

## Knowledge-Guided Autonomous Discovery of Microenvironment-Tuned Metal–Organic Framework Photocatalysts

Yiming Zhao,<sup>⊥</sup> Tao Song,<sup>⊥</sup> Linjiang Chen,<sup>\*,⊥</sup> Yan Huang,<sup>⊥</sup> Kang Sun,<sup>⊥</sup> Wentao Han, Mingyang Shen, Chenwei Mao, Peng Lan, Meng Zhou, Weiwei Shang,<sup>\*</sup> Jun Jiang,<sup>\*</sup> and Hai-Long Jiang<sup>\*</sup>Cite This: *J. Am. Chem. Soc.* 2026, 148, 41390–41402

Read Online

ACCESS |



Metrics &amp; More

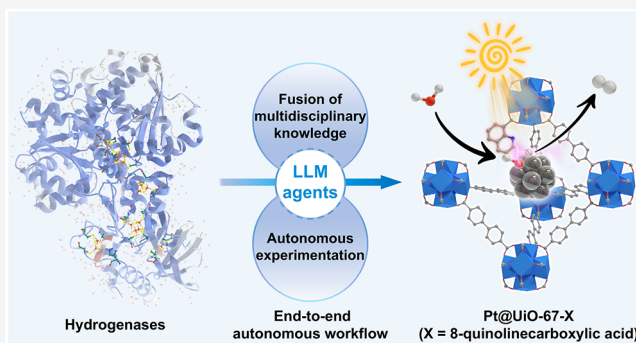


Article Recommendations



Supporting Information

**ABSTRACT:** Designing second-sphere microenvironments that promote proton-coupled electron transfer is central to catalysis yet difficult to achieve in porous solids, such as metal–organic frameworks (MOFs). Here, we report an end-to-end workflow that couples literature-guided large-language-model (LLM) reasoning with real-time experimental feedback to propose, test, and refine microenvironment designs in MOF photocatalysts. The system mined and fused three domains (namely, photocatalytic H<sub>2</sub> production, hydrogenases and enzyme-mimetic catalysis) and deduced the hypothesis that placing basic, hydrogen-bonding groups near catalytic centers would facilitate water activation and proton transfer. The hypothesis was instantiated by postsynthetic modification of UiO-67, generating 31 Pt@UiO-67-X variants and evaluating them across six closed-loop iterations on an automated platform. The search converged on Pt@UiO-67-30 (8-quinolinecarboxylic acid), which delivered 2.33 mmol g<sup>−1</sup> h<sup>−1</sup>, a ~36-fold improvement over the parent material; in a larger, optimally illuminated reactor the same catalyst reached 12.48 mmol g<sup>−1</sup> h<sup>−1</sup> while preserving the library's rank order. Photoluminescence quenching, enhanced photocurrent, and reduced impedance are consistent with faster charge separation, and first-principles calculations are consistent with reduced proton-transfer barriers via N...H hydrogen-bond networks. These results establish a practical microenvironment-engineering strategy in MOFs and show how LLM-guided knowledge fusion with experiment-in-the-loop reasoning can systematize and accelerate targeted discovery of functional materials.



## INTRODUCTION

Enzymes are nature's most efficient catalysts, enabling reactions under mild conditions through precisely organized active sites and second-sphere interactions that stabilize transition states and steer proton/electron flow.<sup>1–4</sup> Yet their structural complexity and susceptibility to inactivation limit broad application.<sup>5–7</sup> Synthetic catalysts offer stability and tunability but typically fall short of enzymatic efficiency. A central strategy has therefore been to mimic enzymatic microenvironments—embedding proton relays, hydrogen-bond networks, and tailored adsorption sites around catalytically active centers.<sup>8–12</sup> In practice, translating enzymatic principles into synthetic catalysts is hindered by fragmented, cross-disciplinary knowledge spanning biocatalysis, coordination chemistry, and materials science. A further obstacle is the vast, multidimensional design space of enzyme-mimetic catalysis, often encompassing active sites, functional moieties, defect chemistry, and interfaces, which makes intuition-driven and trial-and-error exploration intractable or prohibitively demanding.<sup>13–16</sup>

Recent advances in artificial intelligence (AI), especially large language models (LLMs), open a complementary pathway for synthesis-driven discovery.<sup>17–21</sup> Beyond point solutions such as

literature mining or property prediction,<sup>22,23</sup> LLMs can extract and process documented knowledge, generate research ideas and hypotheses, and plan experiments, while robotic platforms execute protocols reproducibly.<sup>24,25</sup> Despite such promising demonstrations, most reported systems remain modular and siloed: LLMs assist single steps, and automation executes predefined protocols, with few platforms closing the loop to couple iterative LLM reasoning with real-time experimental feedback.<sup>19,26–28</sup> Systematically integrating knowledge discovery, hypothesis formulation, candidate selection, autonomous synthesis/testing, and performance-guided refinement remains the key missing capability.

Here we address this gap. We introduce an end-to-end, LLM-driven autonomous framework that discovers and optimizes enzyme-mimetic photocatalysts for hydrogen production

Received: June 28, 2026

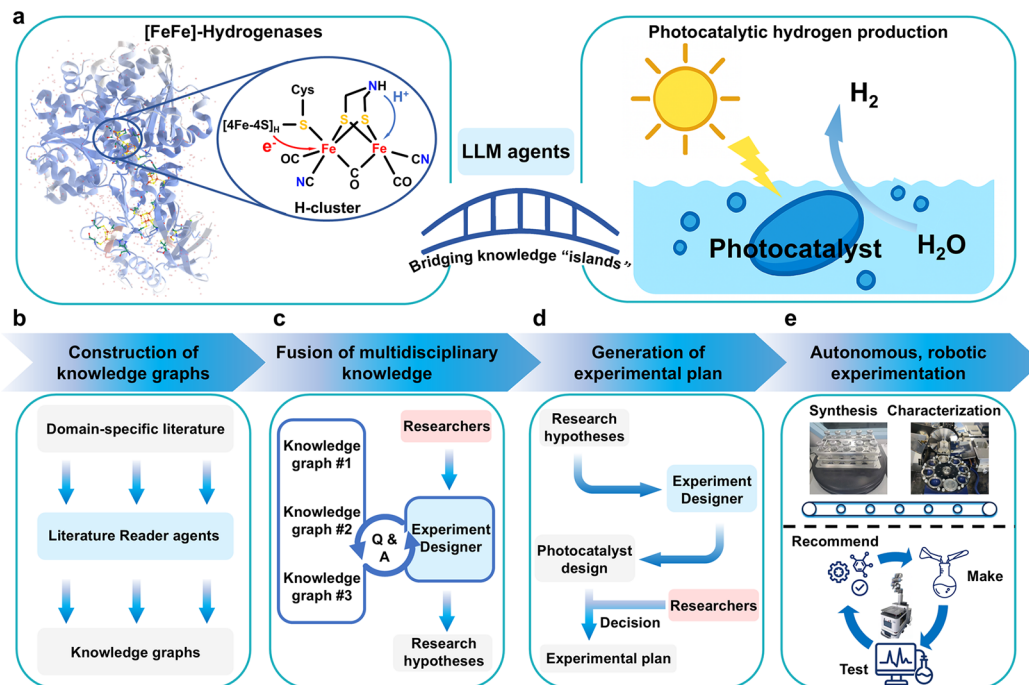
Revised: August 23, 2026

Accepted: August 25, 2026

Published: September 14, 2026



**Scheme 1. Overview of the End-to-End, LLM-Driven Autonomous Research Workflow for Photocatalyst Discovery and Optimization.** (a) LLM Agents Bridge Isolated Knowledge Domains by Connecting Insights from Biological and Materials Sciences, Linking [FeFe]-Hydrogenases to Photocatalytic Water Splitting for Hydrogen Production. (b) Literature Reader Agents Extract Information from Domain-specific Literature to Construct Cross-Domain Knowledge Graphs. (c) Knowledge Graphs are Fused to Enable Multidisciplinary Understanding and Support a Human-AI Q&A Process for Hypothesis Generation. (d) The Experiment Designer Formulates Experimental Plans and Photocatalyst Designs Based on Hypotheses, With Input from Researchers. (e) Guided by LLM-Based Decision-Making, a Robotic Platform Autonomously Performs Synthesis, Characterization, and Testing in a Closed-Loop Workflow for Photocatalyst Optimization



(Scheme 1). Photocatalytic water splitting, a green route to  $H_2$  from sunlight and water,<sup>29–31</sup> is an ideal testbed for its performance is governed by a dense network of interdependent variables.<sup>32–34</sup> Our workflow links: (i) LLM-driven knowledge extraction and graph-based fusion across photocatalysis, hydrogenase enzymology, and enzyme-mimetic catalysis (Scheme 1b); (ii) hypothesis generation centered on embedding basic functional groups as proton relays in the microenvironment around catalytic sites (Scheme 1c); and (iii) actionable experimental plans executed on a modular, high-throughput robotic platform (Scheme 1d). Applied to Zr-MOF scaffolds (UiO-67) with Pt cocatalyst centers, the system designs, synthesizes, evaluates, and iteratively refines postsynthetically modified Pt@UiO-67-X materials, with each round guided by real-time experimental feedback and LLM reasoning (Scheme 1e). This closed-loop approach provides a general blueprint for knowledge-guided, AI-enabled synthesis while directly targeting the microenvironmental features that underpin enzymatic efficiency.

## RESULTS AND DISCUSSION

### Knowledge Extraction and Hypothesis Generation

Building on our previously reported robotic AI chemist driven by a multiagent, LLM-based architecture,<sup>25</sup> we substantially extend the Literature Reader agent to move beyond information retrieval toward autonomous, cross-domain knowledge-graph construction. In this study, three graphs, spanning photocatalytic hydrogen production, hydrogenases, and enzyme-mimetic catalysis, are generated and then fused to expose shared

motifs and mechanistic linkages (Scheme 1a). This graph-based fusion enables traceable, citation-grounded reasoning that surfaces latent connections and prioritizes testable research directions across previously siloed literatures.

We deployed the Literature Reader as three domain-specialized instances to extract and synthesize knowledge relevant to photocatalytic hydrogen production, hydrogenases, and enzyme-mimetic catalysis. Corpora were assembled using the seed terms “photocatalytic hydrogen production”, “hydrogenase” and “enzyme-mimetic catalysis”, yielding 1,261, 348, and 30 publications, respectively (see Supporting Information). Together these corpora span mechanisms, materials, and design strategies—from artificial photocatalysts and performance metrics to enzymatic microenvironments and their synthetic analogues—providing broad coverage for AI-assisted analysis.

Each instance constructed a domain knowledge graph that maps key entities (materials, active-site motifs, functional groups), relationships (e.g., improves, proximal-to, enables PCET), and outcomes (activity, selectivity). In brief.

- Literature Reader #1: analysis of 1261 papers on photocatalytic hydrogen production produced a knowledge graph with 2612 triples (Figures 1a, S1), highlighting three key performance determinants: light absorption range, electron–hole separation efficiency, and the number and spatial distribution of active sites.
- Literature Reader #2: from 348 hydrogenase-related papers, the graph captured how [FeFe]/[NiFe] clusters and basic second-sphere residues (e.g., His, Lys, Arg; also Asp/Cys in local networks) govern natural enzyme

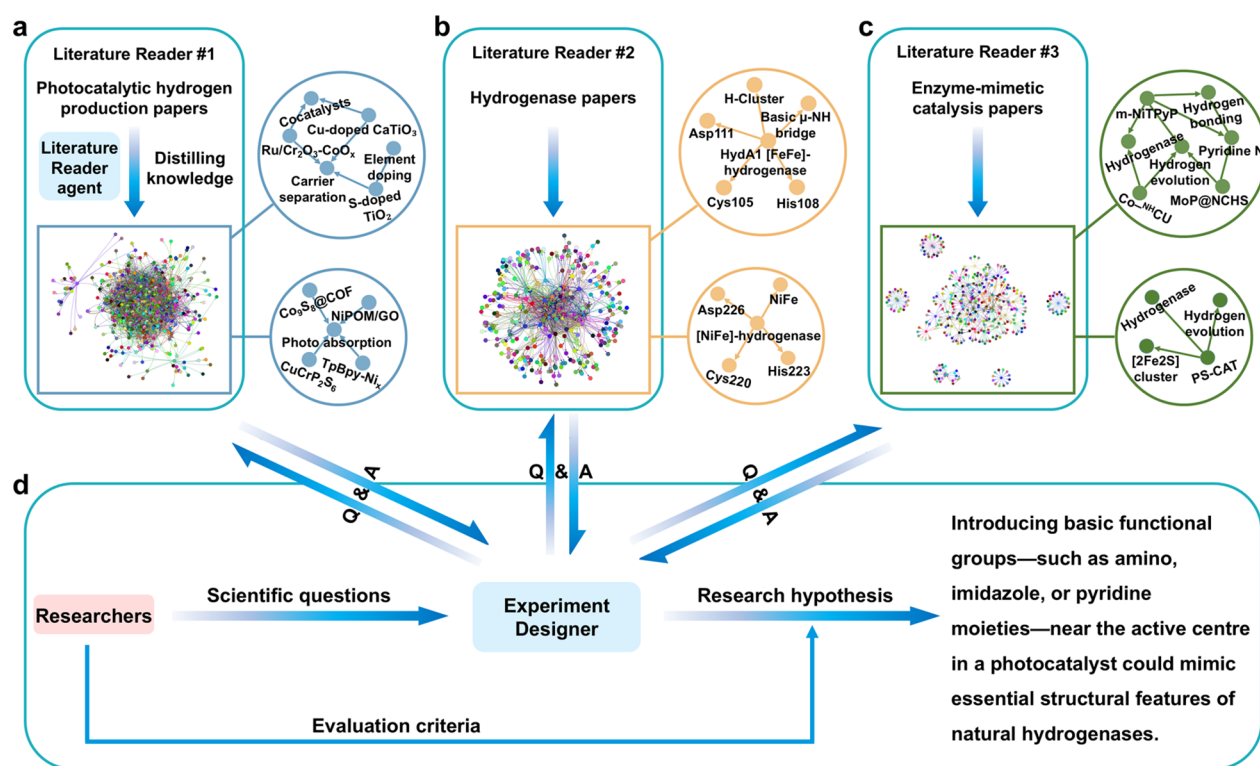


**Figure 1.** LLM-driven knowledge extraction and cross-domain hypothesis generation workflow. Three instances of the Literature Reader agent for analysis and knowledge graph construction from (a) the photocatalytic hydrogen production literature, (b) the hydrogenase literature, and (c) the enzyme-mimetic catalysis literature. (d) Human–AI collaborative workflow: Scientific questions posed by human researchers are addressed by the Experiment Designer agent, which integrates and fuses the knowledge graphs from (a–c) via a dynamic question-and-answer process. This enables the formulation of specific scientific hypotheses, such as the modification of photocatalysts with basic functional groups near the active center, which undergo iterative evaluation and refinement by both AI and human experts.

activity, primarily through hydrogen bonding and electrostatic interactions (Figures 1b, S2).

- Literature Reader #3: from 30 papers on enzyme-mimetic catalysis, the graph distilled structure–function parallels between enzymes and synthetic systems, including mimicry of hydrogenase metal centers and proton-donating/accepting sites such as imidazole, pyridine, and related *N*-heteroaromatics (Figures 1c, S3).

To support traceable reasoning, the three domain-specific Literature Reader instances were coupled to their respective graphs via a graph-enhanced retrieval-augmented generation (RAG) layer,<sup>35</sup> allowing them to answer queries with citations grounded in retrieved graph paths. We then introduced a coordinating agent, the Experiment Designer, to orchestrate these agents. Acting as a top-level controller, it aggregates graph-grounded outputs and fuses cross-domain evidence into research hypotheses (Figure 1d).

Through a human–AI Q&A workflow (Table S1), the Experiment Designer integrated the three domain-specific insights—from photocatalytic  $\text{H}_2$  production (kinetic bottlenecks), hydrogenase (basic-residue proton relays), and enzyme-mimetic catalysis (*N*-heteroaromatic shuttles)—and synthesized them into a central hypothesis: placing basic functional groups (e.g., amino, imidazole, pyridyl/quinolynyl) within the secondary coordination environment of catalytic sites in the photocatalyst should polarize and activate interfacial water and establish proton-relay pathways, thereby improving the coupling between proton-coupled electron transfer (PCET) and charge separation to enhance  $\text{H}_2$ -production efficiency.

In subsequent exchanges (Table S1), the Experiment Designer and researchers noted that slow proton transfer can limit electron utilization and impair charge separation, consistent with the PCET theory, motivating microenvironment engineering to optimize both water adsorption/activation and directional proton migration to catalytic metal sites. Human input was limited to guiding queries, resolving low-confidence items, and approving testable hypotheses, establishing an augmented-intelligence workflow in which cross-domain knowledge fusion directly informed experimental design.

We note that the proposed hypothesis is consistent with established features of [FeFe]/[NiFe] hydrogenases, where basic second-sphere residues (e.g., histidine, lysine, arginine) organ-ize hydrogen-bond microenvironment and pendant amines that relay protons to the H-cluster via reversible protonation.<sup>36–40</sup> By analogy, placing *N*-basic groups (imidazole, pyridyl/quinolynyl) near Pt sites should polarize interfacial water and shorten proton-transfer pathways, improving PCET coupling during  $\text{H}_2$  production. We examine this mechanistic picture experimentally and computationally below.

### Photocatalyst Design

Guided by the cross-domain hypothesis, the Experiment Designer translated mechanistic insight into an actionable catalyst design compatible with the robotic laboratory (Figure 2a). The agent first fixed platinum nanoparticles (Pt NPs) as the catalytic center, reflecting their well-established role as cocatalysts for photocatalytic  $\text{H}_2$  production.<sup>41,42</sup> It then searched for host frameworks and modification strategies that

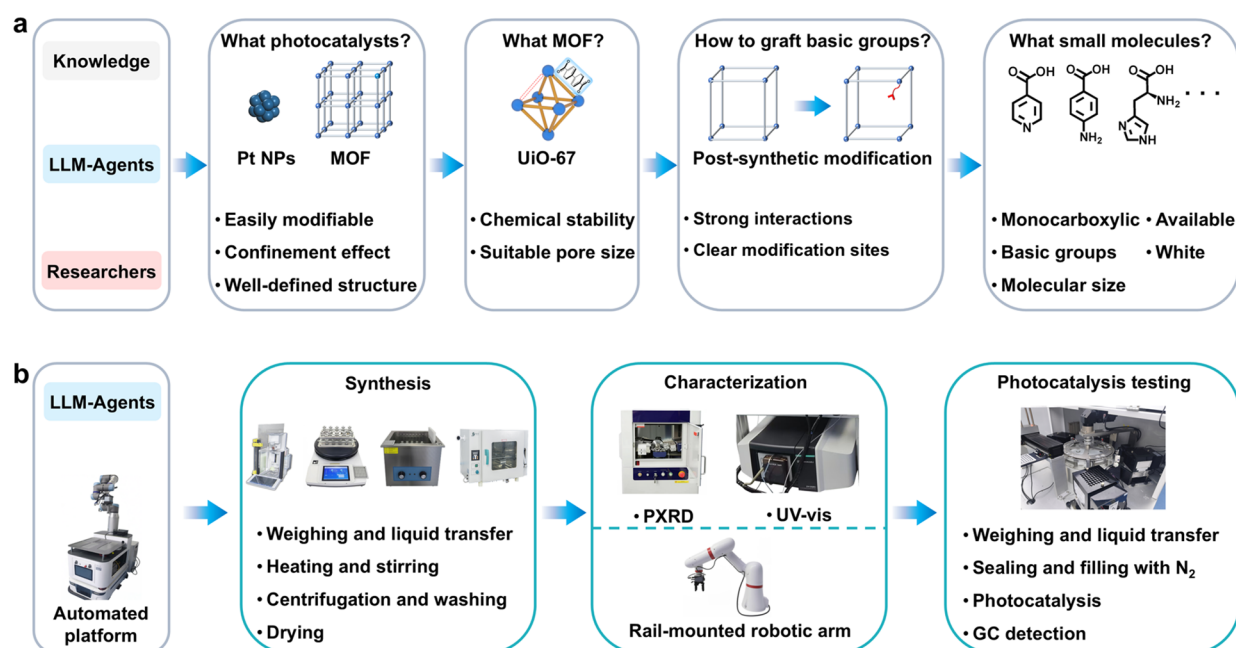


**Figure 2.** LLM-driven catalyst design and automated robotic experimental workflow. (a) Stepwise catalyst design workflow driven by the Experiment Designer agent. The agent, in collaboration with human researchers, systematically guides catalyst design through sequential decision-making—selecting the photocatalyst type (MOF), specific MOF (UiO-67), functionalization strategy (PSM with defect engineering), and finally, the small molecule candidates for basic functional group introduction, all based on rational criteria and knowledge integration. (b) Automated experimental platform for catalyst synthesis, characterization, and activity testing, integrating robotic workflows for material synthesis (encompassing weighing, liquid and solid transfer, heating, stirring, centrifugation, washing, and drying), automated characterization (including powder X-ray diffraction, simply as PXRD, and UV-vis spectroscopy), and high-throughput photocatalytic activity testing (covering sample preparation, light irradiation, and product detection).

would place basic functional groups in the microenvironment of Pt NPs (Table S2).

The Experiment Designer compared representative photocatalyst families, including inorganic semiconductors,<sup>43,44</sup> organic polymers,<sup>45,46</sup> covalent organic frameworks (COFs),<sup>47–49</sup> and metal–organic frameworks (MOFs).<sup>50–52</sup> MOFs were ultimately recommended because of their high scope for programmable architectures, pore accessibility, chemical robustness, and amenability to modular functionalization. Within MOFs, a head-to-head evaluation of Zr-based frameworks, Cu-based frameworks, and MIL series considered linker chemistry, stability under aqueous/illumination conditions, pore metrics, and known defect chemistry (Table S2). This analysis prioritized Zr-MOFs (UiO-66, UiO-67, MOF-808). The research team selected UiO-67, which is composed of biphenyl-4,4'-dicarboxylic acid (H<sub>2</sub>BPDC) linkers, for its larger pore size, exceptional hydrothermal/chemical stability, and extensive precedent for postsynthetic modification (PSM), features that enable precise placement of functional groups around encapsulated Pt NPs.

These design choices, selecting Pt as the cocatalyst, UiO-67 as the host, and defect-engineered PSM with *N*-heteroaromatic monocarboxylates as the modification strategy, were directly informed by the fused cross-domain knowledge: the photocatalytic H<sub>2</sub> production literature justified Pt for its low overpotential and stability; the hydrogenase literature provided the concept of a secondary coordination sphere with basic proton-relay residues; and the enzyme-mimetic catalysis literature supplied the chemical implementation via *N*-heteroaromatic carboxylates grafted onto a stable porous scaffold.

To introduce basic sites without perturbing the primary lattice, the Experiment Designer assessed presynthetic linker functionalization and possible PSM routes, including direct acylation/condensation of amino acids and defect-assisted grafting that exploits unsaturated Zr sites generated by controlled linker deficiency (Table S2). Balancing stability, site-specificity, and operational flexibility, the agent recommended PSM via defect engineering as the most robust route, allowing stable anchoring of monocarboxylates at Zr-oxo nodes while preserving crystallinity, an approach consistent with prior experimental experience in our lab.

To identify suitable small molecules for functionalization, the Experiment Designer was guided by researcher-defined criteria: each molecule should (i) carry a monocarboxylic acid (for coordination), (ii) include at least one nitrogen basic site (amine or aromatic N), (iii) have an effective backbone size smaller than three benzene rings (including fused rings), and (iv) be white/near-white in solid form. These screening criteria, including the basic-nitrogen requirement, were formulated by the research team to render the LLM-generated hypothesis experimentally tractable, and were not generated autonomously by the agent. Applying these filters, the Experiment Designer delineated four principal structural families that mimic amino-acid-like backbones:  $\alpha/\beta$ -alkyl-chain amino acids, *N*-heteroaryl amino acids, aromatic amino acids, and fused-ring aromatic amino acids (Table S2). For specificity control, benzoic acid, which meets the monocarboxylate requirement but lacks nitrogen, was included to test the exclusion logic and establish a nonbasic reference.

Selection within these families considered not only the required chemical functionalities and size, but also color consistency, solubility, salt-forming potential, and overall

compatibility with the MOF's pore environment. By screening a curated database of commercially available compounds from chemical vendors routinely used by our lab (see [Supporting Information](#) for the complete list of 464 entries), the agent shortlisted 31 candidate molecules ([Table 1](#)). The research team

**Table 1. Small Molecules Selected for PSM of Pt@UiO-67 via Defect Engineering, Coordinated through Carboxyl Groups to Zr Nodes, Yielding Pt@UiO-67-X (X = 1, 2, ..., 31, the Number X is Marked below the Molecule)**

| Structural category                | Selected molecules                                                                  |
|------------------------------------|-------------------------------------------------------------------------------------|
| $\alpha$ - or $\beta$ -amino acids |    |
|                                    |    |
|                                    |    |
|                                    |    |
|                                    |    |
|                                    |    |
|                                    |    |
|                                    |    |
|                                    |    |
| N-heteroaromatic carboxylic acids  |    |
|                                    |    |
|                                    |    |
|                                    |    |
|                                    |    |
|                                    |    |
|                                    |   |
|                                    |   |
|                                    |   |
|                                    |   |
|                                    |   |
|                                    |   |
| Aromatic amine-carboxylic acids    |  |
|                                    |  |
|                                    |  |
|                                    |  |
|                                    |  |
|                                    |  |
|                                    |  |
|                                    |  |
|                                    |  |
| Fused N-heteroaromatic amino acids |  |
| Control group                      |  |

then performed molecular visualization and cross-validated each selection against the literature, finalizing the set for subsequent automated synthesis and performance evaluation.

### Catalyst Synthesis and Characterization

Following the catalyst design phase, a sequence of automated tasks implemented the end-to-end process of catalyst synthesis

and characterization: synthesis of Pt NPs, assembly of Pt@UiO-67, postsynthetic modification to give Pt@UiO-67-X, and standardized characterization and photocatalysis. For each task, the Experiment Designer generated stepwise protocols ([Table S3](#)) encoded in XML and the Robot Operator agent compiled these to executable code, enabling fully unattended execution within the ChemAgents-driven laboratory (platform described previously).<sup>25</sup>

The automated workflow encompassed weighing, liquid/solid transfers, heating, stirring, centrifugation, washing and drying, coupled to high-throughput powder X-ray diffraction (PXRD) and UV-vis measurements and an automated photocatalysis line (sample preparation, illumination, product analysis) ([Figure 2b](#)). Apart from the automated operations specified above, the remaining characterization measurements were carried out manually. The robot synthesized Pt nanoparticles with uniform diameters of approximately 3.4 nm, as confirmed by transmission electron microscopy (TEM; [Figure S4](#)). Pt@UiO-67 composites prepared automatically showed PXRD patterns matching simulated UiO-67 ([Figure S5](#)). Thermogravimetric analysis (TGA) of Pt@UiO-67 indicated a 37% ligand-defect content ([Figure S6](#)); Quantitative analyses by nuclear magnetic resonance (<sup>1</sup>H NMR; linker quantitation) and inductively coupled plasma atomic emission spectroscopy (ICP-AES; Zr, Pt quantitation) corroborated these values ([Figure S7](#)). TGA, NMR, and ICP-AES studies were conducted manually.

For postsynthetic modification, small molecules were coordinated to unsaturated Zr sites to afford Pt@UiO-67-X (X = 1–31; [Table 1](#)). Automated PXRD confirmed that crystallinity was preserved across the series ([Figures S8–S10](#)). N<sub>2</sub> sorption (77 K) showed both Pt@UiO-67 and Pt@UiO-67-30 (8-quinolinecarboxylic acid) to be highly porous; the BET area decreased from 1567 m<sup>2</sup> g<sup>−1</sup> (Pt@UiO-67) to 1121 m<sup>2</sup> g<sup>−1</sup> (Pt@UiO-67-30) after modification ([Figure S11](#)), consistent with partial pore occupancy by grafted molecules.

Manual <sup>1</sup>H NMR verified incorporation of all 31 small molecules with an average molar ratio to the H<sub>2</sub>BPDC ligand of approximately 0.4:1 ([Figures S12–S42](#)). ICP-AES indicated a consistent Pt loading of approximately 0.18 wt % across samples ([Figure S43](#)). Scanning electron microscopy (SEM) showed octahedral particles of approximately 200 nm for both Pt@UiO-67 and Pt@UiO-67-30 ([Figure S44](#)), and TEM confirmed uniform, well-dispersed Pt NPs of approximately 3.4 nm in both materials ([Figure S45](#)), indicating that PSM did not perturb Pt size/distribution. To assess optical neutrality of the modifiers, automated UV-vis spectra revealed similar absorption profiles (250–475 nm) for all Pt@UiO-67-X samples ([Figure S46](#)), indicating that the MOF's intrinsic light-absorption window is preserved.

Collectively, these data show that, aside from the intended second-sphere chemistry, the Pt@UiO-67-X series is structurally, compositionally, and optically uniform, providing a controlled platform to isolate the impact of basic-group microenvironment engineering on photocatalytic H<sub>2</sub> production.

### Iterative Catalyst Performance Optimization

To enable fully autonomous discovery and optimization, we extended the Task Manager with a reasoning model (OpenAI's o1-preview). In a closed loop, the agent recommends experiments, analyses results, and updates its selection policy for the next iteration ([Figure 3a](#)). The search space comprised 31 Pt@UiO-67-X candidates ([Table 1](#)). As a baseline under automated

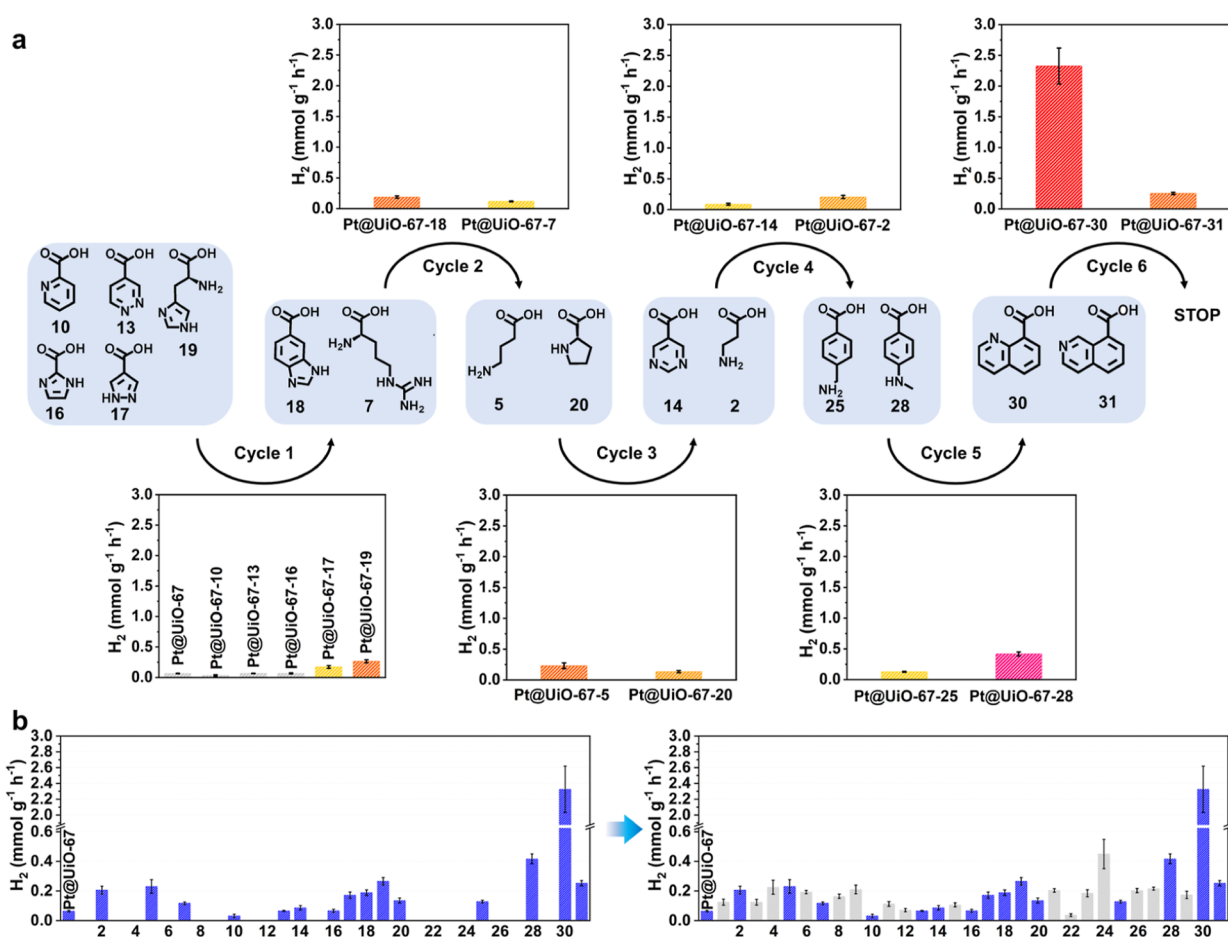


**Figure 3.** Iterative, LLM-guided optimization and validation of photocatalytic hydrogen production performance. (a) Closed-loop experiment recommendation and activity optimization workflow. (b) Comparison of photocatalytic hydrogen production performance for all 31 Pt@UiO-67-X and the parent Pt@UiO-67 catalysts. Error bars represent the mean  $\pm$  standard deviations, calculated from three independent tests.

conditions (full-spectrum illumination, triethylamine as sacrificial agent, H<sub>2</sub>O as hydrogen source), Pt@UiO-67 delivered 0.06 mmol g<sup>-1</sup> h<sup>-1</sup> (Figure 3b).

In the first round, the agent prioritized nitrogen-rich, H-bond-competent motifs expected to aid water activation and proton shuttling, recommending: 2-pyridylcarboxylic acid ( $X = 10$ ), pyridazine-4-carboxylic acid ( $X = 13$ ), imidazole-2-carboxylic acid ( $X = 16$ ), 4-pyrazolecarboxylic acid ( $X = 17$ ), and L-histidine ( $X = 19$ ). After robotic synthesis and testing, Pt@UiO-67-19 and Pt@UiO-67-17 outperformed, with L-histidine reaching 0.27 mmol g<sup>-1</sup> h<sup>-1</sup> (4 $\times$  baseline performance).

In response to the first-round results, the agent analyzed feature importance across the samples, contrasting high performers with low, and updated its working hypothesis to emphasize nitrogen density and geometry near Pt together with conformational flexibility of the basic site. Guided by this refinement, round two prioritized 1H-benzimidazole-5-carboxylic acid ( $X = 18$ ), representing a fused *N*-heteroaromatic with closely spaced nitrogen atoms, and L-arginine ( $X = 7$ ), offering a flexible, multinitrogen (guanidinium) side chain. Both catalysts delivered hydrogen production rates 2–3  $\times$  higher than baseline, supporting the revised structure–activity rationale.

In rounds three through five, the agent further expanded its exploration to flexible aliphatic amino acids, cyclic secondary amines (e.g., L-proline), and rigid aromatic frameworks with amino substituents. Each recommendation was explicitly justified through structure–activity analysis of previous experi-

ments, such as the superior performance of molecules enabling dynamic microenvironments around Pt NPs. Flexible amino acid derivatives (e.g., 4-aminobutyric acid, 3-aminopropionic acid) consistently outperformed more rigid structures, highlighting the importance of adaptability and optimal hydrogen-bonding microenvironment around Pt NPs.

In the final (sixth) round, the agent targeted extended nitrogen-containing aromatic systems such as quinoline derivatives to further enhance electron delocalization and proton mobility. 8-quinolinecarboxylic acid ( $X = 30$ ) emerged as the top performer, achieving a hydrogen production activity of 2.33 mmol g<sup>-1</sup> h<sup>-1</sup>, a  $\sim$ 36-fold improvement over the parent Pt@UiO-67 (Figure 3b).

Because a clear optimum was identified after testing half of the library, the optimization loop was concluded. For completeness, the remaining 16 candidates were synthesized and evaluated under the same automated protocol; all showed activities <0.6 mmol g<sup>-1</sup> h<sup>-1</sup>, confirming that Pt@UiO-67-30 is the highest-activity candidate in this library and that no stronger candidate remained untested (Figure 3b, right panel). Throughout the campaign, the agent's prompts, rationales, and selection criteria were recorded alongside experimental outcomes (Table S4).

### Manual Experimental Optimization

During automated experiments, the photocatalytic H<sub>2</sub> production tests were conducted in 15 mL septum vials, whose small size limited the catalyst's exposure to the full aperture of the

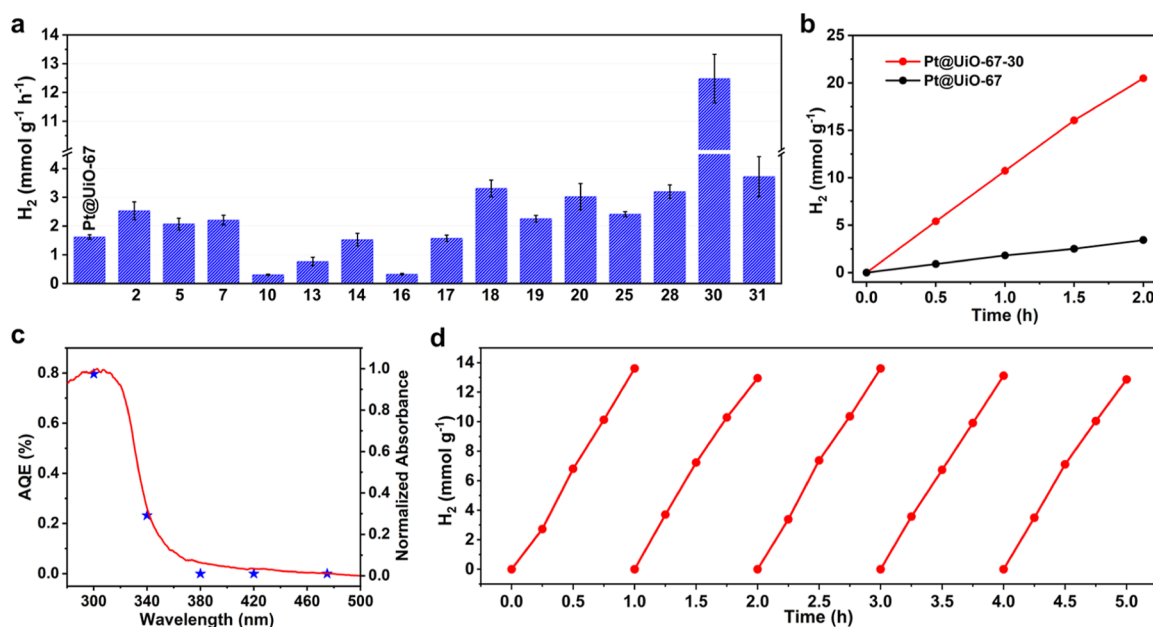


**Figure 4.** Enhanced photocatalytic hydrogen production performance and stability of LLM-optimized catalysts under manual experimental conditions. (a) Hydrogen production rates of Pt@UiO-67 and 15 Pt@UiO-67-X catalysts (selected by ChemAgents) manually measured in a larger photoreactor under optimized light irradiation. Error bars represent the mean  $\pm$  standard deviations, calculated from three independent tests. (b) Time-dependent H<sub>2</sub> production curves for Pt@UiO-67-30 and Pt@UiO-67. (c) Apparent quantum efficiency (AQE, left axis, stars) and UV-vis absorption spectrum (right axis, red curve) of Pt@UiO-67-30 under monochromatic light irradiation at various wavelengths. (d) Cycling stability of Pt@UiO-67-30 over five consecutive photocatalytic runs, demonstrating robust and sustained hydrogen production performance.

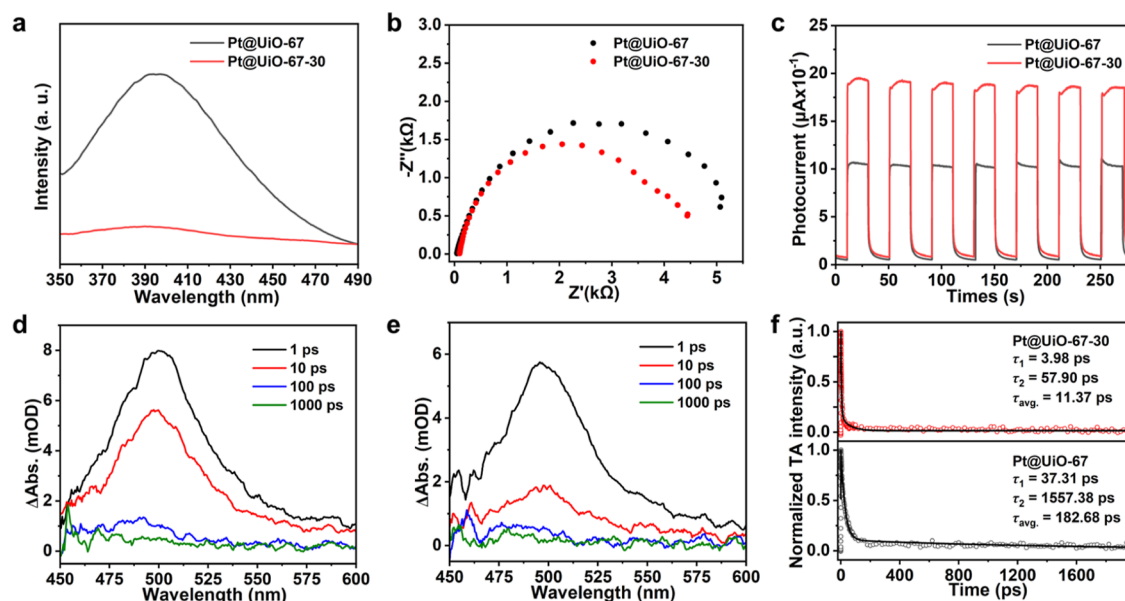


**Figure 5.** Mechanistic studies of Pt@UiO-67 and Pt@UiO-67-30. (a) Steady-state photoluminescence (PL) emission spectra ( $\lambda_{\text{ex}} = 300$  nm), showing quenched emission for Pt@UiO-67-30. (b) EIS Nyquist plots under light irradiation, indicating reduced charge transfer resistance for Pt@UiO-67-30. (c) Transient photocurrent responses under chopped light irradiation, showing enhanced photocurrent for Pt@UiO-67-30. The fs-TA spectra (pump at 300 nm) at selected time delays for (d) Pt@UiO-67 and (e) Pt@UiO-67-30. (f) TA kinetics at 500 nm, fitted with biexponential decay, revealing a markedly shorter average lifetime for Pt@UiO-67-30.

xenon lamp. To address this limitation and maximize light utilization, we performed manual hydrogen production activity tests for the 15 catalysts recommended and tested by the robotic AI Chemist, as well as for the pristine Pt@UiO-67. In these manual experiments, a larger photoreactor was employed, ensuring complete and uniform irradiation of photocatalyst and thus more accurately reflecting the intrinsic photocatalytic performance.

Under these optimized conditions, all tested Pt@UiO-67-X catalysts exhibited significantly enhanced hydrogen production activities compared to the automated setup (Figures 3b, 4a). Most notably, Pt@UiO-67-30 reached a hydrogen production rate as high as 12.48 mmol g<sup>-1</sup> h<sup>-1</sup>. Time-course experiments under continuous xenon lamp irradiation revealed a linear increase in hydrogen production (Figure 4b), while monochromatic light irradiation at 300 nm afforded the maximum

apparent quantum efficiency (Figure 4c). Importantly, the relative activity trends across all catalysts remained consistent with those observed under automated conditions, validating the robustness and reliability of the robotic AI Chemist's approach. Cycling experiments further demonstrated the excellent stability of Pt@UiO-67-30, with its activity remaining unchanged over five cycles (Figure 4d). Postreaction analyses confirmed the structural integrity of the catalyst, as evidenced by unchanged PXRD patterns (Figure S47) and preserved morphology and nanoparticle dispersion in SEM and TEM images (Figure S48).

### Photophysical and Electrochemical Characterization

To probe the origin of the enhanced photocatalytic activity in Pt@UiO-67-30, we performed complementary photo- and electrochemical measurements. Steady-state photoluminescence (PL) spectra revealed significantly lower emission intensity for Pt@UiO-67-30 compared to Pt@UiO-67, indicating suppressed radiative recombination in the former (Figure 5a). Electrochemical impedance spectroscopy (EIS) revealed a reduced charge transfer resistance for Pt@UiO-67-30, as evidenced by a smaller Nyquist arc radius (Figure 5b), and transient photocurrent measurements showed a stronger and more stable photocurrent response for the modified catalyst (Figure 5c). Together, these results point to suppressed charge recombination as the key factor responsible for the enhanced photocatalytic performance of Pt@UiO-67-30.

Femtosecond transient absorption (fs-TA) spectroscopy provided further insights into charge carrier dynamics. Using a 300 nm pump pulse to selectively excite the MOF, similar spectral features before and after small molecule modification can be observed, but with marked differences in the kinetic decay at 500 nm (Figure 5d,e). Biexponential fitting yielded average lifetimes of 182.68 ps for Pt@UiO-67 and 11.37 ps for Pt@UiO-67-30. The shorter lifetime indicates a faster excited-state decay in Pt@UiO-67-30, which may arise from 8-quinolinecarboxylic acid accelerating electron transfer to Pt NPs (Figure 5f). First-principles calculations (see Section S3 in Supporting Information) corroborate this picture by showing that the quinolinyl group stabilizes a hydrogen-bonded water/proton relay near the active site and lowers the proton-transfer barrier relative to unmodified Pt@UiO-67, consistent with the experimentally observed gains in charge separation and H<sub>2</sub> production.

### Validating AI Generated Hypotheses

To test whether the LLM could formulate experimentally testable and generalizable hypotheses beyond literature retrieval and closed-loop optimization within the original library, we conducted a prospective validation study. After the six optimization cycles, the LLM was provided with the structure–activity data for all 31 original catalysts (Figure 3b). It inferred that the leading catalyst, Pt@UiO-67-30 (8-quinolinecarboxylic acid), benefited from the combination of a rigid, fused *N*-heteroaromatic scaffold with a hydrogen-bond-accepting site. The LLM therefore proposed that retaining this rigid heteroaryl proton-relay motif while introducing an additional hydrogen-bond donor or acceptor could produce highly active catalysts. On this basis, it recommended two previously unexplored molecular families—amino-substituted quinolinecarboxylic acids and azaindoles—comprising 10 candidates that were absent from the original library (Table S5).

We synthesized the corresponding catalysts, Pt@UiO-67-*X* (*X* = 32–41), and evaluated them under the same manually

optimized conditions. Their hydrogen-production rates ranged from 1.44 to 7.61 mmol g<sup>−1</sup> h<sup>−1</sup>, exceeding that of the parent Pt@UiO-67 catalyst. Notably, Pt@UiO-67-41, modified with 2-aminoquinoline-6-carboxylic acid, achieved a rate of 7.61 mmol g<sup>−1</sup> h<sup>−1</sup> and outperformed every catalyst in the original library except Pt@UiO-67-30 (Figure S64). The relative performance of the two molecular families also supported the LLM-derived structure–activity rationale, with the aminoquinoline derivatives outperforming the azaindoles.

Together, these prospective results indicate that the LLM extracted a transferable design principle from the original data set and successfully applied it to previously untested molecular space, rather than merely ranking candidates within the predefined library. Full characterization data, including PXRD, <sup>1</sup>H NMR, and UV–vis measurements, are provided in Figures S52–S63.

## CONCLUSIONS

We present an end-to-end autonomous workflow that couples cross-domain knowledge fusion with iterative LLM reasoning from real-time experimental feedback to design, synthesize, and optimize photocatalysts. This work asks whether an LLM-based system can recover a design principle from separate literature and convert it into an executable materials strategy for heterogeneous, porous photocatalysts. Integrated with closed-loop robotic experimentation, the system executed six optimization cycles synthesizing 31 postsynthetically modified Pt@UiO-67-*X* catalysts and converged on Pt@UiO-67-30 (8-quinolinecarboxylic acid), which exhibited markedly enhanced photocatalytic H<sub>2</sub> production (2.33 mmol g<sup>−1</sup> h<sup>−1</sup> under automated conditions; 12.48 mmol g<sup>−1</sup> h<sup>−1</sup> under optimized manual testing). Photophysical and electrochemical measurements, together with supporting theoretical calculations, are consistent with enhanced charge separation and proton transfer, providing a mechanistic rationale that link the knowledge-guided design principle to the observed performance gains.

Beyond this specific system, we frame this study as a proof of concept for a knowledge-guided agentic workflow, not as evidence that automation is necessary for every catalyst library of this size. The value of the approach lies in integrating fragmented chemical knowledge, formulating mechanistically grounded hypotheses that the researchers translate into executable screening criteria, which the Experiment Designer in turn applies to assemble the candidate catalysts, and refining the resulting structure–activity reasoning through feedback from robotic experimentation. The prospective validation with new candidate families further shows that the workflow can generalize learned structure–activity principles beyond the original library.

Looking ahead, scaling this approach will benefit from standardized corpora and ontologies for materials and mechanisms, richer in-line characterization to tighten the feedback loop, and multiobjective reasoning that jointly optimizes performance and practicality (e.g., stability, cost, etc.). With these advances, autonomous, knowledge-guided robotic experimentation can move from single demonstrations to a general engine for function-driven synthesis across chemistry and materials science.

## ■ EXPERIMENTAL SECTION

### Materials and Equipment

Unless otherwise specified, all chemicals were purchased from commercial sources and used without further purification. X-ray diffraction (XRD) patterns were acquired using a Rigaku Miniflex600 diffractometer (Japan) with Cu K $\alpha$  radiation ( $\lambda = 1.54178 \text{ \AA}$ ), incorporating a graphite monochromator. Thermogravimetric behavior was evaluated on a TA Q5000 thermal analyzer. Nitrogen sorption measurements were conducted at 77 K on a Micromeritics ASAP 2020 system, with prior sample activation under vacuum at 150 °C for 12 h. Morphological features were examined by scanning electron microscopy (SEM, Carl Zeiss Supra 40), while the dispersion and size of Pt nanoparticles (NPs) were further investigated using a JEOL JEM-2100F field-emission transmission electron microscope (TEM). Proton NMR spectra were recorded with a Bruker AVANCE III 400 MHz spectrometer. The platinum content was quantified via inductively coupled plasma atomic emission spectroscopy (ICP-AES) using an Optima ICAP 7400 instrument. UV–vis diffuse reflectance spectra (DRS) were collected with a Shimadzu UV-2700 spectrophotometer using BaSO<sub>4</sub> as the baseline. Photoluminescence (PL) measurements were performed with a Hitachi FL-4600 spectrometer under 300 nm excitation. Electrochemical characterizations were carried out on a Zahner Zennium electrochemical workstation and a CHI 760E electrochemical analyzer (Chenhua Instrument, Shanghai, China). Hydrogen production was monitored using a SHIMADZU GC-2014 gas chromatograph equipped with a thermal conductivity detector (TCD).

### Automated Synthesis of Pt@UiO-67

A 15 mL reaction vial was placed onto the automated sample rack. Using the solid handling module, 0.156 g of H<sub>2</sub>BPDC was accurately dispensed into the vial. The system then sequentially added 10.2 g of ZrCl<sub>4</sub> solution in DMF, 0.19 g of Pt@PVP colloidal solution, 1.5 g of acetic acid, and 0.75 g of deionized water using the liquid dispensing station. The mixture was subsequently sonicated at room temperature for 5 min to ensure homogeneous mixing. Following sonication, the vial was sealed using the automated encapsulation unit and transferred to a magnetic stirring and heating module. The reaction was maintained at 120 °C under stirring at 600 rpm for 1 h to facilitate MOF crystallization around the preformed Pt NPs. After cooling to room temperature, the reaction mixture was subjected to a purification sequence. It was first centrifuged at 5000 rpm for 5 min, and the supernatant was aspirated. The resulting solid was washed twice with 10 g of DMF, each time followed by sonication for 5 min, centrifugation at 5000 rpm for 5 min, and removal of the supernatant. Further purification involved two washing cycles with 8 g of acetone, each followed by 5 min sonication, centrifugation at 5000 rpm for 2 min, and aspiration of the supernatant. The final pellet was dried in the drying station at 60 °C for 12 h to yield the Pt@MOF composite as a dry powder.

### Automated Synthesis of Pt@UiO-67-X

A 15 mL reaction vial containing 0.1 g of Pt@UiO-67 was placed onto the automated sample rack. Using the liquid handling station, 8 g of small organic molecules solution was added. The vial was then transferred to the ultrasonic station and sonicated for 5 min to ensure thorough dispersion of the MOF particles in the solution. The vial was sealed via the automated encapsulation module and subjected to continuous stirring at 600 rpm and 40 °C for 12 h using the magnetic stirring station to allow postsynthetic modification or guest molecule incorporation. After the reaction, the vial was centrifuged at 5000 rpm for 5 min, and the supernatant was aspirated. The resulting solid was washed twice with 10 g of DMF, with each wash involving 5 min sonication, centrifugation at 5000 rpm for 5 min, and aspiration of the supernatant to remove unreacted species. Subsequently, the sample was washed twice with 8 g of acetone, each followed by 5 min sonication, centrifugation at 5000 rpm for 2 min, and supernatant removal. The cleaned material was finally dried at 60 °C for 12 h in a drying station to obtain the post-treated Pt@UiO-67-X as a dry powder.

### Automated Photocatalytic Hydrogen Production Experiment

A total of 5 mg of Pt@UiO-67-X photocatalyst was added into a 15 mL serum vial using the automated solid handling module. The liquid handling system then sequentially dispensed 5 g of acetonitrile, 0.21 g of triethylamine (TEA), and 0.15 g of deionized water into the vial. The resulting suspension was transferred to the ultrasonic processing station and sonicated for 5 min to achieve homogeneous dispersion of the catalyst. Following dispersion, the vial was sealed using the encapsulation module. Prior to sealing, five cycles of nitrogen purging and refilling were performed to remove residual oxygen and create an inert atmosphere. A gas-permeable cap was used to allow pressure balance during photocatalysis. The sample was then transferred to the photocatalytic reaction station, where it was irradiated for 1 h using a 300 W full-spectrum xenon lamp as the light source. The reaction temperature was maintained by circulating thermostated cooling water to ensure thermal stability. Upon completion of the photocatalytic test, the vial was automatically delivered to the gas sampling and analysis module. A 400  $\mu$ L volume of gas was extracted from the vial headspace via a headspace-compatible transfer system and injected into a Shimadzu HS-GC (headspace gas chromatography) system to quantitatively analyze the evolved hydrogen (H<sub>2</sub>).

### Manual Photocatalytic Hydrogen Production Experiment

Typically, 10 mg of the photocatalyst Pt@UiO-67-X was added to a 160 mL photoreactor (Beijing Perfectlight Technology Co., LTD, China), then 18 mL of acetonitrile, 1 mL of triethylamine (TEA), and 0.5 mL of water were added, respectively. The mixture was sonicated for 1 min to achieve uniform dispersion. The reactor was purged with nitrogen gas for 15 min to remove any residual air. Use a full-spectrum 300 W Xe lamp (LX-300F, Japan) as the light source to irradiate the sample for 1 h, maintaining a stable reaction temperature with circulating thermostatic cooling water. Finally, 400  $\mu$ L of the gaseous product was injected using a headspace syringe and the hydrogen yield was analyzed using GC (Shimadzu GC-2014, with argon as the carrier gas).

Using a 300 W xenon lamp as the light source, equipped with band-pass filters of 300, 340, 380, 420, and 475 nm, the apparent quantum efficiency (AQE) was measured. The calculation formula for AQE is as follows

$$\text{AQE}(\text{H}_2)\% = \frac{2 \times N(\text{H}_2)}{N(\text{photons})} \times 100\%$$

where  $N(\text{H}_2)$  denotes the number of H<sub>2</sub> molecules produced and  $N(\text{photons})$  denotes the number of photons reaching the surface of the reaction suspension.

### Photocurrent Measurements

Photocurrent response measurements were conducted using a standard three-electrode system on a CHI760E electrochemical workstation (Shanghai Chenhua Instruments, China). First, 2 mg of photocatalyst was dispersed in a mixture of 1 mL ethanol and 10  $\mu$ L of 5 wt % Nafion solution. Then, 100  $\mu$ L of this mixture was uniformly coated onto a 2 cm<sup>2</sup> indium tin oxide (ITO) electrode surface to prepare the working electrode. In the test setup, a platinum electrode and an Ag/AgCl electrode were used as the counter and reference electrodes, respectively. A 0.1 M Na<sub>2</sub>SO<sub>4</sub> solution was used as the electrolyte. A full-spectrum 300 W xenon lamp was used as the light source, and the photocurrent response of the sample was recorded at a bias voltage of +0.5 V.

### Electrochemical Impedance Spectroscopy

Electrochemical impedance spectroscopy (EIS) data were collected using a standard three-electrode system on a Zahner Zennium electrochemical workstation. In the experiment, 2 mg of photocatalyst was dispersed in a mixture of 1 mL ethanol and 10  $\mu$ L of 5 wt % Nafion solution. Then, 20  $\mu$ L of this mixture was coated onto the surface of a glassy carbon electrode with a diameter of 0.5 cm, preparing it as the working electrode. A platinum sheet was used as the counter electrode, and an Ag/AgCl electrode as the reference electrode. A 0.1 M Na<sub>2</sub>SO<sub>4</sub>

solution served as the electrolyte, and EIS measurements were carried out at a bias potential of  $-1.3$  V.

## TA Experiments

The fs-TA spectroscopy was conducted in a Helios ultrafast spectrometer with pump and probe pulses resulting from a regenerative amplified Ti-sapphire laser system (Solstice, 800 nm,  $<100$  fs FWHM, 3.5 mJ/pulse, 1 kHz repetition rate). Typically, 2 mg of catalyst with 0.25 mL of acetone solution containing 50 wt % PMMA in a glass vial was mixed and stirred. Subsequently, the resulting uniformly dispersed system was dropped onto a glass substrate and allowed to dry into a film in air. During the experiment, this film must continuously move to dissipate the heat generated by femtosecond laser pulses, preventing thermal effects that could cause structural damage to the sample.

## Construction of Knowledge Graphs

Corpus assembly and screening: Records were retrieved from Web of Science (accessed November 30, 2024) using three seed queries: “photocatalytic hydrogen production,” “hydrogenase,” and “enzyme-mimetic catalysis,” yielding 35,673 items. To maximize relevance and control corpus size, we applied domain-specific inclusion criteria before full-text processing.

- Photocatalytic hydrogen production and hydrogenase: original research articles in peer-reviewed journals reporting mechanistic or process-level insight (e.g., excited-state dynamics, electron/charge transfer, interfacial steps); non-English, editorials, conference abstracts, patents, and purely device-level reports without mechanistic content were excluded.
- Enzyme-mimetic catalysis: due to the smaller literature base, landmark reviews synthesizing structure–function relationships and representative materials were curated to capture design frameworks and terminology.

After deduplication (by DOI) and screening (title/abstract  $>$  full-text triage), the final corpora comprised 1261 photocatalysis articles, 348 hydrogenase articles, and 30 enzyme-mimetic reviews (see [Supporting Information](#)).

Acquisition and text normalization: References were managed in Zotero; full texts were batch-downloaded via DOI and exported as PDFs. PDFs were converted to UTF-8 plain text with Docling (detailed in [Supporting Information](#)), retaining section headers and figure/table captions when parsable. We performed literature distillation using the GPT-4o-mini API, guided by a domain-specific system prompt (detailed in [Supporting Information](#)) that removed boilerplate content (e.g., author names, affiliations, references) and nonmechanistic information, and normalized chemical entities and symbols (e.g., “H<sub>2</sub>”, “Pt NPs”, “UiO-67”).

Entity and relation extraction: The corpus  $D$  was first segmented into smaller text units  $D_i$  to enable fine-grained retrieval and graph construction

$$D = \{D_1, D_2, \dots, D_n\}, D_i = \text{Segment}(D, \text{window}_{\text{size}})$$

where each chunk  $D_i$  serves as the atomic unit for entity and relation recognition.

Entity and relation candidates were extracted from each chunk using a LLM prompt-based recognizer  $R(\cdot)$ , as defined in LightRAG (detailed in the [Supporting Information](#))

$$(V, E) = \bigcup_{D_i \in D} R(D_i)$$

where  $V$  and  $E$  denote the sets of entities (nodes) and relations (edges) identified across all text segments.

A profiling function  $P(\cdot)$  then generated key–value pairs summarizing each entity and relation

$$P(V, E) = \{(k, v) | k \in \text{labels}(V, E), v$$

$$= \text{LLM\_summary}(D_i[k])\}$$

where each key  $k$  corresponds to an entity or relation label, and its value  $v$  is a compact text summary derived from relevant sentences. The profiling step condenses all context sentences associated with a given

node or edge into a compact textual value  $v$ , providing an interpretable representation for retrieval and reasoning.

Graph Construction and Optimization: the LightRAG graph index was obtained by integrating and deduplicating profiled entities and relations

$$\hat{D} = (\hat{V}, \hat{E}) = \text{Dedupe}(P(V, E))$$

where  $\text{Dedupe}(\cdot)$  merges identical entities and relations across text chunks to minimize redundancy and improve traversal efficiency. The resulting graph  $\hat{D}$  supports LightRAG’s dual-level retrieval mechanism, enabling both entity-level and abstract-level information access for downstream retrieval-augmented generation.

## Human–AI Collaborative Question–Answering and Cross-Domain Fusion

Domain-aware RAG agents: For each domain, we coupled the corresponding knowledge graph to a dedicated GPT-4o-mini instance via a graph-enhanced retrieval-augmented generation (RAG) pipeline: query  $\rightarrow$  dual-level graph retrieval (entity-level + topic-level)  $\rightarrow$  LLM-based answer generation grounded on retrieved graph context. This produced three domain-specific readers (Literature Reader #1–#3) capable of structured, citation-grounded answers and follow-up queries.

Coordinator and dialogue protocol: Cross-domain knowledge synthesis was handled by a coordinating reasoner (“Experiment Designer”) instantiated with GPT-4o. The coordinator.

- formulates a domain-decomposed query plan,
- dispatches subqueries to the appropriate reader(s),
- aggregates retrieved graph paths and rationales, and
- emits testable design rules (hypotheses + constraints + candidate features) with explicit provenance.

Human oversight was limited to (i) posing top-level scientific questions, (ii) resolving low-confidence conflicts, and (iii) approving hypotheses before execution. Representative dialogues are provided in [Table S1](#).

From rules to action: Accepted rules were translated into actionable experiment plans (materials, modifiers, synthesis route, test conditions) with machine-readable steps (XML; [Table S3](#)). Plans included negative controls (e.g., benzoic acid, *N*-free) and constraints derived from the fused evidence (e.g., monocarboxylate for Zr-node binding; backbone size limit for pore compatibility). This ensured that hypotheses were not only mechanistically motivated but executable on the robotic platform.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.6c13252>.

List of DOIs for photocatalytic hydrogen production papers ([XLSX](#))

List of DOIs for hydrogenase papers ([XLSX](#))

List of DOIs for enzyme-mimetic catalysis papers ([XLS](#))

Complete list of 464 compounds screened from the commercial database ([XLS](#))

Additional experimental details, including preparation of chemical solutions, synthesis of platinum nanoparticles, large language model-based studies and details, first-principles studies, and characterization data, including PXRD, <sup>1</sup>H NMR, UV–vis, TEM, SEM, ICP–AES, and N<sub>2</sub> sorption ([PDF](#))

## ■ AUTHOR INFORMATION

### Corresponding Authors

Linjiang Chen — Hefei National Research Center for Physical Sciences at the Microscale, State Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials

Science, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China; School of Chemistry, School of Computer Science, University of Birmingham, Birmingham B15 2TT, U.K.; [orcid.org/0000-0002-0382-5863](https://orcid.org/0000-0002-0382-5863); Email: [linjiangchen@ustc.edu.cn](mailto:linjiangchen@ustc.edu.cn)

**Weiwei Shang** – Department of Automation, State Key Laboratory of Precision and Intelligent Chemistry, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China; Email: [wwshang@ustc.edu.cn](mailto:wwshang@ustc.edu.cn)

**Jun Jiang** – Hefei National Research Center for Physical Sciences at the Microscale, State Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China; Hefei National Laboratory, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China; [orcid.org/0000-0002-6116-5605](https://orcid.org/0000-0002-6116-5605); Email: [jiangjl@ustc.edu.cn](mailto:jiangjl@ustc.edu.cn)

**Hai-Long Jiang** – Hefei National Research Center for Physical Sciences at the Microscale, State Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China; [orcid.org/0000-0002-2975-7977](https://orcid.org/0000-0002-2975-7977); Email: [jianglab@ustc.edu.cn](mailto:jianglab@ustc.edu.cn)

## Authors

**Yiming Zhao** – Hefei National Research Center for Physical Sciences at the Microscale, State Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China

**Tao Song** – Department of Automation, State Key Laboratory of Precision and Intelligent Chemistry, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China

**Yan Huang** – Hefei National Research Center for Physical Sciences at the Microscale, State Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China; [orcid.org/0000-0003-2512-2509](https://orcid.org/0000-0003-2512-2509)

**Kang Sun** – Hefei National Research Center for Physical Sciences at the Microscale, State Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China; [orcid.org/0000-0001-8720-9330](https://orcid.org/0000-0001-8720-9330)

**Wentao Han** – Hefei National Research Center for Physical Sciences at the Microscale, State Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China

**Mingyang Shen** – Hefei National Research Center for Physical Sciences at the Microscale, State Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China

**Chenwei Mao** – Department of Automation, State Key Laboratory of Precision and Intelligent Chemistry, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China

**Peng Lan** – Hefei National Research Center for Physical Sciences at the Microscale, State Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials

Science, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China

**Meng Zhou** – Hefei National Research Center for Physical Sciences at the Microscale, State Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026 Anhui, P. R. China; [orcid.org/0000-0001-5187-9084](https://orcid.org/0000-0001-5187-9084)

Complete contact information is available at: <https://pubs.acs.org/10.1021/jacs.6c13252>

## Author Contributions

<sup>†</sup>Y. Z., T. S., L. C., Y. H., and K. S. contributed equally.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by National Key Research and Development Program of China (2021YFA1500400), National Natural Science Foundation of China (22331009, W2512006), International Partnership Program of CAS (123GJHZ2022028MI), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB0540000; Category B, 2026-IMR). J.J. acknowledges the National Natural Science Foundation of China (Grants 22025304, 22033007), and the CAS Project for Young Scientists in Basic Research (YSBR-005, YSBR-142) for funding. L.C. acknowledges the University of Science and Technology of China (USTC) Startup Program for funding. The AI-driven experiments, simulations and model training were performed on the robotic AI-Scientist platform of Chinese Academy of Sciences (CAS). The calculations in this work are supported by the Supercomputing Center of USTC, Fundamental Research Funds for the Central Universities (WK2060000064).

## REFERENCES

- (1) Radzicka, A.; Wolfenden, R. A proficient enzyme. *Science* **1995**, 267, 90–93.
- (2) Buller, R.; Lutz, S.; Kazlauskas, R. J.; Snajdrova, R.; Moore, J. C.; Bornscheuer, U. T. From nature to industry: harnessing enzymes for biocatalysis. *Science* **2023**, 382, No. eadh8615.
- (3) Bierbaumer, S.; Nattermann, M.; Schulz, L.; Zschoche, R.; Erb, T. J.; Winkler, C. K.; Tinzl, M.; Glueck, S. M. Enzymatic conversion of CO<sub>2</sub> from natural to artificial utilization. *Chem. Rev.* **2023**, 123, 5702–5754.
- (4) Britt, R. D.; Rao, G.; Tao, L. Bioassembly of complex iron–sulfur enzymes: hydrogenases and nitrogenases. *Nat. Rev. Chem.* **2020**, 4, 542–549.
- (5) Jeong, D.; Selverstone Valentine, J.; Cho, J. Bio-inspired mononuclear nonheme metal peroxo complexes: synthesis, structures and mechanistic studies toward understanding enzymatic reactions. *Coord. Chem. Rev.* **2023**, 480, 215021.
- (6) Reed, C. J.; Lam, Q. N.; Mirts, E. N.; Lu, Y. Molecular understanding of heteronuclear active sites in heme–copper oxidases, nitric oxide reductases, and sulfite reductases through biomimetic modelling. *Chem. Soc. Rev.* **2021**, 50, 2486–2539.
- (7) Amaro-Gahete, J.; Pavliuk, M. V.; Tian, H.; Esquivel, D.; Romero-Salguero, F. J.; Ott, S. Catalytic systems mimicking the [FeFe]-hydrogenase active site for visible-light-driven hydrogen production. *Coord. Chem. Rev.* **2021**, 448, 214172.
- (8) Wang, K.-Y.; Zhang, J.; Hsu, Y.-C.; Lin, H.; Han, Z.; Pang, J.; Yang, Z.; Liang, R.-R.; Shi, W.; Zhou, H.-C. Bioinspired framework catalysts: from enzyme immobilization to biomimetic catalysis. *Chem. Rev.* **2023**, 123, 5347–5420.

- (9) Ji, S.; Jiang, B.; Hao, H.; Chen, Y.; Dong, J.; Mao, Y.; Zhang, Z.; Gao, R.; Chen, W.; Zhang, R.; Liang, Q.; Li, H.; Liu, S.; Wang, Y.; Zhang, Q.; Gu, L.; Duan, D.; Liang, M.; Wang, D.; Yan, X.; Li, Y. Matching the kinetics of natural enzymes with a single-atom iron nanozyme. *Nat. Catal.* **2021**, *4*, 407–417.
- (10) Adam, S. M.; Wijeratne, G. B.; Rogler, P. J.; Diaz, D. E.; Quist, D. A.; Liu, J. J.; Karlin, K. D. Synthetic Fe/Cu complexes: toward understanding heme-copper oxidase structure and function. *Chem. Rev.* **2018**, *118*, 10840–11022.
- (11) Helm, M. L.; Stewart, M. P.; Bullock, R. M.; DuBois, M. R.; DuBois, D. L. A Synthetic Nickel Electrocatalyst with a Turnover Frequency Above 100,000 s<sup>-1</sup> for H<sub>2</sub> Production. *Science* **2011**, *333*, 863–866.
- (12) Bediako, D. K.; Solis, B. H.; Dogutan, D. K.; Roubelakis, M. M.; Maher, A. G.; Lee, C. H.; Chambers, M. B.; Hammes-Schiffer, S.; Nocera, D. G. Role of pendant proton relays and proton-coupled electron transfer on the hydrogen evolution reaction by nickel heme-like porphyrins. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 15001–15006.
- (13) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555.
- (14) Mai, H.; Le, T. C.; Chen, D.; Winkler, D. A.; Caruso, R. A. Machine learning for electrocatalyst and photocatalyst design and discovery. *Chem. Rev.* **2022**, *122*, 13478–13515.
- (15) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Big-data science in porous materials: materials genomics and machine learning. *Chem. Rev.* **2020**, *120*, 8066–8129.
- (16) Zhang, W.; Yu, M.; Liu, T.; Cong, M.; Liu, X.; Yang, H.; Bai, Y.; Zhu, Q.; Zhang, S.; Gu, H.; Wu, X.; Zhang, Z.; Wu, Y.; Tian, H.; Li, X.; Zhu, W.-H.; Cooper, A. I. Accelerated discovery of molecular nanojunction photocatalysts for hydrogen evolution by using automated screening and flow synthesis. *Nat. Synth.* **2024**, *3*, 595–605.
- (17) Boiko, D. A.; MacKnight, R.; Kline, B.; Gomes, G. Autonomous chemical research with large language models. *Nature* **2023**, *624*, 570–578.
- (18) Zheng, Z.; Rampal, N.; Inizan, T. J.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. Large language models for reticular chemistry. *Nat. Rev. Mater.* **2025**, *10*, 369–381.
- (19) Bran, A.; Cox, S.; Schilter, O.; Baldassari, C.; White, A. D.; Schwaller, P. Augmenting large language models with chemistry tools. *Nat. Mach. Intell.* **2024**, *6*, 525–535.
- (20) Jablonka, K. M.; Schwaller, P.; Ortega-Guerrero, A.; Smit, B. Leveraging large language models for predictive chemistry. *Nat. Mach. Intell.* **2024**, *6*, 161–169.
- (21) Darvish, K.; Skreta, M.; Zhao, Y.; Yoshikawa, N.; Som, S.; Bogdanovic, M.; Cao, Y.; Hao, H.; Xu, H.; Aspuru-Guzik, A.; Garg, A.; Shkurti, F. ORGANA: a robotic assistant for automated chemistry experimentation and characterization. *Matter* **2025**, *8*, 101897.
- (22) OpenAI. GPT-4 technical report. *arXiv* **2024**, arXiv:2303.08774.
- (23) Polak, M. P.; Morgan, D. Extracting accurate materials data from research papers with conversational language models and prompt engineering. *Nat. Commun.* **2024**, *15*, 1569.
- (24) Ruan, Y.; Lu, C.; Xu, N.; He, Y.; Chen, Y.; Zhang, J.; Xuan, J.; Pan, J.; Fang, Q.; Gao, H.; Shen, X.; Ye, N.; Zhang, Q.; Mo, Y. An automatic end-to-end chemical synthesis development platform powered by large language models. *Nat. Commun.* **2024**, *15*, 10160.
- (25) Song, T.; Luo, M.; Zhang, X.; Chen, L.; Huang, Y.; Cao, J.; Zhu, Q.; Liu, D.; Zhang, B.; Zou, G.; Zhang, G.; Zhang, F.; Shang, W.; Fu, Y.; Jiang, J.; Luo, Y. A multiagent-driven robotic AI chemist enabling autonomous chemical research on demand. *J. Am. Chem. Soc.* **2025**, *147*, 12534–12545.
- (26) Zheng, Z.; Zhang, O.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. ChatGPT chemistry assistant for text mining and the prediction of MOF synthesis. *J. Am. Chem. Soc.* **2023**, *145*, 18048–18062.
- (27) Hocky, G. M. Connecting molecular properties with plain language. *Nat. Mach. Intell.* **2024**, *6*, 249–250.
- (28) Yang, H.; Che, Y.; Cooper, A. I.; Chen, L.; Li, X. Machine learning accelerated exploration of ternary organic heterojunction photocatalysts for sacrificial hydrogen evolution. *J. Am. Chem. Soc.* **2023**, *145*, 27038–27044.
- (29) Fujishima, A.; Honda, K. Electrochemical photolysis of water at a semiconductor electrode. *Nature* **1972**, *238*, 37–38.
- (30) Navalón, S.; Dhakshinamoorthy, A.; Álvaro, M.; Ferrer, B.; García, H. Metal organic frameworks as photocatalysts for solar-driven overall water splitting. *Chem. Rev.* **2023**, *123*, 445–490.
- (31) Hisatomi, T.; Yamada, T.; Nishiyama, H.; Takata, T.; Domen, K. Materials and systems for large-scale photocatalytic water splitting. *Nat. Rev. Chem.* **2025**, *10*, 769–782.
- (32) Feng, C.; Raziq, F.; Huang, H.; Wu, Z.-P.; Alqahtani, H. S.; Alqahtani, R.; Rahman, M. Z.; Chang, B.; Gascon, J.; Zhang, H. Shining light on hydrogen: solar-powered catalysis with transition metals. *Adv. Mater.* **2025**, *37*, 2410387.
- (33) Tao, X.; Zhao, Y.; Wang, S.; Li, C.; Li, R. Recent advances and perspectives for solar-driven water splitting using particulate photocatalysts. *Chem. Soc. Rev.* **2022**, *51*, 3561–3608.
- (34) Wang, Q.; Domen, K. Particulate photocatalysts for light-driven water splitting: mechanisms, challenges, and design strategies. *Chem. Rev.* **2020**, *120*, 919–985.
- (35) Guo, Z.; Xia, L.; Yu, Y.; Ao, T.; Huang, C. LightRAG: simple and fast retrieval-augmented generation. *arXiv* **2025**, arXiv:2410.05779.
- (36) Senger, M.; Eichmann, V.; Laun, K.; Duan, J.; Wittkamp, F.; Knör, G.; Apfel, U.-P.; Happe, T.; Winkler, M.; Heberle, J.; Stripp, S. T. How [FeFe]-hydrogenase facilitates bidirectional proton transfer. *J. Am. Chem. Soc.* **2019**, *141*, 17394–17403.
- (37) Stripp, S. T.; Duffus, B. R.; Fourmond, V.; Léger, C.; Leimkühler, S.; Hirota, S.; Hu, Y.; Jasnowski, A.; Ogata, H.; Ribbe, M. W. Second and outer coordination sphere effects in nitrogenase, hydrogenase, formate dehydrogenase, and CO dehydrogenase. *Chem. Rev.* **2022**, *122*, 11900–11973.
- (38) Cabotaje, P. R.; Walter, K.; Zamader, A.; Huang, P.; Ho, F.; Land, H.; Senger, M.; Berggren, G. Probing substrate transport effects on enzymatic hydrogen catalysis: an alternative proton transfer pathway in putatively sensory [FeFe] hydrogenase. *ACS Catal.* **2023**, *13*, 10435–10446.
- (39) Birrell, J. A.; Rodríguez-Maciá, P.; Reijerse, E. J.; Martini, M. A.; Lubitz, W. The catalytic cycle of [FeFe] hydrogenase: a tale of two sites. *Coord. Chem. Rev.* **2021**, *449*, 214191.
- (40) Reijerse, E. J.; Pham, C. C.; Pelmentschikov, V.; Gilbert-Wilson, R.; Adamska-Venkatesh, A.; Siebel, J. F.; Gee, L. B.; Yoda, Y.; Tamasaku, K.; Lubitz, W.; Rauchfuss, T. B.; Cramer, S. P. Direct observation of an iron-bound terminal hydride in [FeFe]-hydrogenase by nuclear resonance vibrational spectroscopy. *J. Am. Chem. Soc.* **2017**, *139*, 4306–4309.
- (41) Sun, K.; Huang, Y.; Sun, F.; Wang, Q.; Zhou, Y.; Wang, J.; Zhang, Q.; Zheng, X.; Fan, F.; Luo, Y.; Jiang, J.; Jiang, H.-L. Dynamic structural twist in metal–organic frameworks enhances solar overall water splitting. *Nat. Chem.* **2024**, *16*, 1638–1646.
- (42) Liu, Y.; Yang, W.; Chen, Q.; Cullen, D. A.; Xie, Z.; Lian, T. Pt particle size affects both the charge separation and water reduction efficiencies of CdS–Pt nanorod photocatalysts for light driven H<sub>2</sub> generation. *J. Am. Chem. Soc.* **2022**, *144*, 2705–2715.
- (43) Rahman, M. Z.; Raziq, F.; Zhang, H.; Gascon, J. Key strategies for enhancing H<sub>2</sub> production in transition metal oxide based photocatalysts. *Angew. Chem., Int. Ed.* **2023**, *62*, No. e202305385.
- (44) Zhang, J.; Feng, P.; Bu, X.; Wu, T. Atomically precise metal chalcogenide supertetrahedral clusters: frameworks to molecules, and structure to function. *Natl. Sci. Rev.* **2022**, *9*, nwab076.
- (45) Kosco, J.; Gonzalez-Carrero, S.; Howells, C. T.; Fei, T.; Dong, Y.; Sougrat, R.; Harrison, G. T.; Firdaus, Y.; Sheelamantula, R.; Purushothaman, B.; Moruzzi, F.; Xu, W.; Zhao, L.; Basu, A.; De Wolf, S.; Anthopoulos, T. D.; Durrant, J. R.; McCulloch, I. Generation of long-lived charges in organic semiconductor heterojunction nanoparticles for efficient photocatalytic hydrogen evolution. *Nat. Energy* **2022**, *7*, 340–351.
- (46) Ma, D.; Zhang, Z.; Zou, Y.; Chen, J.; Shi, J.-W. The Progress of g-C<sub>3</sub>N<sub>4</sub> in photocatalytic H<sub>2</sub> evolution: from fabrication to modification. *Coord. Chem. Rev.* **2024**, *500*, 215489.

- (47) Côté, A. P.; Benin, A. I.; Ockwig, N. W.; O'Keeffe, M.; Matzger, A. J.; Yaghi, O. M. Porous, crystalline, covalent organic frameworks. *Science* **2005**, *310*, 1166–1170.
- (48) Luo, T.; Gilmanova, L.; Kaskel, S. Advances of MOFs and COFs for photocatalytic CO<sub>2</sub> reduction, H<sub>2</sub> evolution and organic redox transformations. *Coord. Chem. Rev.* **2023**, *490*, 215210.
- (49) Chen, Y.; Jiang, D. Photocatalysis with covalent organic frameworks. *Acc. Chem. Res.* **2024**, *57*, 3182–3193.
- (50) Kong, X.; Deng, H.; Yan, F.; Kim, J.; Swisher, J. A.; Smit, B.; Yaghi, O. M.; Reimer, J. A. Mapping of functional groups in metal-organic frameworks. *Science* **2013**, *341*, 882–885.
- (51) Yang, G.; Wang, D.; Wang, Y.; Hu, W.; Hu, S.; Jiang, J.; Huang, J.; Jiang, H.-L. Modulating the primary and secondary coordination spheres of single Ni(II) sites in metal–organic frameworks for boosting photocatalysis. *J. Am. Chem. Soc.* **2024**, *146*, 10798–10805.
- (52) Stanley, P. M.; Haimerl, J.; Shustova, N. B.; Fischer, R. A.; Warnan, J. Merging molecular catalysts and metal–organic frameworks for photocatalytic fuel production. *Nat. Chem.* **2022**, *14*, 1342–1356.