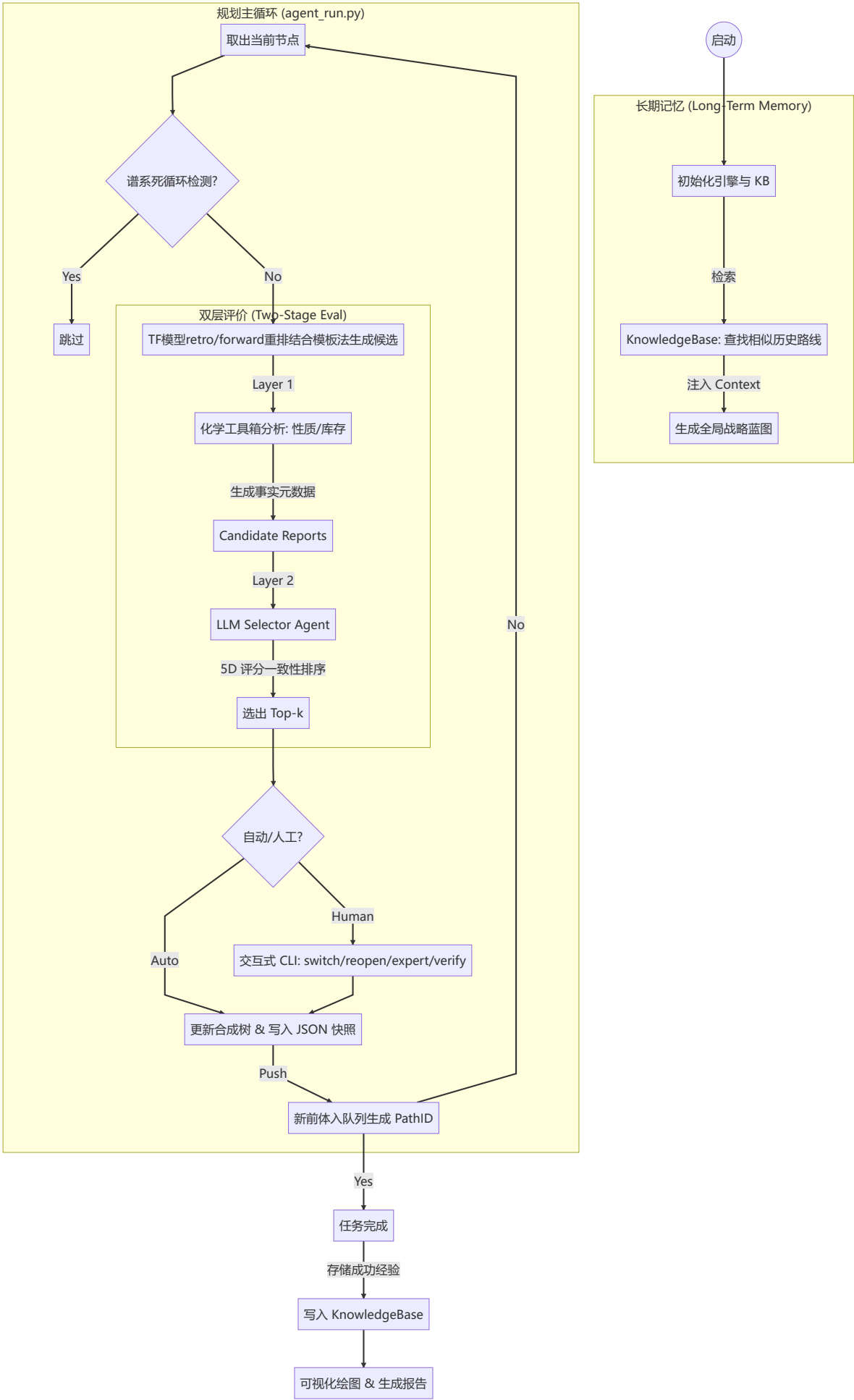
系统具备完整的记忆回路，实现了经验的复用：

- **长期记忆 (Knowledge Base RAG)**
    - **模块:** `knowledge_base.py` & `agent_run.py`
    - **实现:** 在规划开始前，检索 SQLite 数据库中结构相似的历史成功路线 (`retrieve_similar_routes`)，并注入 System Prompt 作为 Few-Shot 示例。成功路线会自动固化回知识库。
    - **价值:** 实现 "越用越聪明" 的进化能力。
  - **短期记忆 (Session Context)**
    - **模块:** `session_logger.py`
    - **实现:** 维护 `cumulative_route` 的 JSON 快照。LLM 始终能感知当前的谱系路径 (`current_lineage`)，防止在同一条死胡同里反复打转 (Loop Detection)。
- 

## 3. 工作流程与生命周期 (Workflow)

整个工作流是一个闭环的 "感知 (Observe) -> 思考 (Think) -> 行动 (Act) -> 记录 (Store)" 循环。

### 3.1 核心流程图



## 3.2 关键步骤说明

- 全局蓝图:** Agent-0 首先识别分子的关键断键点和敏感基团制定全局分析方向。
- 深扫 (Deep Scan):** ReAct 中子 Agent-1先对 Top 候选进行“深层审计” (LLM 自主调用化学工具包, 分析分子详细性质数据、反应条件、毒性以及可买性) 记录信息数据。
- 多agent协助:** Agent-0构建全局战略和目标分子分析; Agent-1执行深层审计; Agent-2进行全面深入分析
- 自愈合 (Self-Healing):** 如果 Agent 发现现有smiles语义错误或者模板不足, 会自动“幻觉”出一条修正路线 (Auto-Spawn Patching), 再自行调用化学工具对其进行全局蓝图匹配和化学校验, 合法则自动转化为有效分支。

## 4. 系统架构与模块实现

本项目采用 **以控制器为中心 (Controller-Centric)** 的星型架构, 确保了状态管理的严谨性。

### 4.1 核心模块概览

| 模块                        | 核心职责          | 技术关键点   |
|---------------------------|---------------|---|
| agent_run.py              | 中央控制器 (Brain) | 维护全局状态机与待解队列。reopen 操作的队列残留问题, 实现了逻辑状态与执行队列的严格同步。         |
| agent_react.py            | ReAct 引擎      | 实现 Level-3 级自主智能体。通过 EXPAND / SWITCH 等指令自主管理目标栈, 实现“无控制”。 |
| handlers/llm_selection.py | 决策与解析         | 内置 JSON 解析器, 能够处理非结构化的 LLM 输出并自动修复格式错误 (Auto-Fix)。        |
| session_logger.py         | 记忆系统          | 实现 JSON-First 的状态持久化, 确保断点或异常中断后可无损恢复现场。                  |
| knowledge_base.py         | 知识库           | 基于 SQLite 的经验存储与检索, 支持由 RAG 驱动的相似路线推荐。                    |

## 5. 项目小结

### 5.1

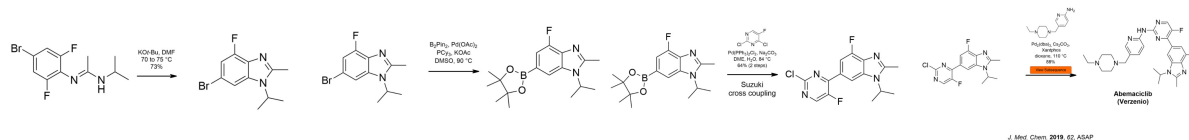
- 解决了“状态-执行分离”难题:** 首次在基于 LLM 的逆合成中实现了支持“无限回溯” (Reopen) 的强健状态机, 彻底解决了队列残留导致的图结构崩溃问题。
- ReAct 战略协同:** 引入 Agent-0 进行全局蓝图规划, 避免了传统算法容易陷入局部最优 (Local Optima) 的缺陷。
- 半人马合成 (Centaur Synthesis):** 并非试图完全取代人, 而是通过精设计的 CLI 交互, 允许让人类专家的直觉引导 AI 的算力, 实现了  $1+1 > 2$  的效果。

4. **自动修补机制 (Auto-Spawn):** 将 LLM 的创造性幻觉转化为经过校验的合法化学路径，变废为宝，MoleReact V3.6 标志着逆合成工具从“单步预测”向“全流程智能规划”的跨越。它不仅是一个代码库，更是一套经受过逻辑审计验证的、具备工业级鲁棒性的 **AI 化学家框架**。
5. **成本低廉极致性价比:** 消费级笔记本即可跑通全流程进行多步逆合成分析，四两拨千斤

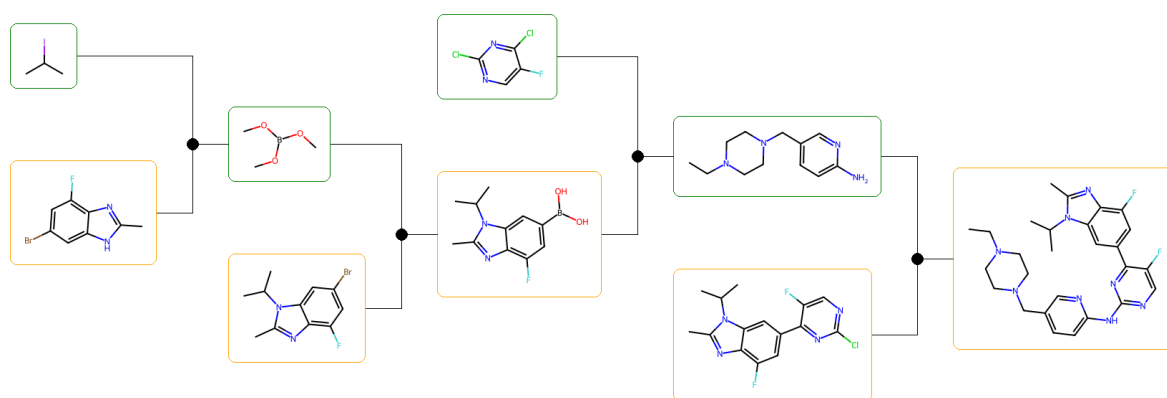
## 6.多步合成路线验证

### 6.1 abemaciclib 2019

合成思路以及骨架拆分：



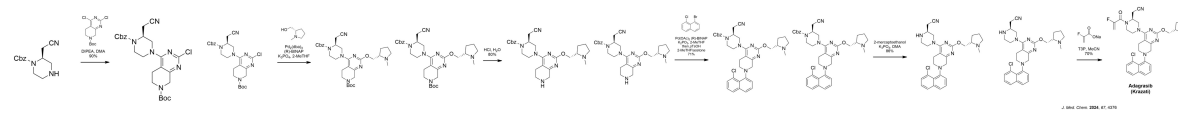
预测思路以及骨架拆分：



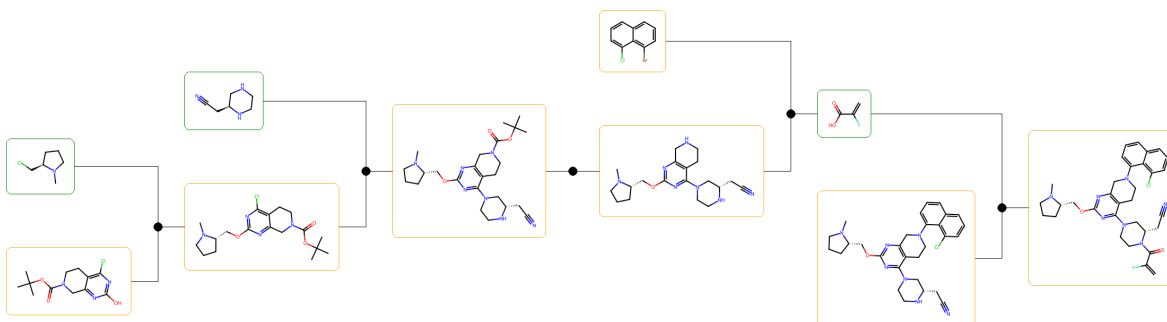
逆合成分析：文献路线采用氯原子取代聚合式合成，主要前体通过硼酸酯与多氯取代得到，相应硼酸酯又基于溴代得到，逆合成最后一步关环；与预测结果一致

### 6.2 adagrasib 2024

合成思路以及骨架拆分：



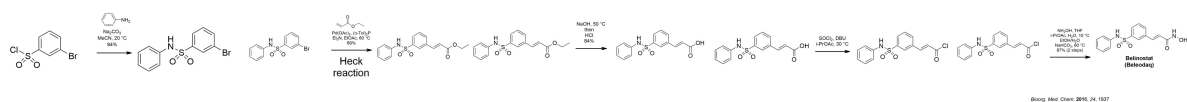
预测思路以及骨架拆分：



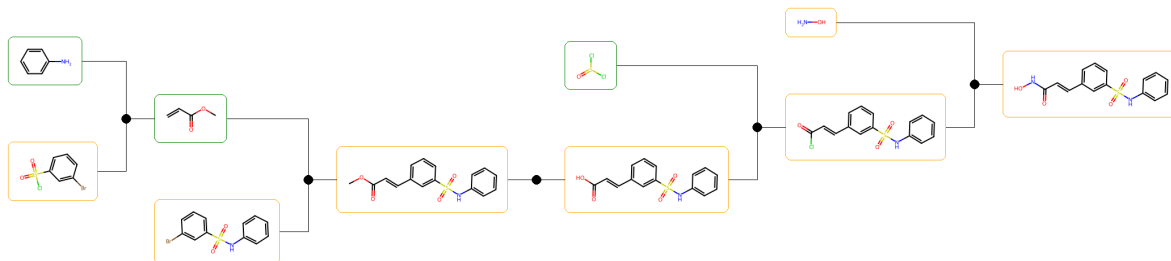
逆合成分析：文献路线采用烯氟酸盐作为逆合成第一步，胺基保护作为第二步，之后卸去二取代溴氯，对暴露的二级氮原子同样boc酸酐保护，并进一步逆向拆分吡咯，最后一步拆分出吡嗪；预测路线拆分思路一致

### 6.3 belinostat 2016:

合成思路以及骨架拆分:



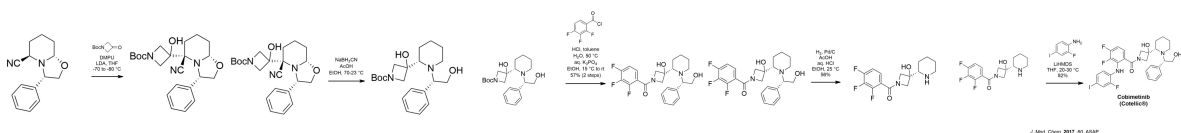
预测思路以及骨架拆分:



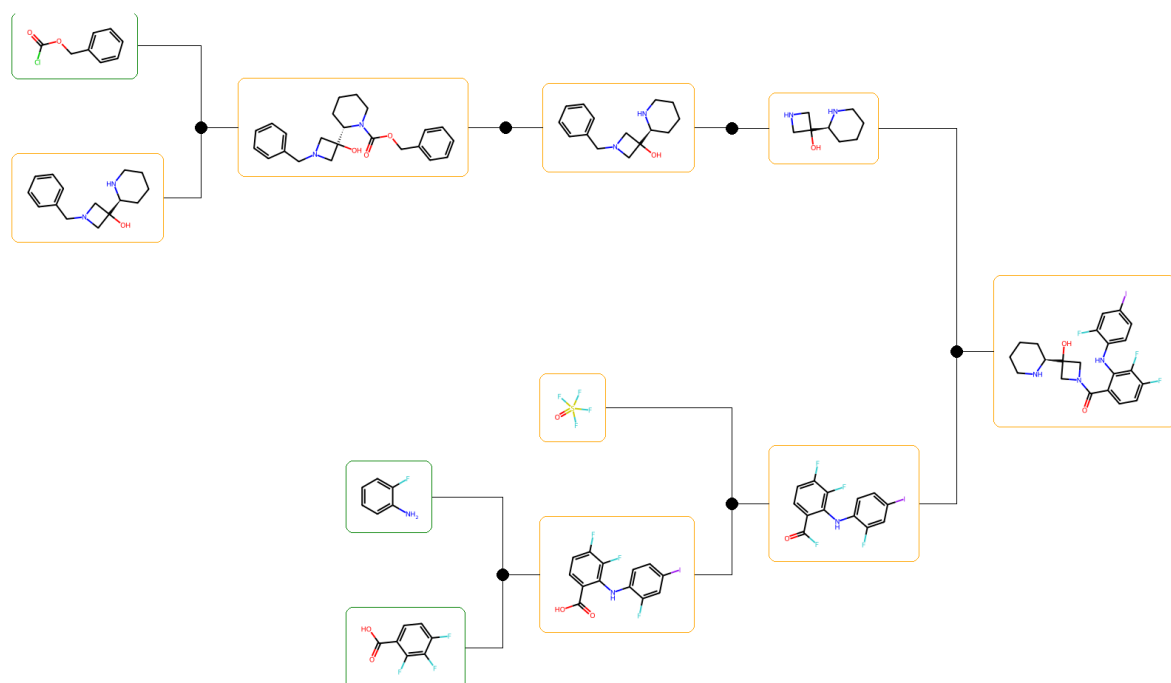
逆合成分析：文献路线采用羟胺与酰氯作为逆合成第一步，在得到酸，进一步为酯，实行迈克尔加成，逆向最后一步为胺基与间溴苯磺酰氯的拆分；预测路线拆分一致

## 6.4 cobimatinib2017:

合成思路以及骨架拆分:



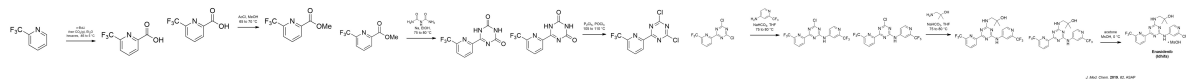
预测思路以及骨架拆分:



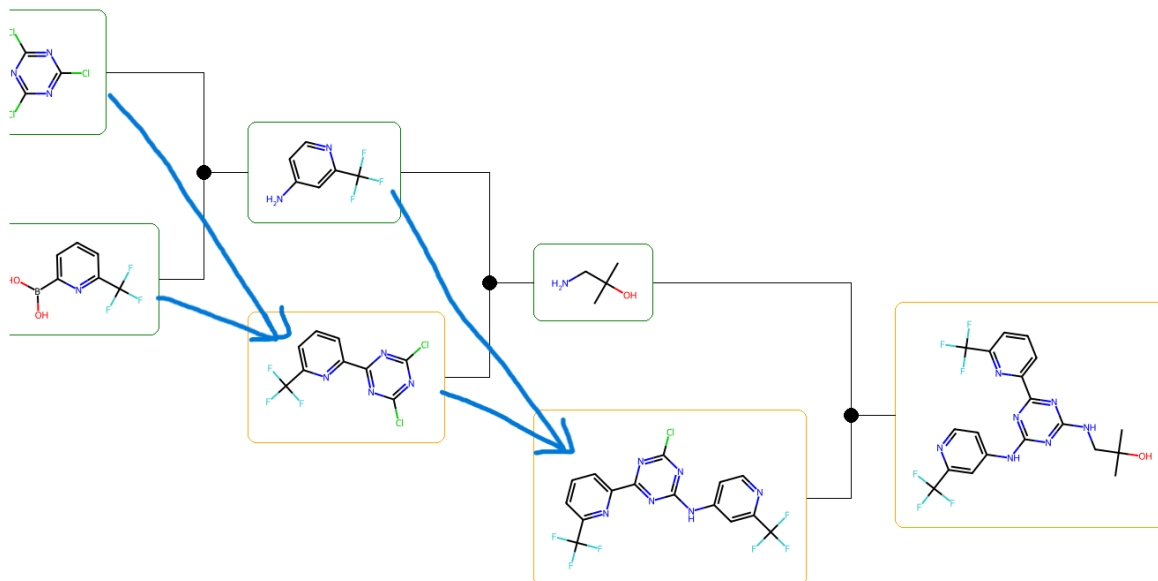
逆合成分析:

## 6.5 enasidenib 2019:

合成思路以及骨架拆分:

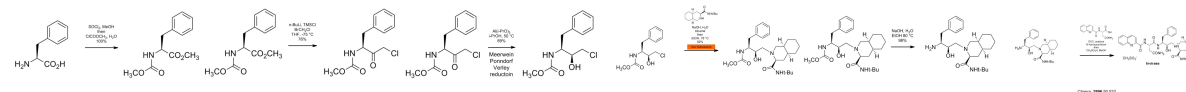


预测思路以及骨架拆分:

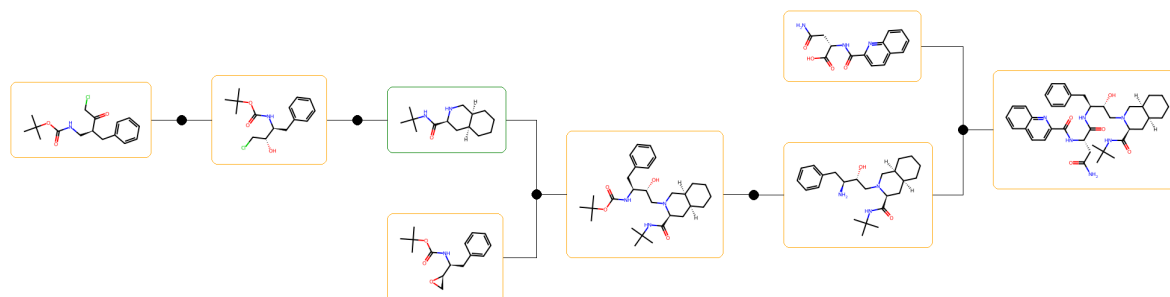


## 6.6 invirase 1996:

合成思路以及骨架拆分:

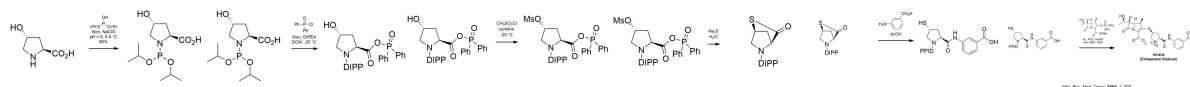


预测思路以及骨架拆分:



## 6.7 ivanz 2004(失败结果)

合成思路以及骨架拆分:



预测思路以及骨架拆分：

