

REINVENT4 配置演进详细对比：Run5 → Run6 → Run7

总体概览

配置	组件总数	核心策略	预期效果	适用场景
Run5	5个	活性优先，基础约束	高活性但化学合理性未知	初步探索活性空间
Run6	10个	活性+稳定性平衡	活性较高且化学合理	平衡探索
Run7	15个	类药性全面优化	活性中等但类药性最优	临床候选物筛选

详细组件对比

1. 核心活性预测

组件	Run5	Run6	Run7	说明
QSARScorer (DENV活性)	<input checked="" type="checkbox"/> weight=2.0	<input checked="" type="checkbox"/> weight=2.0	<input checked="" type="checkbox"/> weight=1.5	Run7降低权重，避免过度优化活性
Transform	double_sigmoid (5.0-9.0)	同Run5	同Run5	所有版本一致

差异总结：

- Run5/6：活性主导（权重2.0，相对其他组件很高）
- Run7：活性与约束更平衡（权重1.5）

2. 类药性评估

组件	Run5	Run6	Run7	说明
QED	<input checked="" type="checkbox"/> weight=0.5	<input checked="" type="checkbox"/> weight=0.5	<input checked="" type="checkbox"/> weight=0.5	所有版本一致
SAScore	<input checked="" type="checkbox"/> weight=0.4 阈值high=7.0	<input checked="" type="checkbox"/> weight=0.5 阈值high=5.0	<input checked="" type="checkbox"/> weight=0.6 阈值high=4.5	逐步严格化合成难度

差异总结：

- Run5：允许SA=7（较复杂）
- Run6：限制SA≤5（中等复杂）
- Run7：限制SA≤4.5（相对简单），权重最高

3. 分子量和大小

组件	Run5	Run6	Run7	说明
MolecularWeight	<input checked="" type="checkbox"/> weight=0.3 范围 200-500	<input checked="" type="checkbox"/> weight=0.4 范围 300-550	<input checked="" type="checkbox"/> weight=0.4 范围 300-500	Run6放宽上限，Run7收紧
NumHeavyAtoms	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> weight=0.3 范围 20-40	<input checked="" type="checkbox"/> weight=0.3 范围 20-35	Run6新增，Run7更严格

差异总结：

- Run5：只控制MW，范围较宽松
- Run6：增加重原子约束，允许更大分子（550 Da）
- Run7：收紧至500 Da以下

4. 立体化学复杂度 ★ 关键差异

组件	Run5	Run6	Run7	说明
NumAtomStereoCenters	✗ 无约束	✓ weight=0.6 上限 ≤3	✓ weight=1.0 上限 ≤2	Run5允许4+手性中心！

差异总结：

- Run5：可能生成4个手性中心（16个异构体）→ 合成地狱
- Run6：限制≤3个（8个异构体）→ 改善但仍复杂
- Run7：限制≤2个（4个异构体）→ 合成友好

5. 分子柔性和环系统

组件	Run5	Run6	Run7	说明
NumRotBond	✗	✓ weight=0.3 上限 ≤8	✓ weight=0.3 上限 ≤7	Run6新增
NumRings	✗	✓ weight=0.3 上限 ≤4	✓ weight=0.3 上限 ≤4	Run6新增
NumAromaticRings	✗	✓ weight=0.2 上限 ≤3	✓ weight=0.2 上限 ≤3	Run6新增

差异总结：

- Run5：无柔性和环系统约束
- Run6/7：控制分子形状和刚性

6. Lipinski规则相关 NEW Run7专属

组件	Run5	Run6	Run7	说明
SlogP	✗ ✗	✓ weight=0.4 范围 1.0-4.0		控制亲脂性
TPSA	✗ ✗	✓ weight=0.3 范围 40-100		控制极性表面积
HBondAcceptors	✗ ✗	✓ weight=0.4 上限 ≤7		氢键受体 (Lipinski ≤10)
HBondDonors	✗ ✗	✓ weight=0.3 上限 ≤3		氢键供体 (Lipinski ≤5)

差异总结：

- Run5/6：不直接控制Lipinski规则
- Run7：全面Lipinski合规性检查

7. 化学稳定性过滤

组件	Run5	Run6	Run7
CustomAlerts	✓ weight=1.0 5个SMARTS	✓ weight=1.0 15个SMARTS	✓ weight=1.0 18个SMARTS

详细SMARTS对比：

SMARTS类别	Run5	Run6	Run7	说明
大环 (r8-r10)	✓ 3个	✓ 3个	✓ 3个	所有版本一致
基础不稳定结构	✓ 2个 ([0][0], [C+])	✓ 2个	✓ 2个	所有版本一致
炔烃 (C#C)	✓	✓	✓	所有版本一致
反应性官能团	✗	✓ 10个 (肼、硫醇、酸酐等)	✓ 10个	Run6新增
吡咯烷特定规则	✗	✗	✓ 3个 (季碳、全取代、双酰胺)	Run7专属

Run6新增的10个稳定性SMARTS：



[NX3][NX3] # 肽类
 [SH] # 硫醇 (易氧化)
 C(=O)OC(=O) # 酸酐
 [N+](=O)[O-] # 硝基
 S(=O)(=O)Cl # 磺酰氯
 [F,Cl,Br,I][C,c][F,Cl,Br,I] # 多卤代
 c1ccc2c(c1)C(=O)c1c(C2=O)cccc1 # 醛类 (PAINS)
 C1=CC(=O)C=CC1=O # 对苯醌
 C(=O)N(C(=O)) # N,N-双酰胺

Run7新增的3个吡咯烷特定SMARTS:



[C;R1]1[C;R1][C;R1][N;R1][C;R1]1([!#1])([#1])([!#1])

吡咯烷C2位季碳 (3个非氢取代)

[C;R1]1[C;R1]([!#1])([#1])[C;R1]([!#1])([#1])[N;R1][C;R1]1

吡咯烷C3和C4同时全取代

C(=O)N([!#1])C(=O)

双酰胺结构 (你的Step 37高分分子有这个)

组件分类统计

按功能分类

功能类别	Run5	Run6	Run7	主要作用
活性预测	1	1	1	QSAR模型打分
类药性	2 (QED, SA)	2	2	综合类药性评估
大小/质量	1 (MW)	2 (MW, HeavyAtoms)	2	控制分子大小
立体化学	0 ❌	1 ★	1 (更严格)	合成可行性关键
柔性/环系统	0	3 ★	3	结构复杂度
Lipinski规则	0	0	4 NEW	口服药物性
化学稳定性	1 (5个)	1 (15个)	1 (18个)	过滤不良结构
总计	5	10	15	-

权重分配对比

总权重分析

配置	活性权重	约束权重	比例	策略倾向
Run5 2.0	2.2	47%	vs 53%	活性主导
Run6 2.0	4.6	30%	vs 70%	平衡偏约束
Run7 1.5	6.8	18%	vs 82%	约束主导

详细权重分布

组件类别	Run5	Run6	Run7	变化趋势
活性	2.0	2.0	1.5	降低
类药性 (QED+SA)	0.9	1.0	1.1	↑ 提高
立体化学	0	0.6	1.0	↑↑ 大幅提高
分子性质	0.3	1.0	1.4	↑ 提高
Lipinski	0	0	1.0	NEW 新增
柔性/环	0	0.8	0.8	→ 持平
稳定性	1.0	1.0	1.0	→ 持平

🎯 实际影响预测

对Step 37问题分子的影响

以这个高分分子为例：



Run5 Score: 0.915

检查项	实际值	Run5	Run6	Run7	结果
立体中心数 4个	✓ 通过	⚠ 惩罚 (>3)	✗ 严重惩罚 (>2)	Run7会大幅降分	
双酰胺结构 有	✓ 通过	✓ 通过	✗ 直接过滤	Run7会拒绝	
重原子数 ~35	✓ 通过	✓ 通过	⚠ 边缘	Run7临界	
HBA ~8个	✓ 通过	✓ 通过	⚠ 惩罚 (>7)	Run7会降分	
SA Score ~3.8	✓ 通过	✓ 通过	✓ 通过	所有版本OK	
MW 461	✓ 通过	✓ 通过	✓ 通过	所有版本OK	

预测最终分数：

- Run5: 0.915 (实际)
- Run6: ~0.65 (立体中心惩罚)
- Run7: 0 或 <0.3 (双酰胺直接过滤 + 立体中心严重惩罚)

迁移建议

从Run5升级到Run6

改进点：

- ✓ 解决立体化学复杂度问题
- ✓ 增加基础稳定性检查
- ✓ 控制分子柔性和环系统

保留点：

- → 活性权重仍然较高 (2.0)
- → 未严格限制Lipinski规则

适合：第一次优化，快速改善化学合理性

从Run6升级到Run7

改进点：

- 全面Lipinski合规性
- 更严格的立体化学控制 (≤ 2 个手性中心)
- 降低活性权重，平衡整体类药性
- 专门针对吡咯烷的不良结构

妥协点：

- 活性预测分数可能略降（权重1.5 vs 2.0）
- 可能生成较少的高活性分子

适合：追求临床候选物质量，愿意牺牲部分活性

推荐决策树



你的目标是什么？

- 快速探索高活性空间
- 使用 Run5（风险：可能不可合成）
- 平衡活性和合成可行性
 - 使用 Run6 ★ 推荐首选
- 获得类药性最优的候选物
 - 使用 Run7（风险：活性可能降低）

建议的实验流程

1. 第一轮：Run5（已完成）→发现高活性但不稳定
2. 第二轮：Run6 → 验证立体化学约束效果
3. 第三轮：
 - 如果Run6结果仍有问题 → Run7
 - 如果Run6结果可接受 → 停止优化，进入实验验证

关键要点

对比维度	Run5	Run6	Run7
哲学	活性第一	活性+稳定性	类药性优先
风险	合成困难	平衡风险	活性可能不足
适用阶段	早期探索	中期优化	后期筛选
实验验证难度	高（可能做不出）	中	低（易合成）
成药概率	低	中	高

核心差异一句话总结：

- Run5： "给我最强的活性，别管能不能做出来"
- Run6： "活性要高，但必须化学合理"
- Run7： "必须是真正的候选药物"