

研究背景与动机

登革热：全球性公共卫生威胁与药物发现挑战



登革热：被忽视的全球威胁

- ▶ **3.9亿人/年感染，遍布129国家**
- ▶ WHO列为**十大全球健康威胁**
- ▶ 气候变化加剧传播：预计2050年**+20亿**风险人口
- ▶ **无特效药**，仅支持性治疗
- ▶ 疫苗效力有限 (Dengvaxia仅60%，有副作用风险)

关键靶点：**NS2B-NS3蛋白酶**

病毒复制必需 | 保守性高 | 成药性好



- ▶ 时
- ▶ 成
- ▶ 成
- ▶ 化
- ▶ 高

DE
• 临
• 已
• 成



为什么选择AI驱动的从头药物设计？

✓ 突破化学空间限制

生成式AI可探索未知化学结构

✓ 多目标同步优化

活性+成药性+可合成性一体化设计

REINVENT4：强化学习驱动的分子设计

从通用工具到靶向设计：LibInvent策略的选择逻辑



REINVENT系列：AstraZeneca的开源贡献

2017

REINVENT 1.0

RNN-based
SMILES生成
基础RL框架

2020

REINVENT 2.0

LibInvent
LinkInvent
骨架装饰



REINVENT4核心优势：为什么选择它？



强化学习引导

DAP/SDAP算法，平衡探索与利用，避免mode collapse



模块化评分

15+组件：活性/ADMET/合成/结构，可自定义组合



多种生成模式

De novo/Scaffold/Linker/R-group，适配不同设计需求



工业验证

AstraZeneca内部使用，多个项目验证



项目策略：为什么用LibInvent而非De Novo？

Reinvent



LibInvent



DENV NS2B-NS3抑制剂AI生成

Run9_T1200 实验进展汇报 | LibInvent + QSAR引导生成

🕒 实验进行中: 2219 / 6000 步 (37%)

项目目标

- 靶点: 登革热病毒(DENV) NS2B-NS3蛋白酶
- 设计策略: 双芳环吡咯烷骨架
- 活性目标: pIC50 **≥ 8.0** (IC50 ≤ 10 nM)
- 成药性: QED ≥ 0.7, SA ≤ 4.0
- 当前规模: **113.6万**候选分子

🕒 预计完成: ~160万分子

当

- 金标
- 活性
- 最优
- 高活
- 极高

✓ 成功

实验进展

当前步数	~2219 / 6000
完成度	37%
已生成分子	113.6万
去重后	109.7万 (96.6%)
R基团多样性	111.9万组合
预计最终	~160万分子
预计金标准	~140个

技

- QSA
- 生成
- 多目
- 显存
- 化学

🚀 生产

方法学与实验设计

多层次评分策略 + QSAR引导优化

评分组件架构 (15 Components)

第一优先级: 生物活性 (Weight: 2.5)

{ QSARScorer - Random Forest pIC50预测
Transform: double_sigmoid (4.0-9.0)

第二优先级: 化学稳定性 (Total: 1.6)

{ CustomAlerts - 15种不稳定substructures
NumAtomStereoCenters

第三优先级: 类药性 (Total: 0.7)

{ QED - Drug-likeness
SAScore - Synthetic Accessibility

第四优先级: 物理化学性质 (Total: 1.9)

{ MW: 250-600 Da
LogP: 0.5-5.0
TPSA: 30-120 μ
HBA \leq 8
HBD \leq 4
Heavy Atoms: 18-45



中期结果与可视化

113.6万分子 (37%进度) | 97.5%高活性 | 96个金标准候选物

Generation Analysis

Generation Analysis

IC50活性分布

0-10 nM (极高)	2,170 (0.2%)
10-50 nM (高)	1,081,138 (98.6%)
>50 nM (低)	13,281 (1.2%)

成功率: 98.8% 分子达到 $\text{pIC50} \geq 7.0$

分子量
LogP
HBA
HBD
TPSA
QED
SA Score

实验进展预测

当前 (37%)

总分子

113.6万

触发警报

[*,r8]



金标准候选物与结构展示

96个极高活性分子 (去重后) | IC50 6.5-10 nM | 无结构警报

◆ 金标准

96

- ✓ pIC50 ≥ 8.0
- ✓ IC50: 6.5-10 nM
- ✓ QED ≥ 0.7
- ✓ SA ≤ 4.0
- ✓ MW: 300-500 Da
- ✓ LogP: 1-4

★ 高标准

289,397

- ✓ pIC50 ≥ 7.5
- ✓ IC50: 5.6-31.6 nM
- ✓ QED ≥ 0.6
- ✓ SA ≤ 4.5

◆ 金标准 Top 10



IC50范围
平均QED
平均SA

6.5-10 nM
0.79
3.63



IC50范围
平均QED
平均SA

下一步工作

- ✓ 等待实验完成 (预计4000步后稳定)
- ✓ 从最终金标准中选择**20-30个**进行合成
- ✓ 体外酶活性测定 (NS2B-NS3 protease)
- ✓ 细胞毒性评估 (CC50)
- ✓ 抗病毒活性测试 (EC50)

中期

- 当前生成11
- 获得96个金
- 98.8%分子
- 化学稳定性
- 预计最终：