arXiv:2603.05188v1 [physics.chem-ph] 5 Mar 2026

逃离水解陷阱：面向耐久光催化共价有机框架逆向设计的智能体工作流

Iman Peivaste[1, 2, *], Nicolas D. Boscher[1], Ahmed Makradi [1], Salim Belouettar [1]

[1] 卢森堡科学技术研究院（LIST），卢森堡埃施-阿尔泽特上炉大道5号，4362

[2] 卢森堡大学物理与材料科学系，卢森堡埃施-阿尔泽特 L-4365

*通讯作者：iman.peivaste@list.lu

2026年3月6日

摘要

共价有机框架（COFs）是极具前景的太阳能制氢光催化剂，然而电子性质最为有利的连接键——亚胺——会在水中迅速水解，从而形成限制实际应用的稳定性–活性权衡。在由节点、连接体、连接键和官能团构成的组合设计空间中进行导航，以识别同时具备活性与耐久性的候选项，仍然是一项艰巨挑战。在此，我们提出**Ara**，一种大语言模型（LLM）智能体，它利用预训练化学知识、给体–受体理论、共轭效应以及连接键稳定性层级，引导搜索同时满足带隙、带边和水解稳定性标准的光催化共价有机框架。在一个由具有不同节点、连接体、连接键和R 基团的候选项组成、并经GFN1-xTB片段流程筛选的空间中，相对于随机搜索和贝叶斯优化（BO）进行评估时，Ara实现了52.7%的命中率（11.5× 随机搜索，$p = 0.006$），在第12次迭代即获得首次命中，而随机搜索为第25次，并且显著优于贝叶斯优化（$p = 0.006$）。对该智能体推理轨迹的检视揭示了可解释的化学逻辑：早期即收敛于亚乙烯基和 β-酮烯胺连接键以确保稳定性，依据吸电子特性进行节点选择，以及系统的R基团优化以将带隙中心置于2.0 电子伏特。对完整搜索空间的穷举评估揭示了该智能体与贝叶斯优化之间互补的利用-探索权衡，表明混合策略或许能够结合两种方法的优势。这些结果表明，大语言模型化学先验能够大幅加速多标准材料发现。

引言

由太阳能驱动的光催化水分解为获取氢燃料提供了一条可持续途径，而寻找高效、稳定的光催化剂仍是材料科学领域的核心挑战之一 [1, 2, 3, 4]。共价有机框架（COFs）是一类由有机结构单元通过可逆共价键组装而成的结晶多孔聚合物，凭借可调节的带隙、高表面积以及模块化合成化学，已成为颇具吸引力的光催化剂候选项 [5, 6, 7]。原则上，通过选择合适的给电子和

吸电子构筑单元，可以合理调控共价有机框架的带隙和导带底（CBM），使其成为逆向设计的天然平台 [8,9]。然而在实践中，电子性能与化学耐久性之间存在根本矛盾：共价有机框架文献中占主导地位的亚胺（C=N）连接键，在光催化运行所处的水相且通常呈酸性的条件下容易发生水解 [10, 11]。这种“水解陷阱”意味着，许多带隙前景良好的共价有机框架在发挥作用之前便会降解。

共价有机框架光催化剂的设计空间十分广阔。将数量适中的七种三角节点、十九种二连接体、四种连接化学和十种芳香族R基取代基组合起来，可得到 820 种受兼容性约束的候选项；而实际的材料库规模可能还会大出几个数量级 [12]。密度泛函理论（DFT）计算虽然准确，但对周期性结构而言，每个候选项都需要耗时数小时，因此穷尽式筛选的成本高得难以承受 [13, 14]。即使是 GFNn-xTB 系列等半经验方法 [15]，每次几何优化也需要数分钟；而该问题具有多准则特性，必须同时满足带隙、带边和稳定性要求，因此无法依据单一描述符进行简单排序。

为探索广阔的材料设计空间，人们提出了多种计算策略。高通量虚拟筛选通过枚举候选项并逐一评估来进行筛选，但当评估次数超过数百次时，其扩展性较差 [16]。贝叶斯优化（BO）基于分子指纹构建高斯过程代理模型，并利用采集函数来平衡探索与利用 [17, 18]；然而，BO 将分子视为特征向量，无法对连接键水解或通过供体–受体工程进行带隙调控等类别化学概念进行推理。机器学习代理模型经训练后可加速筛选 [19, 20, 21, 22] ，但在新材料类别或属性组合缺乏标注数据时，会面临冷启动问题。

大语言模型（LLM）提供了一种本质上不同的方法。大语言模型经过科学文献训练，能够编码化学知识、给体–受体理论、共轭效应和官能团反应性，可作为材料设计的先验知识，而无需针对特定任务的训练数据 [23, 24, 25]。近期研究已探索将大语言模型用作逆合成[26], 、分子性质预测 [27], 、推导设计规则 [28], 和反应优化[29], 的助手，但在具有明确化学推理的多标准材料发现中应用大语言模型，仍鲜有探索。本文介绍 **Ara**，这是一个由 Google 的 Gemini 模型驱动的大语言模型引导智能体 [30] ，通过迭代推理循环探索共价有机框架的设计空间：每一步中，智能体都会收到关于先前选择的反馈（带隙、导带底、稳定性评分），并提出一个新的候选项，同时简要说明其化学依据。我们将该智能体与片段式筛选流程相结合，利用 GFN1-xTB 计算垂直电离势和电子亲和能，并依据 13 种共价有机框架的周期性 DFT 带隙进行校准（斯皮尔曼 $\rho = 0.71$,$p = 0.006$）。综合稳定性指数（S_{CSI}）整合了连接化学、位阻屏蔽和疏水性，用以表征设计问题中的稳定性维度。

我们在五个相互独立的随机种子下，将 Ara 与随机搜索和贝叶斯优化进行比较，每种方法运行 200 次迭代。该智能体的命中率达到 52.7%，是随机搜索的 11.5 倍（$p = 0.006$，曼-惠特尼 U 检验，单侧），并且在累计命中数方面显著优于贝叶斯优化（$p = 0.006$）。对该智能体推理轨迹的分析揭示了可解释的化学逻辑：趋向于采用不可水解的亚乙烯基连接键，避免使用吸电子性过强的节点，并系统地调节 R 基团。对所有候选项进行穷举评估后，确定了 38 个真实命中项，从而能够开展或然分析，揭示出互补的利用-探索权衡：该智能体擅长在预算受限的活动中快速识别高质量候选项，而贝叶斯优化则通过系统探索绘制出更大比例的命中空间。总体而言，这些结果表明，大语言模型化学先验能够显著加速多标准材料发现，并且基于智能体和基于代理模型的策略能够服务于互补的实验目标。

结果

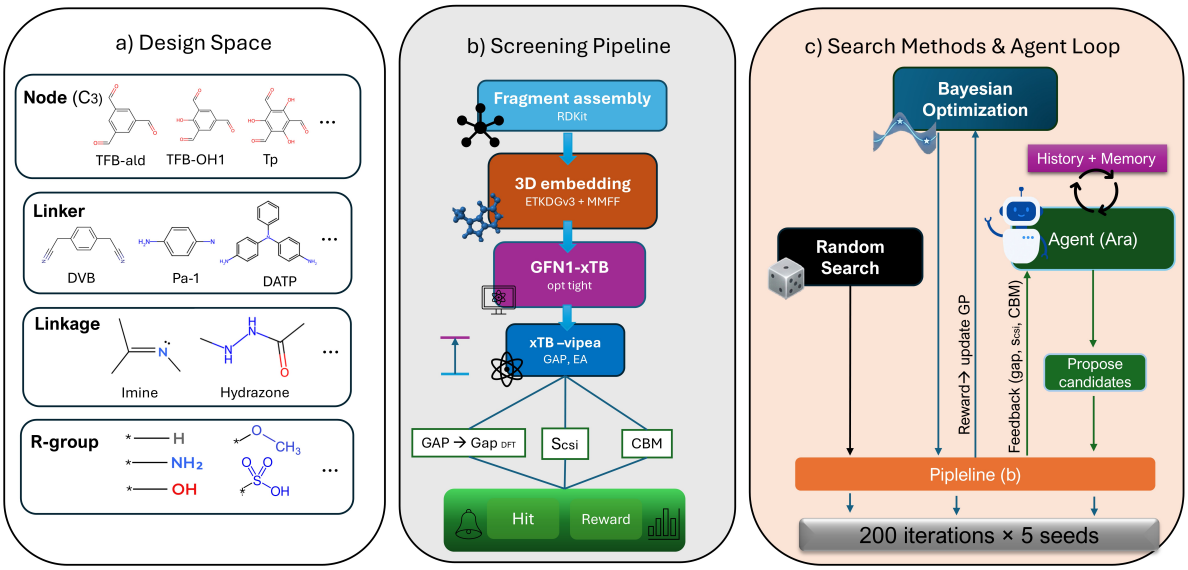


图 1：用于发现 COF 光催化剂的 Ara 智能体工作流概览。(a) 组合设计空间包含 820 个候选项，由 7 个三角节点、19 个二连接体、4 种水解稳定性各异的连接化学以及 10 种芳香族R基取代基构成，并受化学兼容性约束限制。(b) 每个候选项均通过基于片段的筛选流程进行评估：使用 RDKit 组装节点–连接基–节点重复单元，在 3D 空间中嵌入，并通过 GFN1-xTB 进行优化，随后评估带隙（IP–EA，经校准的传递函数映射至 DFT 标度）、导带底（CBM）和综合稳定性指数（S CSI）。只有同时满足全部三项标准的候选项才被归类为命中项。(c) 比较了三种搜索策略：随机采样、基于 Morgan 指纹并使用高斯过程代理模型的贝叶斯优化，以及 Ara 大语言模型智能体。该智能体通过明确的化学推理迭代提出候选项、接收定量反馈，并在 200 次迭代中不断优化其选择。

校准验证了基于片段的筛选

基于片段的基本带隙（IP–EA）由 GFN1-xTB 计算得到，其与六种连接类型下的周期性 DFT 带隙呈显著相关性（斯皮尔曼 $\rho = 0.71$，$p = 0.006$，$n = 13$；补充表 S2），验证了该筛选流程用于候选项比较排序的有效性（图 2）。要获得这一结果，需要进行三次方法上的转变。首先，最初使用 2×2 周期性团簇的尝试颠倒了已知共价有机框架的带隙排序，因此转而采用重复单元片段（∼50–80 个原子），保留了正确的排序 [31]。其次，xTB 单点计算得到的 HOMO–LUMO 带隙与 DFT 数值不相关（$\rho = -0.18$，$p = 0.70$，针对七个共价有机框架的仅含亚胺的子集），而通过 δ-SCF 协议（xtb --vipea）将基本带隙计算为 IP–EA 后 [15]，恢复了预期趋势；这一发现与已知事实一致，即对于有机框架准粒子带隙，Kohn-Sham 本征值差并不适用。第三，五种 β-酮烯胺共价有机框架被排除在校准之外，因为其片段 SMILES 编码的是亚胺互变异构形式，而不是正确的

酮-烯胺结构，导致电子结构计算结果有误。线性传递函数

$$\mathrm{gap}_{\mathrm{DFT}} \approx 0.65 \times \mathrm{gap}_{\mathrm{xTB}} - 0.70 \quad (R^2 = 0.62), \tag{1}$$

将 xTB 基本带隙映射到 DFT 标度，以用于后续筛选。

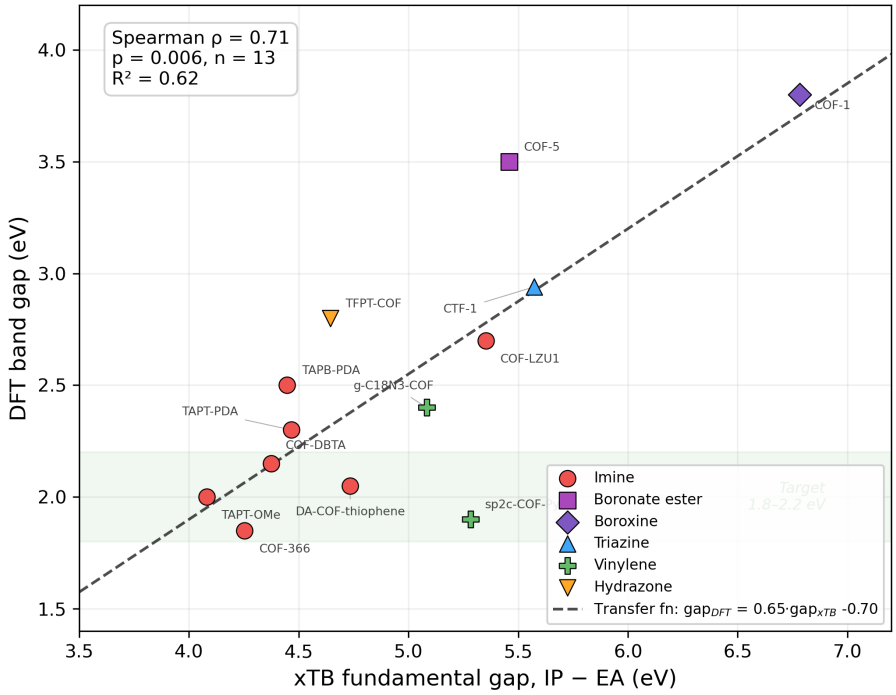


图2：13 种共价有机框架的 xTB（IP–EA）基本带隙与 DFT 带隙的散点图，涵盖六种连接类型，并叠加线性传递函数。校准集包含搜索空间中不存在的硼酸酯、硼氧六环和三嗪连接类型，以拓宽传递函数的适用范围。请注意，CTF-1 采用三嗪环形成作为连接化学，这与含三嗪节点（如 TFPT-ald、TAPT）通过亚胺或其他键型连接的情况不同。

这些结果表明，采用 GFN1-xTB 进行片段级 IP–EA 计算，能够以较低的计算成本可靠地近似周期性 DFT 带隙。传递函数和导带底校准（见方法）构成了后续所有筛选工作的定量基础。

智能体实现了更高的样本效率

我们比较了三种搜索方法——随机搜索、贝叶斯优化（BO）和 Ara（大语言模型智能体）——每种方法均使用五个相互独立的随机种子进行 200 次迭代。如果候选项同时满足以下三个标准，则将其归类为“命中”：校正带隙在 1.8 至 2.2 eV 之间、CBM 低于 0 V，以及综合稳定性指数 $S_{\mathrm{CSI}} \geq 0.7$ （见方法）。表1总结了结果。

表1：比较研究结果（五个随机种子的均值 ± s.d.）。

方法	累计命中次数	命中率（%）	首次命中（迭代次数）	最佳奖励	成功率（%）
随机搜索	9.2 ± 1.9	4.6	25 ± 16	0.903 ± 0.012	82.1
BO	28.2 ± 3.8	14.1	22 ± 31	0.921 ± 0.000	100.0
智能体	105.4 ± 40.9	52.7	12 ± 13	0.895 ± 0.016	95.7

Ara 智能体发现命中的速率是随机搜索的 11.5 倍（$p = 0.006$，单侧曼-惠特尼 U 检验），并且在所有方法中首次命中的中位迭代次数最少（第 12 次迭代；随机搜索为第 25 次，BO 为第 22 次；表1、图3）。累积命中曲线（图3）表明，智能体的优势并非短暂的早期效应：智能体曲线从约第 15 次迭代起快速上升，并在第 200 次迭代前持续领先于两种基线方法。智能体在累积命中数方面也显著优于 BO（$U = 25$、 $p = 0.006$）。然而，BO 获得了最高的单个候选项奖励（0.921），这体现了其通过高斯过程外推进行标量优化的优势；相比之下，智能体的优势在于持续识别同时满足全部三个命中标准的候选项。

尽管由于模型输出解析失败，23% 的智能体迭代退回随机选择，但该智能体的表现仍比纯随机搜索提升了 11.5 倍。这表明，成功获得引导的迭代中， ≈77% 的效果远胜于偶然；即使只有部分迭代得到大语言模型引导，也能带来显著收益。智能体所选候选项的 xTB 评估成功率为 95.7%，高于随机选择的 82.1%，这表明智能体的化学感知选择也避开了容易导致组装或收敛失败的组合。

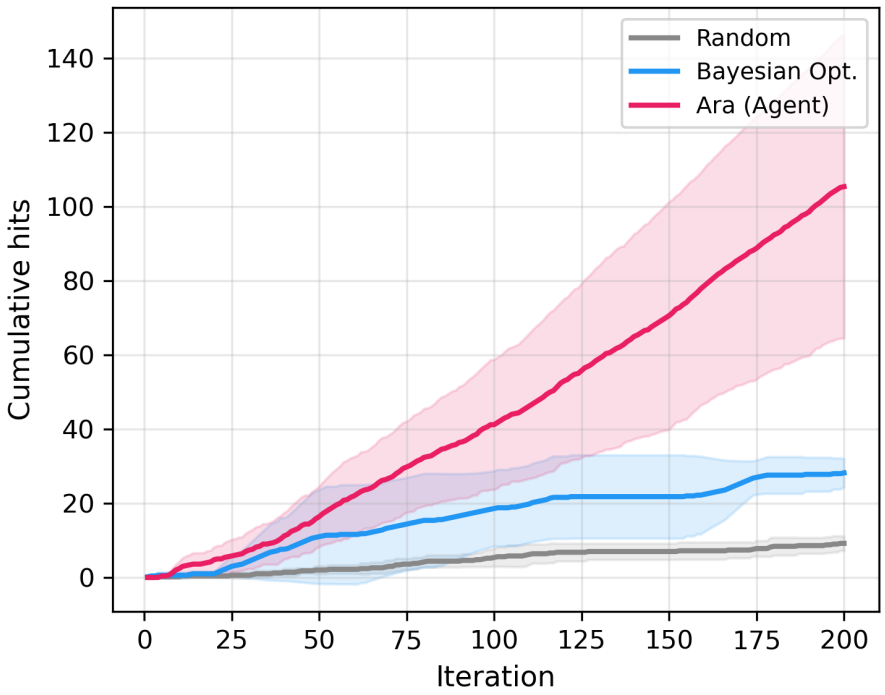


图 3：随机搜索（灰色）、贝叶斯优化（蓝色）和 Ara（智能体，粉色）的累计命中数随迭代次数的变化。阴影区域表示五个随机种子下的均值 ± 标准差（每次运行 200 次迭代）。若候选项满足带隙为 1.8–2.2 eV、CBM < 0 V 和 $S_{CSI} \geq 0.7$，则定义为一次命中。智能体的累计命中数优于随机搜索和贝叶斯优化，并且更早发现了首次命中。

表 2：智能体的部分推理轨迹（20次迭代专用运行，随机种子42）。“命中”列标示符合全部三项标准的候选项。

迭代	选择	带隙（电子伏特）	S_{CSI}	Hit	推理（简要）
0	Tp +BD + β-酮式。 +H	2.61	0.67	—	β-通过不可逆 可逆互变异构化。
1	Tp + BTDA + β-酮式。 + H	2.50	0.64	—	苯并噻二唑受体可缩小 带隙，使其进入目标范围。
2	TFPT-ald +DVB + 乙烯基。 + H	1.53	0.73	—	三嗪受体 + 亚乙烯基，用于非 可水解稳定性；带隙过低。
3	TFB + DVB + 乙烯基。 + H	2.13	0.73	**Hit**	电子中性 TFB 取代强 受体 TFPT-ald，以增宽带隙。
4	TFP-OH1 + DVB + 乙烯基。 + H	2.11	0.73	**Hit**	温和的 e^--供体羟基会使带隙变窄 从 2.13 略微向 2.0 电子伏特靠近。
5	TFP-OH1 +DVB + 乙烯基。 + OMe	2.02	0.75	**Hit**	甲氧基供体会使带隙向 理想的 2.0 电子伏特靠近。
8	TFB + DVB + 乙烯基。 + 甲氧基	2.03	0.73	**Hit**	相同策略应用于 TFB 骨架； 带隙中心位于 2.0 电子伏特。
11	TFB + DVB +乙烯基. + 甲基	2.12	0.73	**Hit**	弱供体；带隙仍处于目标窗口- dow.

智能体在累计命中数方面具有决定性优势，加上更早发现首个命中项，这些结果表明，对于这项多准则设计任务，大语言模型引导的搜索显著优于随机搜索和基于代理模型的策略。

化学推理驱动智能体性能

为理解智能体为何优于贝叶斯优化，我们在提高输出令牌限制的专门20次迭代运行中收集了详细的推理轨迹（见方法）。这些轨迹揭示了一种连贯的化学策略，并在前十余次迭代中逐步展开（表 2）。

推理轨迹表明，该智能体的选择是由明确的化学逻辑、通过供体–受体工程进行的带隙调控、连接键稳定性层级以及基于定量反馈的迭代优化所引导的，而非模式匹配或训练数据记忆。智能体首先测试了 β-酮烯胺连接键（迭代 0–1），明确指出其通过不可逆互变异构化实现的水解稳定性。尽管这些候选项具有足够的稳定性指数，但其带隙超过了 2.2 eV 的上限。在迭代 2 中，智能体转而采用带有强吸电子 TFPT-ald 节点的亚乙烯基连接键，并推断该供体–受体组合能够在提供非水解 C=C 键的同时缩小带隙。当所得带隙（1.53 eV）低于 1.8 eV 的阈值时，智能体在迭代 3 中通过将 TFPT-ald 替换为电子中性的 TFB 节点来修正策略，并解释称此举可降低过强的受体特性，使带隙拓宽至 2.13 eV，从而得到首个命中项。

从第 4 次迭代开始，智能体在亚乙烯基连接的骨架家族内采用了系统优化策略。它首先测试了 TFP-OH1，这是一种带有温和给电子羟基的节点，并观察到带隙略微收窄（2.11 电子伏特）。随后，它探索了 R 基团取代：OMe（带隙 2.02 电子伏特）、NH_2 （2.10 电子伏特）、OH （2.14 电子伏特）和 Me（2.12 电子伏特），利用供体和受体电子效应，将带隙微调至 2.0 电子伏特的最优值附近。每次选择都伴随着一段推理说明，其中引用了前一项结果，并预测带隙变化的方向，而且这些预测往往是正确的。这一迭代优化循环——提出、评估、推理、调整——与随机采样或基于指纹的采集函数相比，体现了截然不同的搜索模式。

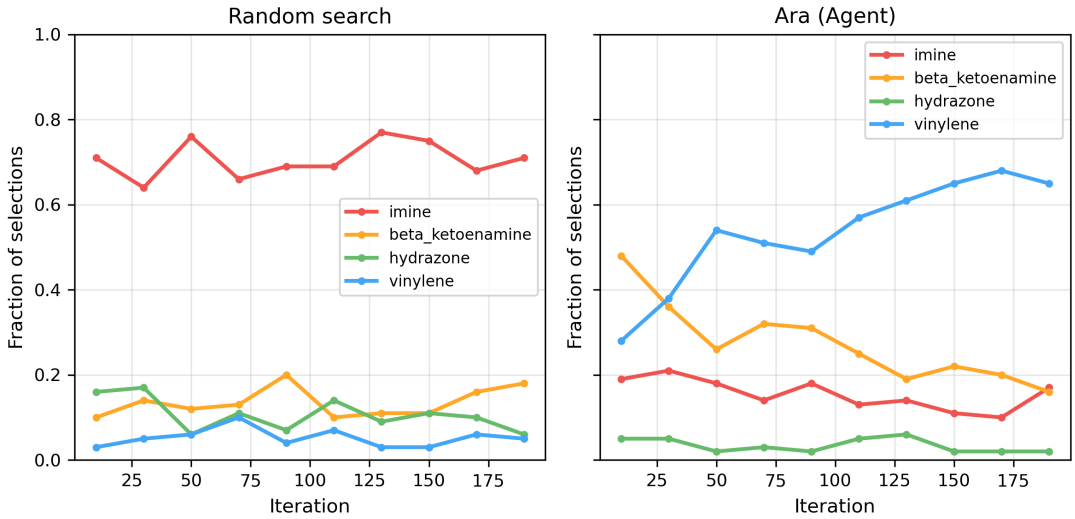


图 4：搜索迭代过程中的连接类型分布。滚动窗口为 20 次迭代，显示各连接类型（亚胺、 β-酮烯胺、腙、亚乙烯基）候选选择所占比例，并对五个随机种子取平均值。左：随机搜索在整个运行过程中，各连接类型的分布大致保持均匀。右：Ara（智能体）在搜索早期呈现混合分布，到第 50 次迭代时则明显偏好亚乙烯基，其次是 β-酮烯胺，这反映出其依据化学感知优先级排序，优先选择耐水解连接键（高 S_{CSI}），同时探索节点和 R 基团的选择，以满足带隙和带边标准。

对完整 200 次迭代运行中连接类型分布的分析，从定量角度证实了这一模式（图 4）：智能体从早期的混合分布转变为在第 50 次迭代时几乎压倒性地偏好亚乙烯基连接键（其次是 β-ketoenamine），而随机搜索始终保持均匀分布。智能体的三级策略——优先选择耐水解连接键、选择电子性质合适的节点，然后通过 R 基团取代调节带隙——体现了人类化学家会采用的层级化推理；如果没有针对连接化学设置明确的类别特征，基于指纹的贝叶斯优化通常无法自然实现这种策略。

搜索空间表征与利用-探索权衡

对全部 820 个候选项进行穷举评估后，得到 670 次成功计算，并确定了 38 个真实命中项，全局命中率为 5.7%（见方法）。命中项集中在两个连接键家族中：30 个为亚乙烯基连接（79%），8 β-酮烯胺连接的共价有机框架占 21%；没有亚胺或腙连接的候选项能同时满足全部三项标准（补充表 S3）。命中项集中于这两个家族，体现了根本性的稳定性-活性权衡：只有亚乙烯基和 β-酮烯胺连接键能实现 $S_{CSI} \geq 0.7$，同时将带隙维持在 1.8–2.2 eV 范围内。在亚乙烯基家族中，命中率因节点而有很大差异：TFB 和 TFP-OH1 在几乎所有 R 基团上均有命中；Tp 部分成功（10 个 R 基团中有 9 个命中，仅 SO_3H 未命中）；TFPT-ald 则仅产生一个命中项（CF_3），因为其强吸电子三嗪核心使其他所有取代基对应的带隙都低于 1.8 eV。

基于这一真实值，我们针对每种方法计算了预言机召回率，即至少发现一次的 38 个唯一命中项所占的比例（补充图 S2）。贝叶斯优化的表现显著优于

智能体，其预言机覆盖率更高，这得益于高斯过程中的不确定性项，它会主动推动探索指纹空间中尚未采样的区域。智能体较低的预言机召回率反映了其对少数高产亚乙烯基连接骨架的强烈利用：一旦发现 TFB $+$DVB $+$ 亚乙烯基是可靠的命中来源，它便专注于探索 R 基团变体，而不是转向相距较远的化学家族。随机搜索提供了介于两者之间的基准多样性。这种利用-探索权衡具有实际意义：在预算受限的实验研究中，每次评估都对应一次密度泛函理论计算或实验室合成，此时智能体 53% 的单次迭代成功率至关重要；而若研究计划旨在全面表征整个设计空间，则贝叶斯优化的系统性探索更为可取。二者优势互补，而非相互竞争。

智能体优势对评分函数权重具有稳健性

综合稳定性指数 S_{CSI} 结合了连接键稳定性（S_{linkage}）、位阻屏蔽（$S_{\mathrm{shielding}}$）和疏水性（$S_{\mathrm{hydrophobicity}}$），权重分别为 0.50、0.30 和 0.20。为评估智能体的优势是否取决于这一特定权重设置，我们针对 30 组权重三元组进行了敏感性分析，其中连接方式权重在 0.3 至 0.7 之间变化，其余权重则以 0.1 为增量分配给位阻屏蔽和疏水性（补充图 S1）。在搜索空间至少包含 10 个命中的 21 种权重组合中（这是比较具有统计意义的范围），智能体与随机搜索的命中率比介于 $2.5\times$ 至 $14.3\times$ 之间，始终高于 $2\times$ 阈值。其余 9 组权重三元组的位阻屏蔽权重（≥ 0.4）极端，导致搜索空间中的命中数少于 3 个，因此命中率比比较缺乏参考意义。

智能体优势在三组权重下均表现出稳健性，这是因为智能体所学习的主要策略——偏好亚乙烯基和 β-酮烯胺连接方式——无论为该组成部分分配何种具体权重，都会产生较高的 S_{linkage} 得分。换言之，智能体对连接化学的类别推理与S_{CSI}内部的具体数值权重相互正交，因此比较结果不会受到这一方法学选择的影响。

讨论

本研究表明，预训练大语言模型编码了足够的化学知识，能够指导对多准则 COF 设计空间的高效探索。Ara 智能体的命中率是随机搜索的 11.5 倍（$p < 0.01$），并且在所有测试方法中最先实现命中，收敛速度最快（表1）。对 30 S_{CSI} 权重组合进行的敏感性分析证实，这一优势在合理的权重空间内始终存在（补充图 S1），排除了其依赖特定评分函数参数化方式的可能性。这些结果表明，大语言模型化学先验，尤其是给体-受体理论、连接键稳定性知识和结构-性质推理，为材料发现智能体提供了可行的基础。

该智能体的表现可归因于一些具体的化学推理策略，这些策略体现在其输出中，也可通过化学直觉加以验证。首先，该智能体迅速收敛到不可水解的亚乙烯基（C=C）连接键，并在无人提示的情况下意识到，不含可水解键便能获得最高的稳定性评分。其次，该智能体学会避开强吸电子的 TFPT-ald 节点，并正确推断出由此产生的给体–受体不匹配会使带隙低于 1.8 电子伏特的阈值；这与穷举评估中观察到的节点依赖性相呼应：TFPT-ald $+$DVB $+$亚乙烯基组合在十种 R 基团中仅有一种命中。第三，该智能体系统地探索了 R 基团取代，

包括 OMe、NH_2, OH、Me 和 tBu，以微调带隙，使其接近 2.0 电子伏特的最佳值。这种三级策略（连接方式 $\rightarrow$ 节点 $\rightarrow$ R 基团）仅凭评估反馈便得以发现，体现了人类化学家可能采用的层级推理。对于仅将分子编码为数值指纹、而未明确表示连接化学类别特征的方法而言，这种策略并非自然可得。不过，该智能体也并非万无一失：尽管亚胺连接候选物的 S_{linkage} 评分较低，它在早期迭代中仍偶尔提出这类候选物（随机种子42的前 50 次迭代中尝试了 8 次亚胺连接，均未命中）；此外，尽管此前已经失败，它仍在第 13 次迭代再次尝试 TFPT-ald 节点（带隙为 1.75 电子伏特，仍低于阈值）。这些情况，加上 23% 的解析失败率，表明当前的大语言模型仍不是完美的化学推理器，不过总体策略是合理的。该智能体的表现可归因于一些具体的化学推理策略，这些策略体现在其输出中，也可通过化学直觉加以验证。首先，该智能体迅速收敛到不可水解的亚乙烯基（C=C）连接键，并在无人提示的情况下意识到，不含可水解键便能获得最高的稳定性评分。其次，该智能体学会避开强吸电子的 TFPT-ald 节点，并正确推断出由此产生的给体–受体不匹配会使带隙低于 1.8 电子伏特的阈值；这与穷举评估中观察到的节点依赖性相呼应：TFPT-ald + DVB + 亚乙烯基组合在十种 R 基团中仅有一种命中。第三，该智能体系统地探索了 R 基团取代，包括 OMe、NH_2, OH、Me 和 tBu，以微调带隙，使其接近 2.0 电子伏特的最佳值。这种三级策略（连接方式 $\rightarrow$节点 $\rightarrow$R 基团）仅凭评估反馈便得以发现，体现了人类化学家可能采用的层级推理。对于仅将分子编码为数值指纹、而未明确表示连接化学类别特征的方法而言，这种策略并非自然可得。不过，该智能体也并非万无一失：尽管亚胺连接候选物的 S_{linkage} 评分较低，它在早期迭代中仍偶尔提出这类候选物（随机种子42的前 50 次迭代中尝试了 8 次亚胺连接，均未命中）；此外，尽管此前已经失败，它仍在第 13 次迭代再次尝试 TFPT-ald 节点（带隙为 1.75 电子伏特，仍低于阈值）。这些情况，加上 23% 的解析失败率，表明当前的大语言模型仍不是完美的化学推理器，不过总体策略是合理的。

对全部 670 个成功完成计算的候选项进行穷举评估后，确定了 38 个真实命中项（补充表 S3），从而能够评估各方法的探索广度。智能体较高的单次迭代命中率，反映出它善于利用少数几个高产化学家族，主要是采用 TFB、TFP-OH1 和 Tp 节点的乙烯基连接共价有机框架，而非广泛覆盖整个命中空间。贝叶斯优化依托其高斯过程不确定性项和内置去重机制，实现了显著更高的独特命中项覆盖率（补充图 S2）。这些互补特性表明，预算受限的实验活动可受益于智能体快速收敛至高质量候选项的能力；而希望全面表征设计空间的项目，则可能更适合采用贝叶斯优化进行系统性探索。混合策略是一种颇具前景的方向：先利用智能体快速实现初步收敛，再通过贝叶斯优化进一步拓展多样性，从而结合两种方法的优势。

约 23% 的智能体迭代所产生的输出无法解析为有效的候选选择，因此被随机抽样替代。这种失效模式并未使比较结果失效，反而引入了保守偏差：智能体的实际自主性约为 77%，近四分之一的预算消耗在随机采样上。尽管如此，它仍实现了纯随机搜索 11.5 倍的提升，这意味着引导式迭代的效率远高于随机碰运气。通过微调、受限解码或结构化输出 API，改进大语言模型指令遵循和输出格式化，将进一步提升性能。

几种锂　制仍需予以承认。首先，本研究仅使用了一个大语言模型（Gemini-3-flas　　h-

preview）；其在不同模型系列（如 GPT-5、Claude）及不同规模模型上的泛化能力仍有待验证。其次，所有评估均基于 GFN1-xTB 半经验方法；尽管传递函数提供了映射至密度泛函理论带隙的校准映射（斯皮尔曼 $\rho = 0.71$），但仍需对智能体发现的顶级候选项进行周期性密度泛函理论验证。第三，搜索空间是预先枚举的（820 个候选项）；在拥有 10^4–10^6个候选项的开放式生成设计场景中，智能体对连接键稳定性的类别推理预计将带来更大优势，但这一点仍有待证明。第四，智能体采用有放回抽样，因此倾向于重新评估已知的优选项，而非进行探索；在保留化学推理的同时实施去重策略，仍是一项开放挑战。最后，综合稳定性指数 S_{CSI} 是一个计算代理指标，可反映连接化学、位阻屏蔽和疏水性，但未考虑连接键附近取代基对水解敏感性的电子影响，也未考虑框架结晶度，而这两者都会影响材料在运行条件下的长期耐久性。需要在光催化条件下对水解稳定性进行实验验证，以确认所发现候选项的实际效用。

这些发现为未来研究开辟了若干方向。对排名靠前的亚乙烯基连接候选项（TFP-OH1 + DVB + 亚乙烯基 + 叔丁基，奖励为 0.921；TFB + DVB +亚乙烯基 + 甲氧基，奖励为 0.900）进行密度泛函理论验证，可确定 xTB 层级的预测能否转化为准确的周期性带结构和能带位置。多模型比较（Claude、GPT）可检验观察到的优势究竟源于普遍的大语言模型化学先验，还是模型特有的能力。大语言模型与贝叶斯优化相结合的混合策略，或可兼顾智能体的快速收敛与贝叶斯优化的系统性探索，从而同时实现较高的样本效率和广泛的覆盖范围。纳入规划器、评判器和记忆组件的多智能体架构，可能支持更复杂的长程推理。将搜索空间扩展至包括拓扑变化（蜂窝状、笛卡尔格、方形）、孔径工程和混合连接体体系，可检验智能体在更高维设计空间中的推理能力。最后，将研究拓展至其他材料类别，包括金属有机框架、沸石和杂化钙钛矿，将有助于验证智能体引导的逆向设计的普适性。

方法

共价有机框架搜索空间定义

候选空间由 820 种共价有机框架构筑单元组合构成，这些组合来自经过筛选的库，其中包含七种三角形（C_3 对称）节点、十九种二连接体、四种连接化学（亚胺、 β-酮烯胺、腙、亚乙烯基）以及十种芳香族R基取代基（H、OH、NH_2, OMe、F、CF_3, NO_2, Me、叔丁基、SO_3H）。兼容性约束缩减了组合总数。亚胺连接键（C=N）通过醛与伯胺官能团之间的席夫碱缩合形成；在我们的库中，这类连接键由醛节点与二胺连接体配对，或由胺节点与二醛连接体配对形成。 β-酮烯胺连接键需要带有可烯醇化 β-二酮基团的 1,3,5-三甲酰基间苯三酚型醛与伯胺配对；在此处，只有 Tp节点满足这一要求，并与二胺连接体配对。腙连接键由醛与酰肼官能团形成；在我们的库中，醛节点与二酰肼连接体配对。亚乙烯基连接键（C=C）通过醛与活性亚甲基之间的克诺维纳盖尔缩合形成；在我们的库中，醛节点与 DVB（1,4-双(氰甲基)苯）配对。

醛节点包括 TFB（1,3,5-三甲酰基苯）、TFPT-ald（三嗪-2,4,6-三甲醛）、TFP-OH1（2-羟基-1,3,5-三甲酰基苯）和 Tp（1,3,5-三甲酰基间苯三酚）；胺

节点包括 TAPB（1,3,5-三(4-氨基苯基)苯）、TAPT（2,4,6-三(4-氨基苯基)-1,3,5-三嗪）和 TAPA（三(4-氨基苯基)胺）。按连接类型划分如下：亚胺（醛 + 二胺）440 个、亚胺（胺 + 二醛）150 个、 β-酮烯胺 110 个、腙 80 个，以及亚乙烯基 40 个候选项。三种始终无法成功进行片段组装的节点–连接体组合（TAPB + Thieno-DA、TAPT + OMe-PDA、TAPA + Thieno-DA）仍保留在枚举中，但在组装失败时不纳入分析。

基于片段的筛选流程

每个候选项均通过基于片段的计算流程进行评估。首先，使用针对各连接化学的 RDKit [32]反应 SMARTS，由 SMILES 构建单元组装节点–连接体–节点重复单元片段：亚胺缩合（醛 + 胺 $\rightarrow$ C=N）、 β-酮烯胺互变异构化（经由 Tp + 胺）、腙形成（醛+ 酰肼）以及克脑文格尔缩合（醛 + 乙烯基 $\rightarrow$ C=C）。当指定的 R 基团不是氢时，基于 SMARTS 的取代会替换连接体上的一个芳氢。组装得到的片段 SMILES 通过 RDKit 的 ETKDGv3 嵌入并随后进行 MMFF94 力场最小化，转换为三维结构，得到通常含 50–80 个原子的初始几何构型。

使用 GFN1-xTB（6.7.1 版） [15]在“tight”收敛级别进行几何优化。基本带隙通过垂直电离势与电子亲和能方案（xtb --vipea）计算；该方案在优化后的中性几何构型上执行 delta-SCF 计算以获得 IP 与 EA 值；基本带隙取为 IP− EA。应用线性传递函数（式 1）将 xTB 基本带隙映射到 DFT 标度，并以文献中涵盖六种连接类型的 13 种共价有机框架的 DFT 带隙进行校准（包括搜索空间中不存在的硼酸酯、硼氧六环和三嗪连接键；补充表 S2）。校正后的带隙被截断至 0.5–6.0 eV 范围，以抑制非物理异常值。相对于标准氢电极的导带底位置估计为

$$\mathrm{CBM}_{\text{vs NHE}} = -\mathrm{EA}_{\mathrm{xTB}} + 2.62\ \mathrm{V}, \tag{2}$$

其中，偏移量是根据三个实验测定带边位置的共价有机框架进行校准的：TpPa-1、TAPB-PDA 和 sp2c-COF [33, 34, 35]。在单个 CPU 核心上，每个候选项进行几何优化和 vipea 的典型计算耗时为 5–10 秒。

综合稳定性指数

为将水解稳定性表示为连续且可微的评分，我们定义了一个综合稳定性指数 S_{CSI}，由三个部分组成:

$$S_{\mathrm{CSI}} = 0.50\, S_{\mathrm{linkage}} + 0.30\, S_{\mathrm{shielding}} + 0.20\, S_{\mathrm{hydrophobicity}}. \tag{3}$$

连接键稳定性评分 S_{linkage} 根据公认的共价有机框架连接键水解稳定性顺序 [10, 36, 37, 38, 39]进行赋值：亚胺 0.40、腙 0.60、 β-酮烯胺 0.80，亚乙烯基 0.95。该顺序反映了其化学键化学的内在差异：亚胺 C=N 键会发生可逆水解；腙连接方式受到相邻酰胺的适度保护；β-酮烯胺键通过不可逆烯醇–酮互变异构和分子内 C=O$\cdots$H–N 氢键得以稳定 [36]；而亚乙烯基 C=C 键不会发生水解 [39]。

位阻屏蔽评分 $S_{\mathrm{shielding}}$ 量化了连接键局部抵御水的亲核攻击的能力。对于每个候选项，通过识别构成连接键的原子

在片段 SMILES 中进行 SMARTS 模式匹配：亚胺对应 [CH]=[N]， β-酮烯胺对应 [NH]-[CX3]=[OX1]（羰基 C 和 O），腙对应 [CH]=[N]-[NH] （C 和第一个 N），乙烯对应 [CH]=[CH] 。在经 xTB 优化的几何结构中，这些原子的中点作为参考点，并将 $S_{\text{shielding}}$ 计算为半径为 4 Å 范围内重原子的数量，再按 clip((count -5)/15，0，1) 进行归一化。当 SMARTS 模式匹配失败时（例如，由于价态异常），则使用分子质心作为备用方案。

疏水性评分 $S_{\text{hydrophobicity}}$ 通过使用 RDKit 计算片段 SMILES 的 Crippen 分配系数（SLogP）获得，并归一化至范围 $[0, 1]$。较高的疏水性不利于水接触连接键，从而有助于提高水解稳定性。

对 30 组权重三元组进行的敏感性分析（S_{linkage} 权重范围为 0.3 至 0.7；其余权重以 0.1 为增量分配于 $S_{\text{shielding}}$ 和 $S_{\text{hydrophobicity}}$ 之间）证实，当搜索空间至少包含 10 个命中项时，在所有权重组合下，智能体相较于随机搜索的优势均超过 $2\times$ （补充图 S1）。

命中标准与奖励函数

如果候选项同时满足以下三个标准，则将其归类为命中项：(1) 校正带隙介于 1.8 至 2.2 eV 之间，目标是覆盖光催化水分解所需的可见光吸收窗口；(2) 相对于标准氢电极，导带底低于 0 V，这是质子还原的热力学要求；以及 (3) $S_{\text{CSI}} \geq 0.7$，以确保足够的水解稳定性。

定义了一个连续奖励函数，用于指导搜索方法：

$$r = 0.35\, g_{\text{score}} + 0.30\, \text{cbm}_{\text{score}} + 0.35\, S_{\text{CSI}}, \tag{4}$$

其中

$$g_{\text{score}} = \exp\left(-\tfrac{1}{2}\left(\frac{g - 2.0}{0.3}\right)^2\right), \tag{5}$$

$$\text{cbm}_{\text{score}} = \frac{1}{1 + \exp(5.0 \times \text{CBM}_{\text{vs NHE}})}. \tag{6}$$

奖励值范围为 0 至约 1，通常大于 0.85 的值对应于命中项。

搜索方法

比较了三种搜索方法，每种方法均使用五个随机种子（42、123、456、789、2024）进行 200 次迭代评估，并使用共享的评估缓存，以确保相同候选项获得相同的计算结果。

随机搜索。 从包含 820 个成员的候选池中均匀随机地（有放回地）抽取候选项。这作为零假设基线。

贝叶斯优化（BO）。使用 RDKit 对组装后的片段 SMILES 计算 2048 位 Morgan 指纹（半径为 2），并据此拟合高斯过程（GP）代理模型。特征使用零均值、单位方差缩放器进行标准化。该 GP 使用 Matérn 核（$\nu = 2.5$，长度尺度为 1.0，边界为 $[10^{-3}, 100]$），目标值经过归一化，并通过 L-BFGS-B 优化器进行五次随机重启拟合 [40]。前 20 轮采用随机初始化；之后

每轮均从尚未评估的候选项中选取期望改进（EI）最高者。系统会跟踪已评估的候选项，并在后续的采集函数最大化过程中将其排除，从而确保每轮评估的都是新候选项。这种内置去重是贝叶斯优化的一项标准功能，有助于提高其探索效率。相比之下，大语言模型智能体采用有放回抽样，保留了其倾向于利用高奖励化学家族的自然特性；讨论部分将探讨这种利用行为与探索广度之间的权衡。

大语言模型智能体（Ara）。该智能体使用 Google 的 Gemini-3-flash-preview 模型 [30]，采用 JSON 输出模式（温度 0.3，最大输出令牌数 1024）。每次迭代时，智能体都会收到一条提示，其中包含：(a) 描述共价有机框架设计任务、可用构建单元、兼容性约束及化学原理（给体–受体理论、共轭效应、稳定性–活性权衡）的系统指令；(b) 最近 10 次评估结果，以及完整历史记录中奖励最高的 5 个结果；(c) 要求从（节点、连接体、连接方式、R 基团）组合中选出一个，并给出 1–3 句化学推理。系统会解析智能体的响应，提取四个构建单元标识符；如果解析失败或响应为空，则以随机候选项替代（这也计入报告的解析失败率）。系统未强制执行去重：智能体可以重新评估先前选过的候选项，而是依赖提示中的软性引导（“如果可能，请选择一个以前未尝试过的新组合”）以及已评估候选项列表来鼓励多样性。讨论部分通过预言机恢复分析考察了这种有放回抽样设计对探索多样性的影响。

统计检验。使用单侧 Mann–Whitney U 检验对不同随机种子的累积命中数进行两两比较，并以 $p < 0.05$作为显著性判定标准。

真实基准评估

为建立真实基准，我们将搜索空间中的全部 820 个候选项提交至基于片段的筛选流程。其中，670 个（81.7%）成功完成评估；其余 150 个在片段组装或 xTB 收敛阶段失败。在完成评估的 670 个候选项中，38 个满足全部三项命中标准（补充表 S3），全局命中率为 5.7%。预言机召回率定义为：在某方法 200 次迭代的预算内，至少发现过一次的这 38 个真实命中项所占的比例（仅计唯一命中项，不重复计算重复评估的结果）。

推理轨迹收集。为获取可解释的推理示例（表 2），我们使用相同的模型（Gemini-3-flash-preview）和系统提示词，另行进行了一次 20 次迭代运行。为避免推理文本被截断，最大输出令牌数上限从 1024 提高至 4096；同时将温度提高至 0.5，以鼓励生成更多样化的推理。其他生成参数（JSON 输出模式、安全设置）均保持不变。此次运行仅用于定性分析，不纳入定量比较。

数据可用性

支持本研究发现的数据可在以下位置获取：

- **构建模块库**（节点、连接体、连接键、R 基团）：已作为“data/building blocks library.json”纳入此代码仓库。
- **校准数据集**（13 种共价有机框架的 xTB 与密度泛函理论带隙）：‘data/calibration dataset.csv‘。
- **导带底校准**（导带底相对于标准氢电极）：‘data/cbm calibration.json‘。
- **比较研究输出**（运行结果、预言机摘要、完整评估）

本研究期间生成或分析的所有其他数据，均已包含在此代码仓库中，或可使用所提供的代码和上述数据文件重现。

代码可用性

支持本研究结论的代码可在 https://github.com/Iman- Peivaste/Ara 获取。代码仓库包含基于片段的筛选流程（主入口为 ‘main.py‘）、搜索方法（随机搜索、贝叶斯优化和大语言模型智能体）、校准与分析脚本，以及重现结果所需的构建模块库和校准数据集。该流程依赖 xtb 程序中的 GFN1-xTB 实现（参见方法）、用于分子处理的 RDKit，以及用于构建共价有机框架的超分子工具包（stk）。代码仓库 README 中提供了环境设置以及重现比较研究和图表的说明。

作者贡献

Iman Peivaste：概念构思、开发初始概念与工作流程、框架设计、撰写初稿、可视化、方法学、调查研究及正式分析。

Nicolas Boscher：撰写、审阅、编辑、资源、方法学及正式分析。

Ahmed Makradi：撰写、审阅、编辑、方法学及正式分析。

Salim Belouettar：监督、经费获取、概念构思（研究方向与框架设定）、撰写、审阅与编辑、方法学、资源。

竞争利益

作者声明，他们不存在任何可能影响本文所报道工作的已知竞争性经济利益或个人关系。

致谢

本研究由卢森堡国家研究基金会（FNR）通过资助项目 PRIDE21/16758661/HYMAT 提供经费。

.

参考文献

[1] Chen, S., Takata, T. & Domen, K. 用于整体水分解的颗粒光催化剂。Nature Reviews Materials 2, 17050 (2017).[2] Wang, Q. & Domen, K. 用于光驱动水分解的颗粒光催化剂：机理、挑战与设计策略。Chemical Reviews 120, 919–985 (2019).[3] Tao, X., Zhao, Y., Wang, S., Li, C. & Li, R. 使用颗粒光催化剂进行太阳能驱动水分解的最新进展与展望。Chemical Society Reviews 51, 3561–3608 (2022).[4] Qi, J., Zhang, W. & Cao, R. 基于水分解的太阳能制氢能量转换。Advanced Energy Materials 8, 1701620 (2018).[5] Ding, S.-Y. & Wang, W. 共价有机框架（COF）：从设计到应用。Chemical Society Reviews 42, 548–568 (2013).[6] Cˆot´e, A. P. et al. 多孔、晶态共价有机框架。science 310, 1166–1170 (2005).[7] Geng, K. et al. 共价有机框架：设计、合成与功能。Chemical reviews 120, 8814–8933 (2020).[8] Stegbauer, L., Schwinghammer, K. & Lotsch, B. V. 一种用于光催化制氢的基于腙的共价有机框架。Chemical Science 5, 2789–2793 (2014).[9] Vyas, V. S. et al. 用于可见光诱导氢气生成的可调谐吖嗪共价有机框架平台。Nature communications 6, 8508 (2015).[10] Lyu, H., Diercks, C. S., Zhu, C. & Yaghi, O. M. 多孔晶态烯烃连接共价有机框架。Journal of the American Chemical Society 141, 6848–6852 (2019).[11] Xu, H., Gao, J. & Jiang, D. 稳定、晶态、多孔的共价有机框架：手性有机催化剂的平台。Nature chemistry 7, 905–912 (2015).[12] Ongari, D., Yakutovich, A. V., Talirz, L. & Smit, B. 构建一致且可复现的共价有机框架吸附评估数据库。ACS central science 5, 1663–1675 (2019).[13] Haase, F. & Lotsch, B. V. 破解 COF 三难困境：迈向兼具结晶性、稳定性与功能性的共价有机框架。Chemical Society Reviews 49, 8469–8500 (2020).[14] Jain, A., Shin, Y. & Persson, K. A. 使用密度泛函理论预测能源材料。Nature Reviews Materials 1, 1–13 (2016).[15] Grimme, S., Bannwarth, C. & Shushkov, P. 一种稳健且精确的紧束缚量子化学方法，用于计算大型分子体系的结构、振动频率和非共价相互作用，并针对所有 spd 区元素（z= 1–86）进行了参数化。Journal of chemical theory and computation 13, 1989–2009 (2017).[16] Hachmann, J. et al. 哈佛清洁能源项目：在 World Community Grid 上大规模计算筛选与设计有机光伏材料。The Journal of Physical Chemistry Letters 2, 2241–2251 (2011).

[17] Shields, B. J. 等. 贝叶斯反应优化作为化学合成工具. Nature 590, 89–96 (2021).[18] Griffiths, R.-R. & Hern´andez-Lobato, J. M. 利用变分自编码器进行自动化学设计的约束贝叶斯优化. Chemical Science 11, 577–586 (2020).[19] Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. 分子与材料科学中的机器学习. Nature 559, 547–555 (2018).[20] Moosavi, S. M., Jablonka, K. M. & Smit, B. 机器学习在材料理解与设计中的作用. Journal of the American Chemical Society 142, 20273–20287 (2020).[21] Peivaste, I. 等. 材料科学与工程中的人工智能：现状、关键挑战与未来发展方向. Composite Structures 372, 119419 (2025).[22] Peivaste, I., Makradi, A., Mercuri, F. & Belouettar, S. Hype-mad：用于分子自动化与数据生成的高产率 Python 环境. Acta Mechanica 1–9 (2025).[23] Jablonka, K. M., Schwaller, P., Ortega-Guerrero, A. & Smit, B. 利用大语言模型开展预测化学. Nature Machine Intelligence 6, 161–169 (2024).[24] Zheng, Z., Zhang, O., Borgs, C., Chayes, J. T. & Yaghi, O. M. 用于文本挖掘和 MOF 合成预测的 ChatGPT 化学助手 . Journal of the American Chemical Society 145, 18048–18062 (2023).[25] Boiko, D. A., MacKnight, R., Kline, B. & Gomes, G. 借助大语言模型开展自主化学研究. Nature 624, 570–578 (2023).[26] Schwaller, P. 等. Molecular Transformer：一种用于不确定性校准的化学反应预测模型. ACS Central Science 5, 1572–1583 (2019).[27] Ramos, M. C., Michtavy, S. S., Porosoff, M. D. & White, A. D. 利用上下文学习进行催化剂贝叶斯优化. arXiv 预印本 arXiv:2304.05341 (2023).[28] Peivaste, I., Makradi, A. & Belouettar, S. ChemNavigator：利用智能体人工智能发现有机光催化剂的设计规则. arXiv 预印本 arXiv:2601.17084 (2026).[29] M. Bran, A. 等. 为大语言模型增添化学工具. Nature Machine Intelligence 6, 525–535 (2024).[30] Team, G. 等. Gemini：一系列能力强大的多模态模型. arXiv 预印本 arXiv:2312.11805 (2023).[31] Onida, G., Reining, L. & Rubio, A. 电子激发：密度泛函方法与多体格林函数方法. Reviews of Modern Physics 74, 601 (2002).[32] RDKit：开源化学信息学工具包. https://www.rdkit.org. 访问日期：2026-02-17.[33] Bi, S. 等. 通过芳基甲基碳原子缩合制备二维半导体共价有机框架. Nature Communications 10, 2467 (2019).[34] Wang, X. 等. 用于水光催化析氢的含砜共价有机框架. Nature Chemistry 10, 1180–1189 (2018).[35] Wei, S. 等. 具有未取代烯烃连接键的半导体二维三嗪核共价有机框架. Journal of the American Chemical Society 141, 14272–14279 (2019).

[36] Kandambeth, S. 等. 通过可逆与不可逆途径相结合，构筑具有卓越化学（耐酸/碱）稳定性的晶态二维共价有机框架。*Journal of the American Chemical Society* **134**, 19524–19527 (2012). URL https://doi. org/10.1021/ja308278w.[37] Banerjee, R.等. 通过机械剥离共价有机框架，制备化学性质稳定的多层共价有机纳米片。 *Journal of the American Chemical Society***135**, 17853–17861 (2013). URL https://doi.org/10.1021/ja408121p.[38] DeBlase, C. R. & Dichtel, W. R. 超越硼：共价有机框架中新型连接化学的兴起。《大分子》**49**, 5297–5305 (2016). URL https://doi.org/10.1021/acs.macromol.6b00891.[39] Zhuang, X. 等. 具有全 sp^2 键合碳骨架的二维共轭聚合物框架。《聚合物化学》**7**, 4176–4181 (2016). URL https://doi.org/10. 1039/C6PY00561F.[40] Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. 一种用于有界约束优化的有限内存算法。*SIAM* 科学计算杂志 **16**, 1190–1208 (1995). URL https://doi.org/10.1137/0916069.