

丘成桐中学科学奖 · 化学奖赛道候选课题 20 则

本方案面向 2027 赛季（2026 赛季提交截止为 2026 年 9 月 15 日，从零启动已来不及）。倒排时间表以 2026 年 9 月为第 0 周、2027 年 9 月 15 日提交为终点，共约 52 周。

学生画像与资源假设：普通重点高中生，有一定编程与数学基础；计算资源为个人笔记本、纯 CPU、无 GPU 无集群；实验资源不超过中学常规实验室；全部使用公开数据；由中学教师即可指导。凡超出这一假设的路线，都在标签里标明了资源门槛。

四条必须先知道的赛制事实（取自官方公告，2020 – 2026 全部公示文章）：指导老师不得来自培训机构或任何谋利机构，违者取消资格；总决赛全程英文（研究报告、PPT、答辩）；入围总决赛的研究报告会在 11 月全网公示；提交材料含论文查重报告与实验视频。规则每年会改，报名开放后须重新核对当年《参赛规则》与本学科《参赛指南》。

这个赛道的评委在奖励什么（以及一个必须先说的坏消息）

先说实话：化学奖是六个赛道里对实验室资源依赖最重的一个。统计 2020 – 2025 共 63 项获奖论文，绝大多数需要中学实验室不具备的表征手段——SEM/TEM 形貌、XRD 物相、XPS 价态、电化学工作站、荧光光谱。金奖清一色是材料合成类：碳化硅微球增强导热复合材料、豆芽衍生过渡金属掺杂催化剂、环保一次性医用口罩、CRISPR 信号放大可视化检测、GFP 索烃的定向进化、海水光催化产氢自支撑膜。这些工作的实际执行地点几乎肯定是高校或科研院所实验室。

这个事实改变了课题设计空间，必须在选题时就正视：普通重点高中生若没有高校合作渠道，直接对标历届金奖题材是不现实的。本方案的应对是两条：

1. C01 – C10 走计算化学与化学信息学——完全绕开仪器依赖，只需一台笔记本。历届此类获奖有 MN4 石墨烯甲烷活化的 DFT 研究、BCN 单层 CO₂ 光催化的理论研究、MOF 的 CO₂ 分离计算筛选、分子对接筛选 PD-1 小分子。注意它们全部止步于优胜奖或铜奖——这一族的天花板确实偏低，但它是资源受限学生唯一能独立完成的路线，且做扎实了能稳定进入分赛区。提高上限的办法写在每条路线里：把课题定位成“评价维度的转换”或“方法学的可复现性贡献”，而不是又一次性质预测。
2. C11 – C20 走中学实验室可执行的实验——限定在比色/电化学/常规合成能覆盖的体系，总试剂预算 ≤ 3000 元。凡是某项表征确实必需的，路线里都写明“须联系高校共享平台送样，属外部依赖，列入 go/no-go 风险项”，并给出不依赖该表征也能成立的主结论。

近年的主题漂移值得注意：2023 年起，绿色化学、废弃物再利用（棉花衍生仿羽绒、牛毛生物炭、电石渣矿化 CO₂）、能源电催化（海水氧化、绿氢、超级电容器）明显上升。这类题材对试剂纯度要求低、原料易得，是低资源学生相对友好的切入点。

另一个高频陷阱：超级电容器、光催化降解染料、MOF 吸附这三类体系已经被中学生和本科生做到烂熟。做这些必须在新颖性边界里点名已有工作并给出真正的差异轴，否则评委三十秒就能问倒。

本赛道 20 条路线的分族逻辑

- C01 – C10 计算化学与化学信息学族：零仪器依赖，纯 CPU 笔记本，风险几乎全在"想法是否已被发表"。每条路线都写明了具体软件的算力边界（xtb 能处理多大体系、PySCF 在什么基组下多少重原子可行）。
- C11 – C20 中学实验室实验族：器材与试剂清单具体到估价，安全与废液处理写进方法路径，每条都有不依赖高端表征的主结论备份。

两族独立排序：C01/C11 最稳，C10/C20 风险与上限最高。

C01 · AqSolDB 水溶解度模型误差的来源分层归因：以测量可靠性等级与官能团亚群分解 MAE

推荐优先级 最高 · 分族 化学信息学 / 性质预测 · 资源需求 纯 CPU, 单核分钟级 · 技能取向 Python 数据处理 + 统计 · 失败风险 极低

1 · 研究问题

在公开水溶解度数据集 AqSolDB 上训练的描述符模型，其骨架划分（scaffold split）测试误差中有多大比例可归因于数据集自身的多来源测量不一致（由 AqSolDB 自带的 SD 列与来源分组标签度量），而非模型泛化能力不足？若按测量可靠性分层重新评估，不同建模方法之间的排名是否发生改变？

2 · 研究背景与空白

水溶解度（aqueous solubility，以 logS 表示）由晶格能、极性、氢键供受体数、疏水表面积共同决定，是定量结构-性质关系（QSPR，quantitative structure – property relationship）的经典试金石。Delaney 于 2004 年的 ESOL 方程仅用 logP、分子量、可旋转键数、芳香原子占比四个描述符，就把平均绝对误差压到约 0.75 log 单位，二十年来一直是“简单模型能做到多好”的参照点。

已有工作走到哪一步：Sorkun 等（Scientific Data, 2019, 6:143）把 9 个来源合并成 AqSolDB，共 9,982 条去重记录，并在数据集里显式保留了两个质量字段——SD（同一化合物在多来源中的测量标准差）与 Group 标签（按来源数量与一致性划分的可靠性分组，G1 – G5，具体判定规则须回原文核实）。后续基准工作（SolTranNet, JCI 2021；TDC 平台的 AqSolDB scaffold split 协议）报告的骨架划分 MAE 落在 0.74 – 1.0 log 单位——例如一项 2026 年的描述符+超参优化工作报 $MAE = 0.739 \pm 0.006$ 、 $R^2 = 0.802 \pm 0.003$ 。随机划分因同一骨架跨越训练/测试集而高估性能已被反复指出；2026 年更有工作（arXiv:2607.10729）指出骨架划分本身也会掩盖“结构前沿”失效。

空白在：几乎所有基准论文都把 AqSolDB 当作一块均质的标签来用，报的是全库单一 MAE，而没有人系统性地问“这 0.74 log 单位里有多少本来就是实验值之间的分歧”。AqSolDB 已经把这个分歧量化在 SD 列里，却几乎无人把它作为分层变量。这属于评价维度的转换：数据全公开、门槛低（RDKit + scikit-learn），且无论误差主要来自数据还是来自模型，结论都成立。

3 · 可检验假设

H1 在 AqSolDB 的高可靠性子集（多来源一致、 $SD \leq 0.5$ log 单位）上，骨架划分 MAE 比全库 MAE 低至少 0.15 log 单位；即全库基准值中至少 20% 的表观误差来自标签噪声而非模型。

H2（备择/机制假设）不同建模方法（岭回归、随机森林、梯度提升、Morgan 指纹 + SVM）在全库上的排名，与在高可靠性子集上的排名不一致：至少有一对方法的相对次序发生翻转，且翻转在自助抽样 95% 置信区间上可分辨。若 H2 被否定（排名完全稳定），则说明现有基准排名对标签噪声稳健，这本身是对社区有用的负结论。

4 · 量化验收标准

1. 方法学校验（硬门槛）：用自建管线在 Delaney 原始 ESOL 分子集（1,1xx 个分子，须以原文为准）上重建其四描述符线性方程，复现的平均绝对误差与文献报告值（约 0.75 log 单位）偏差 ≤ 0.10 log 单位；同时在

TDC 的 AqSolDB scaffold split 协议下复现随机森林基线，MAE 与已发表值偏差 ≤ 0.10 log 单位。此条不过关，后续全部分层结论无效，因为无法区分"分层效应"与"自己管线的 bug"。

- 2. 统计口径预先写死：主结果一律报 MAE 与 RMSE（单位 log 单位），划分方式为 Bemis – Murcko 骨架划分（RDKit MurckoScaffold），并额外做一次随机划分对照，明确报出随机划分相对骨架划分的乐观高估幅度（ Δ MAE）。所有点估计附 1,000 次自助重抽样的 95% 置信区间。不报 R^2 作为唯一指标。
- 3. 分层至少覆盖：可靠性分层（按 SD 分 3 档 + 单来源组）、官能团分层（用 RDKit SMARTS 定义 ≥ 8 类，如羧酸、磺酰胺、硝基、卤代芳环等，每类样本量 ≥ 100 才纳入统计）。每一层给出 MAE $\pm$ 95% CI 与样本量。
- 4. 至少 4 种建模方法 $\times$ 5 个随机种子，全部结果表格化；报告"方法排名在分层前后是否翻转"的具体次序表。
- 5. 全部脚本（数据下载、描述符生成、划分、训练、分层统计、作图）开源到 GitHub，附 environment.yml，第三方一条命令可重跑；随机种子写死。

5 · 数据与工具

用途	来源 / 工具
主数据集	AqSolDB, 9,982 条 logS 记录, 含 SD 与来源分组列。Harvard Dataverse (DOI 10.7910/DVN/OVHAW8) 直接下载 CSV, 约数 MB
方法学校验基准	Delaney ESOL 集 (1,1xx 分子, MoleculeNet delaney.csv 镜像); TDC (Therapeutics Data Commons) Solubility_AqSolDB 的官方 scaffold split。仅用于校验与对比，不计入本项目的数据贡献
描述符	RDKit (开源, pip/conda 装): 内置 200+ 二维描述符、Morgan 指纹、MurckoScaffold 骨架提取、SMARTS 子结构匹配。Mordred 可选 (约 1,800 描述符, 10k 分子约 20–40 分钟单核, 需核实本机实测)
建模与统计	scikit-learn (岭回归/随机森林/SVM)、LightGBM 或 XGBoost (CPU 版)、SciPy 自助抽样
算力	全程纯 CPU。RDKit 二维描述符生成 10k 分子约 1–3 分钟单核; 随机森林在 10k $\times$ 200 矩阵上训练秒级; 4 方法 $\times$ 5 种子 $\times$ 2 划分 $\times$ 若干分层的全量实验, 单机数小时内完成。无任何超出笔记本范围的步骤

6 · 方法路径

- 1. 装环境 (RDKit + scikit-learn)，下载 AqSolDB 与 ESOL，重建 Delaney 四描述符方程并跑通方法学校验硬门槛（第 1–4 周）。
- 2. 核实 AqSolDB 的 Group / SD 字段定义（回 Scientific Data 原文与数据字典），写出明确、可复现的可靠性分层判据，判据一旦写定不得事后调整。
- 3. 构建统一的描述符矩阵与骨架划分/随机划分两套索引，冻结为版本化文件。
- 4. 训练 4 种模型 $\times$ 5 种子，先在全库复现已发表基线数字（第二道校验）。
- 5. 按可靠性等级与官能团亚群做误差分解，逐层给 MAE $\pm$ 自助 95% CI，并画"误差 vs 标签 SD"的散点与分箱曲线。
- 6. 做方法排名稳定性分析：全库排名 vs 高可靠子集排名，用配对自助检验判断次序翻转是否可分辨。
- 7. 独立交叉校验：用一种完全不同的表征（Morgan 指纹 + 最近邻）重算同一套分层结论，确认结论不依赖描述符选择。

7 · 新颖性边界

- 本课题不声称提出新的溶解度预测方法，不声称刷新 AqSolDB 上的 SOTA，不声称"骨架划分优于随机划分"这一已被反复发表的结论为本项目发现。
- 已有工作具体完成了什么：Sorkun 等（2019）构建 AqSolDB 并提供质量字段；SolTranNet（2021）与 TDC 平台给出 scaffold split 下的标准基线；arXiv:2607.10729（2026）指出骨架划分仍会掩盖结构前沿失效。这些工作报告的都是聚合到全库的单一误差数字。
- 本项目的贡献是评价维度的转换，且这是主结论而非附加内容：把 AqSolDB 已有但未被使用的测量可靠性信息作为分层变量，给出"公开基准 MAE 中有多少是标签噪声地板"的定量估计，并检验现有方法排名对该噪声是否稳健。
- 为什么有价值：如果高可靠子集上的误差地板显著低于全库数字，说明社区一直在用一个被噪声抬高的基准衡量新方法的进步，后续方法比较应改用分层报告。
- "差异不显著"同样是有效结论：若分层后 MAE 与排名都不变，则说明 AqSolDB 的标签噪声对基准评估影响可忽略——但必须用自助置信区间证明本研究有能力分辨 $0.1 \log$ 单位量级的差异，否则"不显著"只是样本量不足。

8 · 决策门槛 (go / no-go)

- 第 4 周末：方法学校验硬门槛必须通过（ESOL 复现偏差 $\leq 0.10 \log$ 单位）。若失败，先排查描述符定义与单位；仍失败则降级为只用 MoleculeNet `delaney.csv` 做全流程（分子数少但标注干净），分层轴改为官能团单轴，主结论框架（误差分解 + 排名稳定性）完全保留。
 - 第 8 周末：核实 AqSolDB 高可靠子集（ $SD \leq 0.5$ ）的样本量。若不足 1,500 个分子，骨架划分后测试集过小、置信区间会宽到无法判定，立即降级：其一，把分层阈值放宽到 $SD \leq 1.0$ 并改报三档趋势而非二分对比；其二，改用"来源数量"（单来源 vs 多来源）作为可靠性代理变量。两条降级路径都保留"误差分解 + 排名稳定性"这一主结论框架。
 - 第 12 周末：若发现 AqSolDB 的 `Group` 字段定义在原文中不足以支持可靠性排序（这一点必须在第 8 周前查清，不要边做边发现），则完全弃用该字段，只用 `SD` 数值列。
 - 预算裁剪顺序：先砍 Mordred 大描述符集（改用 RDKit 内置 200 个），再砍模型种类（从 4 种减到 2 种），最后砍官能团分层类别数。砍到只剩"RDKit 描述符 + 2 模型 + `SD` 分层"时主结论仍成立。
 - 选择前提：适合需要一条几乎不可能全盘失败的路线的学生，也适合作为团队保底题与高风险路线并行。
-

C02 · GFN2-xTB 构象能排序在分子内氢键体系上的可靠性审计：以成对反序率与 Boltzmann 权重传播误差为判据

推荐优先级 高 · 分族 半经验量子化学 · 资源需求 纯 CPU, 单分子秒级 · 技能取向 命令行 + 结构文件处理 + 统计 · 失败风险 低

1 · 研究问题

在公开构象能基准集上，GFN2-xTB（Grimme 组的自洽电荷紧束缚半经验方法）相对于高精度参考值的误差，是否随体系中分子内氢键（intramolecular hydrogen bond, IMHB）数量的增加而系统性放大？若用成对反序率（pairwise inversion rate，即被错误排序的构象对占比）而非能量 MAE 作为判据，这种放大在多大的 IMHB 数量上开始使 Boltzmann 加权系综性质失真超过 10%？

2 · 研究背景与空白

分子在室温下是按玻尔兹曼分布加权的构象系综（conformer ensemble）。任何依赖构象的性质——溶解度、光谱、对接打分、NMR 位移——都要先枚举构象再按相对能量加权，精度瓶颈几乎总在构象相对能量而非几何。全 DFT 处理数十个构象在笔记本上不现实，故半经验方法 GFN2-xTB 成为默认工具：H 到 Rn 全参数化，含多极静电与密度依赖色散，单点亚秒级。

已有工作走到哪一步：Grimme 等在 GFN2-xTB 原始论文（JCTC 2019, 15, 1652）中已用构象集（含葡萄糖、麦芽糖等强氢键体系）对照 DLPNO-CCSD(T)/CBS 参考值，结论是 GFN2-xTB "在所有统计指标上优于同类方法"，并指出糖类因氢键网络众多而最难。GMTKN55 的 ACONF / SCONF / PCONF21 / MCONF 子集提供带高精度参考能的构象几何。这些工作报告的核心指标是相对能量的 MAE / RMSE 与相关系数。

空白在：能量 MAE 与"排序对不对"并不是一回事——0.3 kcal/mol 的 MAE 若集中落在能量间隔更小的一批构象对上，就足以把系综中最主要的几个构象顺序完全打乱；而下游应用真正消费的是 Boltzmann 权重，不是能量本身。目前公开文献很少把误差按"IMHB 数量"分层，也很少把误差传播到系综平均性质上给出可操作阈值。这属于评价维度的转换 + 参数维度分层：几何与参考能全公开、xtb 免费且笔记本秒级、结论正负都成立。

3 · 可检验假设

H1 构象相对能量的 MAE 随体系 IMHB 数量单调增大：IMHB ≥ 2 的分组 MAE 至少是 IMHB = 0 分组的 1.8 倍，差异在自助 95% 置信区间上可分辨。

H2 成对反序率与能量 MAE 不成正比：存在一组分子，其能量 MAE 处于全集中位数以下，但在 1 kcal/mol 以内的近简并构象对上反序率 $> 30\%$ ，导致 Boltzmann 加权的系综平均偶极矩相对参考值偏差 $> 10\%$ 。若 H2 被否认（反序率可由 MAE 单调预测），则说明现行以 MAE 为中心的报告方式已足够，这同样是有价值的负结论。

4 · 量化验收标准

1. 方法学校验（硬门槛）：用自建管线在 GMTKN55 的 ACONF（正构烷烃构象）与 SCONF（糖类构象）子集上重算 GFN2-xTB 相对能量，与 Grimme 组已发表的 GFN2-xTB 在这两个子集上的 MAD 数值对比，偏差 ≤ 0.15 kcal/mol（每个子集至少覆盖其全部标准构象）。此条不过关说明输入几何、单位换算或收敛设置有误，后续全部结论无效。

- 统计口径预先写死：能量误差报 MAE 与 RMSE (kcal/mol)，排序误差报成对反序率与 Kendall τ ；所有点估计给 1,000 次分子级自助重抽样的 95% 置信区间（自助单位是分子而非构象对，避免同一分子内构象对的相关性夸大精度）。分层比较用配对自助检验，不用 t 检验。
- IMHB 判据必须明确且可复现：给出 SMARTS + 几何双重判据（如 D-H...A 距离 $\leq 2.5 \text{ \AA}$ 且角度 $\geq 130^\circ$ ，具体数值在第 6 周前写定并冻结），并公布判据的敏感性分析（阈值 $\pm 0.2 \text{ \AA}$ 下结论是否稳定）。
- 样本量下限：纳入统计的分子 ≥ 40 个、构象 ≥ 400 个；每个 IMHB 分层组内分子数 ≥ 10 ，否则合并分层档位。
- 系综性质传播误差至少覆盖 2 个可观测量（如系综平均偶极矩、系综平均回转半径），报"参考权重 vs GFN2-xTB 权重"下的相对偏差分布。
- 全部输入结构、xtb 命令行、后处理脚本开源，一条命令可重跑。

5 · 数据与工具

用途	来源 / 工具
基准几何与参考能	GMTKN55 数据库 (Goerigk 等) 子集 ACONF / SCONF / PCONF21 / MCONF，含笛卡尔坐标与高精度参考相对能 (CCSD(T) 级)。官方网站可下载压缩包。仅用于校验与对比，不计入本项目的数据贡献
半经验计算	xtb (Grimme 组, LGPL, conda-forge 一条命令安装)。内置 GFN0/GFN1/GFN2-xTB 与 GFN-FF；内置几何优化、频率、GBSA/ALPB 隐式溶剂。 注意：xtb 不做 DFT，若需 DFT 参考须另用 PySCF/ORCA
构象生成 (扩展部分)	CREST (基于 xtb 的元动力学构象搜索) ——安装与可用性需核实；降级方案为 RDKit ETKDG + xtb 优化去重
后处理	Python + ASE (读写结构、批量驱动 xtb)、NumPy/SciPy、RDKit (SMARTS 匹配 IMHB)
算力	纯 CPU。GFN2-xTB 单点：30-60 原子分子在单核上 亚秒级 ；几何优化 10-60 秒 ；数百原子体系仍在分钟级。400 个构象的全量优化 + 单点，单机数小时。CREST 元动力学搜索对 30 原子分子约 10-60 分钟/分子 (需本机实测)，是唯一可能超预算的步骤，故列为可选扩展

6 · 方法路径

- 装 xtb 与 ASE，跑通 xtb 自带算例，下载 GMTKN55 子集，完成 ACONF/SCONF 的方法学校验硬门槛（第 1-5 周）。
- 核实工具能力边界：确认 xtb 的相对能量是否需统一到同一收敛阈值（`--opt tight` vs 默认）、是否需要 `--gfn 2` 显式指定、单位是否为 Hartree；确认 CREST 在本机可安装并测得单分子墙钟时间。不确定项在第 8 周前查清。
- 写定 IMHB 判据 (SMARTS + 几何双阈值) 并冻结，对全部构象打标签，公布判据代码。
- 批量跑 GFN2-xTB 单点与优化，构建"参考相对能 vs GFN2 相对能"配对表；按 IMHB 数量分层统计 MAE、反序率、Kendall τ 。
- 把两套相对能分别转为 298 K Boltzmann 权重，计算系综平均偶极矩与回转半径，报相对偏差分布与"权重失真开始超过 10%"的 IMHB 阈值。
- 做判据敏感性分析 (IMHB 几何阈值 $\pm 0.2 \text{ \AA}$ / $\pm 10^\circ$) 与方法敏感性分析 (GFN1-xTB、GFN-FF 各重跑一遍作为对照)。

7. 独立交叉校验：对随机抽取的 10 个分子，用 PySCF 在 B3LYP-D3/6-31G* 水平重算相对能（每分子小时级，仅作抽样校验），确认分层结论方向一致。

7 · 新颖性边界

- 本课题不声称改进 GFN2-xTB，不声称发现该方法的新失效机制，不把"GFN2-xTB 对氢键体系更困难"这一 Grimme 原文已明确指出的定性结论声称为本项目发现。
- 已有工作具体完成了什么：Grimme 等 (JCTC 2019) 在多个构象集上以 DLPNO-CCSD(T)/CBS 为参考评测 GFN2-xTB，报告聚合的 MAD 并定性指出糖类氢键体系最难；GMTKN55 (Goerigk 等 2017) 提供标准化的构象子集与参考能，社区惯例是报子集级 MAD。
- 本项目的贡献是评价维度的转换且构成主结论：把判据从"能量 MAE"换成"成对反序率 + Boltzmann 权重传播到系综可观测量的失真幅度"，并按可复现的 IMHB 计数判据分层，给出一条可操作阈值——"IMHB 数量超过 N 时，GFN2-xTB 权重不再足以支撑系综平均性质的 10% 精度要求"。
- 为什么有价值：下游使用者（对接、光谱、溶剂化计算）真正需要知道的是"什么时候不能只用 xTB"，而现有文献给的是聚合 MAD，无法直接回答这个问题。
- "差异不显著"同样是有效结论：若分层后 MAE 与反序率均无系统趋势，说明 GFN2-xTB 对氢键体系的额外风险被高估了——但必须给出误差棒证明本研究能分辨 1.8 倍的 MAE 比值。

8 · 决策门槛 (go / no-go)

- 第 5 周末：ACONF/SCONF 复现偏差必须 ≤ 0.15 kcal/mol。若失败，先核对参考能的定义（相对最低能构象 vs 相对首个构象）与几何是否被重新优化；仍失败则降为"只做单一子集 SCONF 的深度分析"，样本量下降但反序率 + 权重传播的主结论框架完全保留。
 - 第 10 周末：确认可用构象总数与 IMHB 分层后每组分子数。若 $IMHB \geq 2$ 组的分子数 < 10 ，立即降级：其一，纳入 PCONF21（多肽构象，氢键丰富）与 MCONF（黑色素前体）扩充样本；其二，把分层轴从"IMHB 计数"改为"氢键供体+受体总数"这一连续变量，改做回归而非分组对比。两条路径都保留主结论框架。
 - 第 16 周末：若 CREST 在本机的单分子墙钟时间 > 2 小时，放弃自建构象系综，全部转用 GMTKN55 提供的既有几何。这不影响主结论（主结论建立在既有基准几何上），只是失去"自建系综"的扩展章节。
 - 已知困难即研究内容：IMHB 的定义本身没有唯一标准，不同阈值会给出不同计数。这不是障碍——判据敏感性分析本身就是本项目的一部分交付物，且是评委容易认可的严谨性证据。
 - 预算裁剪顺序：先砍 CREST 构象搜索扩展，再砍 GFN1/GFN-FF 对照，再砍 PySCF 抽样交叉校验（改为引用文献值），最后砍第二个系综可观测量。砍到只剩"GMTKN55 两子集 + 反序率 + 单一系综性质"时主结论仍成立。
-

C03 · ChEMBL 单靶点活性分类的时间外推衰减：以发表年份切分的独立检验与 PR-AUC 口径下的衰减-靶点属性关系

推荐优先级 高 · 分族 化学信息学 / 数据库挖掘 · 资源需求 纯 CPU + 约 50 GB 硬盘 · 技能取向 SQL + Python + 不平衡分类统计 · 失败风险 低-中

1 · 研究问题

在 ChEMBL 数据库中，把训练集与测试集按化合物首次发表年份切分（时间划分，time split）后，单靶点活性二分类模型的 PR-AUC 相对骨架划分与随机划分分别下降多少？这一下降幅度是否可由靶点的可观测属性（该靶点化学空间的年度扩张速率、活性化合物骨架多样性、正样本比例）预测？

2 · 研究背景与空白

QSAR (quantitative structure – activity relationship) 模型在真实药物发现中的用法永远是“用过去预测未来”，而学术基准常用的随机划分 (random split) 把同一化学系列同时放进训练与测试集，系统性高估前瞻性能。Sheridan (JCIM 2013) 最早论证时间划分交叉验证 (time-split cross-validation) 比随机划分与留类法更接近真实前瞻预测的 R^2 。

已有工作走到哪一步：Lenselink 等 (J. Cheminform. 2017) 用 ChEMBL 19 训练、ChEMBL 20 新增活性做时间验证，覆盖数百靶点；后续工作用 ChEMBL 23→24 做同类前瞻验证。Landrum 与 Riniker 的 SIMPD 算法 (J. Cheminform. 2023) 从 ChEMBL 生成 99 个带时间特征的公开数据集。这些工作都确认了时间划分下性能会掉。

空白在：现有工作报告的是“时间划分下性能更低”的平均效应，而没有把“衰减幅度”本身当作因变量去解释——即“哪些靶点的模型会掉得多，哪些几乎不掉，能否事先判断”。同时，前述工作的数据截止在 ChEMBL 19 – 24 (2014 – 2018 年前后)，2023 年之后新增的活性数据构成了一个完全独立的新时间切片。本课题同时占据“换样本”与“换问题方向”两条差异化轴。

3 · 可检验假设

H1 在 2023-01-01 之后首次出现的化合物构成的时间外测试集上，单靶点分类模型的 PR-AUC 中位数比骨架划分低 ≥ 0.05 (绝对值)，比随机划分低 ≥ 0.10 ；差异在靶点级自助 95% 置信区间上可分辨。

H2 衰减幅度 ($\Delta\text{PR-AUC} = \text{骨架划分 PR-AUC} - \text{时间划分 PR-AUC}$) 与“测试期新骨架占比”呈正相关，Spearman $\rho \geq 0.4$ (自助 95% CI 下界 > 0)。若 H2 被否定，说明衰减不可由化学空间扩张速率预测，需另寻解释变量——这同样是有效结论，且直接反驳了一个直觉上很强的假设。

4 · 量化验收标准

1. 方法学校验 (硬门槛)：选取 Lenselink 等 (2017) 或 SIMPD (2023) 公开数据集中至少 3 个靶点，用自建管线 (相同描述符、相同模型类别、相同划分) 复现其报告的性能指标，偏差 ≤ 0.05 (以其报告的主指标绝对值计)。此条不过关意味着数据抽取或建模流程与社区惯例不一致，后续跨靶点比较全部无效。
2. 统计口径预先写死：活性分类天然极不平衡 (正样本常 $< 10\%$)，主指标一律为 PR-AUC，同时报召回率与精确率，明确不报 accuracy (在 5% 正例率下全预测为负即可得 95% accuracy，该指标无信息量)。划分方式三套并行：随机划分、Bemis – Murcko 骨架划分、按化合物首次发表年份的时间划分，三者必须在同一批靶点、同一模型、同一测试集规模下比较。所有点估计给靶点级 1,000 次自助 95% 置信区间；相关性分析 (H2) 给自助置信区间而非仅 p 值。

- 靶点纳入判据写死并冻结：每靶点在时间切点前需 ≥ 300 条带标准活性值 (pChEMBL) 的记录、切点后 ≥ 100 条；活性阈值统一取 pChEMBL ≥ 6 (并做 ≥ 5 与 ≥ 7 的敏感性分析)；只保留 `assay_type = 'B'` (结合实验) 与单蛋白靶点。纳入靶点数 ≥ 30 。
- 报告"时间划分测试集中的新骨架占比"分布，并给出 Δ PR-AUC 对该变量的回归斜率与自助置信区间。
- 全部 SQL 查询、抽取脚本、划分索引、模型代码开源；数据抽取结果以中间 CSV 形式随代码发布 (受 ChEMBL 许可允许范围内)，第三方可跳过数据库直接重跑分析。

5 · 数据与工具

用途	来源 / 工具
主数据	ChEMBL 全库 SQLite dump, EMBL-EBI FTP (ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/latest/)。解压后约 20–40 GB (版本相关, 需核实当年版本实际体积)。年份信息取自 <code>docs.year</code> (文献发表年) 关联 <code>activities - assays - docs</code>
轻量替代 (降级)	Papyrus 数据集 (标准化的 ChEMBL+ExCAPE 生物活性集, 含时间信息, 体积远小于全库, 需核实其年份字段完整性); 或 ChEMBL 网页按靶点导出 CSV (单靶点几 MB)
方法学校验基准	Lenselink 等 (2017) 公开的 ChEMBL 19/20 基准集; SIMPD (2023) 发布的 99 个数据集 (GitHub, rdkit 组织下)。仅用于校验与对比, 不计入本项目的数据贡献
描述符与模型	RDKit (Morgan 指纹 $r=2$, 2048 位 + 二维描述符)、scikit-learn (随机森林、逻辑回归)、LightGBM (CPU)
算力	纯 CPU。SQLite 查询在 SSD 上单靶点秒级、全库聚合查询分钟级; 30 靶点 $\times$ 3 划分 $\times$ 2 模型 $\times$ 5 种子的训练总计单机数小时。瓶颈是硬盘空间与首次下载带宽, 不是 CPU

6 · 方法路径

- 装 RDKit + SQLite, 先用 SIMPD 或 Lenselink 公开集跑通全流程并通过方法学校验硬门槛 (第 1–6 周), 此阶段不需要下载全库。
- 下载 ChEMBL SQLite dump, 写并冻结靶点纳入 SQL (活性类型、单位统一、pChEMBL 阈值、重复记录合并规则), 把抽取结果导出为版本化 CSV。
- 核实关键字段: `docs.year` 的缺失率、同一化合物在多篇文献中出现时"首次年份"的取法、`pchembl_value` 的可用比例。这三项须在第 14 周前查清, 不要边做边发现。
- 构建三套划分索引 (随机 / 骨架 / 时间), 强制三者测试集规模一致 (时间划分决定规模, 另两套按同规模抽样), 否则 PR-AUC 不可比。
- 训练与评估, 逐靶点记录三套划分下的 PR-AUC、召回、精确率与正例率; 汇总为靶点级数据表。
- 计算每个靶点的解释变量 (测试期新骨架占比、年度化合物增长率、正例率、训练集规模), 做 Δ PR-AUC 的相关与多元回归, 全部给自助置信区间。
- 独立交叉校验: 把活性阈值改为 pChEMBL ≥ 5 与 ≥ 7 重跑, 确认衰减幅度与相关性结论方向不变。

7 · 新颖性边界

- 本课题不声称提出新的划分方法或新模型, 不声称发现"时间划分比随机划分更严格" (Sheridan 2013 已确立), 不把 ChEMBL 数据库本身计入学生的数据贡献。

- 已有工作具体完成了什么：Sheridan (2013) 确立时间划分的必要性；Lenselink 等 (2017) 在 ChEMBL 19→20 上做大规模时间验证并比较模型族；Landrum 与 Riniker (2023, SIMPD) 提供模拟时间划分的公开数据集生成算法与 99 个数据集。它们共同的输出形式是跨靶点平均的性能对比。
- 本项目的贡献有两条且都是主结论：其一，换样本——在 2023 年之后的全新时代切片上做一次独立检验，回答"十年前得到的衰减幅度在今天的 ChEMBL 上是否仍然成立"；其二，换问题方向——把衰减幅度本身作为因变量，检验它能否由靶点的可观测属性预测，从而给出"哪些靶点的公开基准数字最不可信"的操作性判据。
- 为什么有价值：如果衰减可预测，使用者在建模前就能估计自己的基准数字有多虚；如果不可预测，则说明必须逐靶点做时间验证，不能靠经验外推。两种结果都改变实践。

8 · 决策门槛 (go / no-go)

- 第 6 周末：方法学校验硬门槛必须通过。若无法复现已发表基准（偏差 > 0.05 ），先不下载全库，排查活性单位换算与去重规则；仍失败则降级为完全基于 SIMPD 的 99 个公开数据集做全部分析（这些数据集自带模拟时间划分），H1 改为"模拟时间划分 vs 骨架划分"的对比，H2 完全保留。
 - 第 14 周末：核实符合纳入判据的靶点数量。若 < 30 个，立即降级：其一，把切点从 2023 提前到 2020，换取更多切点后数据；其二，放宽到 `assay_type` 包含功能实验 (F)，并把实验类型作为协变量纳入回归。两条路径都保留"三划分对比 + 衰减可解释性"的主结论框架。
 - 硬盘风险：ChEMBL 全库解压需数十 GB。若本机空间不足，走上述 Papyrus / 按靶点导出 CSV 的轻量路径——这一项须在第 10 周前确认，不要等到下载失败。
 - 预算裁剪顺序：先砍模型种类（保留随机森林一种），再砍活性阈值敏感性分析，再砍多元回归（只保留单变量 Spearman + 自助 CI），最后把靶点数从 30 降到 15。砍到只剩"15 靶点 $\times$ 3 划分 $\times$ 随机森林 + 单变量相关"时主结论仍成立。
 - 选择前提：适合已经会写 SQL、且能接受"H2 很可能被否定"的学生。若学生不能接受负结论，不要选这条。
-

C04 · Materials Project 钙钛矿带隙模型的化学族外推检验：以留一化学族划分替代随机划分量化可外推边界

推荐优先级 高 · 分族 材料信息学 / 数据库挖掘 · 资源需求 纯 CPU, 分钟级; 需免费 API key · 技能取向 Python + 材料化学基础 + 统计 · 失败风险 中

1 · 研究问题

在 Materials Project (MP) 数据库的 ABX_3 / $A_2BB'X_6$ 钙钛矿子集上, 用组分描述符训练的带隙 (band gap) 回归模型, 在留一化学族划分 (leave-one-chemistry-out, 如留出全部卤化物、或留出全部含 Bi 的 B 位体系) 下的 R^2 与 MAE, 相对文献惯用的随机划分下降多少? 下降幅度最大的化学族有什么共同的结构或电子特征?

2 · 研究背景与空白

钙钛矿 (perovskite) ABX_3 因光伏与发光应用成为材料信息学最热的靶场, 带隙是核心筛选量。经典几何判据是 Goldschmidt 容忍因子 t ($t = 1$ 对应理想立方密堆, 偏离引入畸变)。近年工作把容忍因子、离子半径、电负性、氧化态等组分描述符喂给 CatBoost / 随机森林 / 图网络, 报告 ABX_3 带隙回归 $R^2 \geq 0.88$, 并有工作在 MP 的 2.6 万条材料上训练带隙分类器。

已有工作走到哪一步: ABX_3 与 $A_2BB'X_6$ 带隙的 ML 建模已经被反复做过, 公开报告的 R^2 集中在 0.85 – 0.92 区间, 特征重要性分析普遍指向 X 位阴离子电负性与 B 位阳离子性质。这些报告几乎全部基于随机划分或 K 折交叉验证。

空白在: 材料筛选的真实用途是外推到训练集里没有的化学族 (比如从氧化物外推到卤化物、从含 Pb 外推到无 Pb), 而随机划分下同一化学族的成员同时出现在训练与测试集中, 测得的 $R^2 \geq 0.88$ 对这个用途没有参考价值。留一化学族划分在分子性质领域 (scaffold split) 已是标配, 在钙钛矿带隙上却少见系统报告。这属于评价维度的转换: 数据全公开、算力极低、正负结论都有价值。另需说明: MP 的带隙来自 PBE 泛函计算, 系统性低估实验值约 30 – 50%, 因此本课题的对象是"复现 MP 的 PBE 带隙标签", 不得声称预测实验带隙。

3 · 可检验假设

H1 在留一化学族划分 (按 X 位阴离子分族: O / F / Cl / Br / I / S 等) 下, 带隙回归的 MAE 至少是随机划分下的 2 倍, R^2 下降 ≥ 0.25 (绝对值), 差异在自助 95% 置信区间上可分辨。

H2 外推衰减幅度与"留出族与训练集在描述符空间的距离" (用留出族质心到训练集质心的马氏距离度量) 正相关, Spearman $\rho \geq 0.5$ 。若 H2 成立, 则该距离可作为一个事先可算的"外推可信度"指标; 若被否定, 说明衰减不可由简单的描述符空间距离预警。

4 · 量化验收标准

- 方法学校验 (硬门槛): 用自建管线在与已发表工作相同的 ABX_3 数据切片、相同描述符族、随机划分下复现文献报告的带隙回归性能, R^2 偏差 ≤ 0.05 、MAE 偏差 ≤ 0.10 eV (至少对照 2 篇独立文献)。此条不过关说明数据抽取或描述符实现与文献不一致, 后续外推结论全部无效。
- 统计口径预先写死: 主指标为 MAE (eV) 与 R^2 , 划分方式为留一化学族划分, 并额外做一次随机划分对照, 显式报出随机划分相对留一族划分的乐观高估幅度。所有点估计给 1,000 次自助 95% 置信区间; 因每

个"族"只有一个留出结果，自助单位取族内材料而非族本身，并同时报告族级留一结果的全分布（不做跨族平均后就下结论）。

- 3. 数据纳入判据写死并冻结：只取 MP 中 `energy_above_hull ≤ 0.05 eV/atom`（保证热力学近稳定）、化学式匹配 ABX_3 或 $A_2BB'X_6$ 、带隙 > 0 的条目；记录 MP 数据库版本号与查询日期。样本量 $\geq 1,500$ 条，每个化学族 ≥ 80 条方纳入留一循环。
- 4. 至少 3 类描述符集对照：纯组分统计（matminer `ElementProperty` Magpie）、加入容忍因子/八面体因子的物理描述符、二者并集。二者在两种划分下的表现都要报。
- 5. 明确报告"MP 带隙为 PBE 计算值，系统性低估实验值"的说明，并在结论中不做实验带隙外推声明。
- 6. 全部查询代码（含 MP API 查询语句与版本号）、描述符生成、划分定义、模型与统计脚本开源，一条命令可重跑。

5 · 数据与工具

用途	来源 / 工具
主数据	Materials Project, 通过 <code>mp-api</code> Python 客户端查询（需在 materialsproject.org 注册免费 API key）。字段： <code>band_gap</code> 、 <code>formula_pretty</code> 、 <code>energy_above_hull</code> 、 <code>symmetry</code> 、 <code>structure</code> 。查询结果几 MB—几十 MB
结构与描述符	<code>pymatgen</code> （结构解析、组成解析、离子半径查表）、 <code>matminer</code> （ <code>ElementProperty</code> Magpie 描述符集、 <code>composition</code> 特征器）。容忍因子与八面体因子须自行实现（ <code>pymatgen</code> 是否内置需核实），实现后用手算的 5 个已知体系（如 $SrTiO_3$ 、 $CsPbI_3$ 、 $BaZrO_3$ 等）验证
对照文献值	已发表 ABX_3 / 双钙钛矿带隙 ML 论文报告的 R^2 /MAE（如 Commun. Mater. 2023 双钙钛矿带隙工作等）。 仅用于校验与对比，不计入本项目的数据贡献
建模与统计	<code>scikit-learn</code> （随机森林、梯度提升）、 <code>CatBoost</code> 或 <code>LightGBM</code> （CPU 版）、 <code>SciPy</code> 自助
算力	纯 CPU。API 查询数分钟；Magpie 描述符对 2,000 条材料约 1—3 分钟；单次模型训练秒级；3 描述符集 $\times$ 2 划分 $\times$ 约 8 个留一族 $\times$ 5 种子的全量实验单机 1—2 小时。 无任何 DFT 计算，故完全绕开算力墙

6 · 方法路径

- 1. 注册 MP API key，装 `pymatgen/matminer`，跑通官方示例查询，锁定数据库版本号（第 1—5 周）。
- 2. 抽取钙钛矿子集并冻结为版本化 CSV；自行实现容忍因子与八面体因子并用 5 个手算体系验证实现正确。
- 3. 在随机划分下复现已发表性能，通过方法学校验硬门槛。
- 4. 定义化学族划分判据（按 X 位阴离子为主轴，副轴按 B 位主族/过渡金属），判据须明确、可复现且事先冻结；对每个族做一次留出实验。
- 5. 汇总留一族结果，报每族的 MAE/ R^2 与自助 CI，并与随机划分对照，画"随机 vs 留一族"的成对对比图。
- 6. 计算每个留出族到训练集的描述符空间马氏距离，检验 H_2 ；给 Spearman ρ 的自助置信区间。
- 7. 独立交叉校验：换一种完全不同的模型族（如核岭回归）与一种完全不同的描述符（结构基的 `SineCoulombMatrix` 或 `OrbitalFieldMatrix`）重算，确认外推衰减结论不依赖于模型或描述符选择。

7 · 新颖性边界

- 本课题不声称提出新的带隙预测模型，不声称达到更高的随机划分精度，不声称预测实验带隙（MP 标签为 PBE 计算值），不把 Materials Project 的数据计入学生的数据贡献。
- 已有工作具体完成了什么：多篇 2023 – 2025 年论文在 ABX_3 与 $A_2BB'_2X_6$ 上用组分描述符 + 树模型/图网络做带隙回归与分类，报告随机划分或 K 折下 $R^2 \approx 0.88 - 0.92$ ，并给出以 X 位电负性、容忍因子为主的特征重要性排序；亦有工作在 MP 2.6 万条材料上训练带隙二分类器。
- 本项目的贡献是评价维度的转换且为主结论：把评估协议从随机划分换成留一化学族划分，量化"文献报告的 R^2 有多少来自同族泄漏"，并检验描述符空间距离能否作为事先可算的外推可信度预警指标。
- 为什么有价值：材料筛选的价值全在外推。若留一族划分下 R^2 从 0.90 掉到 0.5 以下，说明当前公开报告的精度对筛选新化学族毫无指导意义，这是一个直接影响使用方式的结论。
- "差异不显著"同样是有效结论：若外推衰减很小，说明组分描述符确实捕捉到了跨族可迁移的物理，这对材料信息学是正面且重要的证据——但必须用自助置信区间证明本研究有能力分辨 0.25 的 R^2 变化。

8 · 决策门槛 (go / no-go)

- 第 5 周末：MP API key 与 `mp-api` 客户端必须跑通并成功导出钙钛矿子集。若 API 访问受限或字段变更（这一项须在第 5 周前查清，MP 的 API 在历史上有过不兼容升级），降级至使用 MP 的静态数据快照（如 `matminer` 内置的 `matbench` 相关数据集或已归档的 MP dump），主结论框架不变。
 - 第 10 周末：方法学校验硬门槛必须通过（随机划分 R^2 偏差 ≤ 0.05 ）。若两篇文献都复现不了，很可能是数据切片定义不同——降级为改用 `matbench` 中的标准化带隙任务（`matbench_expt_gap` 或 `matbench_mp_gap`，定义公开且有官方 leaderboard 可校验），把研究对象从钙钛矿扩大到全无机材料，留一族划分改为"留一化学体系"。主结论框架完全保留。
 - 第 14 周末：确认每个化学族的样本量。若某些族 < 80 条导致可用留一循环少于 4 个，立即降级：其一，合并稀有族（卤化物合并为一族）；其二，把划分轴从 X 位改为 B 位元素周期族。
 - 已知困难即研究内容："化学族"的划分本身没有唯一正确答案。把划分判据的敏感性分析（X 位划分 vs B 位划分 vs 聚类划分）写成一节，本身就是交付物。
 - 预算裁剪顺序：先砍第三个描述符集，再砍结构基描述符的交叉校验，再砍 H_2 的马氏距离分析，最后把模型从 3 种减到 1 种。砍到只剩"Magpie 描述符 + 随机森林 + 两种划分对比"时主结论仍成立。
-

C05 · 笔记本墙钟预算下反应能垒计算的误差-成本前沿：泛函×基组网格在 BH76 上的系统偏置可校正性判定

推荐优先级 中高 · 分族 从头算量子化学 · 资源需求 纯 CPU，单点秒级到分钟级 · 技能取向 量子化学基础 + Python 批量作业管理 · 失败风险 中

1 · 研究问题

在标准反应能垒基准集 BH76 的固定几何上，密度泛函（DFT）方法的"泛函 × 基组"组合构成的误差-墙钟成本前沿在纯 CPU 笔记本上落在哪里？对每个组合，其误差是系统性的（可由一个加性常数或线性标度校正掉大部分）还是随机的（校正后 MAE 几乎不降）？

2 · 研究背景与空白

反应能垒（barrier height）决定反应速率，是计算化学中最难算准也最有用的量之一；化学精度（chemical accuracy）通常定义为 1 kcal/mol。BH76 是 GMTKN55 中的能垒子集，含 76 个正逆向能垒（氢转移、重原子转移、亲核取代等），提供固定参考几何与高精度参考能垒——这一点至关重要：无需过渡态搜索或频率计算，只做单点能量，因此完全落在笔记本算力范围内。

已有工作走到哪一步：Goerigk 与 Grimme 等（PCCP 2017）在 GMTKN55 上评测 200 余个泛函并公布 BH76 上的 MAD，结论是纯 GGA 严重低估能垒（常达 8–10 kcal/mol），杂化泛函随 HF 交换比例改善，双杂化最优。这些评测统一在 def2-QZVP 量级大基组上进行，运行于工作站/集群。

空白在：小型课题组用的是 6-31G 这一级，而在这一级上"哪个泛函最准"的排名与大基组结论未必一致，且没有人以固定墙钟预算为约束报告过误差-成本前沿。更关键的是，现有评测报告 MAD，却没有系统回答"误差有多少是可由一个常数吸收的系统偏置"——若 GGA 的 -9 kcal/mol 误差高度一致，廉价方法加一个经验常数即可达到杂化泛函的实用精度。这属于 换问题方向 + 换评价维度*。

3 · 可检验假设

H1 在 BH76 上，纯 GGA 泛函（如 PBE、BLYP）在 6-31G* 下的平均带符号误差（mean signed error, MSE）为负且量级 ≥ 6 kcal/mol；扣除该常数偏置后，其 MAE 下降 $\geq 50\%$ ，即误差主要是系统性的。

H2 在等墙钟时间口径下（同一时间预算内可完成的组合），基组升级（6-31G* $\rightarrow$ def2-TZVP）对能垒 MAE 的改善幅度小于泛函升级（GGA $\rightarrow$ 杂化）；即在笔记本预算下，把算力花在泛函上比花在基组上更划算。若 H2 被否定，同样给出明确的、可操作的资源分配建议。

4 · 量化验收标准

- 方法学校验（硬门槛）：用自建 PySCF 管线在 BH76 的全部 76 个能垒上重算至少 2 个已被 GMTKN55 评测过的泛函（如 PBE 与 B3LYP）在同一基组下的结果，与 Goerigk 等公布的 MAD 值对比，偏差 ≤ 1.0 kcal/mol（若基组不同，须先在文献所用基组上做一次对照）。此条不过关意味着几何读取、参考能定义（正/逆向、单位 kcal/mol vs Hartree）或 SCF 设置有误，后续全部误差-成本结论无效。
- 统计口径预先写死：误差指标同时报 MAE、RMSE 与 MSE（全部 kcal/mol），并报"扣除 MSE 后的残差 MAE"作为系统偏置可校正性的量化指标。所有点估计给 1,000 次自助（自助单位为反应）95% 置信区间。成本口径采用双口径：等自由度（同基组下比泛函）与等墙钟时间（同时间预算下比一切组合），两种都要报，且墙钟时间必须在同一台机器、同一线程数、连续测量三次取中位数。

3. 网格规模下限： ≥ 6 个泛函（覆盖 GGA / meta-GGA / 杂化 / range-separated 四类） $\times \geq 3$ 个基组（6-31G*、def2-SVP、def2-TZVP），共 ≥ 18 个组合 $\times 76$ 个能垒。
4. 收敛性检查：报告每个组合的 SCF 收敛失败率与 DFT 积分格点敏感性（至少 2 个格点等级对照，能垒变化 ≤ 0.2 kcal/mol 方认为格点收敛）。任何未收敛的算例必须显式列出，不得静默丢弃。
5. 交付一张"MAE vs 单点墙钟时间"的前沿图，并在图上标出 1 kcal/mol 化学精度线与本机 1 分钟/10 分钟预算线。
6. 全部输入几何、批量脚本、原始能量输出、计时日志与统计脚本开源，一条命令可重跑。

5 · 数据与工具

用途	来源 / 工具
基准几何与参考能垒	GMTKN55 的 BH76 子集（Goerigk/Grimme 组官网提供打包下载，含笛卡尔坐标与参考能）。仅用于校验与对比，不计入本项目的数据贡献
主计算引擎	PySCF (Apache-2.0, pip 可装, 纯 Python + NumPy/BLAS 后端)。内置：HF、DFT（大量 LibXC 泛函）、密度拟合、几何优化（需配合 geomeTRIC）、解析梯度。 注意：不同泛函的可用性依赖 LibXC 版本，双杂化泛函与部分 range-separated 泛函是否内置须核实；MP2/CCSD 内置但对本课题非必需
可选对照引擎	ORCA。官方声明对学术与个人用途免费，但需注册并接受 EULA，且条款面向"学术机构"的研究/教学用途——中学生个人是否符合资格须在第 3 周前核实；不符合则完全弃用，PySCF 已足够
结构与驱动	ASE（读写坐标、批量驱动）、geomeTRIC（若做几何优化扩展）、Python 标准库计时
算力	纯 CPU。参考标尺：PySCF 官方基准中苯（6 重原子，6-31G 约 102 个基函数）的 B3LYP 单点约 4 秒。按混合 DFT 约 N^3-N^4 的经验标度外推到本课题： BH76 体系多为 ≤ 6 重原子，6-31G* 单点秒级，def2-TZVP 单点数十秒；18 组合 $\times 76$ 能垒（每个能垒需 3 个单点：反应物、过渡态、产物）$\approx 4,000$ 次单点，单机数小时至一两天，可后台分批跑。超出范围的方案（明确列出）**：过渡态搜索 + 解析频率（10 重原子以上的 Hessian 在笔记本上是数小时到数天/分子）、双杂化泛函在 def2-QZVP 下的全网格、任何 CCSD(T)/CBS 自算参考值——这三项一律不做，参考值直接取 BH76 已发布数值

6 · 方法路径

1. 装 PySCF + ASE，跑通官方算例，下载 BH76，先用 1 个泛函 1 个基组打通"读几何 $\rightarrow$ 单点 $\rightarrow$ 组装能垒 $\rightarrow$ 对比参考"的全链路（第 1-6 周）。
2. 核实工具能力边界：确认 PySCF 中目标泛函的 LibXC 名称与可用性、密度拟合是否会改变能垒（做一次开关对照）、默认积分格点等级是否足够。第 6 周前查清 ORCA 的许可资格问题，结论写进方案（可用/不可用/需核实）。
3. 通过方法学校验硬门槛：复现 PBE 与 B3LYP 在 BH76 上的已发表 MAD。
4. 铺开泛函 $\times$ 基组网格，批量计算并逐个记录墙钟时间（同机、多线程、三次中位数）；未收敛算例单独记录。
5. 做统计分析：MAE / RMSE / MSE、扣除 MSE 后的残差 MAE、按反应类型（氢转移 / 重原子转移 / 亲核取代）分层，全部给自助 95% CI。

6. 绘制误差-成本前沿图（等墙钟口径）与等自由度对比表，判定 H1 与 H2。

7. 独立交叉校验：随机抽 10 个能垒，用 xtb 的 GFN2-xTB 与一个未纳入网格的泛函重算，确认前沿位置与偏置方向的结论不依赖网格的具体选点。

7 · 新颖性边界

- 本课题不声称提出新泛函或新校正方案，不声称发现"GGA 低估能垒"（这是数十年的共识），不自算高精度参考值，不把 GMTKN55/BH76 数据计入学生的数据贡献。
- 已有工作具体完成了什么：Goerigk 与 Grimme 等（PCCP 2017）在 GMTKN55/BH76 上评测 200 余个泛函，统一使用大基组（def2-QZVP 量级）并运行于集群，输出为每个泛函的 MAD 排名与加权总分（WTMAD）。
- 本项目的贡献有两条且都是主结论：其一，换问题方向——把"哪个泛函最准"换成"在纯 CPU 笔记本的固定墙钟预算下误差前沿在哪"，输出一张对小资源使用者可直接照做的前沿图；其二，换评价维度——把 MAD 拆成"系统偏置 + 残差"，定量回答"廉价泛函加一个经验常数能不能顶用"这一从未被系统报告的问题。
- 为什么有价值：绝大多数中学与本科阶段的计算化学工作被迫在 6-31G 这一级做能垒，却只能参考大基组的泛函排名。这份前沿图直接服务于这一使用场景，且方法学层面的可复现性贡献在计算领域是被承认的*——但定位必须讲清楚，否则会被误读为重复 Goerigk 的工作。
- "差异不显著"同样是有效结论：若扣除 MSE 后残差 MAE 几乎不降，说明廉价泛函的误差本质随机、不可用常数校正，这直接否定了一条被广泛使用的民间做法。

8 · 决策门槛 (go / no-go)

- 第 6 周末：方法学校验硬门槛必须通过（PBE/B3LYP 的 BH76 MAD 复现偏差 ≤ 1.0 kcal/mol）。若失败，最可能的原因是参考能定义或几何版本不匹配——降级为改用 BH76 的一个子集（如只做 19 个氢转移反应，参考值在多篇文献中被独立引用，交叉验证更容易），网格规模不变，主结论框架完全保留。
- 第 10 周末：核实本机跑完全网格所需的总墙钟时间。若外推超过 200 小时，立即降级：其一，把基组从 3 个减到 2 个（6-31G* 与 def2-SVP），去掉 def2-TZVP；其二，把能垒数从 76 减到分层抽样的 40 个（每类反应等比例抽取，抽样规则事先写定）。两条路径都保留"误差-成本前沿 + 偏置可校正性"的主结论框架。
- 第 3 周前必须查清（不要边做边发现）：ORCA 的免费学术许可是否覆盖中学生个人使用。若不覆盖，方案中一切 ORCA 相关内容删除，全部改用 PySCF——本课题的设计已保证不依赖 ORCA。
- 已知困难即研究内容：SCF 不收敛在过渡态构型上时有发生。把"哪些泛函/基组组合在哪类过渡态上更容易不收敛"写成一节，本身就是对小资源使用者有价值的交付物，而不是失败。
- 预算裁剪顺序：先砍 def2-TZVP，再砍格点敏感性分析（改为只在 2 个代表组合上做），再砍反应类型分层，最后砍 xtb 交叉校验。砍到只剩"6 泛函 $\times$ 1 基组 $\times$ 76 能垒 + 前沿图 + 偏置分解"时主结论仍成立。
- 选择前提：仅在学生已理解自洽场、基组、交换关联泛函等基本概念时启动，否则先做 C01/C02 打底。

C06 · 反应机理骨架简化误差与速率常数不确定度的可分辨性判据：以公开燃烧机理的点火延迟为探针

推荐优先级 中高 · 分族 反应动力学建模 · 资源需求 纯 CPU, 单次积分亚秒级 · 技能取向 Python + 常微分方程 + 蒙特卡洛 · 失败风险 中

1 · 研究问题

在公开的详细燃烧机理上做骨架简化 (skeletal reduction, 逐步删除物种与反应) 时, 简化引入的点火延迟 (ignition delay time, IDT) 误差, 要到删除多少物种之后才超过速率常数不确定度本身造成的 IDT 分布宽度? 换言之, 在既有的动力学参数不确定度背景下, "简化误差"从哪一点开始才是可分辨的?

2 · 研究背景与空白

详细化学机理动辄数百物种、数千反应 (GRI-Mech 3.0 含 53 物种 325 反应, 更大的燃料机理可达 500+ 物种 2,800+ 反应), 直接耦合进流体计算不可行, 故机理简化是标准工序。主流方法是有向关系图 (directed relation graph, DRG) 及其变体 DRGEP、DRGASA、DRGEP-SA, 判据通常是"点火延迟相对误差 $\leq 5\%$ ", 常见成果形式是"物种减 59%、反应减 62%, IDT 最大相对误差 5%"。

另一条独立线索是动力学参数不确定度: 每个基元反应的 Arrhenius 参数都有不确定因子 (uncertainty factor), 蒙特卡洛传播到 IDT 会得到一个分布, 宽度在很多条件下远超 5%。已有不确定度工作用敏感性分析把参数从 290 降到 32 个, 再用活跃子空间 (active subspace) 构造 IDT 响应面。

空白在: 这两条线几乎是平行发展的——简化工作报"误差 5%", 不确定度工作报"IDT 分布宽度 $\pm X\%$ ", 但很少有人把两者放在同一张坐标图上问"5% 的简化误差在参数不确定度的噪声地板之上还是之下"。文献中甚至有明确表述: DRG 类方法与敏感性分析结合虽能得到紧凑机理, 但无法在考虑不确定度的前提下进行简化。这属于评价维度的转换, 且对学生课题友好: 机理公开、Cantera 免费、0 维积分亚秒级, 正负结论都明确。

3 · 可检验假设

H1 在化学计量比甲烷/空气、初始温度 1000 – 1600 K、压力 1 – 10 atm 的条件网格上, 把 GRI-Mech 3.0 简化到某个物种数 N 时, 简化引起的 IDT 相对偏差首次超过速率常数不确定度传播所得 IDT 分布的 90% 区间半宽; 该 N 显著小于按传统 5% 判据得到的物种数 (即传统判据在噪声地板之下"过度追求精度")。

H2 N 强烈依赖工况: 在负温度系数 (NTC) 区或低温区, N 比高温区大 $\geq 20\%$ (即低温条件下简化更早变得可分辨)。若 H2 被否定 (N^* 对工况不敏感), 则可给出一条跨工况通用的简化深度建议, 同样有价值。

4 · 量化验收标准

1. 方法学校验 (硬门槛): 用自建 Cantera 管线在未简化的 GRI-Mech 3.0 上重算文献中已发表的甲烷点火延迟曲线 (至少 3 个已发表的温度-IDT 数据点或激波管实验对照点), IDT 相对偏差 $\leq 10\%$; 同时同一条件下用两种反应器设定 (定容绝热 vs 定压绝热) 确认自建脚本的 IDT 判据 (温度升高 400 K 或 OH 浓度峰值) 实现一致。此条不过关意味着反应器设置、初始条件或 IDT 判据有误, 后续全部结论无效。
2. 统计口径预先写死: 不确定度传播用蒙特卡洛, 样本数 $\geq 2,000$, 参数扰动采用对数正态分布、以机理文件或文献给出的不确定因子为 2σ (GRI-Mech 3.0 是否随文件提供逐反应不确定因子须核实; 若无, 统一取 A 因子不确定因子 2 并在敏感性分析中检验该假设的影响)。IDT 分布报中位数与 5%/95% 分位数, 不报均值 $\pm$ 标准差 (IDT 分布强烈右偏)。所有" N^* 是多少"的结论给自助 95% 置信区间。

- 数值收敛检查（必需）：报告 ODE 求解器相对/绝对容差扫描（至少 3 档），确认 IDT 在容差收紧 10 倍后变化 $\leq 1\%$ ；报告元素守恒漂移量。凡未通过收敛检查的条件点不得纳入统计。
- 条件网格下限：温度 ≥ 8 点、压力 ≥ 3 点、当量比 ≥ 3 点，共 ≥ 72 个工况；简化序列至少覆盖从全机理到 20 物种的 ≥ 10 个简化深度。
- 简化算法必须自行实现 DRG 并用公开的已发表简化结果校验（例如复现某篇论文报告的“某阈值下保留物种集合”，重合度 $\geq 90\%$ ）；不得直接调用黑箱工具而不验证。
- 全部机理文件、简化代码、蒙特卡洛脚本、原始结果与作图脚本开源，一条命令可重跑。

5 · 数据与工具

用途	来源 / 工具
求解器	Cantera (BSD 许可, conda 一条命令安装)。内置：0 维定压/定容绝热反应器、气相动力学、热力学与输运、YAML 机理格式、敏感性分析接口。 注意：DRG/DRGEP 简化算法 Cantera 不内置，须自行实现 （这正是本课题的方法学工作量所在，也可作为项目贡献之一）
详细机理	GRI-Mech 3.0 (53 物种、325 反应)，随 Cantera 以 <code>gri30.yaml</code> 分发；扩展可选 AramcoMech 或 USC-Mech (官网免费下载，YAML 转换需核实)
实验对照	公开激波管点火延迟数据（如 Petersen、Gardiner 等经典甲烷数据集，多见于文献表格与 respecth 类数据库）。 仅用于校验与对比，不计入本项目的数据贡献
不确定因子	机理文件附带的不确定度信息（ 需核实 GRI-Mech 3.0 是否提供逐反应不确定因子 ）；若无，采用文献惯例的 A 因子不确定因子 2
算力	纯 CPU。GRI-Mech 3.0 的 0 维点火延迟单次积分在笔记本上约 0.05—0.5 秒 （需本机实测）。蒙特卡洛：2,000 样本 $\times$ 72 工况 $\approx$ 144,000 次积分 $\approx$ 2—20 小时 ，可后台分批跑并按工况并行（multiprocessing）。简化序列 10 档 $\times$ 72 工况仅数百次积分，分钟级。 超出范围的方案（明确列出） ：任何一维/多维火焰或 CFD 耦合、500+ 物种机理上的全量蒙特卡洛（积分成本上升一个数量级以上，且 YAML 加载本身即以分钟计）——一律不做

6 · 方法路径

- 装 Cantera，跑通官方 0 维点火算例，用 GRI-Mech 3.0 复现已发表甲烷 IDT 数据，通过方法学校验硬门槛（第 1—6 周）。
- 核实工具能力边界：确认 Cantera 是否提供逐反应不确定度字段、如何以编程方式修改单个反应的 Arrhenius 参数（`Reaction.rate` 接口）、以及如何构造删除物种后的合法机理对象。这三项须在第 10 周前查清。
- 自行实现 DRG 简化：定义耦合系数、阈值扫描、生成简化序列；用公开已发表的简化结果做交叉校验（保留物种集合重合度 $\geq 90\%$ ）。
- 做数值收敛检查（容差扫描 + 元素守恒漂移），冻结求解器设置与 IDT 判据（判据须明确、可复现，例如“温度较初值升高 400 K 的时刻”）。

5. 跑不确定度蒙特卡洛，得到每个工况下的 IDT 分布与 90% 区间半宽（噪声地板）。
6. 跑简化序列，得到每个简化深度的 IDT 偏差；把两条曲线画在同一坐标上，读出交点 N ，给自助置信区间；按温度/压力/当量比分层报 N 。
7. 独立交叉校验：换一个可观测量（如层流火焰速度需一维求解，成本过高，故改用"最大热释放速率时刻"或"峰值 OH 摩尔分数"）重算 N^* ，确认结论不依赖于单一目标量。

7 · 新颖性边界

- 本课题不声称提出新的简化算法，不声称构建新机理，不声称改进不确定度量化方法，不把 GRI-Mech 3.0 或公开实验数据计入学生的数据贡献。
- 已有工作具体完成了什么：DRG/DRGEP/DRGASA/DRGEPSA 系列方法（Lu & Law 及后续大量工作）以"IDT 相对误差 $\leq 5\%$ "为判据给出骨架机理，典型成果为"物种减 59%、反应减 62%"；独立地，敏感性分析 + 活跃子空间的不确定度工作把参数从 290 降到 32 维并构造 IDT 响应面。文献中已明确指出 DRG 类简化"无法在考虑不确定度的前提下进行"。
- 本项目的贡献是评价维度的转换且构成主结论：把"简化误差"与"参数不确定度"放在同一坐标系里，给出一条可操作的判据——"简化到 N^* 物种以下，误差才在动力学参数噪声地板之上；在此之上继续追求 5% 精度是把算力花在不可分辨的差异上"。附带贡献是一个自行实现并经公开结果校验的 DRG 实现。
- 为什么有价值：这直接改变简化工序的停止准则，把"越准越好"换成"准到噪声地板即可"，对算力受限的使用者尤其有意义。
- "差异不显著"同样是有效结论：若在全部工况下简化误差始终远低于噪声地板，说明现行 5% 判据确实过严；若始终远高于，则说明简化才是主导误差源、传统判据合理。两种结果都必须用蒙特卡洛区间证明有能力分辨。

8 · 决策门槛 (go / no-go)

- 第 6 周末：GRI-Mech 3.0 的 IDT 必须复现已发表数据到 $\leq 10\%$ 。若失败，先核对当量比、稀释比与 IDT 判据定义；仍失败则降级为放弃与实验对照，改以"Cantera 官方教程中的标准算例"为唯一校验基准，并把校验门槛改为"两种反应器设定与两种 IDT 判据下结果自洽 $\leq 5\%$ "。主结论框架不变（本课题的核心是相对比较，不依赖绝对准确性）。
 - 第 10 周末：核实是否能拿到逐反应不确定因子。若拿不到，**立即降级**：统一取不确定因子 2（文献惯例），并把"不确定因子取值本身的影响"做成敏感性分析（取 1.5 / 2 / 3 各跑一遍）——这个降级反而增加了一节内容，主结论 N 的存在性不受影响，只是 N 的绝对数值带上一层假设。
 - 第 16 周末：核实自建 DRG 实现能否复现公开简化结果（重合度 $\geq 90\%$ ）。若不能，降级为改用"随机删除 + 重要性排序"的更简单简化策略（按物种在敏感性分析中的排名逐个删除），主结论"简化误差 vs 噪声地板的交点"完全保留，只是简化路径不再是标准 DRG。
 - 算力风险与裁剪顺序：若蒙特卡洛外推超过 40 小时，先把样本数从 2,000 降到 800（分位数估计精度下降但 90% 区间仍可用），再把工况从 72 降到 24（保留温度轴、砍压力轴），再砍第二个目标量的交叉校验。砍到只剩"800 样本 $\times$ 24 工况 $\times$ 单目标量"时主结论仍成立。
 - 选择前提：适合数学基础较好、能接受结果以数字表格为主的学生；偏好可视化与化学直觉者选 C04。
-

C07 · 笔记本算力预算下水中小分子扩散系数的力场可分辨性：有限尺寸校正后两个力场差异何时才在统计上成立

推荐优先级 中 · 分族 分子动力学 · 资源需求 纯 CPU，多核后台连跑数周 · 技能取向 命令行 MD + 时间序列统计 · 失败风险 中高

1 · 研究问题

在纯 CPU 笔记本可承受的模拟预算内（盒子 $\leq 3,000$ 原子、累计轨迹 ≤ 300 ns），经 Yeh – Hummer 有限尺寸校正后，两个常用力场组合（如 GAFF2/TIP3P 与 OPLS-AA/TIP4P-2005）给出的小分子水中自扩散系数之差，需要多长的累计轨迹与多少条独立重复才能在 95% 置信区间上与零区分开？该阈值与分子大小、盒子尺寸如何标度？

2 · 研究背景与空白

自扩散系数 D 由均方位移（mean squared displacement, MSD）长时线性段斜率给出： $D = \lim MSD/(6t)$ ，是最容易与实验对照的液相输运性质，也是力场验证的常规靶标。它有两个众所周知的陷阱：其一，周期边界下的流体力学自相互作用使 D 系统性偏小，需 Yeh – Hummer 校正 $D_\infty = D_{\text{PBC}} + k_B T \xi / (6\pi\eta L)$ （ $\xi \approx 2.837297$ ， L 为盒长， η 为剪切黏度），小盒子里可达 20 – 40%；其二，MSD 斜率的统计误差只以 $\sqrt{t}$ 收敛，故“两个力场差 10%”在短轨迹上常常是噪声。

已有工作走到哪一步：Yeh 与 Hummer（2004）确立有限尺寸校正公式；大量力场评测论文报告 GAFF/OPLS/CHARMM 配不同水模型下小分子 D 与实验的偏差（典型 20 – 50%）；TIP3P 水自扩散系数比实验高约 2 倍是公认缺陷。

空白在：这些工作几乎都在集群上跑几百纳秒到微秒，报告的是“力场 A 比力场 B 更接近实验”这类结论，没有人从相反方向问：在一台笔记本的算力上限内，这类结论到底还能不能成立。这是把算力预算显式写进研究设计的“可分辨性/统计功效”问题，属于换问题方向：资源约束不是妥协而恰是研究对象，正负结论都对同类小资源使用者有直接指导意义。

3 · 可检验假设

H1 在 2,500 原子量级的盒子中，单条 20 ns 轨迹给出的溶质 D 的相对标准误差 $\geq 15\%$ ；要把两个力场组合 20% 的 D 差异在 95% 置信区间上分辨开，需要累计 ≥ 100 ns（ ≥ 5 条独立重复）。

H2 有限尺寸校正项在本课题可及的盒子尺寸（ $L \approx 3 - 4$ nm）上贡献 $\geq 15\%$ 的 D ，其量级大于同一体系两个力场之间的差异；即不做有限尺寸校正的力场对比结论会被系统性污染。若 H2 被否认（校正项远小于力场差异），说明小盒子对力场排序无害，这对小资源使用者是一个正面结论。

4 · 量化验收标准

- 方法学校验（硬门槛）：用自建管线模拟纯 TIP3P 水（约 2,000 – 3,000 原子），计算其自扩散系数并与该水模型的公认文献值对比（TIP3P 在 298 K 的 D 约为 $5.5 - 6.0 \times 10^{-5}$ cm²/s 量级，须以原文核实），经有限尺寸校正后偏差 $\leq 10\%$ ；同时复现同一模型的密度（偏差 $\leq 1\%$ ）。此条不过关意味着系综设置、时间步长、静电处理或 MSD 拟合区间有误，后续全部力场对比结论无效。
- 统计口径预先写死： D 由 MSD 在明确指定的线性拟合区间（例如 20 – 80% 轨迹长度，区间选择规则事先冻结并公布）上最小二乘得到；误差棒来自 ≥ 5 条独立初速度重复的样本标准误，并额外用块平均（block

averaging) 交叉验证误差估计; 所有力场差异的显著性用自助重抽样给 95% 置信区间, 不用单条轨迹的拟合残差冒充误差棒 (这是该领域最常见的错误)。

- 3. 收敛性检查 (必需): 报告能量守恒/温度控制的漂移量、NPT 平衡阶段的密度时间序列、以及 D 对拟合区间选择的敏感性 (至少 3 种区间)。报告 MSD 的对数斜率是否达到 1.0 (确认进入扩散区而非亚扩散区)。
- 4. 有限尺寸校正必须自行实现并验证: 至少用 2 个不同盒长 (如 $L \approx 3.0\text{ nm}$ 与 3.8 nm) 做外推对照, 检验 D_PBC 对 $1/L$ 的线性关系斜率与 Yeh – Hummer 公式预测值一致 (偏差 $\leq 25\%$); 黏度 η 由同一模拟的 Green – Kubo 或文献值给出 (取法事先写定)。
- 5. 交付一张"可分辨的 D 相对差异 vs 累计轨迹长度"的功效曲线, 并在图上标出本机 1 周 / 4 周预算线。
- 6. 全部拓扑文件、mdp / 输入文件、分析脚本与原始 MSD 数据开源, 一条命令可重跑。

5 · 数据与工具

用途	来源 / 工具
MD 引擎 (主)	GROMACS (LGPL, conda-forge 或源码编译)。CPU 并行成熟, 是本课题 唯一推荐的主引擎
MD 引擎 (不推荐作主力)	OpenMM。官方与社区反复指出其为 GPU 设计、CPU 平台性能不佳 (有报告: 48 核节点上 95,000 原子体系仅约 0.2 ns/day)。本课题仅在体系搭建/力场转换环节使用其生态 (如 openff-toolkit), 不用它跑生产轨迹
力场与拓扑	GAFF2 (经 acpype /AmberTools 或 OpenFF 工具链生成)、OPLS-AA (LigParGen 在线服务, 可用性与许可须核实)、OpenFF Sage (开源, openff-toolkit 直接生成)。水模型 TIP3P / TIP4P-2005 均随 GROMACS 分发
实验对照	公开的小分子水中自扩散系数实验值 (如甲醇、乙腈、尿素等, 见物性手册与文献表格)。仅用于校验与对比, 不计入本项目的数据贡献
分析	GROMACS gmx msd (内置 MSD 与 Einstein 拟合)、MDAnalysis 或 MDTraj (Python 侧独立重算, 用于交叉校验)、自写 Yeh–Hummer 校正与块平均
算力	纯 CPU, 这是全部路线中算力最紧的一条, 必须在第 4 周实测本机数字。可用的公开标尺: 8 核 CPU 上 41,175 原子体系约 0.661 ns/day。按原子数近似线性外推, 2,500–3,000 原子体系在 8 核上约 9–11 ns/day (此外推必须由本机基准测试确认, 不得直接采信)。据此, 4 周后台连跑可得约 250–300 ns 累计轨迹, 恰好落在"5 条 $\times$ 20 ns $\times$ 两个力场 + 校验"这一预算内。超出范围的方案 (明确列出): 任何 > 10,000 原子的盒子、微秒级轨迹、自由能微扰/伞形采样、蛋白-配体体系——一律不做

6 · 方法路径

- 1. 装 GROMACS (优先用 conda-forge 预编译包, 避免编译链问题), 跑通官方水盒子教程, 在本机实测 ns/day 并据此重新确定盒子大小与轨迹长度 (第 1–4 周, 此步的输出直接决定后续方案规模)。
- 2. 做纯 TIP3P 水的方法学校验硬门槛: 密度 + 自扩散系数 (含有限尺寸校正) 对照文献值。
- 3. 核实工具能力边界: 确认 GROMACS 的 gmx msd 拟合区间默认取法、是否内置有限尺寸校正 (几乎肯定不内置, 须自行实现)、Green – Kubo 黏度是否可直接输出。这三项须在第 8 周前查清。自行实现的校正必须用双盒长外推验证——这一步本身可作为项目的方法学贡献写入。
- 4. 搭建溶质体系 (1 个小分子 + 水, 两套力场组合), 做 NPT 平衡与生产轨迹, 每套 ≥ 5 条独立重复。
- 5. 计算 D_PBC、施加有限尺寸校正、给出块平均与重复间误差棒; 做拟合区间敏感性分析。
- 6. 构造功效曲线: 对轨迹长度做子采样 (取前 2/5/10/20 ns), 画"可分辨的相对差异 vs 累计轨迹长度"。

7. 独立交叉校验：用 MDAnalysis 在 Python 侧完全独立地重算 MSD 与 D，确认与 `gmx msd` 结果一致（偏差 $\leq 3\%$ ）；再用一个盒长完全不同的体系确认标度关系。

7 · 新颖性边界

- 本课题不声称提出新力场或新校正方法，不声称发现 TIP3P 扩散过快（这是公认缺陷），不声称给出比文献更准的 D 值——本课题的算力预算注定做不到这一点，任何相反的声称都会被评委在三十秒内问倒。
- 已有工作具体完成了什么：Yeh 与 Hummer (J. Phys. Chem. B, 2004) 给出周期边界下的有限尺寸校正公式；大量力场评测论文在集群上以几百纳秒到微秒的轨迹报告小分子 D 与实验的偏差（典型 20 – 50%）并对力场排序。这些工作的算力前提与本课题完全不同。
- 本项目的贡献是换问题方向且为主结论：把算力预算显式写进研究设计，回答"在纯 CPU 笔记本上，力场对比结论的可分辨性下限在哪里"，输出一条功效曲线与一条可操作判据（"累计轨迹短于 X ns 时，不要报告力场之间 Y% 以内的 D 差异"）。附带贡献是自行实现并双盒长验证的有限尺寸校正流程。
- 为什么有价值：小资源使用者（中学、本科课程、小课题组）大量在做这类对比，却普遍用单条短轨迹的拟合残差当误差棒，从而报出不存在的差异。本课题给出的是他们可以直接照用的门槛。
- "差异不显著"同样是有效结论——而且在本课题中是预期的主结论之一：明确写出"在笔记本预算内两个力场的 D 差异不可分辨"，并用重复间误差棒证明这不是因为做得不够仔细，而是统计精度的物理上限。

8 · 决策门槛 (go / no-go)

- 第 4 周末（最关键的一道门）：本机 GROMACS 基准测试必须给出 ≥ 5 ns/day @ 2,500 原子。若低于此值，立即降级：其一，把体系缩到 1,200 – 1,500 原子（约 400 个水分子 + 溶质），代价是有限尺寸校正项更大——但这反而强化了 H2 的检验，主结论框架不受损；其二，把研究对象从"溶质扩散"换成"纯水自扩散"（水分子数多，统计量天然充足，同样长度的轨迹误差棒小得多），功效曲线的主结论完全保留。
 - 第 8 周末：纯 TIP3P 水的方法学校验必须通过（D 偏差 $\leq 10\%$ 、密度偏差 $\leq 1\%$ ）。若失败，先查静电处理（PME 参数）、时间步长与约束设置；仍失败则降级为只做"同一力场下不同盒长的有限尺寸标度检验"（这部分不需要与文献 D 值对齐，只需内部自洽），H2 完全保留，H1 改为"同一力场两个独立实现之间的可分辨性"。
 - 第 12 周前必须查清（不要边做边发现）：OPLS-AA 拓扑的获取途径（LigParGen 的可用性与使用条款）。若不可用，改用 OpenFF Sage 与 GAFF2 作为两个对比力场——方案的主结论不依赖于具体是哪两个力场。
 - 已知困难即研究内容：MSD 线性拟合区间的选择带有主观性。把区间选择的敏感性分析与"如何给出可复现的区间判据"写成一节，本身就是交付物。
 - 预算裁剪顺序：先砍第二个盒长（放弃双盒长外推验证，改引用 Yeh – Hummer 原文系数），再砍重复次数（5 $\rightarrow$ 3，误差棒变宽但功效曲线仍可画），再砍第二个力场（只做单力场的功效分析），最后砍 Green – Kubo 黏度（改用文献 η 值）。砍到只剩"纯水 + 单力场 + 3 重复 + 功效曲线"时主结论仍成立。
 - 选择前提：仅在学生有一台可连续后台运行数周的机器时启动。若机器需频繁携带、休眠、断电，改选可分批可中断的 C05 或 C06。
-

C08 · 预训练机器学习势在域外化学基元上的可迁移性审计：以扭转能垒与非共价二聚体扫描为探针、以不确定度指标能否预警失效为判据

推荐优先级 中 · 分族 机器学习势 / 模型评估 · 资源需求 纯 CPU, 推理为主; 需数 GB 内存 · 技能取向 PyTorch 推理 + 量子化学参考计算 · 失败风险 中高

1 · 研究问题

在训练集覆盖薄弱的化学基元上（如高价态硫/磷、卤键、两性离子、罕见杂环），预训练机器学习原子间势（machine-learned interatomic potential, MLIP）MACE-OFF23 与 ANI-2x 的扭转能垒与二聚体相互作用能误差，比其在训练分布内的报告误差放大多少倍？模型自带或可廉价构造的不确定度指标（如集成方差、近邻描述符距离）能否在不看参考值的前提下预警这种失效？

2 · 研究背景与空白

MLIP 以经典力场的成本交付量子化学精度，是近年计算化学最实质的进展之一。MACE-OFF23（Kovács、Csányi 等，JACS 2025）是面向有机分子的短程可迁移力场，基于高阶 E(3) 等变消息传递网络，训练在 SPICE 数据集（药物样分子，含 H/C/N/O/F/P/S/Cl/Br/I 十种元素）上；ANI-2x 是更早的基于原子环境对称函数的模型。MACE-OFF 原文报告其多项性质优于 ANI-2x 与 AIMNet（如分子液体密度 MAE 0.09 g/cm³，ANI-2x 约为其两倍），并展示对未见分子二面角扫描的准确描述。2026 年的工作（arXiv:2605.00640, [Knowing when to trust machine-learned interatomic potentials](#)）开始系统讨论 MLIP 的可信度判定问题。

空白在：模型作者的自评天然倾向于覆盖良好的化学空间；而使用者真正需要知道的是在自己那个可能没被覆盖的角落里，误差会放大到什么程度，以及有没有一个不需要参考值就能算的预警指标。SPICE 的元素虽覆盖十种，但对高价态硫（磺酰、亚磺）、三价/五价磷、卤键给体、两性离子等基元的采样密度差异很大（具体覆盖情况须回 SPICE 原文与数据统计核实）。本课题属于换样本切片 + 换评价维度：切片是刻意选取的域外化学基元，评价维度是“不确定度指标的预警能力”而非平均精度。算力可控，正负结论都明确。

3 · 可检验假设

H1 在刻意构造的域外基元子集上，MACE-OFF23 的扭转能垒 MAE 相对其在分布内基元上的 MAE 放大 ≥ 3 倍（放大倍数的自助 95% 置信区间下界 > 2 ）。

H2 廉价可算的不确定度代理指标（模型集成方差，或测试构型的原子环境描述符到训练分布的最近邻距离）与实际误差的 Spearman 相关系数 $\rho \geq 0.5$ （自助 95% CI 下界 > 0.3 ）；即失效可以被事先预警。若 H2 被否定，则给出“现有廉价不确定度指标不足以预警域外失效”的负结论，这对 MLIP 的实际部署是重要的警示。

4 · 量化验收标准

- 方法学校验（硬门槛）：用自建推理管线在 MACE-OFF 原文或 SPICE 公开测试集中的已发表算例上重算能量/力，与原文报告的能量 MAE 对比，偏差 ≤ 0.5 kcal/mol（或原文单位下的等价阈值）；同时对至少 3 个分子做“MLIP 扭转扫描 vs 原文图中扫描曲线”的目视与数值对照，峰位偏差 $\leq 5^\circ$ 、峰高偏差 ≤ 0.5 kcal/mol。此条不过关意味着模型权重、单位换算（eV vs kcal/mol）或元素映射有误，后续全部可迁移性结论无效。
- 参考值口径预先写定：域外基元的参考能垒由 PySCF 在明确写定的水平上自算（如 ω B97X-D/def2-SVP 或 B3LYP-D3/6-31G，具体水平在第 8 周前冻结），并对其中 ≥ 10 个点用更大基组做一次收敛检查（能垒变

化 ≤ 0.3 kcal/mol 方接受)。参考水平低于 SPICE 的训练水平这一事实必须显式声明^{*}，并把由此引入的系统偏置作为独立的误差项报告，不得混入 MLIP 误差。

- 统计口径预先写死：误差报 MAE 与 RMSE (kcal/mol)，分层为"域内基元 / 域外基元"两组，放大倍数与所有相关系数给 1,000 次自助（自助单位为分子）95% 置信区间。域外/域内的划分判据必须明确、可复现且事先冻结（基于 SMARTS 定义的基元清单 + SPICE 中该基元出现频次的统计阈值）。
- 样本量下限：域外基元 ≥ 5 类、每类 ≥ 8 个分子；扭转扫描每分子 ≥ 24 个角度点；非共价二聚体 ≥ 30 对（可取自 S22/S66 公开集的相关子集）。
- 计时口径：报告每个模型在本机 CPU 上的单帧推理时间（同机、同线程、三次中位数），以及"MLIP 推理 vs PySCF 参考计算"的成本比——这是使用者关心的第二维信息。
- 全部模型版本号、权重来源、扫描结构、参考计算输入输出、不确定度指标实现与统计脚本开源，一条命令可重跑。

5 · 数据与工具

用途	来源 / 工具
预训练模型	MACE-OFF23 (GitHub ACESuit/mace，权重按其许可发布——许可条款与是否允许中学生个人非商业使用须在 第4周前核实)；ANI-2x (torchani，MIT 许可，权重随包分发)。两者均可通过 ASE calculator 接口调用
参考计算	PySCF (见 C05 的能力说明)。域外基元分子须控制在 ≤ 12 重原子，否则参考计算成本失控
非共价对照集	S22 / S66 公开二聚体基准集 (含高精度参考相互作用能)。仅用于校验与对比，不计入本项目的数据贡献
训练分布统计	SPICE 数据集 (公开，HDF5 格式，体积较大， 是否需全量下载须核实 ；仅需其化学基元频次统计，可考虑用已发布的统计表或按需下载子集)
结构生成	RDKit (ETKDG 生成构象、SMARTS 定义基元、约束二面角扫描)
算力	纯 CPU。MLIP 推理：MACE-OFF23(S) 对 20—40 原子体系单帧约 数十毫秒到 1 秒 (强依赖线程数与 PyTorch 版本， 必须本机实测)；40 分子 $\times$ 24 点 $\times$ 2 模型 $\approx$ 2,000 次推理， 分钟到小时级 。参考 DFT： ≤ 12 重原子在 def2-SVP 下单点约 10—60 秒 ，2,000 个点约 6—30 小时 ，可后台分批。内存：MACE 模型加载需数 GB， 须确认本机内存 ≥ 16 GB 。 超出范围的方案 (明确列出) ：用 MLIP 跑长时 MD 生产轨迹、自行训练或微调 MLIP、在 SPICE 训练水平 (ω B97M-D3/def2-TZVPPD) 上自算参考值——三者一律不做

6 · 方法路径

1. 装 PyTorch (CPU 版)、`mace-torch`、`torchani`、ASE，加载权重跑通官方示例，实测单帧推理时间与内存占用（第 1–5 周）。若内存或速度不达标，立即走降级路径。
2. 通过方法学校验硬门槛：复现原文算例的能量 MAE 与 3 条扭转扫描曲线。
3. 核实工具能力边界与许可：MACE-OFF23 权重的使用条款；模型对形式电荷/自由基的支持情况（MACE-OFF23 与 ANI-2x 均为闭壳层中性分子设计，自由基与带电体系很可能不在支持范围——这一点须核实并据此界定域外基元清单）。
4. 冻结域外基元清单（SMARTS + SPICE 频次阈值），用 RDKit 生成分子与约束扭转扫描结构。
5. 跑 MLIP 推理与 PySCF 参考计算，组装能垒表；同时计算两类不确定度代理指标（多模型/多 size 集成方差、描述符最近邻距离）。
6. 做分层统计：域内 vs 域外 MAE、放大倍数、不确定度-误差相关性，全部给自助 95% CI；并把参考水平引入的系统偏置单列一项。
7. 独立交叉校验：对 S22/S66 子集用 MLIP 算相互作用能，与该集的高精度参考值直接对比（不经自算 DFT），确认"域外放大"的结论在完全独立的参考体系上方向一致。

7 · 新颖性边界

- 本课题不声称训练或改进任何 MLIP，不声称发现 MACE-OFF23 的新架构缺陷，不声称自己的参考计算精度高于 SPICE 的训练水平，不把 SPICE、S22/S66 的数据计入学生的数据贡献。
- 已有工作具体完成了什么：Kovács、Csányi 等 (JACS 2025) 发布 MACE-OFF23 系列并在分子液体密度、扭转扫描、分子晶体等性质上自评，报告优于 ANI-2x 与 AIMNet（如液体密度 MAE 0.09 g/cm³，ANI-2x 误差约两倍）；arXiv:2605.00640 (2026) 开始讨论 MLIP 的可信度判定框架。
- 本项目的贡献有两条且都是主结论：其一，**换样本切片**——以训练分布频次统计为依据、用可复现判据刻意构造域外基元集，量化误差放大倍数；其二，**换评价维度**——不评"模型准不准"，而评"廉价不确定度指标能不能在没有参考值时预警失效"，这是使用者实际部署时唯一能依赖的信息。
- 为什么有价值：MLIP 正在被大量非专业使用者当作黑箱调用。一份"哪些基元上会失效、失效能否被预警"的独立审计，正是这类模型从论文走向使用最缺的一环。
- "差异不显著"同样是有效结论：若域外基元上误差并未显著放大，说明 MACE-OFF23 的可迁移性优于预期，这对社区是正面且有分量的独立验证——但必须给出自助置信区间证明本研究有能力分辨 2 倍以上的放大。

8 · 决策门槛 (go / no-go)

- 第 5 周末（关键门）：本机必须能加载 MACE-OFF23 权重并在 ≤ 2 秒/帧完成 30 原子体系推理，内存占用在可承受范围。若不达标，立即降级：其一，只做 ANI-2x（`torchani` 更轻、CPU 推理更快），把研究对象从"两模型对比"改为"单模型的域外审计 + 不确定度预警能力"，主结论框架完全保留；其二，把体系规模压到 ≤ 20 原子。
- 第 4 周前必须查清（不要边做边发现）：MACE-OFF23 权重的许可条款是否允许本项目的使用与结果发布。若不允许，全部改用 ANI-2x (MIT 许可) 与其他明确开源的 MLIP。
- 第 10 周末：方法学校验硬门槛必须通过。若原文算例复现不了（偏差 > 0.5 kcal/mol），最可能是单位或模型 size 变体选错——降级为放弃与原文数值对齐，改以 S22/S66 公开参考值作为唯一校验基准（该集参考值

权威且无歧义），主结论框架不变。

- **第 14 周末：**核实域外基元清单能否凑齐 5 类 $\times$ 8 分子且参考计算成本可控 (≤ 12 重原子)。若不能，降级：其一，把域外/域内的划分从"基元类别"改为"SPICE 频次连续变量"，改做回归而非分组对比；其二，把探针从扭转能垒换成成本更低的"平衡键长/键角偏差"（只需单点优化对比，参考计算量下降一个数量级）。两条路径都保留"误差放大 + 预警能力"的主结论框架。
 - **预算裁剪顺序：**先砍第二个模型，再砍非共价二聚体部分（只做扭转），再砍第二类不确定度指标，最后把每类基元的分子数从 8 降到 5。砍到只剩"ANI-2x + 3 类域外基元 $\times$ 5 分子 + 单一不确定度指标"时主结论仍成立。
 - **选择前提：**仅在学生已完成过一次 PyTorch 环境配置、且机器内存 ≥ 16 GB 时启动。这条路线的失败模式主要在工程环境而非科学内容，第 5 周的环境门必须严格执行，不得拖延。
-

C09 · 小数据 Δ -机器学习把 GFN2-xTB 校正到 DFT 的可用性边界：以化学相似度分层的增益消失阈值为判据

推荐优先级 中低 · 分族 机器学习 + 半经验量子化学 · 资源需求 纯 CPU, xtb 批量数小时 · 技能取向 量子化学 + 机器学习 + 学习曲线分析 · 失败风险 高

1 · 研究问题

在只有 $10^2 - 10^3$ 个训练分子的小数据条件下（这是一台笔记本 + 一学年真实能达到的规模），把 GFN2-xTB 能量校正到 DFT 水平的 Δ -机器学习（ Δ -machine learning）模型，其相对于未校正 xTB 的增益，在测试分子与训练集的化学相似度低到什么程度时完全消失？该“增益消失阈值”是否随训练集规模系统性外移？

2 · 研究背景与空白

Δ -机器学习（Ramakrishnan 等, JCTC 2015）不直接学高精度能量，而是学“低精度与高精度方法之差”；这个差比能量本身平滑得多，同样的数据能换来大得多的增益。以 GFN2-xTB 为基线、DFT 为目标是最实用的搭配：xTB 单点亚秒级、DFT 单点分钟级。

已有工作走到哪一步：Zhao 与 Savoie 的 Δ^2 模型（Chem. Sci. 2023, 14, 13392；代码见 [GitHub zhaoqy1996/Delta2ML](#)）以 GFN2-xTB 几何与能量为低阶输入、B3LYP-D3/TZVP 为高阶目标，训练在约 167,000 个反应上，并在多个外部集上验证近化学精度的可迁移性。相关方向还有用核岭回归（kernel ridge regression, KRR）做 PBE $\rightarrow$ 耦合簇校正，以及 2026 年用基础 MLIP 隐特征做 Δ 学习校正 DFT 生成能。

空白在：这些工作全部处在大数据 + 集群的体制里（十万量级训练样本），其结论“ Δ 学习可迁移”对一个只能拿到几百个 DFT 参考点的使用者没有直接指导意义。小数据体制下的真问题不是“能否达到化学精度”，而是“离训练集多远时这个校正还比不做校正强”——超过那个距离， Δ 模型可能反而把 xTB 本来合理的能量拉偏。这属于换问题方向 + 参数维度扫描。风险在于结论可能是“小数据下增益本来就很小”，故本课题必须从设计上保证该负结论也是完整可交付的成果。

3 · 可检验假设

H1 以 Tanimoto 相似度（Morgan 指纹）度量测试分子到训练集的最近邻距离， Δ 模型相对未校正 xTB 的 MAE 增益随相似度单调下降，并在某个阈值 T^* 处降为零（增益的自助 95% 置信区间跨零）；在 1,000 分子训练规模下 $T^* \leq 0.4$ 。

H2 T^* 随训练集规模（100 / 300 / 1,000）系统性外移，且外移速率呈对数标度；即扩大训练集主要买到的是“适用范围”而非“分布内精度”。若 H2 被否定（ T^* 对训练规模不敏感），则说明小数据 Δ 学习的适用范围由化学表征而非样本量决定，这是一条更强、更有用的结论。

4 · 量化验收标准

1. 方法学校验（硬门槛）：用自建管线在 QM9 数据集的一个标准子集上重算 GFN2-xTB 原子化能，并复现文献中报告的“直接用 KRR/GBDT 学 QM9 原子化能”的基线学习曲线，在 1,000 训练样本处的 MAE 与已发表数值偏差 ≤ 1.0 kcal/mol；同时验证自建原子化能定义（DFT 与 xTB 各自的孤立原子参考能）在 20 个分子上自洽（残差无系统漂移）。此条不过关意味着能量参考基准或单位换算有误，后续全部增益结论无效——这是本课题最容易出错的地方。

- 统计口径预先写死：误差报 MAE 与 RMSE (kcal/mol)；划分必须用骨架划分与相似度分层双重口径，并额外做一次随机划分对照以量化其乐观高估幅度。学习曲线在每个训练规模上跑 ≥ 10 次重抽样，报中位数与四分位区间。"增益"定义为 $\text{MAE}(\text{xTB 原始}) - \text{MAE}(\Delta \text{ 模型})$ ，其置信区间由 1,000 次自助（自助单位为测试分子）给出。相似度分箱至少 5 档，每档测试分子 ≥ 50 。
- 训练规模至少 4 档（100 / 300 / 1,000 / 3,000，最后一档视算力）；模型至少 2 种（核岭回归 + 梯度提升），确认 T^* 不是单一模型的产物。
- 明确报告"DFT 参考值取自 QM9 (B3LYP/6-31G(2df,p))，本项目不自算 DFT 参考"，以及由此带来的适用范围限制（QM9 仅含 ≤ 9 个重原子的 C/H/O/N/F 分子）。
- 交付一张"增益 vs 相似度 $\times$ 训练规模"的二维热图与 T^* 的置信区间表。
- 全部 xtb 批处理脚本、特征生成、学习曲线代码与原始能量数据开源，一条命令可重跑。

5 · 数据与工具

用途	来源 / 工具
DFT 参考与几何	QM9 数据集 (134k 个 ≤ 9 重原子的有机分子，含 B3LYP/6-31G(2df,p) 优化几何与内能)。公开下载 (Figshare / MoleculeNet / DeepChem 镜像)，体积数百 MB。仅用于校验与对比，不计入本项目的数据贡献——DFT 参考值全部来自该数据集，本项目不自算
低阶方法	xtb (GFN2-xTB 单点，直接在 QM9 提供的几何上算，不重新优化，以保证 Δ 定义为纯粹的方法差而非几何差)
特征	RDKit (Morgan 指纹用于相似度与骨架划分)；能量学习特征用组成计数 + 键类型计数 + 简单几何描述符 (如 Coulomb 矩本征值或 BoB 的轻量实现，须自行实现并用小算例验证)
模型	scikit-learn (核岭回归，高斯/拉普拉斯核)、LightGBM (CPU)
对照文献	Ramakrishnan 等 (JCTC 2015) 的 Δ -ML 原始工作；Zhao 与 Savoie (Chem. Sci. 2023) 的 Δ^2 模型；QM9 上公开的学习曲线基线。仅用于校验与对比，不计入本项目的数据贡献
算力	纯 CPU。xtb 单点在 QM9 分子 (≤ 9 重原子) 上约 0.05–0.3 秒；20,000 个分子约 20 分钟–2 小时 (可 multiprocessing 并行)。KRR 在 3,000 $\times$ 3,000 核矩阵上训练秒级，10 次重抽样 $\times$ 4 规模 $\times$ 2 模型 $\times$ 5 相似度分箱仍在小时级。这条路线的算力完全可控——真正的风险是科学结论而非算力。超出范围的方案 (明确列出)：自算任何 DFT 参考值、训练规模 $> 20,000$ 、深度神经网络势的训练——一律不做

6 · 方法路径

- 下载 QM9，装 xtb + RDKit，在 QM9 提供的几何上批量跑 GFN2-xTB 单点，构建 $\Delta = E_{\text{DFT}}^{\text{atomization}} - E_{\text{xtb}}^{\text{atomization}}$ 的配对表（第 1–6 周）。
- 核实关键定义（第 8 周前必须查清，不要边做边发现）：QM9 内能的单位与参考态、孤立原子的 xTB 参考能取法、是否有已知的 QM9 "不可靠条目"清单（QM9 有一批未通过几何一致性检查的分子，须剔除，剔除清单的来源须注明）。
- 通过方法学校验硬门槛：复现 QM9 直接学习的基线学习曲线。
- 实现特征与相似度度量，构建骨架划分 + 相似度分层的评估协议，协议一旦冻结不得事后调整。
- 跑学习曲线：4 个训练规模 $\times$ 2 个模型 $\times$ 10 次重抽样，同时记录 xTB 原始误差作为无校正基线。
- 按相似度分箱计算增益与置信区间，定位每个训练规模下的 T^* ；画二维热图；检验 H_2 的对数标度关系。
- 独立交叉校验：换一种完全不同的特征 (Morgan 指纹计数向量替代几何描述符) 重算 T^* ，确认阈值的存在与量级不依赖特征选择。

7 · 新颖性边界

- 本课题不声称提出新的 Δ -学习架构或新特征，不声称达到化学精度，不自算 DFT 参考值，不把 QM9 数据计入学生的数据贡献。
- 已有工作具体完成了什么：Ramakrishnan 等 (JCTC 2015) 确立 Δ -ML 范式；Zhao 与 Savoie (Chem. Sci. 2023) 以 GFN2-xTB 为基线、约 167,000 个反应为训练集构建 Δ^2 模型，并在多个外部集上验证近化学精度的可迁移性；相关工作用 KRR 做 PBE $\rightarrow$ 耦合簇的 Δ -XC 校正。这些工作的共同前提是大训练集，且报告的是聚合精度而非适用范围。
- 本项目的贡献是换问题方向且为主结论：在 $10^2 - 10^3$ 的小数据体制下，把研究对象从"精度能到多少"换成"适用范围有多大"，给出一条以化学相似度为坐标的增益消失阈值 T^* ，以及它随训练规模的标度关系。
- 为什么有价值：任何一个资源受限的使用者在决定"要不要花时间做 Δ 校正"时，需要的正是"我的目标分子离我能拿到的参考数据有多远"这个判断，而现有文献只回答"数据足够多时它管用"。
- "增益本来就很小"同样是有效结论，且必须诚实报告：若小数据下 Δ 模型在所有相似度区间上的增益置信区间都跨零，则本项目的结论是"小数据 Δ 学习在 QM9 化学空间上不具备实用价值"——这是一条明确的、可被后续工作引用的负结论，但必须用学习曲线与自助区间证明这不是因为模型没调好（这正是要求 2 种模型 + 10 次重抽样的原因）。

8 · 决策门槛 (go / no-go)

- 第 8 周末 (关键门)：原子化能定义必须自洽、方法学校验硬门槛必须通过。若 QM9 内能与 xTB 能量的参考基准无法对齐 (残差有系统漂移)，立即降级：其一，把学习目标从原子化能改为 HOMO-LUMO 能隙 (QM9 提供，xTB 也直接输出，无参考态问题， Δ 定义干净得多)；其二，改为学习"偶极矩"的 Δ (同样两端都有)。两条降级路径都完整保留"小数据增益 vs 相似度阈值"的主结论框架，只是物理量换了。
 - 第 14 周末：确认在最大训练规模下 Δ 模型相对原始 xTB 的分布内增益是否 ≥ 1.0 kcal/mol。若连分布内都没有可分辨的增益，说明特征或目标量选择有问题——降级为上述 HOMO-LUMO 能隙路径；若换目标量后仍无增益，则如实把"小数据 Δ 学习无增益"作为主结论交付，并把学习曲线、模型对照、特征对照作为"这不是调参失败"的证据链。此时研究问题改写为"小数据 Δ 学习的增益下限在哪"，八块结构其余部分不变。
 - 第 20 周末：确认相似度分箱后每箱测试分子 ≥ 50 。若不足，把分箱从 5 档减到 3 档，或把相似度处理为连续变量做回归而非分箱。
 - 已知困难即研究内容： Δ 学习在小数据下失效是一个很可能发生的结果。本课题从第一天起就把这种可能性写进设计 (两个目标量、两个模型、完整学习曲线)，因此"失效"是被度量的对象而非事故。
 - 预算裁剪顺序：先砍最大训练规模档 (3,000 $\rightarrow$ 1,000)，再砍第二个模型，再砍特征交叉校验，最后把重抽样次数从 10 降到 5。砍到只剩"3 个训练规模 $\times$ KRR $\times$ 3 个相似度分箱"时主结论仍成立。
 - 选择前提：仅在学生兼具 C05 级量子化学理解与 C01 级机器学习经验、且明确接受结论可能为负时启动；建议与 C01/C02 并行，不作为唯一课题。
-

C10 · 从构象系综到实验 $\lambda_{\max}$ 的端到端预测链：误差在构象采样、能量权重、激发态方法三层上的方差分解

推荐优先级 低（最高风险） · 分族 激发态计算 + 化学信息学 · 资源需求 纯 CPU，半经验为主，DFT 仅小子集 · 技能取向 量子化学 + 统计建模 + 大规模批处理 · 失败风险 最高

1 · 研究问题

在把有机染料的最大吸收波长 ($\lambda_{\max}$) 从结构预测到实验值的完整计算链中，端到端误差有多少来自构象采样不足（只取最低能构象 vs 取玻尔兹曼系综）、多少来自能量权重方法（GFN-FF vs GFN2-xTB 给出的权重）、多少来自激发态方法本身？在纯 CPU 笔记本可及的方法组合下，构象层的贡献是否大到足以被分辨？

2 · 研究背景与空白

预测染料 $\lambda_{\max}$ 是激发态计算最经典的应用之一，标准做法是：生成构象系综 → 用某种能量方法给玻尔兹曼权重 → 逐构象算激发能 → 加权得系综光谱。文献报告的 TD-DFT $\lambda_{\max}$ 误差通常在 0.2 – 0.4 eV，归因于泛函选择、溶剂模型、振动效应等。

已有工作走到哪一步：泛函层面的评测最充分（多篇工作比较 BMK、CAM-B3LYP、M06-2X、MN15 在特定染料族上的 $\lambda_{\max}$ 误差）；构象系综的重要性被反复定性强调，已有工具链（CREST 生成系综 + 多泛函 TD-DFT + 高斯展宽合成光谱）实现该流程，也有关于玻尔兹曼加权自身误差的分析；构象生成 workflow 差异对下游性质的影响近年开始被系统比较（如 COSMO-RS 构象 workflow 对比工作）。

空白在：这三层误差几乎从未被放在同一实验对照集上做方差分解——使用者不知道该把有限算力花在“多找几个构象”还是“换个更贵的激发态方法”上。这属于评价维度的转换。之所以排在最后，是因为它有一个真实的失败模式：若激发态方法层的误差（0.3 – 0.5 eV）远大于构象层贡献（可能仅 0.05 eV），构象层就会被淹没在噪声里而无法分辨。本课题的设计必须让“淹没”本身成为可交付的定量结论，否则不应启动。

3 · 可检验假设

H1 在 ≥ 200 个有机染料的实验 $\lambda_{\max}$ 对照集上，端到端误差的方差分解中，激发态方法层贡献 $\geq 60\%$ ，构象采样层贡献 $\leq 15\%$ ；即“多找构象”对精度的边际收益远低于“换更好的激发态方法”。

H2 构象层的贡献在柔性分子子集（可旋转键数 ≥ 5 ）上显著高于刚性子集，比值 ≥ 2.5 （自助 95% 置信区间下界 > 1.5 ）；即存在一个可用可旋转键数事先判断的分子类别，对它们构象采样才值得投入。若 H2 被否定，则给出“构象层贡献与柔性无关”的结论，同样直接改变实践。

4 · 量化验收标准

1. 方法学校验（硬门槛）：用自建管线在文献中已有 $\lambda_{\max}$ 计算报告的 ≥ 15 个染料上，用相同的激发态方法与溶剂模型重算 $\lambda_{\max}$ ，与已发表计算值对比，偏差 ≤ 0.1 eV（约 0.1 eV 对应可见区约 15 – 25 nm，须在报告中给出该换算）；同时在 ≥ 5 个分子上确认自建的玻尔兹曼加权实现与手算一致（权重相对误差 $\leq 1\%$ ）。此条不过关，后续全部方差分解无效。
2. 统计口径预先写死：误差统一以 eV 为单位报告（波长的误差分布强烈非对称，不得只用 nm，但须同时给出 nm 换算以便与实验文献对读）；报 MAE、RMSE 与平均带符号误差。方差分解用嵌套设计（每个分子 $\times$ 构象策略 $\times$ 权重方法 $\times$ 激发态方法的全因子或部分因子设计），用方差分量估计（如 ANOVA 或线性

混合模型) 给出各层贡献比例, 所有比例给 1,000 次自助 (自助单位为分子) 95% 置信区间。溶剂作为协变量或分层变量处理, 不得忽略。

3. 样本量下限: 分子 ≥ 200 个 (覆盖 ≥ 4 个染料骨架族), 每个分子构象 ≥ 10 个 (柔性子集 ≥ 30 个); 柔性/刚性分层各 ≥ 60 个分子。
4. 必须报告"可分辨性": 明确给出本设计能分辨的最小方差分量 (基于自助区间宽度), 并在结论中声明哪些层的贡献低于该分辨阈值——这是本课题诚实性的核心。
5. 至少 2 个激发态方法层级对照 (sTDA-xTB 全量 + 真 TD-DFT 在 ≤ 30 个小分子子集上), 用于估计"廉价方法层"本身引入的偏置。
6. 全部结构、构象系综、激发态计算输入输出、加权与方差分解脚本开源, 一条命令可重跑。

5 · 数据与工具

用途	来源 / 工具
实验 $\lambda_{\max}$	公开的有机化合物光学性质实验数据库 (如 Joung 等, Scientific Data 2020 发布的有机分子吸收/发射光谱数据集, 约 2 万条记录, 含溶剂信息; Figshare 公开下载)。仅用于校验与对比, 不计入本项目的数据贡献
构象生成	RDKit ETKDG + MMFF/xtb 优化去重 (主路径, 成本可控); CREST 元动力学 (可选, 成本高, 见 C02 的实测要求)
能量权重	xtb 的 GFN2-xTB 与 GFN-FF (两套权重构成"权重方法层"), 含 ALPB/GBSA 隐式溶剂
激发态 (主)	sTDA / sTD-DFT (Grimme 组的 stda 程序, 配合 xtb4stda 或 xtb 的相应接口)。极廉价: 30—50 原子染料每构象秒级。其精度低于 TD-DFT (典型偏差 0.3—0.5 eV), 这一点必须显式声明并作为"激发态方法层"的一个水平纳入设计, 而非隐瞒
激发态 (对照)	PySCF 的 TDDFT (pyscf.tddft) 在 ≤ 30 个小染料 (≤ 15 重原子) 上做对照。成本警示: 15 重原子染料在 def2-SVP 下的基态 SCF + TDDFT 单点约 10—60 分钟/构象, 30 分子 $\times$ 5 构象 $\approx$ 25—150 小时, 必须后台分批并严格限制子集规模
算力	纯 CPU。主链 (RDKit 构象 + xtb 优化 + sTDA) 对 200 分子 $\times$ 10 构象 $\approx$ 2,000 个构象, 估计 10—40 小时 (xtb 优化是主要开销, 需本机实测)。超出范围的方案 (明确列出): 全量 TD-DFT、显式溶剂 MD 采样、激发态几何优化与荧光发射预测、任何 20 重原子以上分子的 TD-DFT——一律不做

6 · 方法路径

1. 装 xtb + stda + RDKit + PySCF, 跑通各自的官方算例, 实测单构象 sTDA 与单构象 TD-DFT 的墙钟时间, 据此确定分子数与构象数上限 (第 1—6 周)。
2. 下载实验 $\lambda_{\max}$ 数据集, 写并冻结分子筛选判据 (溶剂限定为 1—2 种常用溶剂、分子量上限、元素限定、去除含金属与自由基体系、去除多重实验值冲突条目)。
3. 通过方法学校验硬门槛: 复现 ≥ 15 个染料的已发表计算 $\lambda_{\max}$; 验证玻尔兹曼加权实现。
4. 核实工具能力边界 (第 10 周前查清): stda 与 xtb 的接口版本兼容性、sTDA 是否支持所需的溶剂模型、PySCF 的 TDDFT 是否支持所需的隐式溶剂 (很可能需要额外模块或不支持, 须核实)。若溶剂模型不可用, 把溶剂降为固定单一体系并在结论中限定适用范围。
5. 按全因子/部分因子设计跑主链: 2 种构象策略 (最低能单构象 vs 系综) $\times$ 2 种权重方法 $\times$ 1 种廉价激发态方法, 覆盖全部 200 分子。
6. 在小子集上加跑 TD-DFT, 估计激发态方法层的偏置量级。

7. 做方差分解与分层分析（柔性 vs 刚性），给全部分量的自助置信区间，并显式报告本设计的最小可分辨分量。
8. 独立交叉校验：对随机 20 个分子把构象数从 10 提到 30，确认构象层的方差分量估计不随构象数上限系统漂移（若漂移，说明 10 个构象不足，须在结论中声明）。

7 · 新颖性边界

- 本课题不声称提出新的激发态方法或新泛函，不声称给出比文献更准的 $\lambda_{\max}$ 预测，不声称发现"构象系综对光谱重要"（这是被反复强调的既有认识），不把公开实验光谱数据库计入学生的数据贡献。
- 已有工作具体完成了什么：多篇工作在特定染料族上系统评测 TD-DFT 泛函（BMK、CAM-B3LYP、M06-2X、MN15 等）的 $\lambda_{\max}$ 误差，典型量级 0.2 – 0.4 eV；已有工具链用 CREST 生成构象系综、按吉布斯自由能做玻尔兹曼加权、以高斯展宽合成溶剂特异的系综光谱；玻尔兹曼加权自身的误差与构象生成工作流的差异也已有专门讨论（如 COSMO-RS 构象工作流对比工作）。
- 本项目的贡献是评价维度的转换且为主结论：把三层误差放进同一个嵌套实验设计做方差分解，回答"有限算力应该花在哪一层"，并给出一个可用可旋转键数事先判断的分子分类判据。
- 为什么有价值：现行做法是把三层都尽量做好，但小资源使用者做不到。一份带置信区间的方差分解直接告诉他们该砍哪一层。
- "构象层贡献低于可分辨阈值"同样是有有效结论，且是本课题很可能出现的主结论：它意味着在廉价激发态方法的误差水平下，投入算力去扩大构象系综是浪费。但这个结论只有在报告了最小可分辨分量之后才成立——否则它与"没做出来"无法区分。这是本课题验收标准第 4 条存在的全部理由。

8 · 决策门槛 (go / no-go)

- 第 6 周末（关键门）：`stda` + `xtb` 工具链必须在本机跑通并给出合理的 $\lambda_{\max}$ （对 3 个已知染料，误差 ≤ 0.6 eV）。若跑不通或安装受阻，立即降级：其一，把激发态方法换成 PySCF 的 TDDFT，同时把分子集缩到 ≤ 60 个小染料（ ≤ 12 重原子）——分子数下降，方差分解的置信区间变宽，但三层设计与主结论框架完全保留；其二，把预测目标从 $\lambda_{\max}$ 换成 HOMO-LUMO 能隙（`xtb` 直接输出，无需任何激发态程序），对照数据改用 QM9 或公开的实验氧化还原电位集，三层分解改为"构象层 / 权重层 / 电子结构方法层"。
- 第 12 周末：方法学校验硬门槛必须通过（ ≥ 15 个染料复现偏差 ≤ 0.1 eV）。若无法与任何已发表计算对齐，降级为放弃与文献计算值对照，改以"两条独立实现的内部一致性"为校验基准（例如 `stda` 命令行 vs Python 侧重算），并在论文中如实说明校验强度下降。
- 第 20 周末（决定成败的一道门）：先用 60 个分子做一次预实验，估计三层方差分量与本设计的最小可分辨分量。若预实验显示构象层与权重层的分量都低于可分辨阈值，则主动把研究问题改写为"在廉价激发态方法的误差水平下，构象与权重层的贡献低于 X eV 的可分辨阈值"，并把剩余算力转投到"激发态方法层的分子族依赖性"（哪些骨架族上 sTDA 相对 TD-DFT 偏得最多）。这不是放弃，而是把预先设计好的第二主结论提为第一主结论——因此本课题从一开始就必须把这两个结论都写进论文提纲。
- 必须提前核实而非边做边发现（第 10 周前）：实验数据集中同一分子多来源 $\lambda_{\max}$ 的分歧幅度。若分歧本身就达 0.1 eV 量级，它构成一个不可消除的噪声地板，必须作为方差分解中的独立一层（实验层）显式建模，否则会被错误地归入方法误差。
- 预算裁剪顺序：先砍 TD-DFT 对照子集（改为引用文献给出的 sTDA-vs-TDDFT 偏差量级），再砍第二种权重方法，再把构象数从 10 降到 5，最后把分子数从 200 降到 100。砍到只剩"100 分子 $\times$ 5 构象 $\times$ 单权重方

法 $\times$ sTDA + 两层分解"时仍能交付主结论。

- **选择前提：**全部十条中风险最高，须同时满足：已完成过至少一个 C02/C05 级量子化学项目；有可长时间后台运行的机器；接受"主结论可能是一个阈值声明而非漂亮的预测模型"；团队另有一条更稳路线（建议 C01 或 C02）并行。不建议单人单课题启动。

C11 · 手机比色法测定天然水体正磷酸盐：色度校正策略在真实基质下的失效边界与偏差预算

推荐优先级 ★★★★★（最稳） · 分族：分析化学 · 比色法 · 预算 ≈ 400 元 · 表征依赖：零外部平台 · 技能取向：定量分析 + 图像数据处理

1 · 研究问题

在含浊度与腐殖质底色的真实水样中，手机 RGB 比色法配合色卡内标校正测定正磷酸盐（orthophosphate），相对国标钼锑抗分光光度法的偏差能否稳定控制在 $\pm 10\%$ 以内？若不能，该偏差能否由手机自身可测的两个量（样品空白区的明度 L^* 与色相角 h ）提前预警，从而给出「何时必须离心或稀释」的可操作判据？

2 · 研究背景与空白

正磷酸盐是水体富营养化（eutrophication）的关键控制指标。国标《水质 总磷的测定 钼酸铵分光光度法》（GB 11893-89）的化学基础是：磷酸根与钼酸铵在酸性条件下生成磷钼杂多酸（phosphomolybdic acid），经抗坏血酸还原为钼蓝（molybdenum blue），于 700 nm 测吸光度，线性范围约 0.01 – 0.6 mg/L（以 P 计）。手机比色的物理本质，是用相机三通道的宽带积分响应替代单波长吸光度：红通道（约 600 – 700 nm）恰好覆盖钼蓝的吸收带，因此 $-\log(R/R_0)$ 可作为 A_{700} 的替代量。但这个替代关系成立有前提——样品中其他吸收与散射在该通道内可忽略。真实水样恰恰不满足这个前提。

已有工作沿“色卡内标 + 色空间变换 + 机器学习”这条主线推进得相当远。2026 年 2 月 [Analytica Chimica Acta](#) 的一篇工作用嵌入式色卡参考集成学习做亚硝酸盐/铵离子的手机比色，把跨设备 R^2 从 0.76 – 0.89 提升到 0.89 – 0.95、MSE 下降 25 – 75%；[Biosensors and Bioelectronics](#)（2024）系统比较色空间，指出把 gamma 编码的 sRGB 线性化到 CIEXYZ 再转 CIELAB 可显著改善跨平台一致性；[PLOS ONE](#)（2020）则论证了受控光箱下“设备无关”（device-independent）校正的可行性。这些工作的共同形态是：自变量是设备与光源，因变量是拟合优度，基质基本是纯水标准溶液。

空白在——没有工作把评估维度换成真实基质下的失效边界：给定基质底色与浊度，各校正策略在哪个点上开始失效，以及能否用手机自身的读数事先判断该失效。这个空白适合一年期中学课题：仪器只需可见分光光度计与手机；参考方法有国标可核；且结论正负都成立——若校正策略在真实基质下依然稳健，是对现有方法适用范围的一次独立扩展检验；若失效，产出的失效判据本身就是可用成果。

3 · 可检验假设

H1 在色卡归一化的 CIELAB 色差（ ΔE_{ab} ）特征下，当样品浊度替代量 $A_{660} \leq 0.05$ 且基质底色 $b \leq 5$ 时，手机法与国标法在 0.05 – 0.60 mg/L(P) 区间的相对偏差 $\leq 10\%$ （Bland – Altman 95% 一致性界限宽度 ≤ 0.12 mg/L）；越过任一阈值后偏差随该量单调上升，在 $A_{660} = 0.20$ 处 $> 25\%$ 。

H2 抗光源鲁棒性与抗基质鲁棒性不能由同一特征同时提供：HSV 的色相 H 对光源色温变化最不敏感（三种光源间偏差 $\leq 5\%$ ）但对基质底色最敏感，色卡归一化 ΔE^*_{ab} 则相反。若 H2 被否认（存在同时最优的特征），该特征即为本课题推荐的默认特征，结论同样成立。

4 · 量化验收标准

1. 方法学校验（硬门槛）：按 GB 11893-89，用磷酸二氢钾基准试剂配制 0.00 / 0.05 / 0.10 / 0.20 / 0.40 / 0.60 mg/L(P) 系列，分光光度法标准曲线 $R^2 \geq 0.995$ ；对自来水与河水各做 3 个浓度水平的加标回收，回收率

95 – 105%、RSD ≤ 5%。此门槛不过，后续全部手机数据无效。

- 2. 重复口径写死：每个浓度点 ≥ 3 次独立配制并独立显色（不是同一比色皿连拍三次），报均值与 RSD；日内 RSD ≤ 5%、日间（≥ 3 个不同日期）RSD ≤ 8%。
- 3. 检出限按 3σ/斜率法计算，σ 取 11 份独立空白的标准差；定量限按 10σ/斜率。分光光度法与手机法分别报告，并明确写「手机法 LOD 高于国标法不是失败，而是本研究要量化的量」。
- 4. 实验设计：3 台不同品牌手机 × 3 种光照（D65 自制灯箱 / 教室荧光灯 / 窗边散射日光）× 4 种基质（去离子水、自来水、河水、腐殖酸+高岭土合成干扰基质，浊度 5 级梯度）全因子，单元内 n = 3。
- 5. 统计口径：方法一致性一律用 Bland – Altman 偏差与 95% 一致性界限，明确写「不用相关系数 r 代替一致性，因为 r 对系统偏差不敏感」；因子效应用三因素方差分析（ANOVA），前提检查预先写死（Shapiro – Wilk 检验残差正态性、Levene 检验方差齐性；任一不满足改用 Kruskal – Wallis 并做 Bonferroni 校正）。失效阈值用分段线性回归（breakpoint regression）估计，断点置信区间由 1000 次自助抽样（bootstrap）给出。
- 6. 可复现性：全部原始照片（含色卡）、ROI 提取与色空间变换脚本（Python + Pillow/NumPy，纯 CPU）、拟合与作图脚本开源，第三方可一键重跑。

5 · 数据与工具

用途	来源 / 工具
参考方法	GB 11893-89《水质 总磷的测定 钼酸铵分光光度法》全文（国家标准全文公开系统 openstd.samr.gov.cn；生态环境部标准库）。仅用于校验与对比，不计入本项目的数据贡献
显色试剂	钼酸铵（AR 25 g，≈60 元）、酒石酸锑钾（AR 5 g，≈40 元）、抗坏血酸（AR 25 g，≈25 元）、浓硫酸（教学装，≈20 元）
基准物	磷酸二氢钾（基准试剂 25 g，≈45 元）
干扰基质	腐殖酸钠（100 g，≈35 元）、高岭土（500 g，≈20 元），用于配制可控浊度/底色的合成基质
真实样品	校园周边河湖水、自来水、含磷洗衣粉稀释液（自采，记录采样点坐标与时间）
定值仪器	可见分光光度计（721/722 型，学校常规配置），700 nm；浊度替代量用同机 660 nm 消光
成像	3 台学生自有手机（记录型号与系统版本，关闭 HDR 与画质增强，锁定 ISO 与曝光，RAW 优先；无 RAW 则记录 JPEG 参数）；24 色标准色卡（国产仿 ColorChecker，≈50 元）；纸箱自制灯箱 + D65 LED（≈60 元）
图像处理	Python 3 + Pillow + NumPy + scikit-image（sRGB → 线性 RGB → CIEXYZ → CIELAB）。纯 CPU，单张 < 0.1 s
统计	Python + SciPy + statsmodels（ANOVA、Bland–Altman、分段回归）；自助法自写并开源
方法对照文献	Anal. Chim. Acta （2026）参考校正 + 集成学习； Biosens. Bioelectron. （2024）最优色空间比较； PLOS ONE （2020）设备无关手机比色。仅作参考对照与新颖性界定，不计入本项目贡献

6 · 方法路径

- 1. 装配试剂并跑通参考方法：按 GB 11893-89 完成完整标准曲线与加标回收，通过验收标准第 1 条后方可进入下一步。

2. 建立成像标准作业程序 (SOP)：固定拍摄距离、锁定曝光与白平衡、色卡与样品同框、逐张记录光源与时间。ROI (感兴趣区) 选取须由脚本按色卡定位自动裁切，不允许人工目视框选，判据须可复现。
3. 核实工具能力边界 (第 8–10 周完成，不要边做边发现)：确认每台手机能否导出 RAW、能否锁定 ISO/曝光、厂商色彩增强能否关闭。不可关闭的项写入不确定度预算，不剔除该机型。
4. 在纯水基质上建立三套校正特征 (原始 R 通道消光、HSV 的 H、色卡归一化 ΔE^*_{ab})，各自建标准曲线并与分光光度法做 Bland–Altman 对照。
5. 加入合成干扰基质做浊度/底色梯度扫描，用分段线性回归定位每套特征的失效断点，检验断点能否由 L^* 、 h 两个手机自测量预测。
6. 在真实河水 / 自来水 / 洗衣粉废水盲样上做外部验证：先用判据预测「是否需前处理」，再实测验证判据命中率 (报灵敏度与特异度)。
7. 独立交叉校验：对同一批盲样用标准加入法 (standard addition) 在真实基质上独立核定回收率，作为不依赖外部仪器的第二条验证通道。

7 · 新颖性边界

- 本课题不声称提出新的显色化学、新的手机比色装置或新的机器学习模型。钼锑抗法、色卡内标、色空间变换均为已发表方法。
- 已有工作完成到：内标色卡 + 集成学习把亚硝酸盐/铵的跨设备手机比色 R^2 提到 0.89–0.95 ([Anal. Chim. Acta](#), 2026)；系统比较色空间并推荐线性化后的 CIELAB ([Biosens. Bioelectron.](#), 2024)；受控光箱下设备无关校正的可行性 ([PLOS ONE](#), 2020)。它们的自变量是设备与光源，因变量是拟合优度，基质基本是纯水标准溶液。
- 本项目的贡献是评价维度的转换，且构成主结论：自变量换成真实基质的浊度与底色，因变量换成相对国标法的系统偏差与一致性界限，交付物是一条可操作的失效判据 (用手机自身读数预警何时必须前处理) 与一份完整的不确定度预算。
- 价值：手机比色的实际使用场景 (野外、课堂、家庭) 恰恰是有基质干扰的场景，而现有文献的适用范围声明基本停在受控条件。给出失效边界，比再提高一次 R^2 更能决定这类方法能否被真的用起来。
- 若结果显示「各策略在真实基质下均稳健、差异不显著」，这同样是有效结论，但必须用误差棒证明本研究有能力分辨 10% 量级的偏差 (由验收标准第 2、3 条保证)。

8 · 决策门槛 (go / no-go)

- 第 4 周末：验收标准第 1 条必须通过 ($R^2 \geq 0.995$ 且加标回收 95–105%)。未通过 → 降级路径 A：把分析物换成亚甲基蓝或市售合成食用色素的水溶液 (显色稳定、无显色反应变量)，研究问题、假设结构与统计口径全部保留，只是从体系中移除“化学显色不确定度”；主结论 (色度校正策略的基质失效边界) 不受影响。
- 第 10 周末：确认至少 2 台手机可锁定曝光与白平衡。3 台全部不可控 → 降级路径 B：改为「固定灯箱 + 单机多日重复」设计，把“跨设备”因子换成“跨日期 / 跨批次试剂”因子，全因子设计与断点分析框架不变。
- 第 20 周末：合成干扰基质上必须已观察到至少一套特征出现可定位的失效断点 (断点自助置信区间宽度 $\leq$ 梯度跨度的 40%)。若所有特征都不失效 → 主动加大干扰强度 (腐殖酸至 20 mg/L、浊度至 50 NTU)，并把「在中学可及浓度范围内未观察到失效」写成定量结论 (附最小可分辨偏差)。

- 第 34 周末：真实盲样验证完成，判据灵敏度应 ≥ 0.80 。低于 0.80 则不修补判据，如实报告适用范围并做失败案例分析——这个困难本身就是研究内容。
 - 第 44 周数据冻结，转入英文报告与英文 PPT；第 48 周完成实验视频剪辑。
 - 实验视频要求：固定机位记录三段——(a) 国标法标准曲线制备与比色皿读数；(b) 灯箱内一次完整的手机拍摄 SOP；(c) 真实水样前处理与盲样测定。每段含操作者出镜、试剂标签特写与仪器读数特写，累计 ≥ 8 分钟，原始素材保留至答辩结束。
 - 安全与废液：酒石酸锑钾属有毒品，须由教师配制母液并锁柜管理，学生只接触稀释后工作液，全程手套 + 护目镜；浓硫酸稀释遵循「酸入水」，在通风橱或开窗处操作。含钼、锑的显色废液严禁排入下水道，按「含重金属显色废液」单独收集、贴签注明成分与日期，交学校危废渠道。不涉及剧毒品、易制毒/易制爆化学品、氢氟酸、高压反应釜与明火有机溶剂操作。
 - 预算裁剪顺序：先砍手机台数（3 $\rightarrow$ 2，跨设备因子降为两水平，H1 仍可检验）；再砍标准色卡（改自印色块并用分光光度计标定反射率，精度下降但框架不变）；灯箱与显色试剂不可砍。
 - 适配学生：适合愿意做大量重复性定量操作、能写基础 Python 图像处理的学生。不适合追求"造一个新器件"的学生。
-

C12 · 乙酸异戊酯三条酯化路线的绿色度实测对比：六项绿色化学指标的排序一致性与反转条件

推荐优先级 ★★★★★ · 分族：绿色化学 · 合成 · 预算 ≈ 500 元 · 表征依赖：零外部平台（皂化滴定替代 GC） · 技能取向：有机合成 + 物料衡算

1 · 研究问题

对同一目标产物（乙酸异戊酯，isoamyl acetate）的三条中学可执行酯化路线，六项常用绿色化学指标——原子经济性（atom economy, AE）、反应质量效率（reaction mass efficiency, RME）、E-因子（E-factor）、过程质量强度（process mass intensity, PMI）、EcoScale、单位产物电耗——给出的路线优劣排序是否一致？若不一致，排序反转由哪一项核算边界（是否计入洗涤水、是否计入催化剂回收、是否计入溶剂回收率）驱动，该驱动关系能否写成一条可预测的判据？

2 · 研究背景与空白

绿色化学指标的作用，是把“哪条路线更绿”从口号变成可计算量。AE（Trost, 1991）只看化学计量式；E-因子（Sheldon, 1992）= 废物质量 / 产物质量，把溶剂与水洗计入；PMI = 投入总质量 / 产物质量；RME 介于两者之间；EcoScale（Van Aken, 2006）是包含收率、成本、安全、能耗、纯化难度的罚分制综合评分。指标之间的关键差别不在公式，而在核算边界：AE 的边界由化学方程式唯一确定，E-因子与 PMI 的边界却是操作者自己划的。

已有工作到哪一步：Sheldon 反复指出 E-因子因概念简单而被广泛采用，而 PMI、RME 未获同等接受度；Processes 10(7):1274（2022）的绿色指标综述系统整理了各指标定义与适用性；教学文献已把 AE/RME/PMI/EcoScale 写入本科有机实验（含香蕉油 Fischer 酯化的绿色化改造，World J. Chem. Educ., 2024），EcoScale 也被提出作为本科绿色化学教学与评分框架（2018）。这些工作的共同形态是：给学生一套指标去算已知路线，或对文献路线做纸面核算。

空白在——很少有工作把“指标之间是否一致”本身当作被检验的对象，用同一实验者、同一台天平、同一套物料衡算口径实测三条路线的全部质量流，再定量回答「指标排序在什么核算边界下会反转」。这个空白适合一年期中学课题：三条路线都是经典酯化，安全可控；核心工作量是称量与记账而非高端表征；且“排序一致”与“排序反转”都是有效结论。

3 · 可检验假设

H1 三条路线在 AE 上的排序与在 E-因子上的排序不一致（Kendall $\tau \leq 0.34$ ，即三条路线中至少有一对发生反转）；反转的充分条件是「两条路线 AE 差 ≤ 5 个百分点，而水洗用水量差 ≥ 3 倍」。

H2 把溶剂/催化剂回收率 r 作为连续变量，E-因子对 r 的敏感度 $|dE/dr|$ 显著大于 EcoScale 对 r 的敏感度；且存在临界回收率 r_c ，越过 r_c 后均相浓硫酸路线在 E-因子上反超非均相树脂路线。 r^* 由实测质量流外推并给出自助置信区间。

4 · 量化验收标准

1. 方法学校验（硬门槛）：无 GC 条件下用皂化滴定法（saponification value）测产物中酯的质量分数。门槛为：用分析纯乙酸异戊酯与异戊醇配制 5 个已知质量分数（60/70/80/90/100%）的合成样，皂化滴定测得值对配制值的线性 $R^2 \geq 0.995$ 、各点相对误差 $\leq 3\%$ 、每点 3 次平行 $RSD \leq 2\%$ ；同时以折光率（阿贝折光

仪；无则用密度瓶测密度）作第二独立通道，两通道对同一样品的纯度差 ≤ 5 个百分点。此门槛不过，所有收率及由收率导出的指标全部无效。

- 2. 每条路线 ≥ 3 次独立完整重复（独立投料、独立后处理），报每项指标的均值与 RSD；指标 RSD $> 15\%$ 的路线须补做至 5 次。
- 3. 物料衡算闭合度：每次实验的质量闭合（投入总质量 - 回收总质量 - 产物质量 - 记录损失）偏差 $\leq 3\%$ ，否则该次作废并记录原因。所有质量用同一台 0.1 mg 分析天平称量，含空瓶皮重、洗涤水、干燥剂。
- 4. 能耗：用插座式功率计记录加热与搅拌实际电能（kW · h），换算为单位产物电耗（kW · h/mol），每条路线报 3 次重复的均值与 RSD；功率计须先用已知功率电阻标定。
- 5. 统计口径：路线间差异用配对设计（同批试剂、同周内完成一组三路线）+ 重复测量方差分析；前提检查预先写死（残差 Shapiro – Wilk 正态性、Mauchly 球形性检验，不满足用 Greenhouse – Geisser 校正；重复数 < 5 时改用 Friedman 检验）。指标排序一致性用 Kendall's τ 与 Spearman ρ 双报，置信区间由 2000 次自助抽样给出。敏感性分析必须同时报单因素扫描与全局蒙特卡洛（对回收率、洗涤水量、催化剂用量三个参数各设合理分布， 10^4 次抽样，报排序反转发生的概率）。
- 6. 可复现性：物料衡算电子表格模板、蒙特卡洛与统计脚本开源；每次实验的原始称量记录（照片 + 表格）随论文附上。

5 · 数据与工具

用途	来源 / 工具
指标定义与公式	Sheldon 的 E-factor 系列综述（ Green Chem. ）； Processes 10(7):1274（2022）绿色指标综述；Van Aken EcoScale 原文（ Beilstein J. Org. Chem. 2006, 2:3，开放获取）。 仅用于定义与对照，不计入本项目贡献
路线 A（均相酸催化）	冰乙酸（500 mL， ≈ 30 元）、异戊醇（500 mL， ≈ 60 元）、浓硫酸（教学装， ≈ 20 元）、饱和碳酸氢钠、无水硫酸镁
路线 B（非均相酸催化）	强酸性阳离子交换树脂（001 $\times$ 7 或同类，100 g， ≈ 80 元），可回收复用并称重
路线 C（无溶剂过量醇驱动）	对甲苯磺酸一水合物（100 g， ≈ 45 元）
产物定量	皂化滴定（KOH-乙醇标准溶液 + 盐酸回滴，酚酞指示剂）；阿贝折光仪；无折光仪时用 10 mL 密度瓶 + 分析天平测密度，须先用纯品建立密度-纯度曲线
能耗	插座式功率计（ ≈ 40 元，精度 $\pm 1\%$ ）
称量	0.1 mg 分析天平（学校常规配置）
加热	电加热套或恒温油浴（ 禁止明火 ）+ 球形冷凝管回流
计算与统计	Python + NumPy / SciPy / pandas（蒙特卡洛与统计）；Excel 物料衡算模板
外部依赖（可选）	若需结构确证，须联系高校共享平台送 ^1H NMR 或 GC-MS 样品—— 属外部依赖，列入 go/no-go 风险项 。不依赖该表征的主结论：指标排序一致性只依赖质量流与酯含量，皂化滴定 + 折光/密度双通道已足以支撑

6 · 方法路径

- 1. 建立并校验定量方法：配制合成样，完成皂化滴定与折光/密度双通道的验收标准第 1 条；未通过不得进入合成阶段。

2. 建立物料衡算 SOP 与记账模板：明确每一项质量（试剂、催化剂、洗涤水、干燥剂、回收溶剂、废液、产物）的称量时点与归属规则；规则须书面化且三条路线一致，任何主观归属（如"损失"）都要有明文判据。
3. 逐条路线做条件摸索（各 2-3 次预实验），固定温度、时间、醇酸比，使三条路线收率落在可比区间。若某条路线收率 < 40%，如实记录为该路线的真实表现，不得"调优到齐平"。
4. 正式重复实验：每周完成一组三条路线（配对设计），共 ≥ 3 组，同步记录电能。
5. 计算六项指标，并做核算边界敏感性分析（三种边界：仅反应物 / 反应物+溶剂 / 反应物+溶剂+水洗+催化剂），得到 3×6 指标矩阵 $\times 3$ 种边界。
6. 做全局蒙特卡洛，估计排序反转概率与临界回收率 r^* ，给出自助置信区间。
7. 独立交叉校验：用第二种独立方式核定收率（对反应液做酸值滴定，测残余乙酸并反推转化率），与皂化滴定结果比较，偏差 $\leq 5\%$ 方可采信。

7 · 新颖性边界

- 本课题不声称提出新反应、新催化剂或新指标；三条酯化路线与六项指标均为教科书与文献已有内容。
- 已有工作完成到：Sheldon 等确立并推广 E-因子与 PMI 并指出各指标接受度差异；Processes (2022) 综述完成指标的系统分类；本科教学文献（含香蕉油 Fischer 酯化绿色化改造）已把多指标核算引入实验课，EcoScale 已被提出作为教学评估框架。这些工作要么是纸面核算，要么是让学生"算一遍"，没有把指标间的排序一致性当作被检验的假设。
- 本项目的贡献是问题方向的转换，且构成主结论：从"哪条路线更绿"转为"不同指标是否给出同一答案、在什么核算边界下答案反转"，并用同一实验者的实测质量流 + 蒙特卡洛给出反转概率与临界回收率 r^* 。
- 价值：绿色度指标的实际用途是决策依据。若不同指标对同一组路线给出相反排序，则任何只报单一指标的"绿色化"声明都不可靠——这是可以被中学实验直接检验、且有实际后果的命题。
- 若结果是"六项指标排序完全一致"，同样是有效结论（对指标稳健性的一次独立支持），但必须由验收标准第 2、5 条的误差棒证明本研究有能力分辨排序反转。

8 · 决策门槛 (go / no-go)

- 第 6 周末：皂化滴定 + 折光/密度双通道的方法学校验必须通过。未通过 $\rightarrow$ 降级路径 A：目标产物换成乙酸乙酯（沸点 77°C ，密度-折光双通道对纯度更灵敏，标准品更便宜），三条路线的化学骨架与全部指标框架不变。
- 第 12 周末：三条路线各完成 ≥ 2 次成功合成且质量闭合偏差 $\leq 3\%$ 。若树脂路线收率 < 20% 且两轮条件摸索无改善 $\rightarrow$ 降级路径 B：路线 B 换成分子筛脱水 / 微量浓硫酸路线（催化剂用量降一个数量级），"三路线 $\times$ 六指标"的对比结构不变。
- 第 26 周末：3 组配对重复完成，指标矩阵成型。若六项指标排序完全一致且蒙特卡洛显示反转概率 < 5% $\rightarrow$ 不换课题，把重心转到 H_2 （临界回收率 r^* 的定位与外推），主结论框架保留。
- 第 36 周末：敏感性分析与交叉校验完成。若酸值滴定与皂化滴定给出的收率偏差 > 5%，须把偏差来源（水分、残余醇、副产物醚）作为独立一节写入，不得掩盖。
- 外部表征风险项：若指导教师要求结构确证，须在第 30 周前联系高校共享平台预约 ^1H NMR 或 GC-MS 送样（1-2 个代表性样品即可）。联系不上则主结论不变，论文中明确写"产物身份由沸点、折光率、密度、皂化值四项一致性判定，未做波谱确证"。

- **实验视频要求：**三条路线各录一段完整流程（称量投料 → 回流 → 后处理 → 产物称量），含天平读数与功率计读数特写；另录一段皂化滴定终点判定近景。累计 ≥ 12 分钟。
 - **安全与废液：**浓硫酸与对甲苯磺酸具强腐蚀性，全程护目镜 + 耐酸手套；回流必须用电加热套或油浴，**严禁明火**；异戊醇与乙酸异戊酯易燃，单次用量 ≤ 30 mL，操作区禁明火并备灭火毯与砂桶。有机废液按"含酯/醇有机废液"单独收集；酸性水洗液先用碳酸氢钠中和至 pH 6 – 8 再交危废；离子交换树脂回收后单独存放，不入下水道。不涉及剧毒品、易制毒/易制爆化学品、氢氟酸与高压反应釜。
 - **预算裁剪顺序：**先砍阿贝折光仪（改密度瓶，增加标定工作量）；再砍路线 C（三条降为两条，Kendall τ 失效，改报成对反转的符号检验，H2 仍完整成立）；分析天平与功率计不可砍。
 - **适配学生：**适合细致、能坚持记账与重复实验的学生。不适合期待"合成出一个新化合物"的学生——本课题的成果形态是数据与判据，不是新物质。
-

C13 · 中学可得电极 + 千元级恒电位仪的电化学计量学边界：哪些伏安参数可自校验、哪些必须弃报

推荐优先级 ★★★★★ · 分族：电化学 · 方法学 · 预算 ≈ 1900 元（含一次性仪器） · 表征依赖：零外部平台 · 技能取向：电化学基础 + 误差分析

1 · 研究问题

在没有商用恒电位仪可作对照的条件下，用千元级国产恒电位仪（或 Arduino 自制机）配合中学可得的电极（铅笔芯工作电极、自制 Ag/AgCl 参比、石墨对电极），哪些循环伏安参数（峰电流 i_p 、峰电位差 ΔE_p 、峰电流比 i_{pa}/i_{pc} 、由 Randles – Sevcik 方程反推的扩散系数 D 与有效电极面积 A ）能通过仅依赖内部一致性与文献公认常数的协议实现自校验？无法自校验的参数，其系统偏差能否用一台哑元电池（dummy cell）预先标定并扣除？

2 · 研究背景与空白

循环伏安法（cyclic voltammetry, CV）是电化学最基础的定量工具。对扩散控制的可逆体系，Randles – Sevcik 方程给出 $i_p = 2.69 \times 10^5 \text{ n}^{3/2} A D^{1/2} C v^{1/2}$ (25°C)，即峰电流正比于扫速平方根。铁氰化钾/亚铁氰化钾体系 ($[\text{Fe}(\text{CN})_6]^{3-/4-}$, 0.1 mol/L KCl 支持电解质) 是公认的可逆内标探针，其扩散系数有被反复引用的文献值（约 $7.6 \times 10^{-6} \text{ cm}^2/\text{s}$ ）。但低成本仪器有两类固有缺陷：未补偿溶液电阻造成的 iR 降会把 ΔE_p 从理论的 $59/n \text{ mV}$ 撑开；有限的 ADC 采样率与 DAC 分辨率会在高扫速下削平峰形。这两类误差都不随机、不能靠多次重复消除。

已有工作到哪一步：DIY 恒电位仪的验证已经相当成熟。CheapStat ([PLOS ONE](#), 2011, 约 80 美元)、基于 ESP32 的手机控恒电位仪 (2022)、Arduino 平台机 ([J. Chem. Educ.](#) 系列, 2021)、UWED 无线机 ([Anal. Chem.](#), 2018)、支持阻抗谱的 HELPStat ([ACS Electrochem.](#), 2024)——几乎全部采用同一套验证范式：用专业三电极（玻碳/铂/商品 Ag-AgCl）跑铁氰化钾，做 $i_p - v^{1/2}$ 线性拟合，并与一台商用恒电位仪并排对比；UWED 甚至给出了“峰幅度一致低于商用机约 10%”的量化结论。

空白在——这些验证的前提是“手边有一台商用仪器和一套专业电极”，而这恰恰是中学实验室没有的两样东西。没有工作回答：在没有金标准仪器、且电极本身也是自制的情况下，哪些参数还能被可靠地报告。这个空白适合一年期中学课题：所有耗材便宜，公认常数公开，且结论正负都成立——能自校验的参数清单是成果，不能自校验的参数清单同样是成果（而且对后来者更有用）。

3 · 可检验假设

H1 采用“双探针互标”协议——用 $\text{K}_3[\text{Fe}(\text{CN})_6]$ 的文献 D 值反推有效电极面积 A ，再用该 A 盲测 $\text{K}_4[\text{Fe}(\text{CN})_6]$ 的 D ——可在 10 – 200 mV/s 扫速窗口内把 D 的回收偏差控制在 $\leq 20\%$ （相对文献值），且该偏差在 3 支独立制备的铅笔芯电极之间的 $\text{RSD} \leq 15\%$ 。

H2 ΔE_p 与 i_{pa}/i_{pc} 这两个“可逆性判据”在低成本装置上被未补偿电阻系统性污染： ΔE_p 相对 59 mV 的超出量随 $v^{1/2}$ 近似线性增长，其斜率可由已知电阻的哑元电池预先测得；扣除该 iR 贡献后， ΔE_p 应回落至 $59 \pm 15 \text{ mV}$ 。若扣除后仍不回落，则 ΔE_p 被判为该装置不可靠报的参数，并给出该判定的定量依据。

1. 方法学校验（硬门槛，两级）：(a) 哑元电池级：用已知阻值的金属膜电阻（1 k Ω 、10 k Ω ， $\pm 1\%$ ）串联已知电容搭建 dummy cell，跑线性扫描。纯电阻元件上 $i-E$ 曲线须为直线，拟合斜率给出的电阻值与标称值偏差 $\leq 3\%$ ；电位输出用万用表核对，设定值与实测值偏差 ≤ 5 mV。(b) 化学体系级：5 mmol/L $K_3 [Fe(CN)_6]$ / 0.1 mol/L KCl，扫速 10/20/50/100/200 mV/s， i_p 对 $v^{1/2}$ 线性 $R^2 \geq 0.995$ ；由文献 $D = 7.6 \times 10^{-6}$ cm²/s 反推的有效面积与铅笔芯几何面积之比落在 0.7 – 1.5 之间（越界须给出电极表面粗糙度或活性面积不足的定量解释）。两级中任一不过关，后续全部电化学数据无效。
2. 重复口径写死：每个条件 ≥ 3 支独立制备的电极 $\times$ 3 次独立扫描，报电极间 RSD 与电极内 RSD，两者分开报告——这是区分“仪器噪声”与“电极制备重现性”的唯一手段。
3. 检出限：在应用演示环节（果汁中抗坏血酸的差分脉冲伏安定量）按 3σ /斜率法给出 LOD、 10σ /斜率给出 LOQ， σ 取 11 份独立空白支持电解质的峰电流标准差。
4. 参数可靠性判定预先写死：某参数被判为“可自校验”须同时满足——(i) 独立电极间 RSD $\leq 15\%$ ；(ii) 与文献公认值的偏差 $\leq 20\%$ ；(iii) 该偏差在扫速窗口内无系统趋势（用残差对 $v^{1/2}$ 的线性回归斜率检验， $p > 0.05$ ）。三条中缺一即判为“不可靠报”，并给出不可靠的量化描述。
5. 统计口径：参数偏差的显著性用单样本 t 检验对照文献值，前提检查写死（Shapiro – Wilk 正态性； $n < 8$ 或不满足正态性时改用 Wilcoxon 符号秩检验）。不确定度预算按 GUM 框架逐项列出（浓度配制、面积标定、温度、 iR 降、采样率），报合成标准不确定度与 $k = 2$ 的扩展不确定度。
6. 可复现性：全部原始伏安数据（CSV）、哑元电池图纸、扫描控制与拟合脚本开源；恒电位仪若为自制，须给出完整电路图与固件。

5 · 数据与工具

用途	来源 / 工具
恒电位仪	国产入门恒电位仪 (≤ 2000 元档, 须核实其是否开放原始电流-电位数据导出、最高扫速与采样率——第 6 周前核实清楚); 备选 Arduino/ESP32 自制方案 (≈ 300 元, 电路参考 CheapStat / J. Chem. Educ. 2021 开源设计)
参考文献值	[Fe(CN) ₆] ³⁻ 在 0.1 mol/L KCl 中的 $D \approx 7.6 \times 10^{-6} \text{ cm}^2/\text{s}$ (电分析化学教科书与 Konopka–McDuffie 经典测定)。仅作校验基准, 不计入本项目贡献
探针试剂	铁氰化钾 (AR 100 g, ≈45 元)、亚铁氰化钾 (AR 100 g, ≈40 元)、氯化钾 (AR 500 g, ≈25 元)
工作电极	2B 绘图铅笔芯 0.5/0.7 mm (≈10 元); 备选: 商品丝网印刷碳电极 SPCE (10 片装, ≈200 元)
参比电极	银丝 (0.5 mm, ≈60 元) + 次氯酸钠漂白液氧化自制 Ag/AgCl, 饱和 KCl 盐桥; 须用商品参比或标准氢电位对照核定电位漂移
对电极	石墨棒 / 铂丝 (若校内有)
哑元电池	金属膜电阻 (±1%, 1 kΩ / 10 kΩ) + 薄膜电容 (≈20 元) + 面包板
应用演示样品	市售橙汁/柠檬汁; 抗坏血酸标准品 (AR 100 g, ≈30 元); 2,6-二氯靛酚滴定法 (GB 5009.86) 作独立比对方法
数据处理	Python + NumPy / SciPy / pandas (基线扣除、峰提取、线性拟合、不确定度传播); 纯 CPU
对照文献	CheapStat (PLOS ONE , 2011)、ESP32 potentiostat (2022)、UWED (Anal. Chem. , 2018)、HELPStat (ACS Electrochem. , 2024)。仅作参考对照与新颖性界定, 不计入本项目贡献

6 · 方法路径

1. 采购/组装仪器, 先做哑元电池校验 (验收标准第 1(a) 条), 未过关不得接触化学体系。同期核实仪器能力边界: 最高扫速、采样率、原始数据导出格式、是否有 iR 补偿功能。
2. 建立电极制备 SOP: 铅笔芯的封装长度、每次实验前的打磨/活化步骤、浸没深度的固定方式, 全部书面化并可复现; 几何面积由游标卡尺测量并记录。
3. 完成化学体系级校验 (第 1(b) 条), 得到有效电极面积 A 及其不确定度。
4. 执行"双探针互标": 以 $K_3 [\text{Fe}(\text{CN})_6]$ 标定的 A 盲测 $K_4 [\text{Fe}(\text{CN})_6]$ 的 D, 检验 H1。
5. 用哑元电池在多个已知电阻下测定 iR 降对 ΔE_p 的贡献斜率, 回代扣除真实体系的 ΔE_p , 检验 H2。
6. 建立参数可靠性清单: 对 i_p 、 ΔE_p 、 i_{pa}/i_{pc} 、D、A 逐项按验收标准第 4 条判定, 并给出完整不确定度预计算表。
7. 独立交叉校验与应用演示: 用清单里判为"可靠"的参数做果汁抗坏血酸的 DPV 定量, 结果与 2,6-二氯靛酚滴定法 (国标方法) 做 Bland – Altman 比对, 验证清单的实际有效性。

7 · 新颖性边界

- 本课题不声称设计新的恒电位仪、新电极或新方程。Randles – Sevcik 方程、铅笔芯电极、自制 Ag/AgCl 参比、哑元电池校验都是已有技术。

- 已有工作完成到：CheapStat、UWED、ESP32/Arduino 平台机、HELPStat 等均已用专业电极 + 商用仪器并排对照证明低成本机可做定量 CV，并给出诸如“峰幅度低约 10%”与台式机统计可比”的结论。
- 本项目的贡献是问题方向的转换，且构成主结论：把问题从“低成本机能否逼近商用机”改成“在没有商用机可对照、电极也是自制的现实条件下，哪些参数还配被报告”，交付物是一份带判定规则的参数可靠性清单与一份 GUM 不确定度预算，以及一套只依赖公认常数与内部一致性的自校验协议。
- 价值：中学与欠资源实验室是低成本电化学的真正使用者，而现有验证范式恰恰假设了他们没有的东西。一份“哪些数字可以信”的清单，直接决定这类装置在教学与初级研究中能否被负责任地使用。
- 若结果显示所有参数都通过自校验，这同样是有效结论（对低成本电化学可靠性的一次独立支持），但必须由验收标准第 2、4 条的误差棒证明本研究有能力分辨 20% 量级的系统偏差。

8 · 决策门槛 (go / no-go)

- 第 6 周末：仪器能力核实完成，且哑元电池校验通过（电阻拟合偏差 $\leq 3\%$ ）。未通过 → 降级路径 A：改用 Arduino/ESP32 开源方案自制恒电位仪（电路公开、可自行修 iR 与采样率），全部研究框架保留，且“自制机与商品廉价机的哑元电池表现对比”成为额外一节。
- 第 12 周末：铁氰化钾体系上 $i_p - v^{1/2}$ 线性 $R^2 \geq 0.995$ 。未达标 → 降级路径 B：工作电极从铅笔芯换成商品丝网印刷碳电极（SPCE， ≈ 200 元），把“电极自制”这一变量移出体系，参数清单的判定框架不变，并在论文中把“铅笔芯电极不满足线性判据”作为定量结论报告。
- 第 22 周末：H1 的双探针互标必须给出明确结论（D 回收偏差是否 $\leq 20\%$ ）。若偏差 $> 50\%$ 且原因不可定位 → 降级路径 C：把研究对象从“扩散系数的绝对准确度”降为“相对测量的重现性”（同一装置在不同日期、不同电极批次上的 i_p 重现性），主结论框架（参数可靠性清单 + 不确定度预算）完全保留，只是清单里“绝对量”一栏改判为不可靠。
- 第 32 周末：果汁抗坏血酸的 DPV 定量与国标滴定法的 Bland – Altman 比对完成。若两法偏差 $> 30\%$ ，如实报告并把偏差归因分析写成独立一节——这个困难本身就是研究内容。
- 第 44 周数据冻结；第 48 周完成实验视频。
- 实验视频要求：录制四段——(a) 哑元电池校验全过程含万用表读数特写；(b) 铅笔芯电极制备 SOP；(c) 一次完整 CV 扫描与软件界面录屏；(d) 果汁样品的 DPV 测定与滴定法对照。累计 ≥ 10 分钟。
- 安全与废液：铁氰化钾/亚铁氰化钾中氰基以稳定络合形式存在，常温常压下低毒，但严禁与酸混合、严禁加热蒸干、严禁与强氧化剂共存（可能释放 HCN）。所有含氰络合物废液必须保持碱性（ $pH > 9$ ）单独收集、贴签，交学校危废处置，绝对禁止倒入下水道或与酸性废液混合。银丝氧化用家用漂白液，操作在通风处，含银废液单独回收。不涉及剧毒品、易制毒/易制爆化学品、氢氟酸、高压反应釜与明火有机溶剂操作。
- 预算裁剪顺序：先砍丝网印刷电极备选（只用铅笔芯，风险上升但主结论仍成立）；再砍商品恒电位仪改为 Arduino 自制（省 1500 元，但需增加 4 – 6 周电子调试，须在第 0 周就决定，不可中途改）；哑元电池元件与探针试剂不可砍。
- 适配学生：适合对仪器原理与误差分析有兴趣、能忍受“结论可能是‘这个参数不能报’”的学生。仅在学生已具备欧姆定律级电路基础并明确接受“可能测不准”这一风险时才启动。

C14 · 总酚含量能否代表抗氧化力：六种植物基质中非酚还原物的当量拆解与替代关系失效判据

推荐优先级 ★★★★★ · 分族：天然产物 · 分析化学 · 预算 ≈ 900 元 · 表征依赖：零外部平台 · 技能取向：分光光度法 + 滴定分析 + 相关性统计

1 · 研究问题

在六种中国常见植物基质中，Folin – Ciocalteu 法（FC 法）测得的"总酚含量"（total phenolic content, TPC）与自由基清除活性之间的相关性，有多大比例是由非酚类还原物（抗坏血酸、还原糖）贡献的？扣除这部分贡献后，纯酚 TPC 是否仍能作为抗氧化力的有效替代指标；若不能，能否用一个可独立测得的量（抗坏血酸当量占 FC 总响应的比例 f_{AA} ）给出"何时不得用 TPC 代替抗氧化力"的操作判据？

2 · 研究背景与空白

FC 试剂是磷钼酸 – 磷钨酸混合物，在碱性条件下被还原剂还原生成蓝色络合物，760 nm 有吸收。Singleton 与 Orthofer 的经典文献早已把话说明白：FC 法测的不是多酚含量，而是样品的还原能力。DPPH（2,2-二苯基-1-苦基肼基自由基）清除法与铁氰化钾还原力法（Prussian blue 法）测的同样是还原/供氢能力。三者高度相关几乎是必然的——问题在于这种相关性能否支撑"TPC 高 $\Rightarrow$ 富含多酚 $\Rightarrow$ 抗氧化好"这条被大量文献默认使用的推理链。

已有工作到哪一步：干扰的存在是公认的。文献明确记录 FC 法会与还原糖（果糖、蔗糖）、芳香胺、二氧化硫、有机酸（尤其抗坏血酸）及其他烯二醇反应；蜂蜜 TPC 因还原糖而被系统高估，已有改良 FC 法（[Applied Sciences](#), 2023）；绿茶与商品果汁中葡萄糖与抗坏血酸对 FC 与 Fast Blue BB 法的影响已被比较（2019）；提高 FC 特异性的方案（固相萃取柱、按抗坏血酸还原当量做校正、氧化预处理）也已提出。这些工作的形态是：在单一基质上证明"有干扰"并给出一种校正手段。

空白在——没有工作把"TPC 作为抗氧化力替代指标"这一命题本身当作被检验的对象，跨多个基质定量拆解 FC 响应的组成，再回答"拆解之后替代关系是否还成立、在哪类基质上崩溃"。这个空白适合一年期中学课题：全部方法是分光光度法与滴定，无需色谱；每步都有国标或经典方法可校；且"替代关系成立"与"替代关系崩溃"都是有效结论。

3 · 可检验假设

H1 在六种基质的全集上，未校正 TPC 与自由基清除活性（以 $1/IC_{50}$ 表征）的 Pearson $r \geq 0.85$ ；但用 PVPP（聚乙烯聚吡咯烷酮）选择性吸附扣除酚类、并用 2,6-二氯酚滴定法扣除抗坏血酸当量之后，纯酚 TPC 与清除活性的 r 下降 ≥ 0.20 （即表观相关性被非酚组分显著"借用"）。

H2 存在一个可独立测得的判据量 $f_{AA} = \text{抗坏血酸当量} / \text{FC 总响应}$ ：当 $f_{AA} > 0.30$ 时，用未校正 TPC 预测自由基清除活性的相对误差 $> 25\%$ ； $f_{AA} \leq 0.30$ 时相对误差 $\leq 15\%$ 。若 H2 被否认（不存在这样的阈值），则须报告非酚干扰的实际分布并说明为何无法用单一判据刻画——该负结论同样有价值。

4 · 量化验收标准

- 方法学校验（硬门槛，三通道）：(a) 没食子酸标准曲线（0 – 200 mg/L）线性 $R^2 \geq 0.995$ ；(b) 合成混合样拆解校验：配制 5 个已知比例的没食子酸 + 抗坏血酸 + 葡萄糖三元混合样，PVPP 吸附法对多酚部分的回收

- 率 95 – 105%、RSD $\leq$ 5%；(c) 抗坏血酸用 2,6-二氯酚酚滴定法（GB 5009.86）测定，对配制值的相对偏差 $\leq$ 5%。三通道任一不过关，全部拆解结果与由其导出的相关性结论无效。
2. 样本量与重复：6 种基质（绿茶、柑橘、山楂、枸杞、蓝莓/桑葚、鲜枣，具体以当地可得为准），每种 $\geq$ 3 个独立批次样品，每批次 $\geq$ 3 次独立提取，每次提取 3 次平行测定；报批次间 RSD 与提取间 RSD，两者分开报告。
 3. 检出限：FC 法与 DPPH 法各按 3σ /斜率法给出 LOD (σ 取 11 份独立空白)， 10σ /斜率给出 LOQ。
 4. 对照与空白预先写死：每块微孔板/每批比色测定必须含——试剂空白、提取溶剂空白、没食子酸质控点（中浓度，允许偏差 $\pm 10\%$ ，越界整批重做）、PVPP 空白（PVPP 本身对 FC 响应的贡献须 $\leq$ 检出限）。DPPH 法另设 Trolox 或抗坏血酸阳性对照。
 5. 统计口径：相关性同时报 Pearson r 与 Spearman ρ （因基质数少、可能非线性），置信区间由 2000 次自助抽样给出；两个相关系数之差（校正前 vs 校正后）的显著性用基于自助的配对比较，不使用 Fisher z 变换的独立样本近似（因两组来自同一批样品，非独立）。H2 的阈值用分段线性回归定位，断点置信区间自助给出。所有检验前先做 Shapiro – Wilk 与残差诊断，不满足则改用秩方法并说明。
 6. 可复现性：全部原始吸光度、滴定记录、提取参数与统计脚本（Python + SciPy/statsmodels）开源；基质的采购/采集信息（产地、批号、日期）随论文附上。

5 · 数据与工具

用途	来源 / 工具
参考方法	GB 5009.86《食品中抗坏血酸的测定》(2,6-二氯酚酚滴定法部分)；Singleton—Orthofer FC 法经典程序 (Methods in Enzymology , 1999)。仅用于校验与对比，不计入本项目贡献
FC 试剂	Folin—Ciocalteu 试剂 (100 mL, $\approx$ 180 元)、无水碳酸钠 (AR 500 g, $\approx$ 25 元)
标准品	没食子酸 (AR 25 g, $\approx$ 60 元)、抗坏血酸 (AR 100 g, $\approx$ 30 元)、葡萄糖 (AR 500 g, $\approx$ 25 元)
活性通道	DPPH (1 g, $\approx$ 300 元，用量极小)；铁氰化钾还原力法 (Prussian blue 法) 试剂：铁氰化钾、三氯化铁、磷酸缓冲 ($\approx$ 60 元)，作第二独立活性通道
拆解试剂	PVPP 聚乙烯聚吡咯烷酮 (100 g, $\approx$ 120 元)；2,6-二氯酚酚钠 (5 g, $\approx$ 90 元)；草酸 (AR, $\approx$ 20 元)
基质样品	本地市售绿茶、柑橘、山楂、枸杞、桑葚、鲜枣，各 $\geq$ 3 批次（不同产地/批号），记录采购日期与批号
仪器	可见分光光度计 (760 nm / 517 nm / 700 nm)、分析天平、恒温水浴、离心机、pH 计——全部为中学常规配置
统计	Python + SciPy / statsmodels / pandas；自助法自写并开源；纯 CPU
对照文献	Applied Sciences 13(4):2135 (2023) 蜂蜜改良 FC 法；2019 年葡萄糖/抗坏血酸对绿茶与果汁 TPC 估计的影响研究；FC 法特异性改良综述。仅作方法对照与新颖性界定，不计入本项目贡献

6 · 方法路径

1. 建立并校验三通道定量方法：完成验收标准第 1 条 (a)(b)(c) 全部三项，未通过不得进入真实样品。
2. 固定提取 SOP：溶剂（60% 乙醇水）、料液比、温度、时间、离心条件全部书面固定；六种基质用同一套 SOP，不为单个基质单独优化（否则跨基质比较失效），此约束须在论文中明确声明。

3. 逐基质测定四个原始量：未校正 TPC、DPPH 清除活性 (IC_{50})、铁氰化钾还原力、抗坏血酸含量（滴定法）。
4. 做 PVPP 拆解：测定 PVPP 处理前后的 FC 响应差，得到"酚类贡献"与"非酚残余贡献"；须先用合成混合样确认 PVPP 对抗坏血酸与葡萄糖的吸附可忽略（作为拆解成立的前提条件，写入方法学校验）。
5. 计算 f_{AA} 与纯酚 TPC，重算相关性，用自助配对比较检验 H1。
6. 用分段回归定位 H2 的阈值，并做留一交叉验证 (leave-one-matrix-out)：每次留出一种基质，用其余五种拟合判据再预测被留出者，报预测相对误差——这是防止 6 个基质点被过拟合的必要步骤。
7. 独立交叉校验：用第二活性通道（铁氰化钾还原力法）重做全部相关性分析。若两个活性通道给出相反结论，须作为独立一节报告并分析原因，不得只报支持假设的那一个。

7 · 新颖性边界

- 本课题不声称发现 FC 法的干扰、不提出新的校正试剂、不声称测得任何基质的"真实多酚含量"。FC 法的非特异性、PVPP 吸附多酚、2,6-二氯靛酚滴定抗坏血酸，均为已有方法。
- 已有工作完成到：Singleton – Orthofer 已明确 FC 法测的是还原能力；蜂蜜中还原糖导致 TPC 高估并有改良法（2023）；绿茶与果汁中葡萄糖/抗坏血酸对 FC 与 Fast Blue BB 的影响已被定量比较（2019）；固相萃取、抗坏血酸当量校正、氧化预处理等提高特异性的方案已被提出。这些工作在单一基质上证明干扰存在并给出校正手段。
- 本项目的贡献是评价维度的转换，且构成主结论：被检验的对象不是"FC 法准不准"，而是"TPC 作为抗氧化力替代指标这一推理链在哪类基质上崩溃"；交付物是六基质的 FC 响应组成拆解表、校正前后相关性的自助配对比较，以及一条可用独立滴定测得的失效判据 f_{AA} 。
- 价值：食品与保健品文献中"TPC 与抗氧化活性显著相关"是极常见的结论句式，而这条相关性若主要由抗坏血酸驱动，则该句式的推理是错的。用中学可执行的方法把这条推理链拆开检验，是有实际后果的。
- 若结果是"扣除非酚贡献后相关性不下降"，同样是有效结论（对 TPC 替代关系的一次独立支持），但必须由验收标准第 2、5 条的自助置信区间证明本研究有能力分辨 0.20 量级的相关系数变化——若 6 个基质不足以分辨，须如实报告统计功效不足并给出所需基质数的估计。

8 · 决策门槛 (go / no-go)

- 第 8 周末：三通道方法学校验必须全部通过。PVPP 拆解回收率不达 95 – 105% → 降级路径 A：拆解手段换成碱性 H_2O_2 选择性氧化抗坏血酸（处理前后 FC 响应差 = 抗坏血酸贡献），拆解目标从"分离酚/非酚"退为"分离抗坏血酸/其余"，H2 完整保留，H1 改述为"扣除抗坏血酸当量后的相关性变化"。
- 第 14 周末：六种基质必须全部完成 ≥ 3 批次采集且提取 SOP 稳定（同基质批次间 TPC RSD $\leq 20\%$ ）。某基质 RSD $> 40\%$ 且原因不可控 → 替换为更均质的基质（如袋泡茶、市售冻干粉），基质数不得低于 5，否则统计功效不足须在论文中声明。
- 第 26 周末： f_{AA} 在六种基质上的跨度必须 ≥ 0.30 （即至少覆盖 0.1 – 0.4 的范围）。若所有基质的 f_{AA} 都聚集在窄区间 → 降级路径 B：改用梯度加标设计，向低 f_{AA} 基质中定量加入抗坏血酸人工构造 f_{AA} 梯度（0.05 – 0.60，6 个水平），把研究从"自然基质对比"改为"受控加标的剂量 – 响应"，H2 的阈值定位反而更精确，主结论框架保留。
- 第 34 周末：留一交叉验证完成。若预测相对误差 $> 50\%$ ，如实报告判据不可外推，并给出所需基质数的功效估算——这个负结论必须写成主结论的一部分，不得降级为"局限性"一句带过。

- 第 44 周数据冻结；第 48 周完成实验视频。
 - 实验视频要求：录制四段——(a) 提取 SOP 全流程（称样、超声/水浴、离心、定容）；(b) FC 显色与 760 nm 读数；(c) DPPH 褪色反应与 517 nm 时序读数；(d) 2,6-二氯靛酚滴定终点判定近景（粉红色 15 s 不褪）。累计 ≥ 10 分钟，每段含试剂标签与仪器读数特写。
 - 安全与废液：FC 试剂含钼、钨的杂多酸并有强酸性，须戴手套护目镜，稀释在通风处；DPPH 为自由基试剂，粉末勿吸入，甲醇/乙醇溶液易燃，配制远离热源、单次 ≤ 50 mL；2,6-二氯靛酚与草酸有毒，草酸溶液勿接触皮肤伤口。含钼钨的 FC 废液与含铁氰化物的还原力法废液分别单独收集（后者须保持碱性、严禁与酸混合），贴签交学校危废渠道；醇类废液单独收集。不涉及剧毒品、易制毒/易制爆化学品、氢氟酸、高压反应釜与明火操作。
 - 预算裁剪顺序：先砍第二活性通道（铁氰化钾还原力法，省 60 元但失去交叉校验，不推荐）；再砍 DPPH（改用铁氰化钾还原力法为唯一活性通道，省 300 元，H1/H2 仍可检验但活性口径单一，须在论文中声明）；FC 试剂、PVPP 与滴定试剂不可砍。
 - 适配学生：适合喜欢做统计推断、能接受"结论是一条否定性判据"的学生。不适合期望"发现某种植物抗氧化力最强"的学生——那种结论本课题明确不声称。
-

C15 · 果皮生物炭的 $\text{Ca}^{2+}/\text{Mg}^{2+}$ 二元交换选择性：热解温度对"容量—选择性权衡"的调控与再生衰减

推荐优先级 ★★★★★ · 分族：生物质材料 · 吸附 · 预算 $\approx$ 700 元 · 表征依赖：零外部平台（滴定表征替代 SEM/XRD） · 技能取向：材料制备 + 络合滴定 + 等温线拟合

1 · 研究问题

果皮生物炭（fruit-peel biochar）的热解温度在提高或降低阳离子交换容量（cation exchange capacity, CEC）的同时，是否同步改变其对 Ca^{2+} 相对 Mg^{2+} 的二元交换选择性系数 $K_{\text{Ca/Mg}}$ ？是否存在一个热解温度，使"总硬度去除容量 $\times$ 选择性"的乘积达到最优；该最优点能否由 Boehm 滴定测得的表面酸性含氧官能团密度预测？

2 · 研究背景与空白

生物炭是生物质在限氧条件下热解得到的多孔炭材料。它对阳离子的吸附主要来自表面羧基、内酯基、酚羟基的去质子化位点以及炭灰中 $\text{Ca}/\text{Mg}/\text{K}$ 的离子交换。硬水软化是这类材料最直观的应用场景： Ca^{2+} 与 Mg^{2+} 电荷相同、水合半径不同（ Mg^{2+} 的水合半径更大、脱水能更高），因此二元交换体系的选择性系数是一个能直接反映位点性质的物理量，而不只是一个工程指标。

已有工作到哪一步：热解温度对 CEC 的影响已有清晰的定量图景—— 250°C 生物炭 CEC 平均 $51.9 \pm 15.3 \text{ cmol}_\text{c}/\text{kg}$ ， 400°C 与 650°C 分别降至 16.2 ± 6.0 与 $21.0 \pm 17.2 \text{ cmol}_\text{c}/\text{kg}$ ， $> 800^\circ\text{C}$ 则低于 $5 \text{ cmol}_\text{c}/\text{kg}$ （*Front. Soil Sci.*, 2024 等）；机理层面也已明确：低温保留含氧官能团、高温热解破坏之；另有工作定量给出" $\text{Ca}^{2+}/\text{Mg}^{2+}$ 离子交换对污染物去除的贡献率随热解温度从 300°C 升至 600°C 由 37.4% 降至 11.7%"。此外还有用 FTIR + 偏最小二乘预测生物炭 CEC 的工作（*Sci. Total Environ.*, 2021）。

空白在——这些工作报告的都是"总容量"，几乎没有工作报告 $\text{Ca}^{2+}/\text{Mg}^{2+}$ 的二元交换选择性系数，也没有把"容量与选择性是否同向变化"当作被检验的命题。这个空白适合一年期中学课题：全部定量靠 EDTA 络合滴定（国标方法，精度高、成本极低），表面官能团靠 Boehm 滴定（同样是滴定），完全绕开 SEM/XRD/FTIR；且"容量与选择性同向"与"存在权衡"都是有效结论。

3 · 可检验假设

H1 在 $300 / 450 / 600^\circ\text{C}$ 三个热解温度上，生物炭的总硬度吸附容量 q_{max} （Langmuir 拟合）随温度升高单调下降，但 $\text{Ca}^{2+}/\text{Mg}^{2+}$ 选择性系数 $K_{\text{Ca/Mg}}$ 随温度升高而上升（即存在容量—选择性权衡）； 450°C 样品的 ($q_{\text{max}} \times K_{\text{Ca/Mg}}$) 乘积显著高于另外两个温度（单因素方差分析 $p < 0.05$ ）。

H2 $K_{\text{Ca/Mg}}$ 与 Boehm 滴定测得的羧基密度占总酸性位点的比例呈单调关系（Spearman ρ 的绝对值 ≥ 0.7 ）；若该关系成立，则羧基比例可作为不依赖任何仪器表征的选择性预测量。若 H2 被否认，须报告哪一类位点（羧基/内酯基/酚羟基）与选择性无关——负结论同样构成对机理的约束。

4 · 量化验收标准

- 方法学校验（硬门槛，两级）：
 - 滴定级：EDTA 标准溶液用基准碳酸钙标定，3 次平行 $\text{RSD} \leq 0.3\%$ ；用已知浓度的 $\text{Ca}^{2+}/\text{Mg}^{2+}$ 合成硬水（5 个水平， $\text{Ca}:\text{Mg} = 4:1$ 至 $1:4$ ）按 GB/T 7477 测总硬度、按钙指示剂法测 Ca^{2+} ，两者测得值对配制值的线性 $R^2 \geq 0.999$ ，各点相对误差 $\leq 2\%$ ；差减法得到的 Mg^{2+} 相对误差 $\leq 5\%$ 。
 - 文献复现级：用同一批果皮在 450°C 制备生物炭，测其 CEC（乙酸铵交换法），结果须落在文献同温

度区间（约 10 – 30 cmol_c/kg）内；若越界须给出原料灰分差异的定量解释。两级任一不过关，后续全部吸附数据无效。

- 2. 重复口径写死：每个热解温度 ≥ 3 个独立批次生物炭（独立热解），每批次的等温线实验 ≥ 3 次独立平行；报批次间 RSD 与批次内 RSD，两者分开报告。批次间 q_max 的 RSD > 25% 时须补做至 5 批。
- 3. 等温线与选择性口径预先写死：单组分等温线同时用 Langmuir 与 Freundlich 拟合，两者都报，并用调整后 R² 与 AIC 双口径比较，不允许只报拟合更好的那一个；二元选择性系数按离子交换的质量作用定律 $K_{Ca/Mg} = (q_{Ca} \cdot C_{Mg}) / (q_{Mg} \cdot C_{Ca})$ 计算，在 ≥ 5 个不同 Ca:Mg 初始比下测定并报均值与 95% 置信区间（自助抽样 2000 次）。
- 4. 再生：吸附 – 脱附（1 mol/L NaCl 再生）循环 5 次，报每次的容量保留率与选择性保留率；衰减用线性回归给出每循环衰减率及其置信区间，明确报告"容量衰减与选择性衰减是否同步"。
- 5. 统计口径：温度对 q_max 与 K_Ca/Mg 的影响用单因素方差分析 + Tukey HSD 事后检验，前提检查预先写死（Shapiro – Wilk 正态性、Levene 方差齐性；不满足改用 Kruskal – Wallis + Dunn 检验并做 Bonferroni 校正）。H2 的关联用 Spearman ρ（因样本点只有 3 个温度 × 3 批次 = 9 点，明确声明这是探索性关联而非因果推断）。
- 6. 可复现性：热解程序（升温速率、保温时间、限氧方式）、Boehm 滴定原始数据、全部等温线原始数据与拟合脚本（Python + SciPy）开源。

5 · 数据与工具

用途	来源 / 工具
参考方法	GB/T 7477 《水质 钙和镁总量的测定 EDTA 滴定法》；Boehm 滴定经典程序（Boehm, Carbon, 2002）；乙酸铵交换法测 CEC（土壤农化分析标准方法）。仅用于校验与对比，不计入本项目贡献
原料	柚子皮 / 橘皮 / 香蕉皮（食堂或家庭废弃物，免费；须记录品种、来源、干燥方式），洗净 → 60 °C 烘干 → 粉碎过 40–60 目筛
热解	学校马弗炉（须核实最高温度与控温精度，第 4 周前核实清楚）；限氧方式：带盖瓷坩埚 + 外层石英砂覆盖，或坩埚内衬铝箔并压实（须在预实验中确认产率重现性 RSD ≤ 10%）
滴定试剂	EDTA 二钠（AR 500 g, ≈70 元）、基准碳酸钙（≈50 元）、铬黑 T 指示剂、钙指示剂、氨–氯化铵缓冲液、NaOH / HCl 标准溶液（≈80 元合计）
Boehm 滴定	NaHCO₃、Na₂CO₃、NaOH（分别滴定羧基 / 羰基+内酯基 / 全部酸性位点），AR 级共 ≈60 元
硬水配制	无水氯化钙、六水氯化镁（AR 各 500 g, ≈70 元）
表面性质	pH_pzc 用 pH drift 法（pH 计 + NaCl 溶液，零成本）；灰分用马弗炉 815 °C 灼烧法；产率、元素比例不测（明确声明）
再生	氯化钠（AR 500 g, ≈20 元）
仪器	分析天平、马弗炉、恒温振荡器或磁力搅拌加热台、离心机、pH 计、酸式/碱式滴定管——全部为中学常规配置
外部依赖（可选）	若需确证孔结构或晶相，须联系高校共享平台送 BET 或 XRD 样品——属外部依赖，列入 go/no-go 风险项。不依赖该表征的主结论：容量–选择性权衡关系只依赖滴定测得的 q、K 与 Boehm 位点密度，与孔结构表征无关
统计	Python + SciPy / statsmodels / pandas；纯 CPU
对照文献	Front. Soil Sci. (2024) 生物炭 CEC 与热解温度；Sci. Total Environ. (2021) FTIR–PLS 预测 CEC；热解温度对 Ca/Mg 交换贡献率的定量研究。仅作参考对照与新颖性界定，不计入本项目贡献

6 · 方法路径

1. 校验滴定体系：完成验收标准第 1(a) 条，标定 EDTA 并在合成硬水上验证 Ca/Mg 分测的准确度；未通过不得进入制备。
2. 建立热解 SOP 并核实马弗炉能力：固定升温速率、保温时间、限氧方式与装填量；用产率重现性 ($RSD \leq 10\%$) 作为 SOP 稳定的判据。完成 450°C 样品的 CEC 文献复现（第 1(b) 条）。
3. 三个温度各制 ≥ 3 独立批次生物炭，测灰分、 pH_{pzc} 与 Boehm 三档酸性位点密度。
4. 单组分等温线：分别对 Ca^{2+} 、 Mg^{2+} 做 6 点浓度梯度等温线，Langmuir 与 Freundlich 双口径拟合。
5. 二元竞争实验：固定总硬度、变化 Ca:Mg 比（5 个比例），测平衡后液相与固相分配，计算 $K_{\text{Ca/Mg}}$ 及其自助置信区间。
6. 再生循环 5 次，跟踪容量与选择性的同步/非同步衰减。
7. 独立交叉校验：对同一批平衡液，除 EDTA 滴定外，另用 **电导率法**测总离子强度变化作为质量衡算的独立通道；固相吸附量由液相差减法与固相酸解回收法两条路径分别计算，两者偏差 $\leq 10\%$ 方可采信。

7 · 新颖性边界

- 本课题不声称制备新材料、不声称发现新的吸附机理、不声称生物炭软化硬水优于商品树脂。果皮生物炭、热解温度调控、Boehm 滴定、EDTA 滴定均为成熟方法。
- 已有工作完成到：生物炭 CEC 随热解温度的定量变化 ($250 \rightarrow 650^\circ\text{C}$ 的 CEC 数值区间) 已被系统报道；含氧官能团热降解是 CEC 下降的机理已被确认；Ca/Mg 离子交换对污染物去除的贡献率随温度的变化 ($37.4\% \rightarrow 11.7\%$) 已被量化；FTIR-PLS 预测 CEC 已实现。这些工作的因变量都是容量或贡献率。
- 本项目的贡献是评价维度的转换，且构成主结论：因变量从"容量"换成"二元交换选择性系数 $K_{\text{Ca/Mg}}$ "，被检验的命题是"提高容量是否以牺牲选择性为代价"，并给出一条不依赖任何仪器表征的选择性预测量（Boehm 羧基比例）。
- 价值：吸附材料文献长期以 q_{max} 为唯一优劣标尺，而实际水处理中选择性往往比容量更决定成败。用最廉价的滴定手段把权衡关系测出来，对材料评价范式是一次具体的补充。
- 若结果显示"容量与选择性同向变化、不存在权衡"，这同样是有效结论（对现有以容量为标尺的评价方式的一次独立支持），但必须由验收标准第 3 条的自助置信区间证明本研究有能力分辨 $K_{\text{Ca/Mg}}$ 的 20% 差异。

8 · 决策门槛 (go / no-go)

- 第 5 周末：EDTA 滴定校验 ($R^2 \geq 0.999$ 、相对误差 $\leq 2\%$) 必须通过，且马弗炉控温精度已核实 ($\pm 20^\circ\text{C}$ 以内)。控温不达标 $\rightarrow$ 降级路径 A：把三个温度点改为"高/低"两个可稳定实现的档位（如 400°C 与 600°C ），H1 从"存在最优"退为"存在单调权衡趋势"，H2 与全部方法框架不变。
- 第 12 周末：三温度各 ≥ 3 批次生物炭制备完成，批次间产率 $RSD \leq 10\%$ 、CEC 落入文献区间。若产率或 CEC 无法稳定 $\rightarrow$ 降级路径 B：把原料从混合果皮换成单一来源、单一批次的柚子皮（消除原料变异），并把"原料变异对生物炭性质的贡献"作为独立一节定量报告。
- 第 24 周末：二元竞争实验必须给出 $K_{\text{Ca/Mg}}$ 的自助 95% 置信区间宽度 $\leq$ 均值的 40%。若置信区间过宽 $\rightarrow$ 增加 Ca:Mg 比例点数 ($5 \rightarrow 8$) 并加大平衡时间，同时把"选择性系数难以精确测定"这一发现连同其原因（滴定精度、差减法误差传播）定量写入——这个困难本身就是研究内容。
- 第 32 周末：5 次再生循环完成。若第 1 次再生后容量即衰减 $> 50\%$ ，不掩饰，如实把"低再生性"作为该材料的定量结论报告，并把研究重心转到"衰减的是容量还是选择性"这一更细的问题上，主结论框架保留。

- **外部表征风险项：**若指导教师要求孔结构或晶相证据，须在第 28 周前联系高校共享平台预约 BET/XRD 送样（3 个温度各 1 个代表样即可）。联系不上则主结论不变，论文中明确写"表面化学由 Boehm 滴定与 pH_{pzc} 表征，未做孔结构与晶相表征"。
 - **第 44 周数据冻结；第 48 周完成实验视频。**
 - **实验视频要求：**录制四段——(a) 果皮预处理与坩埚限氧装填；(b) 马弗炉程序升温与出炉（含控温表读数特写）；(c) Boehm 滴定与 EDTA 滴定的终点判定近景（铬黑 T 由酒红变纯蓝）；(d) 一次完整的二元竞争吸附与取样过滤。累计 ≥ 10 分钟。
 - **安全与废液：**马弗炉高温操作须戴隔热手套、用长柄坩埚钳，出炉后置于耐火板冷却，炉口不得正对面部；热解烟气须在通风橱或强制排风条件下进行，操作间禁止无人值守。铬黑 T 与钙指示剂含偶氮结构，避免皮肤接触；氨-氯化铵缓冲液有刺激性氨味，在通风处配制。滴定废液（含 EDTA 络合物与少量指示剂）中和至 pH 6-9 后按"含络合剂废液"收集交学校处置；生物炭残渣可作固废收集，不得随意倾倒。不涉及剧毒品、易制毒/易制爆化学品、氢氟酸、高压反应釜与明火有机溶剂操作。
 - **预算裁剪顺序：**先砍再生循环次数（5 → 3，衰减率精度下降但趋势仍可判）；再砍热解温度点（3 → 2，见降级路径 A）；EDTA 滴定与 Boehm 滴定试剂不可砍。
 - **适配学生：**适合手稳、耐得住重复滴定的学生。滴定精度直接决定本课题成败——若学生无法把平行滴定 RSD 稳定在 0.5% 以内，不建议启动这条路线。
-

C16 · 加碘食盐在家庭烹饪条件下的碘损失动力学：加碘剂形态 × 酸度 × 共存还原剂的因子设计与半衰期测定

推荐优先级 ★★★☆☆ · 分族：食品分析 · 反应动力学 · 预算 ≈ 600 元 · 表征依赖：零外部平台 · 技能取向：碘量法滴定 + 因子设计 + 动力学拟合

1 · 研究问题

加碘食盐在模拟家庭烹饪条件下的碘损失，能否用一级动力学（first-order kinetics）描述并给出速率常数 k 与半衰期 $t_{1/2}$ ？加碘剂形态（碘酸钾 KIO_3 vs 碘化钾 KI ）、介质酸度与共存还原剂（抗坏血酸）三个因子对 k 的主效应与交互效应各占多大比例，能否由此解释已发表文献中“烹饪碘损失 14 – 66%”这一极宽跨度？

2 · 研究背景与空白

食盐加碘是全球最成功的公共营养干预之一。中国现行加碘剂主要为碘酸钾（ KIO_3 ），部分国家仍用碘化钾（ KI ）。碘的损失机理与形态直接相关： I^- 易被空气氧化并以 I_2 形式挥发； IO_3^- 相对稳定，但在酸性且存在还原剂时会被还原为 I^- 乃至 I_2 ，从而进入挥发通道。烹饪过程恰好同时提供了这三个条件——高温、食醋/番茄的酸度、蔬菜中的抗坏血酸。

已有工作到哪一步：结论方向明确但数值极其分散。有研究报告加碘盐单纯受热到 350°C 最大仅损失 18.5%（[Indian J. Pediatr.](#)）；多项工作一致指出 KIO_3 盐在常规烹饪温度下损失很小而 KI 盐损失可观；HPLC/ICP – MS 研究证实烹饪过程中 IO_3^- 会被还原为 I^- （[Food Chem.](#), 2016）；综述性数据则给出“烹饪中碘损失 14 – 66%”这样一个跨度达 4 倍以上的区间（[PMC](#), 2013 等）。这些工作几乎全部报告的是单点百分比损失，条件（温度、时间、pH、共存物、盐种）各不相同且未统一口径。

空白在——没有工作把碘损失参数化为动力学速率常数，也没有用因子设计定量拆分“形态 × 酸度 × 还原剂”的主效应与交互项。这正是文献数值跨度巨大的最可能原因，而这个原因本身可以被检验。这个空白适合一年期中学课题：定量方法是经典碘量法滴定（国标方法）；实验装置是家用锅具与恒温水浴；结论正负都成立——若一级动力学成立，得到一组可迁移的 k 与 $t_{1/2}$ ；若不成立，得到“损失不是简单一级过程”这一同样重要的结论。

3 · 可检验假设

H1 在 $90 \pm 2^\circ\text{C}$ 恒温、开放体系下，加碘盐溶液中的总碘浓度对时间的对数呈线性（一级动力学），线性拟合 $R^2 \geq 0.95$ ，且 KI 型样品的速率常数 k 至少是 KIO_3 型的 3 倍（对应半衰期缩短至 $1/3$ 以下）。

H2 在 2^3 全因子设计（形态 × 酸度 × 抗坏血酸）中，“酸度 × 还原剂”交互项对 $\ln k$ 的贡献显著（ $p < 0.05$ ）且其效应量不小于任一主效应的 50%——即碘损失不能由单因子加和预测。若 H2 被否认（交互项不显著），则说明文献数值跨度须由条件差异之外的因素（如取样与测定口径）解释，该负结论须与已发表数值一并讨论。

4 · 量化验收标准

- 方法学校验（硬门槛，两级）：
(a) 滴定级：硫代硫酸钠标准溶液用基准重铬酸钾以外的途径标定——采用基准碘酸钾直接标定（避免引入铬），3 次平行 $\text{RSD} \leq 0.3\%$ ；用已知碘含量的模拟盐（ $\text{NaCl} + \text{KIO}_3$ ，配制 5 个水平：10 / 20 / 30 / 40 / 50 mg I/kg）按 GB 5009.267 氧化还原滴定法测定，测得值对配制值线性 $R^2 \geq 0.995$ 、加标回收率 95 – 105%、每点 3 次平行 $\text{RSD} \leq 3\%$ 。
(b) 市售样品复现级：至少 3 个批次市售加碘

盐的实测碘含量须落在包装标示值 $\pm 30\%$ 内（国标允许波动范围内），否则须先排查方法偏差再进入动力学实验。两级任一不过关，全部动力学数据无效。

- 动力学采样口径写死：每条动力学曲线 ≥ 6 个时间点（0 / 5 / 10 / 20 / 40 / 60 min），每个时间点 3 次独立取样（不是同一样品测三次），每条曲线在 ≥ 3 个独立实验日重复；报 k 的均值、RSD 与 95% 置信区间（自助抽样 2000 次）。
- 检出限：碘量法按 3σ /斜率法给出 LOD（ σ 取 11 份空白盐溶液的滴定体积标准差）， 10σ /斜率给出 LOQ；须明确证明实验末期（损失最多时）的碘浓度仍高于 LOQ，否则该时间点作废。
- 对照与空白预先写死：每个实验日必须含——不加热对照（同温放置的密闭样）、纯 NaCl 空白（排除基质干扰）、标准加入质控点（允许偏差 $\pm 5\%$ ，越界整日数据作废）；蒸发损失用称重法逐点补水校正，补水量须记录并计入不确定度。
- 统计口径：动力学级数用零级 / 一级 / 二级三种模型同时拟合并以调整后 R^2 与 AIC 双口径比较，不允许只报一级；因子效应用 2^3 全因子设计的方差分析（含全部二阶与三阶交互项），前提检查预先写死（残差正态性 Shapiro – Wilk、方差齐性 Levene；不满足则对 $\ln k$ 做 Box – Cox 变换后重做并说明）。效应量报偏 η^2 ，不只报 p 值。
- 可复现性：全部滴定原始记录、温度记录、补水记录与拟合脚本（Python + SciPy）开源；市售盐的品牌、批号、生产日期随论文附上。

5 · 数据与工具

用途	来源 / 工具
参考方法	GB 5009.267《食品中碘的测定》氧化还原滴定法部分；GB 26878《食品安全国家标准 食用盐碘含量》（标示值与允许波动）。 仅用于校验与对比，不计入本项目的数据贡献
样品	市售加碘食盐（KIO ₃ 型， ≥ 3 品牌 $\times$ 3 批次， ≈ 60 元）；KI 型模拟盐由分析纯 NaCl + KI 自配（配制值已知，正好用于方法学校验）
试剂	硫代硫酸钠（AR 500 g， ≈ 40 元）、基准碘酸钾（ ≈ 70 元）、碘化钾（AR 100 g， ≈ 60 元）、可溶性淀粉（ ≈ 15 元）、磷酸/硫酸（教学装， ≈ 25 元）、氯化钠（AR 500 g， ≈ 25 元）、抗坏血酸（AR 100 g， ≈ 30 元）、冰乙酸/柠檬酸（ ≈ 25 元）
酸度控制	pH 计 + 柠檬酸/磷酸盐缓冲体系（pH 3.0 与 pH 6.0 两水平）
加热	恒温水浴锅（ 90 ± 2 °C，学校常规）；高温档用可控温电陶炉 + 温度计（须 核实 控温波动 $\leq \pm 5$ °C）
滴定	10 mL 酸式滴定管或 5 mL 微量滴定管（ ≈ 80 元）；分析天平 0.1 mg
备用比色通道	碘–淀粉蓝比色（590 nm，可见分光光度计），用于低浓度点的独立校核
统计	Python + SciPy / statsmodels / pandas（动力学拟合、因子设计 ANOVA、自助法）；纯 CPU
对照文献	Indian J. Pediatr. 加碘盐受热碘损失； Food Chem. 209 (2016) 烹饪中 IO ₃ ⁻ $\rightarrow$ I ⁻ 的 HPLC/ICP–MS 证据； PMC (2013) 不同烹饪方式碘损失综述（14–66%）。 仅作参考对照与新颖性界定，不计入本项目贡献

6 · 方法路径

- 标定硫代硫酸钠并完成模拟盐的方法学校验（验收标准第 1(a) 条），未通过不得进入动力学实验。
- 测定 ≥ 3 批次市售加碘盐的碘含量并与标示值比对（第 1(b) 条），同时自配 KI 型模拟盐。

3. 建立动力学实验 SOP：固定液固比、加热功率、搅拌速率、敞口面积、取样体积与补水规则；敞口面积与搅拌速率必须固定并写入 SOP，因为挥发是损失的主要通道，这两项是最容易被忽略的隐变量。
4. 单因子预实验：先在 KIO_3 / pH 6 / 无还原剂条件下测一条完整动力学曲线，确定合适的时间窗口（使总损失落在 20 – 70% 区间，避免"太快测不到"或"太慢无变化"）。
5. 执行 2^3 全因子设计（8 个条件 $\times$ 3 个独立实验日），每条件测完整动力学曲线，拟合三种级数模型。
6. 做方差分析，拆分主效应与交互效应，报偏 η^2 ；用得到的 k 反算文献报道条件下的预期损失率，检验能否覆盖 14 – 66% 这一跨度。
7. 独立交叉校验：对同一批样品的部分时间点，用碘 – 淀粉比色法独立测定总碘，与滴定法做 Bland – Altman 比对，偏差 $\leq 10\%$ 方可采信；另做一次质量衡算（密闭体系 vs 敞口体系的同条件对照），定量证明损失确经挥发通道。

7 · 新颖性边界

- 本课题不声称发现碘的新损失机理、不声称提出新的加碘剂、不对膳食碘摄入做健康风险结论（那超出本项目的数据范围）。碘量法、 KIO_3 稳定性优于 KI、烹饪造成碘损失，均为已知事实。
- 已有工作完成到：加碘盐单纯受热的损失上限（ 350°C 下约 18.5%）； KIO_3 型显著优于 KI 型的稳定性对比；HPLC/ICP – MS 证实烹饪中 IO_3^- 被还原为 I^- （[Food Chem.](#), 2016）；不同烹饪方式的碘损失百分比综述（14 – 66%）。这些工作报告的是单点百分比损失，条件口径不统一。
- 本项目的贡献是评价维度的转换，且构成主结论：把因变量从"损失百分比"换成"动力学速率常数 k 与半衰期 $t_{1/2}$ "，把研究设计从"单点比较"换成" 2^3 全因子的效应拆分"，并用得到的 k 去检验它能否解释已发表数值的 4 倍跨度。
- 价值：百分比损失依赖具体时间与条件因而不可迁移，速率常数可迁移。如果一组 k 能把文献中看似矛盾的数值统一解释，那是对该领域数据可比性的实质贡献；如果不能，则指出问题出在测定口径而非条件差异，同样有价值。
- 若结果是"交互项不显著、损失可由主效应加和预测"，同样是有效结论，但必须由验收标准第 2、5 条的置信区间与效应量证明本研究有能力分辨 50% 量级的效应差异。

8 · 决策门槛 (go / no-go)

- 第 6 周末：方法学校验第 1(a) 条必须通过 ($R^2 \geq 0.995$ 、回收率 95 – 105%)。若因碘含量太低导致滴定体积过小、RSD 超标 $\rightarrow$ 降级路径 A：加大取样量至 50 g 盐 / 次并改用 0.002 mol/L 稀硫代硫酸钠 + 5 mL 微量滴定管；仍不达标则改用碘 – 淀粉比色法为主通道（滴定法退为校核），全部动力学与因子设计框架不变。
- 第 12 周末：单因子预实验必须找到使总损失落在 20 – 70% 的时间窗口。若在 90°C 、60 min 内 KIO_3 型损失 $< 5\%$ $\rightarrow$ 降级路径 B：把研究对象聚焦到"在什么条件下 KIO_3 才会显著损失"，即把酸度与还原剂浓度扩展为连续梯度（各 4 水平）做剂量 – 响应，H1 改述为" KIO_3 的损失存在阈值条件"，动力学与统计框架完全保留。
- 第 24 周末： 2^3 因子设计的 8 个条件必须全部完成 ≥ 2 个独立实验日。若某条件的日间 k 的 $\text{RSD} > 40\%$ $\rightarrow$ 优先排查敞口面积与搅拌速率的控制，而非增加重复；排查结论须写入论文的不确定度分析。
- 第 32 周末：交叉校验（比色法比对 + 密闭/敞口质量衡算）完成。若密闭体系也观察到同等损失，说明损失并非经挥发通道 $\rightarrow$ 不改课题，把这一发现作为独立一节报告并重新讨论机理归属——这个意外本身就是研究内容。

- 第 44 周数据冻结；第 48 周完成实验视频。
 - 实验视频要求：录制四段——(a) 硫代硫酸钠标定与基准碘酸钾称量特写；(b) 一次完整的动力学实验（恒温水浴、定时取样、补水称重）；(c) 碘量法滴定终点判定近景（蓝色恰好消失）；(d) 密闭 vs 敞口对照实验的装置全景。累计 ≥ 10 分钟。
 - 安全与废液：碘量法涉及酸化释放 I_2 ，须在通风橱或开窗强通风条件下进行，严禁密闭容器内酸化后剧烈加热；碘蒸气刺激呼吸道与眼睛，操作全程戴护目镜。硫酸/磷酸稀释遵循“酸入水”。碘化钾与碘酸钾均为普通化学试剂，不属剧毒或管制品，但须避免与强氧化/还原剂混存。滴定废液含碘化物与硫代硫酸盐，先用少量硫代硫酸钠还原剩余 I_2 至无色，中和至 pH 6–9 后按“含碘废液”收集交学校处置。热水浴与电陶炉操作须防烫伤，禁止无人值守。不涉及剧毒品、易制毒/易制爆化学品、氢氟酸、高压反应釜与明火有机溶剂操作。
 - 预算裁剪顺序：先砍市售盐品牌数（3 $\rightarrow$ 2，仍可覆盖批次变异）；再砍比色备用通道（省 0 元但失去交叉校验，不推荐）；再砍高温档（只做 90 ° C 水浴，去掉温度因子， 2^3 设计仍完整，因温度不是本设计的三个因子之一）；基准碘酸钾与硫代硫酸钠不可砍。
 - 适配学生：适合喜欢做因子设计与数据建模、能接受“结论是一组速率常数而非一个新材料”的学生。本课题的实验操作重复量大但技术门槛低，适合两人小队分工（一人滴定、一人控温取样）。
-

C17 · 常压微波法碳点的批次间重现性：化学品前驱体与生物质前驱体的方差分解及"绿色路线的重现性代价"

推荐优先级 ★★★☆☆ · 分族：纳米材料 · 合成重现性 · 预算 ≈ 800 元 · 表征依赖：低（相对荧光，无 TEM/XRD；绝对量子产率须外部平台，非必需） · 技能取向：材料合成 + 嵌套方差分析

1 · 研究问题

在名义完全相同的合成条件下重复制备碳点（carbon dots, CDs），其表观相对量子产率与发射峰位的批次间变异有多大？以精确称量的化学品前驱体（柠檬酸 + 尿素）为基准，改用生物质前驱体（柑橘汁 / 柚子皮提取液）这一"更绿色"的路线，究竟要付出多少倍的重现性代价——该代价能否用前驱体本身的成分变异（以总固形物与总酚含量表征）解释？

2 · 研究背景与空白

碳点是尺寸 $< 10\text{ nm}$ 的荧光碳基纳米颗粒。与传统无机量子点不同，碳点的发光并非单一能带跃迁，而更接近"纳米尺度的荧光团集合体"——其荧光性质取决于组装单元的分子结构与表面缺陷，而这些又由合成路径决定。这一物理图像直接预示了一个后果：合成条件的微小变化会显著改变发光性质。

已有工作到哪一步：这个后果已被反复观察到。有工作明确以"Small variations in reaction conditions tune carbon dot fluorescence"为题（PMC, 2022），把条件敏感性当作调控手段；反应时间对量子产率的影响被描述为"先升后降的非单调关系"；文献报道的水相量子产率跨度极大——从常见的 $< 5\%$ 到宣称的 50% 、乃至 $90 \pm 5\%$ ；关于植物源碳点的综述也指出水热路线"产率有限、产物质量参差、纯化繁琐，最终削弱合成的效率与重现性"。

空白在——条件敏感性一直被当作"可调控"来庆祝，几乎没有工作把它当作"重现性问题"来量化：没人报告"同一实验室、同一条件、重复 N 次，批次间 RSD 是多少"，更没人比较化学品前驱体与生物质前驱体在这一指标上的差距。这个空白适合一年期中学课题：常压微波法与油浴热解法均可在中学完成（不使用高压反应釜）；核心工作量是重复合成与统计，而非高端表征；且"生物质路线重现性显著更差"与"两者无显著差异"都是有效结论——后者反而是对绿色路线的有力支持。

3 · 可检验假设

H1 在名义相同条件下重复 8 次合成，柠檬酸 + 尿素路线的表观相对量子产率批次间 $RSD \leq 15\%$ ，而柑橘汁路线的批次间 $RSD \geq 35\%$ （即至少 2 倍的重现性代价）；发射峰位的批次间标准差前者 $\leq 5\text{ nm}$ 、后者 $\geq 12\text{ nm}$ 。

H2 嵌套方差分析（nested ANOVA：批次 $\rightarrow$ 纯化重复 $\rightarrow$ 测量重复）显示，生物质路线的方差主要来自"批次"层（占总方差 $\geq 60\%$ ），而化学品路线的方差主要来自"测量"层；且生物质路线的批次层方差与前驱体总固形物含量的批次变异呈正相关（Spearman $\rho \geq 0.6$ ）。若 H2 被否认（方差主要来自纯化层），则说明重现性瓶颈在后处理而非前驱体——该结论对改进路线更直接有用。

4 · 量化验收标准

- 方法学校验（硬门槛，两级）：
(a) 荧光定量级：以硫酸奎宁（ $0.1\text{ mol/L H}_2\text{SO}_4$ 中，文献量子产率 0.54）为参比，用"吸光度通道（可见分光光度计， $360 - 365\text{ nm}$ ， $A \leq 0.05$ 以避免内滤效应） $\times$ 发射通道（ 365 nm UV 灯 + 手机定量成像，与参比同框）"计算表观相对量子产率。硬门槛为：对 5 个已知浓度梯度的硫酸奎宁溶液，发射强度对吸光度的线性 $R^2 \geq 0.995$ ；用第二种荧光标准品（荧光素钠，碱性介质）做交叉验证，

两标准品给出的同一碳点样品的相对量子产率之差 $\leq 20\%$ 。(b) 合成复现级：柠檬酸 + 尿素路线的产物必须呈现文献一致的特征——365 nm 激发下蓝色发射、发射峰随激发波长红移（激发依赖性）、350 nm 附近有吸收肩。缺任一特征须先排查合成，不得进入重现性统计。两级任一不过关，全部重现性数据无效。

- 2. 重复口径写死（这是本课题的核心）：每条路线 ≥ 8 个独立批次（独立称量、独立加热、独立纯化），每批次做 2 次独立纯化重复，每次纯化产物做 3 次独立稀释测量。这构成三层嵌套结构，任何一层的重复都不得省略，否则方差分解无效。
- 3. 检出限与线性范围：报告本测量体系能分辨的最小相对量子产率差异（由参比线性拟合的残差标准差按 3σ /斜率折算），须证明该值 $<$ 假设中要比较的效应量（15% vs 35%），否则功效不足须先改进测量。
- 4. 对照与空白预先写死：每个测量日必须含——溶剂空白、硫酸奎宁质控点（允许偏差 $\pm 10\%$ ，越界当日数据作废）、上一日留存的碳点样品（作为跨日漂移监测）。手机成像沿用固定 SOP（锁定曝光与白平衡、固定距离、参比同框）。
- 5. 统计口径：方差分解用嵌套随机效应模型（restricted maximum likelihood, REML），报各层方差分量及其自助置信区间（2000 次）；两条路线 RSD 的差异用 Levene 检验或 Fligner – Killeen 检验（比较方差而非均值），前提检查写死；峰位变异用标准差而非 RSD（因峰位是位置量而非比例量）。明确写"不用单次合成的最高量子产率作为路线优劣依据"及其原因。
- 6. 可复现性：全部合成记录（含微波实际工作时间、功率档、容器规格）、原始照片、光谱数据与统计脚本（Python + statsmodels / pymr4 或 R lme4）开源。

5 · 数据与工具

用途	来源 / 工具
合成方法 (常压)	家用微波炉 (700–800 W, 须核实其占空比式功率调节的实际通断周期, 第 6 周前用温升法标定); 或硅油浴 180 °C 常压加热。明确不使用高压水热反应釜
化学品前驱体	柠檬酸一水合物 (AR 500 g, ≈ 35 元)、尿素 (AR 500 g, ≈ 25 元)
生物质前驱体	本地柑橘 (≥ 8 个独立采购批次, 记录品种/产地/日期, ≈ 60 元); 榨汁后过滤定容, 测总固形物 (烘干称重) 与总酚 (FC 法, 可复用 C14 的方法)
纯化	透析袋 (MWCO 1000 Da, 10 m, ≈ 180 元) 或 0.22 μm 滤膜 + 硅胶柱简易分离; 纯化方式两条路线必须一致
荧光参比	硫酸奎宁 (AR 5 g, ≈ 120 元)、荧光素钠 (AR 5 g, ≈ 60 元)
激发与成像	365 nm UV 手电 (≈ 60 元, 须测其光强稳定性)、暗箱 (纸箱自制, ≈ 40 元)、手机 (锁定参数, 沿用 C11 的成像 SOP)
吸收测定	可见分光光度计 (学校常规; 若无紫外档, 用 400 nm 处吸收作替代量并明确声明该替代的局限)
统计	Python + statsmodels (混合效应/嵌套方差分量) + SciPy; 纯 CPU
外部依赖 (可选)	若需颗粒尺寸或绝对量子产率证据, 须联系高校共享平台送 TEM 或积分球荧光光谱——属外部依赖, 列入 go/no-go 风险项。不依赖该表征的主结论: 重现性与方差分解只需相对量子产率与峰位, 二者由本方案的相对测量即可获得; 论文中不得声称任何绝对量子产率数值
对照文献	PMC (2022) "Small variations in reaction conditions tune carbon dot fluorescence"; 植物源碳点综述 (PMC, 2023) 关于重现性的论述; 各类高量子产率报道 (50%、 $90 \pm 5\%$)。仅作参考对照与新颖性界定, 不计入本项目贡献

6 · 方法路径

1. 标定微波炉：用已知热容的水样测温升，标定各功率档的实际有效功率与通断周期；不完成此步不得开始合成（这是本课题最大的隐变量）。
2. 完成荧光定量方法学校验（验收标准第 1(a) 条），建立并固定成像 SOP。
3. 跑通柠檬酸 + 尿素路线并确认特征（第 1(b) 条），固定配比、体积、容器、加热时间与纯化流程，写成书面 SOP。
4. 执行三层嵌套重复：化学品路线 8 批次 $\times$ 2 纯化 $\times$ 3 测量；同样流程做柑橘汁路线 8 批次（每批次用不同采购批次的柑橘），同步测前驱体总固形物与总酚。
5. 做嵌套方差分解，报各层方差分量与置信区间，检验 H1、H2。
6. 做前驱体变异与批次层方差的关联分析；若关联成立，尝试用“总固形物归一化投料”重做 4 个批次，检验归一化能否把批次层方差压下来（这是一次干预性验证，比相关分析更有说服力）。
7. 独立交叉校验：对同一批样品的相对量子产率，另用紫外 - 可见吸收峰强度归一化的荧光比值法独立计算一次；两条路径结果偏差 $\leq 20\%$ 方可采信。另请一位未参与实验的同学按 SOP 独立合成 2 批，检验 SOP 的可迁移性。

7 · 新颖性边界

- 本课题不声称制备了新型碳点、不声称任何绝对量子产率数值、不声称提出新的合成或纯化方法。柠檬酸 + 尿素路线、生物质前驱体路线、微波常压合成、相对量子产率法均为已发表内容。
- 已有工作完成到：明确报道反应条件的微小变化会显著改变碳点荧光并将其用作调控手段（[PMC, 2022](#)）；系统研究反应时间对量子产率的非单调影响；综述指出植物源水热路线在产率与重现性上的不足；文献量子产率报道跨度从 $< 5\%$ 到 90% 。
- 本项目的贡献是评价维度的转换，且构成主结论：把“条件敏感性”从调控优点重新表述为可测量的重现性缺陷，用三层嵌套设计给出方差分解，并定量回答“绿色前驱体路线要付出多少重现性代价、代价出在哪一层”。
- 价值：碳点文献以“我做出了高量子产率”为主流叙事，而重现性数据几乎缺席。一份诚实的批次间 RSD 与方差分解，对判断这类材料能否走出实验室是直接相关的证据；而“代价出在哪一层”则直接指向改进方向。
- 若结果是“两条路线重现性无显著差异”，这同样是有效结论（且是对绿色路线的有力支持），但必须由验收标准第 3 条证明本研究有能力分辨 2 倍的 RSD 差异——若 8 批次不足以分辨，须报告所需批次数的功效估算。

8 · 决策门槛 (go / no-go)

- 第 6 周末：微波炉功率标定完成且荧光定量方法学校验通过（参比线性 $R^2 \geq 0.995$ 、双标准品交叉偏差 $\leq 20\%$ ）。未通过 $\rightarrow$ 降级路径 A：加热方式从微波改为硅油浴 180°C 常压加热（控温更稳、无占空比问题），全部研究框架不变，且“微波法功率不可控”本身作为一节定量报告。
- 第 12 周末：柠檬酸 + 尿素路线必须稳定产出具备激发依赖发射特征的碳点，且 3 次预实验的表观相对量子产率 $RSD \leq 30\%$ 。若完全无荧光或无法重复 $\rightarrow$ 降级路径 B：把体系换成更易成功的柠檬酸 + 乙二胺（常压回流）路线，前驱体对比改为“化学品 vs 柚子皮提取液”，假设与统计框架完全保留。
- 第 24 周末：两条路线各 ≥ 6 批次完成。若批次层方差分量的自助置信区间宽度 $>$ 分量估计值本身 $\rightarrow$ 增加批次数至 12，或降级路径 C：把三层嵌套简化为两层（批次 $\rightarrow$ 测量，取消纯化层），用节省下来的工作量

增加批次：方差分解框架保留，只是纯化层的贡献改为在讨论中定性说明。

- 第 32 周末：干预性验证（总固形物归一化投料）完成。若归一化后批次层方差未下降，如实报告并把"重现性瓶颈不在投料量"作为主结论的一部分——这个否定结论比一个正面结论更能指导后续工作。
 - 外部表征风险项：若指导教师要求颗粒尺寸证据，须在第 28 周前联系高校共享平台预约 TEM 送样（每路线 1–2 个代表样）。联系不上则主结论不变，论文中明确写"未做形貌与尺寸表征，全部结论基于相对光学性质，且不声称任何绝对量子产率"。
 - 第 44 周数据冻结：第 48 周完成实验视频。
 - 实验视频要求：录制五段——(a) 微波炉功率标定（温升法）；(b) 一次完整合成（称量、加热、冷却、纯化）；(c) 暗箱内 365 nm 激发下的样品与参比同框成像 SOP；(d) 分光光度计吸光度读数；(e) 不同批次样品在同一 UV 灯下的并排对比（这段最能直观呈现批次差异）。累计 ≥ 12 分钟。
 - 安全与废液：微波加热含水有机体系存在过热暴沸风险，容器必须敞口或加透气盖、装量 $\leq$ 容器容积 1/3、分段短时加热（每次 ≤ 60 s）并在每段之间取出观察，严禁密闭容器微波加热、严禁无人值守；油浴须用带控温的电加热套并配温度计，油温不得超过硅油闪点，操作区备灭火毯。365 nm 紫外光会伤害眼睛与皮肤，须戴防紫外护目镜，暗箱须闭合后再开灯。硫酸奎宁溶液含稀硫酸，按酸性废液处理；碳点废液与透析废液按"含有机纳米材料废液"单独收集交学校处置，不得排入下水道（纳米材料的环境行为未知，这一处置理由须写入论文的安全章节）。不涉及剧毒品、易制毒/易制爆化学品、氢氟酸与高压反应釜。
 - 预算裁剪顺序：先砍透析袋（改用 0.22 μm 滤膜 + 离心超滤，纯化程度下降但两条路线一致，比较仍成立）；再砍第二荧光标准品（省 60 元但失去交叉验证，不推荐）；UV 灯、参比品与前驱体不可砍。
 - 适配学生：适合能接受"整整一年在重复同一个合成"的学生，且必须明确接受本课题的成果是一组变异数据而非一个高性能材料。不适合追求"做出更亮的碳点"的学生。
-

C18 · 普鲁士蓝类似物纳米酶类过氧化物酶活性的口径依赖性：同一组原始数据在四种报告惯例下的活性排序是否反转

推荐优先级 ★★★☆☆ · 分族：催化 · 纳米酶 · 预算 ≈ 1200 元 · 表征依赖：低（动力学不依赖形貌表征；组成确证须外部平台，非必需） · 技能取向：酶动力学 + 分光光度法 + 元分析式统计

1 · 研究问题

用同一台分光光度计、同一批底物、同一实验者测得三种普鲁士蓝类似物（Prussian blue analogue, PBA: Fe-Fe、Co-Fe、Ni-Fe）的类过氧化物酶（peroxidase-like）稳态动力学原始数据后，改变四项报告口径——氧化型 TMB 的摩尔吸光系数取值、催化剂用量的归一化基准（质量 / 过渡金属摩尔数 / 亚甲基蓝可及表面积）、底物浓度扫描范围、缓冲液 pH——三种材料的“活性排序”是否会发生反转？若会，反转由哪一项口径主导？

2 · 研究背景与空白

纳米酶（nanozyme）是具有类酶催化活性的纳米材料。普鲁士蓝及其类似物是由过渡金属离子与氰根桥连构成的框架结构，被公认具有类过氧化物酶活性：在 H_2O_2 存在下催化 3,3',5,5'-四甲基联苯胺（TMB）氧化生成蓝色产物，652 nm 处有强吸收。其催化循环被描述为“TMB 先把纳米酶还原为普鲁士白、 H_2O_2 再把普鲁士白氧化回去”的乒乓机制。活性通常以米氏方程的 K_m 与 $V_{\max}$ （或催化效率 $V_{\max}/K_m$ ）报告。

已有工作到哪一步：这一体系已被反复研究。PB 纳米颗粒负载 MoS_2 用于多巴胺比色检测（[Biosensors](#), 2022）；刻蚀的 CoFe PBA 展现“更优的类过氧化物酶活性”，归因于电子与结构性质改变带来的底物亲和力和提升（[Int. J. Biol. Macromol.](#), 2024）；聚乙烯亚胺介导的不同类型 PB 纳米颗粒的类过氧化物酶活性对比（2025）；PB 纳米酶的尺寸调控与 ELISA 类免疫分析应用（2022）；甚至已有用停流光谱做前稳态动力学的工作（[J. Phys. Chem. Lett.](#), 2024）。

空白在——这些工作全都在报告并互相比对 K_m 与 $V_{\max}$ ，却各自采用不同的 $\epsilon(652\text{ nm})$ 取值、不同的催化剂归一化基准、不同的底物浓度范围与 pH；没有工作检验过“这些口径差异是否足以改变谁比谁更活跃的结论”。这个空白适合一年期中学课题：合成是室温共沉淀（常温常压，无反应釜）；测定只需可见分光光度计与恒温水浴；核心贡献来自设计与统计而非仪器；且“排序稳健”与“排序反转”都是有效结论。

3 · 可检验假设

H1 三种 PBA 在统一口径下的催化效率（ $V_{\max}/K_m$ ）排序，在改变归一化基准（质量 → 过渡金属摩尔数）后发生至少一对反转；而改变 $\epsilon(652\text{ nm})$ 取值（文献常见的 3.9×10^4 与 $3.55 \times 10^4\text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$ 等）不改变排序（因其对三者是同一乘性因子）——即口径的影响是结构性的而非数值性的，只有改变“每份催化剂算多少”的口径才会反转排序。

H2 底物浓度扫描范围的选择会系统性偏置 K_m ：当扫描上限低于真实 K_m 的 2 倍时，拟合得到的 K_m 相对全范围拟合值低估 $\geq 25\%$ ；该偏置幅度随材料不同而不同，因而在窄范围扫描下三种材料的 K_m 排序也可能反转。若 H2 被否定，则说明该体系的米氏拟合对范围不敏感，可作为对现有报告实践的一次背书。

4 · 量化验收标准

1. 方法学校验（硬门槛，两级）：(a) 体系级：以商品辣根过氧化物酶（horseradish peroxidase, HRP）为金标准酶，在本实验室条件下测其对 TMB 的 K_m ，结果须与文献报道的同条件值偏差 $\leq 30\%$ ；同时 TMB 氧化产

- 物在 652 nm 的标准曲线（用已知浓度的氧化型 TMB，或用 HRP 完全转化法生成）线性 $R^2 \geq 0.995$ 。(b) 动力学级：所有反应初速率必须取自吸光度 - 时间曲线的线性段（转化率 $< 10\%$ ），线性段拟合 $R^2 \geq 0.99$ ；同一条条件的 3 次独立平行反应初速率 $RSD \leq 8\%$ 。两级任一不过关，全部动力学与排序结论无效。
2. 重复口径写死：每种材料 ≥ 3 个独立合成批次，每批次做完整的双底物米氏扫描（TMB 与 H_2O_2 各 ≥ 6 个浓度点），每点 3 次独立平行。报批次间与批次内 RSD ，两者分开。
 3. 归一化基准的三条通道必须全部测定：(i) 质量（真空/烘箱干燥后称重）；(ii) 过渡金属摩尔数（用 EDTA 络合滴定或邻菲罗啉/丁二酮肟比色法测定 Fe/Co/Ni 含量，须先用标准溶液校验回收率 95 - 105%）；(iii) 亚甲基蓝吸附法测得的可及表面积（作为 BET 的低成本替代量，须明确声明其为替代量而非 BET 面积）。
 4. 统计口径：米氏参数拟合一律用非线性最小二乘直接拟合 $v - [S]$ 曲线，明确写“不用 Lineweaver - Burk 双倒数图求参数”及其原因（双倒数变换会放大低浓度点误差并给出有偏估计），但可附双倒数图作为可视化。参数置信区间由 2000 次自助抽样给出。排序反转的判定：两材料的参数自助分布的差值分布中，0 是否落在 95% 区间内。多重比较用 Holm 校正。
 5. 口径敏感性分析：对同一组原始数据，按 4 项口径 $\times$ 各 2 - 3 个常见取值做全组合重算（ ≥ 24 种组合），报“三材料排序”在这些组合下的分布（即多少比例的口径组合给出哪种排序），并用 Kendall's W（一致性系数）量化排序的稳健程度。
 6. 可复现性：全部原始吸光度 - 时间曲线、合成记录、金属含量滴定记录与重算脚本开源，第三方可用同一组原始数据复现任意口径下的结论。

用途	来源 / 工具
合成（常温常压共沉淀）	铁氰化钾（AR，≈45 元）、氯化亚铁/硫酸亚铁铵、氯化钴、氯化镍（AR 各 100 g，合计 ≈150 元）、柠檬酸（稳定剂）。室温滴加共沉淀 + 陈化 + 离心洗涤， 无反应釜、无高温
底物	TMB（3,3',5,5'-四甲基联苯胺，1 g，≈250 元，用量极小）、30% H ₂ O ₂ （≈20 元）、二甲基亚砜（溶解 TMB，≈30 元）
缓冲液	乙酸-乙酸钠缓冲（pH 3.5 / 4.0 / 4.5），≈40 元
金标准酶	辣根过氧化物酶 HRP（冻干粉，最小包装，≈200 元）。 仅用于方法学校验，不计入本项目贡献
金属含量测定	EDTA 标准溶液（可复用 C15 试剂）、邻菲罗啉（测 Fe）、丁二酮肟（测 Ni）、亚硝基 R 盐或 EDTA 直接滴定（测 Co），合计 ≈120 元
表面积替代量	亚甲基蓝（AR 25 g，≈25 元）吸附法
仪器	可见分光光度计（652 nm，须 核实 是否支持时间扫描/定时读数模式——第 4 周前核实；不支持则用手动定时读数，须在方法中说明并评估其对初速率精度的影响）、恒温水浴（37 °C）、离心机、pH 计、分析天平—— 全部为中学常规配置
统计	Python + SciPy（curve_fit 非线性拟合）、statsmodels、自写自助法；纯 CPU
外部依赖（可选）	若需确证 PBA 的晶相与金属配比，须联系高校共享平台送 XRD 或 ICP 样品—— 属外部依赖，列入 go/no-go 风险项 。不依赖该表征的主结论：口径依赖性的检验只需同一组原始动力学数据与本地可测的归一化量，与晶相确证无关；但论文中不得声称材料的确切化学式，只能以"按投料比命名的 Fe-Fe / Co-Fe / Ni-Fe 类似物"表述
对照文献	Biosensors 12(5):260 (2022) PB/MoS ₂ 纳米酶； Int. J. Biol. Macromol. (2024) 刻蚀 CoFe PBA；PEI 介导 PB 纳米颗粒 POD 活性对比 (2025)； J. Phys. Chem. Lett. (2024) 前稳态动力学。 仅作参考对照与新颖性界定，不计入本项目贡献

6 · 方法路径

- 核实分光光度计的时间扫描能力并完成 HRP 金标准校验（验收标准第 1(a) 条），未通过不得进入纳米酶测定。
- 建立共沉淀 SOP：滴加速率、陈化时间、洗涤次数、干燥条件全部书面固定，三种材料使用完全一致的 SOP（这是可比性的前提，须在论文中显式声明）。
- 每种材料制 ≥ 3 独立批次，测三条归一化通道（质量、金属摩尔数、亚甲基蓝可及面积），金属含量测定先用标准溶液验证回收率。
- 在统一口径（pH 4.0、37 °C、全范围底物扫描、ε 取 3.9×10⁴）下测完整双底物米氏动力学，非线性拟合得 K_m 与 V_{max} 及自助置信区间。
- 做口径敏感性重算：4 项口径 × 各 2–3 取值的全组合，输出排序分布与 Kendall's W。其中"底物范围"一项须用同一组原始数据的子集重拟合（截取低浓度段），以确保比较的是口径而非实验差异。
- 检验 H1（归一化基准导致反转）与 H2（范围导致 K_m 偏置）。
- 独立交叉校验：用第二种显色底物（ABTS 或邻苯二胺 OPD）在统一口径下重测活性排序，检验排序是否随底物改变；并对同一批材料在另一台分光光度计（若校内有第二台）上复测一条曲线，量化仪器间差异相对口径差异的大小。

7 · 新颖性边界

- 本课题不声称发现新纳米酶、不声称任何材料"活性最高"、不提出新的活性单位定义、不做任何生物传感应用声明。PBA 的类过氧化物酶活性、TMB 显色体系、米氏动力学分析均为已发表内容。
- 已有工作完成到：PB/PBA 的类过氧化物酶活性被反复证实并用于多巴胺、 H_2O_2 等的比色检测；刻蚀 CoFe PBA 通过改变电子与结构性质提升催化效率（*Int. J. Biol. Macromol.*, 2024）；不同 PB 纳米颗粒类型的活性对比（2025）；尺寸调控与 ELISA 类应用（2022）；前稳态动力学的停流研究（2024）。这些工作都在报告并比较 $K_m / V_{\max}$ ，但各自的 ϵ 取值、归一化基准、底物范围与 pH 并不统一。
- 本项目的贡献是问题方向的转换，且构成主结论：被检验的对象不是"哪种 PBA 更活跃"，而是"在现行报告惯例的差异下，'哪种更活跃'这个问题是否有稳定答案"；交付物是同一组原始数据在 ≥ 24 种口径组合下的排序分布与一致性系数。
- 价值：纳米酶文献的横向比较几乎全部建立在跨实验室的 $K_m / V_{\max}$ 数值之上。如果排序会因归一化基准而反转，那么这类比较的可信度需要被重新评估；如果排序稳健，则现行实践获得一次量化的背书。两种结果都直接作用于该领域的数据可比性。
- 若结果是"排序在所有口径下稳健（Kendall's $W \geq 0.9$ ）"，同样是有效结论，但必须由验收标准第 4 条的自助置信区间证明本研究有能力分辨会导致反转的效应量。

8 · 决策门槛 (go / no-go)

- 第 4 周末：核实分光光度计能否做定时/时间扫描读数。不能 → 降级路径 A：改用固定时间法（fixed-time assay，反应 3 min 后加终止液读单点吸光度），米氏拟合仍成立但精度下降；须在方法学校验中额外证明固定时间点仍落在线性段内（转化率 $< 10\%$ ）。
- 第 10 周末：HRP 金标准校验必须通过（ K_m 与文献偏差 $\leq 30\%$ ）。未通过 → 降级路径 B：放弃绝对参数的准确性诉求，把研究彻底定位为相对比较与口径敏感性（这本身就是主结论），并在论文中明确声明"本工作报告的 K_m 为体系内相对值，不与跨实验室数值直接比较"——这一声明本身正是本课题要论证的观点，因此降级不损伤主结论。
- 第 18 周末：三种 PBA 必须各有 ≥ 2 个批次成功合成且呈现明确的 TMB 显色活性（同条件下 652 nm 吸光度增长速率至少为空白的 5 倍）。若某种材料无活性 → 材料数由 3 减为 2，排序问题退化为"两两比较是否反转"，Kendall's W 改用符号检验，主结论框架保留。
- 第 28 周末：金属含量测定的回收率必须达 95 – 105%。达不到 → 降级路径 C：归一化基准从三条减为两条（质量 + 亚甲基蓝可及面积），H1 改述为"质量归一 vs 面积归一是否导致反转"，口径敏感性框架不变。
- 第 36 周末：口径全组合重算与第二底物交叉校验完成。若第二底物给出完全不同的排序，如实报告并把"排序还依赖底物选择"作为主结论的强化证据——这不是失败，而是结论的加强。
- 外部表征风险项：若指导教师要求组成确证，须在第 30 周前联系高校共享平台预约 XRD 或 ICP 送样（每种材料 1 个代表样）。联系不上则主结论不变，材料一律以投料比命名，论文中明确写"未做晶相与元素定量确证"。
- 第 44 周数据冻结；第 48 周完成实验视频。
- 实验视频要求：录制五段——(a) 三种 PBA 的共沉淀合成（同一 SOP 并排展示）；(b) 离心洗涤与干燥；(c) 一次完整的动力学测定（加样、恒温、定时读数，含分光光度计屏幕特写）；(d) 三种材料在同一条件下的显色速率并排对比（最直观的一段）；(e) 金属含量滴定终点近景。累计 ≥ 12 分钟。

- **安全与废液：**铁氰化钾及全部普鲁士蓝类似物严禁与酸接触、严禁加热蒸干（可能释放 HCN），全部合成与洗涤在中性至弱碱性条件下进行，废液保持 $\text{pH} > 9$ 单独收集。TMB 虽被认为是联苯胺的低毒替代品，仍须视为可疑致突变物：称量在通风橱内进行、全程戴丁腈手套、禁止用嘴吸移液管，其 DMSO 母液密封避光冷藏并单独收集废液。30% H_2O_2 具强氧化性与腐蚀性，避光冷藏，避免与金属粉末或有机物大量混合。含钴、镍盐属有毒有害物质，须由教师管理，含钴镍废液单独收集交危废渠道。不涉及剧毒品、易制毒/易制爆化学品、氢氟酸、高压反应釜与明火有机溶剂操作。
 - **预算裁剪顺序：**先砍 HRP 金标准（省 200 元，但方法学校验降级为"仅用氧化型 TMB 标准曲线 + 线性段判据"，可行但说服力下降）；再砍第二底物交叉校验（省 60 元，不推荐）；再砍材料种数（3 $\rightarrow$ 2）；TMB 与 PBA 前驱体不可砍。
 - **适配学生：**适合能理解"米氏方程只是一个拟合模型"、愿意做元分析式统计的学生。仅在学生已具备基本的化学动力学概念、且明确接受"结论可能是'现有比较方式不可靠'这种得罪人的说法"时才启动。
-

C19 · 水系锌—二氧化锰电池的自放电：24 小时开路电压衰减能否预测 7 天容量损失，以及电解液添加剂对"循环性能"与"储存性能"的解耦

推荐优先级 ★★☆☆☆ · 分族：电化学储能 · 预算 ≈ 1900 元（含一次性仪器，可与 C13 共用） · 表征依赖：低（无需 SEM/XRD；电极形貌变化以外部平台为可选项） · 技能取向：电化学测试 + 时间序列拟合

1 · 研究问题

水系锌—二氧化锰（ $\text{Zn}-\text{MnO}_2$ ）电池搁置过程中的开路电压（open-circuit voltage, OCV）衰减，能否作为 7 天自放电容量损失的短时替代指标（即用 24 h 的 OCV 曲线预测一周后的容量损失）？其次，常用电解液添加剂（ Mn^{2+} ）在改善循环库仑效率的同时，是否同步改善自放电——还是"循环性能"与"储存性能"这两个维度是解耦的？

2 · 研究背景与空白

水系锌离子电池以其安全性（水系电解液不可燃）、低成本与环境友好被广泛研究。 $\text{Zn} \parallel \text{MnO}_2$ 体系的典型电解液是 ZnSO_4 水溶液，正极侧的 Mn^{2+} 溶出是主要失效机制之一。自放电则来自另一组过程：锌负极在酸性水溶液中的析氢腐蚀、正极的自还原、以及穿梭型副反应。

已有工作到哪一步：这一领域的添加剂研究非常密集。 Mn^{2+} 添加剂被证明可提高初始库仑效率与电解液离子导电，并通过补偿正极 Mn^{2+} 溶出、在正极表面形成多孔纳米结构 MnO_x 膜而改善循环（[Nat. Commun.](#) 8:405, 2017 等）；含磷添加剂被用于抑制自放电、腐蚀与析氢，从而提升库仑效率、循环寿命与货架期（美国专利 12218314）； $\text{Zn}(\text{CF}_3\text{SO}_3)_2$ 类电解液可实现接近 100% 的库仑效率；调控 Zn^{2+} 溶剂化结构与构筑（准）固体电解质界面膜以抑制析氢与枝晶已成为主流策略（[Int. J. Extreme Manuf.](#), 2024 综述）。

空白在——绝大多数工作的因变量是"循环寿命 / 容量保持率"，自放电即使被提及，也是以"搁置 X 天后容量保持 Y%"的单点形式出现；没有工作把 OCV 衰减曲线参数化并检验其对自放电的预测能力，也没有工作系统检验"改善循环"与"改善储存"是否解耦。这个空白恰好适配中学的设备条件：长循环测试需要专业电池测试仪与数月机时，而 OCV 记录只需一台高阻抗电压表或廉价恒电位仪 + 数据记录，是典型的"绕开设备墙"的课题形态。

3 · 可检验假设

H1 搁置期间的 OCV 衰减在 $\log(t)$ 坐标上近似线性（拟合 $R^2 \geq 0.95$ ）；其斜率 $|dV/d\log t|$ 与 7 天后实测容量损失率之间的相关 $R^2 \geq 0.80$ ，且用 24 h 数据外推预测 7 天容量损失的相对误差 $\leq 25\%$ 。

H2 加入 0.1 mol/L MnSO_4 添加剂后，循环库仑效率显著提升（前 20 次循环平均库仑效率提高 ≥ 3 个百分点， $p < 0.05$ ），但 7 天自放电率无显著改善（差异的 95% 置信区间包含 0）——即两个性能维度解耦。若 H2 被否定（自放电同步改善），须给出改善幅度并讨论其与文献中含磷添加剂机理的异同。

4 · 量化验收标准

- 方法学校验（硬门槛，两级）：
 - 仪器级：恒流充放电通道用已知阻值的精密电阻替代电池做恒流放电校验，实测电流与设定值偏差 $\leq 2\%$ ；电压测量用万用表比对，偏差 $\leq 5\text{ mV}$ ；OCV 记录通道的输入阻抗须 $\geq 10\text{ M}\Omega$ （否则测量本身就在放电，须核实并记录）。
 - 体系级：在标准配方（2 mol/L ZnSO_4 + 0.1 mol/L MnSO_4 ）下组装的电池，首次放电比容量须落在文献报道的同体系区间内（约 200 – 300 mAh/g，以 MnO_2 计），偏差超出须先排查装配再进入实验。两级任一不过关，全部电池数据无效。

2. 重复口径写死：每个电解液配方 ≥ 5 个独立组装的电池（不同日期、不同电极片），每个电池的自放电实验独立进行；报电池间 RSD。RSD $> 30\%$ 时必须补做至 8 个——电池组装的重现性是本领域最大的噪声源，不得用“取最好的三个”的方式处理数据，此禁令须写入方法。
3. 自放电口径预先写死：定义为“满充后开路搁置 t 天，再以同一倍率放电所得容量 / 搁置前同电池的放电容量”。搁置时间点取 1 / 3 / 7 天，全部在同一恒温环境 ($25 \pm 2^\circ \text{C}$ ，须记录实际温度曲线)。OCV 每 10 min 记录一次，连续 7 天。
4. 对照与空白预先写死：每批实验含——空白电解液 (2 mol/L ZnSO_4 无添加剂) 对照、同批次组装但立即放电的“零搁置”对照（用于扣除组装批次差异）、以及一个只连接测量仪器而不搁置的“仪器漏电”对照（量化测量本身造成的电荷损失）。
5. 统计口径：自放电率与库仑效率的组间比较用混合效应模型（电池为随机效应，配方为固定效应），前提检查预先写死（残差正态性与方差齐性；不满足则对损失率做 logit 变换后重做）。预测能力用留一交叉验证（leave-one-cell-out）报预测相对误差，不用同一批数据的拟合 R^2 冒充预测能力。时间序列拟合给自助置信区间（2000 次）。
6. 可复现性：电极配方与涂布参数、组装图纸、全部 OCV 时间序列与充放电原始数据、拟合脚本开源。

5 · 数据与工具

用途	来源 / 工具
正极材料	电解二氧化锰 EMD 或商品 MnO_2 (AR 100 g, ≈ 60 元)；导电剂乙炔黑 (≈ 40 元)；水系粘结剂 CMC/SBR (≈ 40 元，明确不使用 NMP/PVDF 体系以避免有毒有机溶剂)
负极	锌箔 (0.1 mm, ≈ 50 元)，使用前用稀酸短时清洗并记录处理 SOP
电解液	七水硫酸锌 (AR 500 g, ≈ 40 元)、一水硫酸锰 (AR 500 g, ≈ 40 元)
隔膜与壳体	玻璃纤维隔膜 (≈ 60 元)；CR2032 扣式壳 + 手动封口钳 (≈ 180 元)，或亚克力自制夹具式电池 (≈ 80 元，可拆卸、便于观察)
集流体	石墨纸 / 钛箔 / 不锈钢网 (≈ 60 元)
测试	千元级国产恒电位仪（可与 C13 共用；须 核实 是否支持恒电流充放电与长时 OCV 记录，第 4 周前核实）；备选 Arduino + 精密电阻恒流放电 + ADS1115 高精度 ADC 数据记录 (≈ 250 元)
恒温	恒温培养箱或恒温水浴 ($25 \pm 2^\circ \text{C}$)，配温度记录仪
统计	Python + SciPy / statsmodels（混合效应模型、时间序列拟合、自助法）；纯 CPU
外部依赖 (可选)	若需锌负极枝晶或正极形貌证据，须联系高校共享平台送 SEM 样品—— 属外部依赖，列入 go/no-go 风险项 。不依赖该表征的主结论：OCV 预测能力与添加剂的解耦检验只需电学量，与形貌无关
对照文献	Nat. Commun. 8:405 (2017) Zn– MnO_2 高能量密度水系电池；US 12218314 含磷添加剂抑制自放电； Int. J. Extreme Manuf. (2024) 水系锌电解液添加剂综述。 仅作方法对照与新颖性界定，不计入本项目贡献

6 · 方法路径

1. 核实并校验测试通道（验收标准第 1(a) 条）：用精密电阻做恒流与 OCV 通道校验，测定测量系统自身的漏电流。未通过不得组装电池。
2. 建立电极制备与组装 SOP：浆料配比、涂布厚度（用刮刀与厚度规控制）、干燥条件、活性物质载量（每片称重，控制在 $\pm 10\%$ 内）、装配压力，全部书面固定。
3. 完成体系级复现（第 1(b) 条），确认首次比容量落入文献区间。

4. 两种配方（含/不含 0.1 mol/L MnSO_4 ）各组装 ≥ 5 个电池，做 20 次循环测库仑效率。
5. 满充后进入自放电实验：恒温搁置，连续记录 OCV 7 天，在 1 / 3 / 7 天节点取部分电池放电测残余容量。
6. 拟合 $\text{OCV} - \log(t)$ 关系，建立斜率与容量损失的回归，用留一交叉验证评估 24 h 外推预测 7 天的能力（H1）。
7. 独立交叉校验：用混合效应模型分别检验添加剂对库仑效率与自放电率的效应（H2）；另设一组只做 OCV 记录、从不放电的电池，7 天后一次性放电，与"中途多次放电"组比较，量化"测量过程本身是否扰动了自放电"——这是本课题最容易被评委问到的问题，必须预先设计对照。

7 · 新颖性边界

- 本课题不声称提出新电极材料、新电解液配方或新添加剂，不声称任何性能纪录。 $\text{Zn} - \text{MnO}_2$ 水系电池、 Mn^{2+} 添加剂、自放电现象均为已发表内容。
- 已有工作完成到： Mn^{2+} 添加剂提升初始库仑效率与循环稳定性并给出机理（[Nat. Commun.](#), 2017）；含磷添加剂被专利化用于抑制自放电与析氢（US 12218314）；三氟甲磺酸锌体系达到接近 100% 库仑效率；溶剂化结构调控与界面膜构筑成为主流策略（2024 综述）。这些工作的主因变量是循环寿命与容量保持率，自放电多以单点百分比出现。
- 本项目的贡献是评价维度的转换与问题方向的转换，且构成主结论：一是把自放电从"单点百分比"改为"OCV 衰减的可外推参数化"，并用留一交叉验证给出预测误差；二是把"添加剂好不好"改为"添加剂对循环维度与储存维度的作用是否解耦"。
- 价值：自放电测试的时间成本是电池评价中最贵的一项，一个经过交叉验证的短时替代指标有直接的方法学价值；而"循环好 $\neq$ 储存好"若被证实，则意味着大量只报循环数据的工作对货架期没有发言权。
- 若结果是"OCV 斜率无法预测容量损失（预测误差 $> 50\%$ ）"，这同样是有效结论——它说明自放电由 OCV 之外的过程主导，须报告并讨论可能的机制归属；此时必须由验收标准第 2、5 条证明本研究有能力分辨该量级的相关性，而不是被电池组装噪声淹没。

8 · 决策门槛（go / no-go）

- 第 4 周末：恒电位仪的恒流充放电与长时 OCV 记录能力核实完成，且电阻校验通过（电流偏差 $\leq 2\%$ ）。不支持 $\rightarrow$ 降级路径 A：改用 Arduino + 精密电阻 + 高精度 ADC 自建记录系统（ ≈ 250 元），恒流放电用恒流源模块或恒电阻放电（后者须在数据处理中做电流校正），全部研究框架保留。
- 第 14 周末：电池组装重现性必须达标——同配方 5 个电池的首次放电容量 $\text{RSD} \leq 25\%$ 。超标 $\rightarrow$ 降级路径 B：从扣式电池改为可拆卸亚克力夹具式电池（组装压力可控、可反复观察、单个成本更低），并把"组装重现性"作为独立一节定量报告；若仍超标，则把研究对象从"配方间比较"改为"同一电池内的时间序列自相关"，即 H1（OCV 预测能力）成为唯一主结论，H2 降为探索性分析——H1 不依赖电池间比较，因此这条降级路径保住了主结论。
- 第 24 周末：7 天自放电实验至少完成 2 轮。若 7 天容量损失 $< 5\%$ （信号太弱无法分辨） $\rightarrow$ 把搁置时间延长至 21 天，或提高环境温度至 40°C 做加速自放电，并明确声明加速条件下的结论不直接外推至室温货架期。
- 第 32 周末：留一交叉验证完成。若预测相对误差 $> 50\%$ ，如实报告"OCV 不是自放电的充分预测量"，并把"扰动对照组"（只记录不放电 vs 中途放电）的结果作为解释性证据——这个否定结论必须写成主结论的一部分。

- **外部表征风险项：**若指导教师要求枝晶或形貌证据，须在第 28 周前联系高校共享平台预约 SEM 送样。联系不上则主结论不变，论文中明确写"未做形貌表征，全部结论基于电学测量"。
 - **第 44 周数据冻结；第 48 周完成实验视频。**
 - **实验视频要求：**录制五段——(a) 浆料配制与电极涂布、称重；(b) 电池组装全流程（含隔膜、电解液滴加、封口）；(c) 恒流充放电曲线的实时采集界面；(d) 7 天 OCV 记录装置的全景（含恒温箱温度显示）；(e) 拆解一枚搁置后的电池并展示锌负极外观变化。累计 ≥ 12 分钟。
 - **安全与废液：**本体系为水系电解液，不可燃、无有机溶剂，这是选择该体系的重要理由之一。但须注意：锌在酸性电解液中析氢，电池严禁完全密封后大电流过充（可能鼓包），自制夹具须留泄压路径；短路会瞬间发热，组装与搬运时正负极须分隔，工作台面铺绝缘垫，禁止金属工具跨接两极。硫酸锌与硫酸锰属有害物质，含锌锰废液单独收集交学校危废渠道，不得排入下水道；废弃锌箔与电极片作为固废单独收集。稀酸清洗锌箔须戴手套护目镜。不涉及剧毒品、易制毒/易制爆化学品、氢氟酸、高压反应釜与明火有机溶剂操作。
 - **预算裁剪顺序：**先砍扣式壳与封口钳（改亚克力夹具，省 100 元且更便于观察）；再砍恒电位仪（改 Arduino 自建，省 1500 元，但须在第 0 周决定并预留 4–6 周电子调试）；恒温设备与电极材料不可砍。
 - **适配学生：**这是一条设备风险较高的路线。仅在学校已确有恒温设备、且学生（或指导老师）具备基本电子测量能力时才启动。学生须明确接受两个风险：电池组装重现性可能吃掉全部信号；以及主结论可能是"短时指标不可用"。
-

C20 · 生物质磺化炭固体酸的结构—活性替代指标：在无谱学表征条件下，滴定酸密度与简易疏水性指数谁能预测酯化 TOF

推荐优先级 ★☆☆☆☆（最高风险） · 分族：绿色催化 · 生物质材料 · 预算 ≈ 900 元 · 表征依赖：中（无 FTIR/XRD/TGA；确证磺酸基须外部平台，列为风险项） · 技能取向：材料制备 + 滴定表征 + 催化动力学

1 · 研究问题

在完全没有谱学表征（FTIR / XRD / TGA / 元素分析）的条件下，仅用中学可测的两个替代量——离子交换滴定得到的磺酸位点密度与一个自建的简易疏水性指数（水/正己烷两相分配 + 润湿下沉时间）——能否预测生物质磺化炭催化乙酸酯化反应的转换频率（turnover frequency, TOF）？两个替代量中哪一个的预测力更强，二者的组合是否显著优于单独任一个？

2 · 研究背景与空白

生物质衍生的磺化炭（sulfonated carbon）是一类廉价的固体酸催化剂：把秸秆、果壳、糖等在限氧条件下部分碳化，再用磺化试剂在芳香碳骨架上引入 $-\text{SO}_3\text{H}$ ，即得可回收的布朗斯特酸催化剂，被广泛用于酯化与生物柴油制备。

已有工作到哪一步：结构—活性关系已被反复研究且结论相当一致。磺酸密度与酯化效率正相关，且高木质素含量的生物质有利于磺酸基负载；不同原料的酸密度差异显著（如 NaOH 滴定给出 2.64 mmol/g、稻壳基催化剂达 4.24 mmol/g，另一些体系为 0.493–1.35 mmol H^+ /g）；酸密度普遍用离子交换滴定测定，而催化剂则用 FTIR、XRD、SEM-EDS、TGA 综合表征；文献同时强调酸位点周围的疏水性是获得高 TOF 的关键因素之一（水是酯化的产物，亲水表面会富集水而抑制反应）；一步磺化催化剂的失活机理（磺酸基流失）也已被研究。

空白在——“酸密度”与“疏水性”这两个被公认重要的因素，几乎从未被放在同一组样品上做预测力的直接对比；而疏水性的表征在文献中依赖接触角仪或吸附等温线，从未有人检验过低成本替代量是否够用。这个空白的价值在于：如果一个用滴定与秒表就能测的指数确实有预测力，那么欠资源实验室就能做固体酸的理性筛选。但这条路线同时也是本批课题中风险最高的一条——磺化步骤本身就是主要障碍。

3 · 可检验假设

H1 在同一原料、不同磺化条件（温度 × 时间，≥ 6 个样品）制得的磺化炭系列上，酯化初始 TOF 与磺酸位点密度的相关 $R^2 \leq 0.6$ （即酸密度单独不足以解释活性），而加入简易疏水性指数后的二元线性模型的留一交叉验证 R^2 提升 ≥ 0.15。

H2 疏水性指数与磺化程度呈非单调关系：磺化过度会同时引入羧基/酚羟基等亲水基团，使疏水性指数在中等磺化条件下达到极大；因此存在一个使 TOF 最优的中间磺化条件，而非“磺化越深越好”。若 H2 被否认（TOF 随酸密度单调上升），则本体系中疏水性不是限制因素，须报告该边界条件。

4 · 量化验收标准

1. 方法学校验（硬门槛，三级）：
(a) 滴定级：离子交换滴定（NaCl 交换 + NaOH 滴定）测总酸密度，用商品强酸性阳离子交换树脂（标称交换容量已知）做正确度校验，测得值与标称值偏差 ≤ 10%，3 次平行 RSD ≤ 5%；Boehm 三档滴定用于区分强酸位（ $-\text{SO}_3\text{H}$ ， NaHCO_3 可中和）与弱酸位。
(b) 反应级：以浓硫酸均相催化的同一酯化反应为参照，本实验室测得的转化率—时间曲线须与已发表的同条件结果偏差

- $\leq 20\%$ ；酸值滴定法测转化率的准确度用已知配比的酸/酯混合样校验，相对误差 $\leq 3\%$ 。(c) 疏水性指数级：自建指数须先在已知疏水性梯度的对照物（活性炭、氧化石墨、商品磺酸树脂、硅胶）上验证其能给出与预期一致的排序（Spearman $\rho \geq 0.8$ ），否则该指数不得用于后续分析。三级任一不过关，全部结构-活性结论无效。
2. 样品与重复： ≥ 6 个磺化条件（如 2 温度 $\times$ 3 时间），每条件 ≥ 2 个独立制备批次；每个样品的酸密度、疏水性指数、TOF 各做 ≥ 3 次独立测定，全部报 RSD。
 3. TOF 口径预先写死：TOF = 初始转化速率（取转化率 $< 20\%$ 的线性段）/ 酸位点摩尔数 / 时间，单位 h^{-1} 。必须同时报以质量归一的比活性，并明确说明两种口径可能给出不同排序（与 C18 同源的口径问题）。
 4. 对照与空白预先写死：每批催化实验含——无催化剂空白（量化非催化酯化的贡献，须扣除）、未磺化炭对照（量化炭骨架本身的贡献）、商品磺酸树脂正对照。热滤实验（hot filtration test）作为均相流失对照必须做：反应中途滤去固体后继续反应，若转化率继续上升则说明存在磺酸基流失导致的均相催化，此时全部 TOF 须重新解释——这一对照是本题可信度的生命线。
 5. 统计口径：预测力比较用留一交叉验证 R^2 （不用拟合 R^2 ），模型比较用 AIC 与交叉验证误差双口径；由于样本量小（6–12 个样品），必须明确声明这是探索性关联分析，回归系数的置信区间由自助法（2000 次）给出，且不做因果表述。非单调关系用二次项显著性检验，前提检查写死。
 6. 可复现性：磺化条件记录、滴定原始数据、疏水性指数的测定 SOP 与视频、转化率曲线与全部脚本开源。

用途	来源 / 工具
原料	花生壳 / 稻壳 / 蔗糖（低成本、木质素或糖含量差异明确，≈30 元）
碳化	学校马弗炉，限氧碳化 400 °C（同 C15 的坩埚限氧方案）
磺化（主路线，高风险）	浓硫酸，150—160 °C 常压回流磺化， 单批 ≤ 30 mL 酸、必须在通风橱内、教师全程在场 。若学校无通风橱， 直接走降级路线，不得强行开展
磺化（降级路线，低风险）	硫酸氢钠（NaHSO ₄ ）固相加热磺化，或对甲苯磺酸浸渍—热固定法；两者均无热浓硫酸操作
酸密度	氯化钠（交换）、NaOH 与 HCl 标准溶液、NaHCO ₃ / Na ₂ CO ₃ （Boehm 三档），≈90 元
校验对照	商品强酸性阳离子交换树脂（标称容量已知，≈80 元）；活性炭、硅胶（≈40 元）
疏水性指数	正己烷（≈60 元，用量小，密闭操作）、秒表、量筒；指数定义：固定质量粉末置于水/正己烷两相界面，记录 60 s 内进入水相的质量分数 + 粉末在水面完全下沉所需时间，二者合成一个无量纲指数（定义须在方法中写死并公开）
催化反应	冰乙酸 + 正丁醇（或异戊醇），电加热套回流， 禁止明火 （≈90 元）
转化率	酸值滴定（KOH—乙醇标准液），可复用 C12 的滴定体系
仪器	分析天平、马弗炉、电加热套、恒温水浴、离心机、滴定管、pH 计—— 全部为中学常规配置
外部依赖（可选但重要）	磺酸基的存在 只能由滴定间接推断 ；若需直接证据，须联系高校共享平台送 FTIR（1030 / 1180 cm ⁻¹ 的 S=O 伸缩）或元素分析（S 含量）—— 属外部依赖，列入 go/no-go 风险项 。不依赖该表征的主结论：本课题问的是“两个可测替代量谁的预测力强”，该问题只需替代量与 TOF，不需要磺酸基的谱学确证；但论文中 必须明确写 “—SO ₃ H 的存在由 NaHCO ₃ 可中和的强酸位推断，未经谱学确证”
统计	Python + SciPy / scikit-learn（留一交叉验证）/ statsmodels；纯 CPU
对照文献	Bioresour. Technol. (2013) 生物质炭基固体酸的制备与表征； Renew. Energy (2023) 水热碳化磺化炭球； Renew. Energy (2021) 一步磺化催化剂的失活机理； Carbon (2019) 纳米炭催化剂的表面官能团协同效应。 仅作方法对照与新颖性界定，不计入本项目贡献

6 · 方法路径

1. 先做安全评估与路线选择：核实学校是否具备合格通风橱与教师全程监督条件。不具备则直接采用 NaHSO₄ / 对甲苯磺酸降级路线，且此决定必须在第 2 周前作出并书面记录。
2. 完成三级方法学校验（验收标准第 1 条）：树脂标定滴定、浓硫酸均相参照反应复现、疏水性指数在已知梯度对照物上的排序验证。任一未过不得继续。
3. 制备磺化炭样品系列（≥ 6 条件 × ≥ 2 批次），逐样测酸密度（含 Boehm 分档）与疏水性指数。
4. 测催化性能：固定催化剂用量、醇酸比、温度，取初始线性段算 TOF，同时报质量比活性。**每个样品都必须做热滤对照。**
5. 建立单变量与二元回归模型，用留一交叉验证比较预测力（H1）；检验疏水性指数与磺化程度的非单调关系及 TOF 的极值（H2）。
6. 做失活/回收实验：催化剂回收 3 次，跟踪酸密度与 TOF 的同步衰减，检验流失机理与文献报道的一致性。

7. 独立交叉校验：用第二种酯化底物组合（乙酸 + 异戊醇，可与 C12 共用体系）重测排序，检验替代指标的预测力是否随底物改变；另用 pH 滴定曲线的等当点分布作为酸强度分布的第二独立通道，与 Boehm 分档比对。

7 · 新颖性边界

- 本课题不声称制备新催化剂、不声称任何活性纪录、不声称磺酸基的直接结构证据、不做生物柴油相关的应用性声明。生物质磺化炭、离子交换滴定测酸密度、Boehm 滴定、酯化 TOF 均为已发表方法。
- 已有工作完成到：磺酸密度与酯化效率正相关且木质素含量促进磺酸负载已被定量报道；多种原料的酸密度数值区间（0.493 – 4.24 mmol/g）已被记录；酸位点周围的疏水性对 TOF 的重要性已被明确指出；一步磺化催化剂的失活（磺酸基流失）机理已被研究。这些工作的表征体系都建立在 FTIR / XRD / SEM-EDS / TGA 之上，且从未把“酸密度”与“疏水性”放在同一组样品上做预测力的直接比较。
- 本项目的贡献是评价维度的转换，且构成主结论：不追求做出更活跃的催化剂，而是回答“在只有滴定与秒表的条件下，哪个可测量最能预测活性”，交付物是一个带留一交叉验证的预测力比较与一个公开定义的低成本疏水性指数。
- 价值：固体酸的理性筛选目前依赖高端表征，这把欠资源实验室排除在外。若一个廉价指数确有预测力，其方法学价值直接；若没有，则明确划出了“低成本表征的能力边界”，同样有用。
- 若结果是“酸密度单独就足以预测 TOF ($R^2 > 0.8$)”，同样是有效结论（对现有以酸密度为主要指标的实践的一次独立支持），但必须由验收标准第 2、5 条证明本研究有能力分辨 0.15 量级的交叉验证 R^2 提升——若 6 – 12 个样品不足以分辨，须如实报告统计功效不足并给出所需样品数估计。

8 · 决策门槛 (go / no-go)

- 第 2 周末：安全条件评估完成并书面记录。无合格通风橱或无法保证教师全程在场 → 强制走 NaHSO_4 / 对甲苯磺酸降级路线，不得例外。降级后酸密度可达范围变窄（预计 0.3 – 1.0 mmol/g），须在第 8 周确认该范围内 TOF 仍有可分辨的变化；若无变化则见下一条。
- 第 8 周末：三级方法学校验必须全部通过，且样品系列的酸密度跨度 ≥ 3 倍（如 0.3 → 1.0 mmol/g）。跨度不足 → 降级路径 A：把样品系列从“同原料不同磺化条件”扩展为“不同原料（花生壳/稻壳/蔗糖/活性炭）× 统一磺化条件”，用原料差异制造酸密度与疏水性的跨度，H1/H2 与全部统计框架保留（H2 的“非单调”改述为“跨原料的酸密度 – 疏水性权衡”）。
- 第 16 周末：热滤实验必须给出明确结论。若滤液继续反应（存在严重磺酸基流失），不得继续按非均相催化解释 TOF → 降级路径 B：把课题重心转为“磺酸基流失的定量表征：流失速率与哪个可测替代量相关”，仍是“替代指标预测力”这一主结论框架，只是被预测的对象从 TOF 换成流失率。这条降级路径保住了主结论的形式与全部方法。
- 第 24 周末：疏水性指数的测定重现性必须达 $\text{RSD} \leq 20\%$ 。超标 → 降级路径 C：疏水性指数改为更粗但更稳的二分/三分等级判定（完全下沉 / 部分下沉 / 漂浮），H1 中的连续变量回归改为等级变量的方差分析，预测力比较仍成立但分辨率下降，须在论文中声明。
- 第 32 周末：留一交叉验证与第二底物校验完成。若两个替代量都无预测力（交叉验证 $R^2 < 0.3$ ），如实报告“低成本替代表征不足以预测本体系活性”，并给出所需的最低表征清单——这个否定结论是对欠资源实验室最有用的成果，必须写成主结论而非局限性。

- **外部表征风险项（本课题最大的外部依赖）：**磺酸基的直接证据只能来自 FTIR 或元素分析。须在第 20 周前联系高校共享平台预约送样（2–3 个代表样即可，成本通常 ≤ 300 元）。联系不上则主结论不变，但论文全文必须一致地使用“强酸位点（推断为磺酸基，未经谱学确证）”这一表述，不得在任何一处写成已确证。
- **第 44 周数据冻结；第 48 周完成实验视频。**
- **实验视频要求：**录制六段——(a) 原料磺化与出炉；(b) 磺化操作全过程（通风橱内，含防护装备与教师在场画面；若走降级路线则录 NaHSO_4 固相磺化）；(c) 离子交换滴定与 Boehm 滴定；(d) 疏水性指数测定的完整 SOP（这段最需要清晰记录，因为该指数是自建的）；(e) 一次完整的酯化反应与酸值滴定；(f) 热滤对照实验。累计 ≥ 15 分钟。
- **安全与废液（本批课题中风险最高的一条，务必逐条执行）：**热浓硫酸磺化是本课题的主要危险源——须在通风橱内进行、单批酸量 ≤ 30 mL、使用带温度计的电加热套（严禁明火、严禁油浴超温）、全程佩戴护目镜 + 面屏 + 耐酸手套 + 实验服、教师必须全程在场且不得由学生单独操作、操作台旁常备大量碳酸氢钠固体与洗眼器。反应后的酸浆必须冷却至室温后缓慢倒入大量冰水中稀释（严禁向浓酸中加水），洗涤至中性后滤饼方可干燥。正己烷易燃，用量 ≤ 20 mL、密闭容器、远离热源。马弗炉高温与烟气防护同 C15。全部酸性废液中和至 pH 6–9 后按危废收集；含炭固废单独收集。不涉及剧毒品、易制毒/易制爆化学品、氢氟酸、高压反应釜。
- **预算裁剪顺序：**先砍第二底物交叉校验（省 50 元，不推荐）；再砍原料种类（3 $\rightarrow$ 1）；再砍外部 FTIR 送样（省 300 元，但须全文一致使用“未经确证”的表述）；滴定试剂与树脂标准品不可砍。
- **适配学生：**这是全批风险最高的一条路线。仅在下列三个条件同时满足时才启动——(i) 学校有合格通风橱且教师能全程陪同高危操作，或已明确接受低风险降级磺化路线；(ii) 学生已在其他课题中证明过滴定操作的稳定性（平行 RSD $\leq 1\%$ ）；(iii) 学生与家长明确接受“最终结论可能是这套低成本表征不够用”这一结果。若三条中缺任何一条，建议改选 C11–C15 中的任一条。

附录 A · 倒排时间表（52 周，通用骨架）

终点是 2027 年 9 月 15 日提交截止（按 2026 赛季日程外推，须以当年官方公告为准）。第 0 周为 2026 年 9 月第一周。每个阶段都有可测的硬门槛，不是任务描述。

周次	阶段	硬门槛（可测，不达标即触发降级）
0—2	选题定稿	从本方案选定 1 条主路线 + 1 条备选路线；确认指导老师；核对当年官方参赛指南的学科研究范围
2—6	环境与方法学校验	跑通该路线验收标准第 1 条的方法学校验。 这是全程最重要的门槛：不过关不得进入下一阶段
6—10	数据/器材到位	数据完成下载并通过完整性检查；或器材到齐且装置重复性达标（实验类 $RSD \leq$ 阈值）
10—14	第一次 go/no-go	按各路线写明的阈值判定。不达标立即切换到该路线的具名降级路径， 不要拖到第 20 周
14—28	主体研究	完成主实验/主计算的全部参数点；中期产出可展示的核心图表 3 张
28—34	第二次 go/no-go	主结论方向已明确（含“差异不显著”这一有效结论）；误差棒已能证明有分辨该差异的能力
34—40	交叉校验	用第二种独立方法重算核心量，两者一致或差异有可解释来源
40—46	中文初稿	全文完成；图表定稿；代码与数据整理为可一键重跑的仓库
46—50	英文稿与查重	完成英文研究报告与查重报告 （总决赛强制英文，且提交材料含查重报告）
50—52	提交与答辩准备	系统提交（9 月 15 日截止，逾期不可修改）；英文 PPT 与答辩问答演练

两个容易被忽略的时间陷阱：

其一，英文化不是最后一周的事。总决赛全程英文（研究报告、PPT、答辩），从第 0 周起就要用英文记录关键术语与方法描述，否则第 46 周会成为瓶颈。

其二，11 月 3—10 日的公示期会把研究报告全网公开。这意味着新颖性问题会暴露在公众视野，查重与文献边界必须在提交前就处理干净。

附录 B · 资源清单（全赛道通用底座）

用途	来源 / 工具	说明
文献检索	Google Scholar、arXiv、PubMed、Semantic Scholar、Connected Papers	必须用英文检索 ；中文检索会系统性漏掉近三年工作
全文获取	arXiv、bioRxiv、PMC、作者主页、通过学校图书馆或指导老师获取	不要使用非法下载渠道，公示期会暴露
代码与数据托管	GitHub（公开仓库）+ Zenodo（存档并获取 DOI）	可复现性是评审加分项，也是自证独立完成的证据
计算环境	Python 3.11+（conda/uv 管理）、Jupyter	环境须导出为 <code>environment.yml</code> 或 <code>requirements.txt</code>
统计与作图	numpy / scipy / statsmodels / matplotlib	图表须含误差棒；配色对色盲友好
实验记录	纸质或电子实验记录本，逐日、带日期、不得回填	评委可能追问原始记录；实验类还需按要求提交实验视频
写作	LaTeX（Overleaf）或 Word；参考文献用 Zotero 管理	英文稿建议请母语者或老师润色，但 内容必须学生本人撰写
版本管理	Git，从第 0 周开始提交	连续的提交历史是“学生本人完成”最有力的旁证

附录 C · 红线清单

以下每一条，触碰即可能导致取消参赛/获奖资格，或在评审中被直接判负。

资格红线 1. 指导老师不得来自公司、培训机构或任何谋利机构。组委会**有理由相信学生受培训机构指导完成研究报告，即保留取消资格的权利**。2. 不得跨校组队；同一学科每人只能报名一次；同一篇论文不得报两个学科。3. 参赛信息须据实填写；报告提交截止后不得更改团队与指导老师信息。4. 曾参加或拟参加其他竞赛、已投稿或发表的，**必须在报名时如实申报**。

学术诚信红线 5. 不得将他人（含指导老师、高校课题组）已完成的工作写成本人贡献。研究报告须明确区分“他人已发表成果”“公开数据库/事件表”与“本项目贡献”。6. 公开数据库、已发表事件表、预训练模型只能作为对照与输入，**不得计入学生的数据或方法贡献**。7. 不得给出未标注为推测的物理/因果外推。讨论部分可以推测，但必须明示。8. 提交材料含查重报告；公示期报告全网公开，抄袭与过度重合会被公开发现。9. AI 使用政策以当年官方参赛规则为准（公开公告中未见明文，须核实）。无论政策如何，研究报告的科学内容必须由学生本人完成；若使用 AI 辅助，按当年规则要求致谢。

方法学红线（不违规，但会被评委问倒） 10. 没有方法学校验的结果不可信——所有路线的验收标准第 1 条都是这个门槛。11. 不平衡分类报 accuracy、时间序列随机划分、幂律只做双对数最小二乘、模拟不做收敛性检查：这四项是评审最常见的追问点，方案里已逐条写死正确口径。12. “差异不显著”是有效结论，但必须有误差棒证明有能力分辨该差异；没有误差棒的“无差异”等于没有结论。

附录 D · 本方案的使用方式

1. 先读赛道导语，判断这个赛道的评委口味是否匹配学生的能力结构。跨赛道选题（例如数学好的学生去做经济金融建模的理论建模族）常常比在本赛道硬拼更有胜算。
2. 按可行性顺序从上往下读，不要从最精彩的那条开始。编号越小越稳。
3. 选定一条主路线后，必须再选一条备选路线，且两条最好落在不同的族——这样第一次 go/no-go 触发时有真正的退路。
4. 第 2-6 周的方法学校验是全案的生死线。这一步跑不通，说明学生的技能栈与该路线不匹配，此时换题的成本远低于第 20 周换题。
5. 本方案的赛制信息取自官方公告（yau-awards.com，2020-2026 全部公示文章），但规则每年会改。报名开放后必须重新核对当年的《参赛规则》与六个学科各自的《参赛指南》，尤其是研究范围、评审标准与 AI 使用政策。