

乳酸对映体与乙醇：性质、指纹相似度与三维构象 方法报告

科研工作台

报告日期：未知（本轮未提供明确日期；以工作区工具执行记录为准）。所有数值均来自本轮实际工具运行产物，来源在相应章节标明。

1 目的与范围

比较 C[C@H](O)C(=O)O ((S)-乳酸)、C[C@@H](O)C(=O)O ((R)-乳酸) 与 CCO (乙醇) 三个分子：(1) 用 RDKit 描述符计算真实理化性质；(2) 计算 Morgan 位指纹 Tanimoto 相似度矩阵，并明确指纹是否编码手性；(3) 为两个乳酸对映体各生成固定种子、MMFF94s 优化的三维构象 (ETKDGv3 坐标生成 + MMFF94s 力场优化，非二维坐标)，并以独立解析验证 SDF 可读、坐标确为三维、立体化学在三维坐标中得到保留。

2 分子与输入

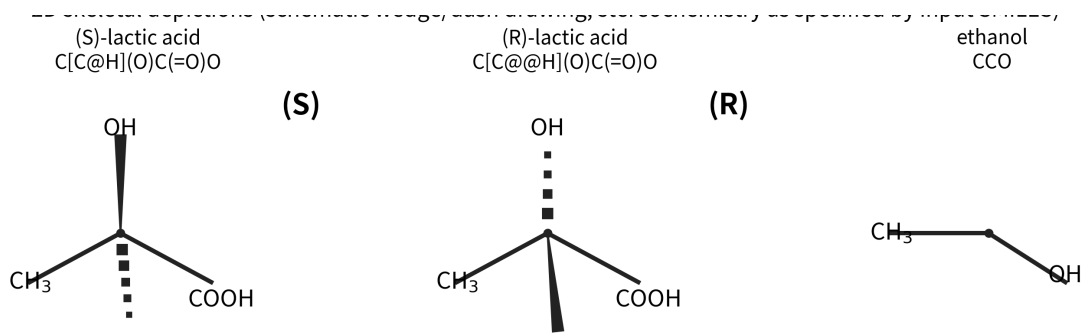


图 1：2D 骨架示意图（楔键/虚线楔按输入 SMILES 指定的立体化学绘制；此为手绘式示意，非 RDKit 自动 2D depiction；原子间相对朝向为示意，不代表优化坐标）。RDKit 渲染版另存于 out/structures_rdkit.svg。

3 理化性质 (RDKit 描述符)

来源：molecular-analysis__properties。两对映体所有标量描述符完全相同 (C₃H₆O₃, MW 90.078, exact mass 90.0317, cLogP -0.5482, TPSA 57.53 Å², HBD/HBA 各 2, 可旋转键 1, 重原子 6, fraction csp³ 0.667, Lipinski 违反数 0)；乙醇 C₂H₆O, MW 46.069, cLogP

-0.0014, TPSA 20.23 Å², HBD/HBA 各 1。注意: RDKit 标量描述符不区分手性; pKa、logD7.4、ADMET 明确未计算。

4 Morgan 指纹 Tanimoto 相似度

来源: molecular-analysis__fingerprint_similarity。参数: Morgan 位指纹, radius = 2, 2048 bits, **chirality = True** (明确编码手性)。

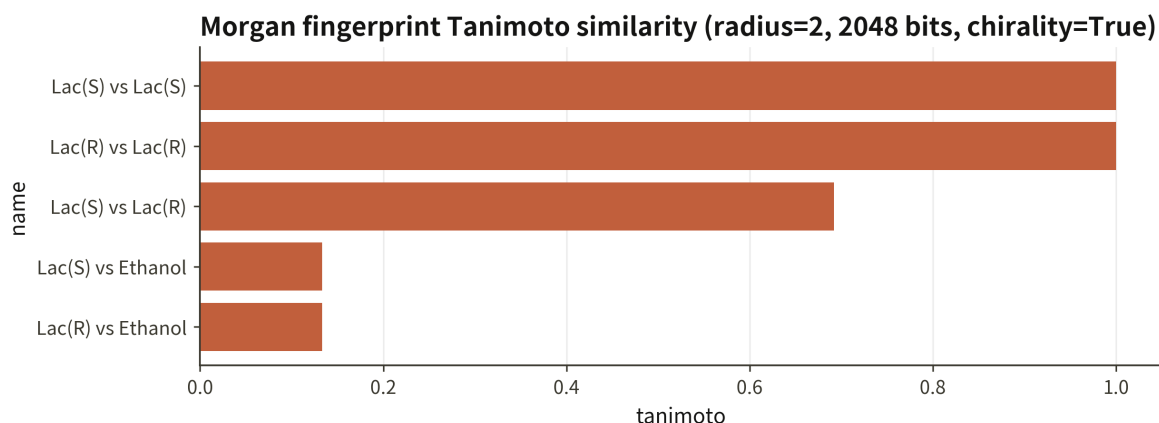


图 2: Tanimoto 相似度 (去重对, 共 5 个有效比较)。两个对映体之间 0.692 (小于 1, 证明手性确被编码); 对映体各自 vs 乙醇均为 0.133。

关键结论: 在手性编码的指纹下, 两个对映体相似度为 0.692 而非 1.0 ——手性被真实编码; 若 chirality=False, 对映体对会退化为 1.0。乙醇与两对映体均仅 0.133 (结构差异大)。

限制: 相似性不是药效、结合亲和力或任何生物效应的证据; 位指纹存在哈希碰撞; 本报告不以相似度推断药效。

5 三维构象生成与验证

来源: molecular-analysis__conformers 与独立解析验证。方法: ETKDGV3 生成 + MMFF94s 优化, max iterations = 500, 固定随机种子 seed = 42, requested count = 10 (两个对映体各 1 条收敛构象)。

5.1 SDF 可读性与三维性验证

对两个 conformers.sdf 做独立逐行解析 (不经 RDKit, 纯 Python/NumPy):

- 12 原子 / 11 键, SDF V2000 块解析成功;
- 坐标协方差三特征值均大于 0.37 Å², z 轴标准差 0.80-1.14 Å, z 跨度 2.60-4.01 Å ⇒ 非平面、确为三维 (非二维坐标冒充);
- 用四面体行列式法从 3D 坐标直接重算手性中心 CIP: 第一个对映体构象得 S、第二个得 R, 两构象行列式符号相反 (+8.136 / -8.140) ⇒ 立体化学在三维坐标中保留且互为对映。

Ball-and-stick view of MMFF94s-optimized 3D conformers
 (S)-lactic acid C[C@H](O)C(=O)O (R)-lactic acid C[C@@H](O)C(=O)O
 3D conformer (ETKDGv3+MMFF94s, seed=42) 3D conformer (ETKDGv3+MMFF94s, seed=42)
 CIP from 3D coords: S CIP from 3D coords: R

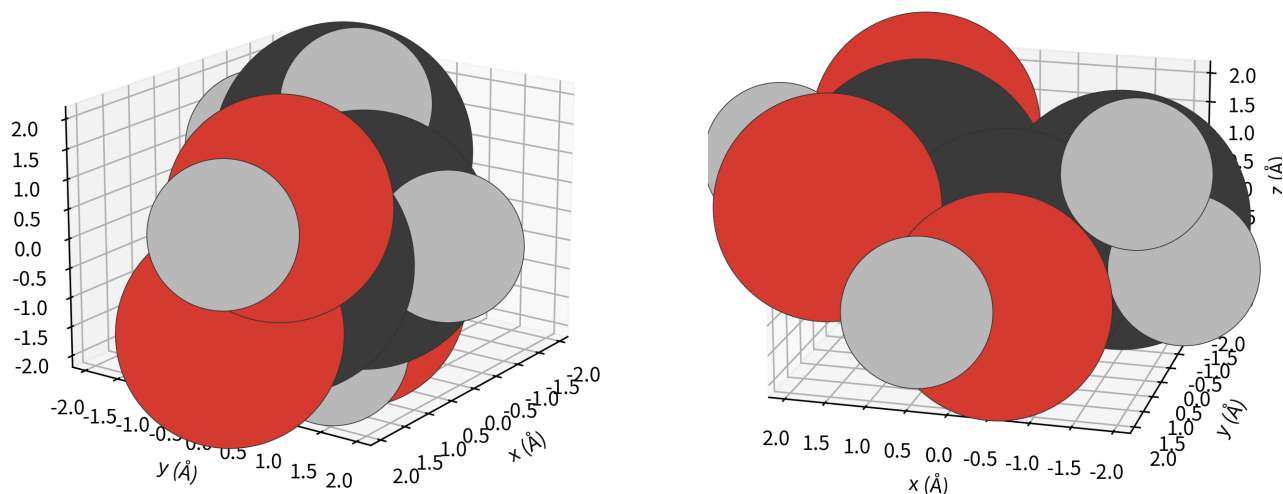


图 3: 由 SDF 实际三维坐标绘制的球棍图 (matplotlib, 深度着色)。左: 第一个对映体 (C@H), 从 3D 坐标重算 CIP = S; 右: 第二个对映体 (C@@H), CIP = R。构象能量: S 型 10.8035 kcal/mol, R 型 10.9099 kcal/mol (MMFF94s, 力场相对能量非绝对值)。

验证明细见工作区 out/sdf_validation.json; 构象产物为 out/molecular_conformers_f8b116155057/conformers.sdf (S 型, 能量 10.8035 kcal/mol) 与 out/molecular_conformers_cbd2b7b0f5c0/conformers.sdf (R 型, 能量 10.9099 kcal/mol)。

5.2 构象能量对比

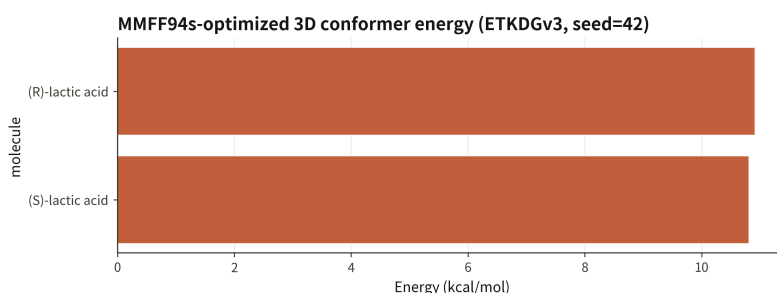


图 4: MMFF94s 收敛构象能量 (kcal/mol)。差值 0.106 kcal/mol 在力场精度内, 不构成构象偏好的统计证据。

6 汇总数据表

汇总 CSV (含名称、SMILES、性质、相似度、构象能量、3D 验证) 见 out/lactic_acid_ethanol_summary.csv (3 行 × 20 列)。表头与每行数值一一对应, SMILES 长字符串在 CSV 中完整保存 (未截断)。主要列摘录如下:

name	formula	MW	cLogP	TPSA	Tan.
(S)-lac	C3H6O3	90.078	-0.5482	57.53	1.0
(R)-lac	C3H6O3	90.078	-0.5482	57.53	0.692
ethanol	C2H6O	46.069	-0.0014	20.23	vs. ethanol 0.133

Tan. 列：上一行是 vs (S)-乳酸，第二行是 vs (S)-乳酸（对映体对），第三行单独注明 vs ethanol。完整相似度见第 4 节图与 CSV。

7 文件清单

- out/lactic_acid_ethanol_summary.csv ——汇总 CSV (名称/SMILES/性质/相似度/构象能量/3D 验证)
- out/molecular_conformers_f8b116155057/conformers.sdf ——(S)-乳酸三维构象 (seed=42, 收敛)
- out/molecular_conformers_cbd2b7b0f5c0/conformers.sdf ——(R)-乳酸三维构象 (seed=42, 收敛)
- out/sdf_validation.json ——SDF 可读性/三维性/CIP 重算验证明细
- out/similarity_tanimoto.png、out/conformer_energy.png、out/conformers_3d_structure.png/.svg、out/structures_2d_schematic.png/.svg ——各图
- out/structures_rdkit.svg ——RDKit 渲染的 2D 结构（性质工具自动产出）
- out/molecular_properties_3e0c13bdf19f/provenance.json 等——各工具的 provenance 记录

8 限制与不做的声明

- 相似度数值不用于推断药效、结合或生物效应（工具 limitation 亦如此声明）；
- MMFF94s 能量为力场应变能量，非结合自由能；
- 描述符不含 pKa/logD/ADMET（本轮明确未计算）；
- 2D 骨架示意图为手绘式表达，立体化学信息来自输入 SMILES，不代表任何优化坐标；优化三维坐标以 conformers.sdf 的真实坐标为准；
- RDKit 描述符层面无法区分对映体（标量描述符相同），区分只能依赖手性指纹（0.692）或三维坐标（CIP 重算）。