

Can We Automate Scientific Reasoning in Closed-Loop Experiments using Large Language Models?

Abdoulatif Cissé, Max E. Cooper, Mengjia Zhu,
Xenophon Evangelopoulos and Andrew I. Cooper*

Materials Innovation Factory and Department of Chemistry,
University of Liverpool. Liverpool, L7 3NY, United Kingdom

Email: aicooper@liverpool.ac.uk

Supporting Information

Contents

1. Asymptotic Convergence of BORA to BO (pg. 2)
2. Opsight Software Platform and BORA Implementation (pg. 3)
3. Supporting Figures and Tables (pp. 4–31)

1. Asymptotic Convergence of BORA to BO

The following theorem proves the asymptotic convergence of the BORA algorithm (ref. 71, main text).

Theorem 1. Let $GP(0, k)$ be a fixed Gaussian process prior over a Lipschitz continuous function $f: \mathcal{X} \rightarrow \mathbb{R}$ defined on a compact set $\mathcal{X} \subset \mathbb{R}^d$, with a continuous and bounded kernel $k(\mathbf{x}, \mathbf{x}) = \sigma_t^{GP}(\mathbf{x}) > 0$, potentially with i.i.d observation noise $\epsilon \sim \mathcal{N}(0, \eta)$. When a GP surrogate with an expected improvement (EI) acquisition strategy are used in the BORA optimiser, then with probability 1, there is a finite step T' such that:

$$\sigma_{t,mean}^{GP} < \beta \sigma_{t,max}^{GP} \quad \forall t \geq T', \quad (1)$$

where σ_t^{GP} represents the posterior standard deviation of the GP at step t , and $\beta \in (0, 1)$.

Proof. Let $\{\mathbf{x}_{mon}^{(1)}, \mathbf{x}_{mon}^{(2)}, \dots, \mathbf{x}_{mon}^{(q)}\} \subset \mathcal{X}$ be a fixed set of q ‘monitoring points’ of uncertainty in the GP in the Bayesian optimisation framework. According to the BORA algorithm (ref. 71, main text) we have:

$$\sigma_{t,mean}^{GP} = 1/q \sum_{i=1}^q \sigma_t(\mathbf{x}_{mon}^{(i)}) \quad (2)$$

$$\sigma_{t,max}^{GP} = \max\left(\sigma_{t-1,max}^{GP}, \max_{1 \leq i \leq q} \sigma_t(\mathbf{x}_{mon}^{(i)})\right), \quad (3)$$

$$\text{where } \sigma_{0,max}^{GP} = \max_{1 \leq i \leq q} \sigma_t(\mathbf{x}_{mon}^{(i)}). \quad (4)$$

By the no-empty-ball (NEB) property of k (valid for a large class of kernels – Matérn, exponential, etc.)¹ we have that the solution sequence $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(T)}\}$ generated by a sufficiently large number of actions a_1 in BORA (i.e., ‘vanilla’ BO using a GP and EI) is dense in \mathcal{X} and thus no persistently high-uncertainty region is ignored with probability 1. Subsequently, and due to epistemic uncertainty, which ensures $\|\mathbf{x}^{(t)} - \mathbf{x}_{mon}^{(t')}\| < \epsilon \implies |\sigma_t^{GP}(\mathbf{x}^{(t)}) - \sigma_t^{GP}(\mathbf{x}_{mon}^{(t')})| = 0$ we get

$$\sigma_{t,mean}^{GP} \rightarrow 0. \quad (5)$$

From Eq. (3) we have that $\sigma_{t,max}^{GP}$ is upper bounded by a finite value $\sigma_{t,max}^{GP} \leq M^* < \infty$ and from Eq. (5) we get

$$\sigma_{t,mean}^{GP} / \sigma_{t,max}^{GP} \rightarrow 0. \quad (6)$$

Therefore, there exists a finite iteration $T' > T$ for which

$$\sigma_{t,mean}^{GP} < \beta \sigma_{t,max}^{GP} \quad \forall t \geq T'. \quad (7)$$

Corollary 1. Given that $\sigma_{t,mean}^{GP} < \beta \sigma_{t,max}^{GP} \quad \forall t \geq T'$, the BORA optimisation framework¹ reduces to a ‘vanilla’ Bayesian optimiser for the remaining iterations, thus inheriting its asymptotic convergence guarantees.^{1,2}

Remark 1. While the previous result ensures convergence for a fixed choice of prior hyperparameters, in practice these are estimated from data using maximum likelihood estimates which can cause EI to fail to converge³. Following a modified EI acquisition strategy by Bull³ ensures the desired convergence with estimated prior parameters.

Remark 2. In the presence of i.i.d observation noise $\epsilon \sim \mathcal{N}(0, \eta)$, Theorem 1 still holds with the only difference that $\sigma_{t,mean}^{GP} \rightarrow \sqrt{\eta}$ instead.

References (section 1)

1. Vazquez, E. and Bect, J., *Journal of Statistical Planning and Inference*, **2010**, 140(11), 3088.
2. Bull, A.D., *Journal of Machine Learning Research*, **2011**, 12(10).

2. Opsight Software Platform and BORA Implementation

2.1. Overview and Architecture

This study's optimisation campaigns were executed using Opsight (<https://opsightlab.com/>), a cloud-native platform we developed to deploy hypothesis-driven optimizers such as BORA (ref. 71) and HypBO (ref. 50). Opsight employs a three-layer architecture separating user interaction (Django web application),³ optimisation logic (BoTorch,⁴ LangGraph agentic framework⁵), and distributed execution (Azure Container Apps⁶). This separation enabled the large-scale benchmarking studies presented here, where for each experimental configuration, we ran 10–20 independent optimisation campaigns concurrently while maintaining full reasoning traces for post-hoc analysis. The platform's distributed architecture, automatic horizontal scaling, and detailed logging capabilities allowed us to capture all optimisation runs with complete LLM commentary, tool usage, and hypothesis evolution. The web-based interface provided real-time visualisation and collaborative monitoring accessible to the entire research team, which proved valuable for identifying outlier runs and understanding failure modes during this study, which would have been cumbersome to replicate with traditional scripting approaches in a multi-user study such as this.

2.2. BORA Implementation within Opsight

The BORA optimiser deployed here leverages the core algorithmic framework from our original publication (ref. 71), where LLM reasoning guides Bayesian optimisation through plateau-aware interventions. Operationalising BORA at the scale required here (624 LLM optimisations in total) demanded production-grade infrastructure that Opsight provided via an agentic architecture with autonomous tool ecosystem and persistent memory management. While the original BORA implementation demonstrated viability through simple LLM prompt engineering, Opsight wraps this with these more enhanced capabilities, allowing deployment in campaigns where domain knowledge evolves and a plethora of literature is available.

Concretely, Opsight's infrastructure additions to BORA (and subsequently to LLM-only optimisations, sections 3.2 and 3.4 in main text) include: (i) agentic framework with autonomous tool calling, where the LLM can invoke web search for literature scanning, Python code execution for data analysis (*e.g.*, correlation matrices), resource retrieval for human input, and memory operations without human oversight; (ii) RAG-based memory integration through a vector database that allows the LLM to query past hypotheses, experimental outcomes, and memories. The system uses OpenAI's text-embedding-3-large or Google's gemini-embedding-exp-03-07 embeddings (depending on LLM provider) with each optimisation campaign assigned a dedicated namespace for knowledge isolation.

References (section 2)

3. <https://www.djangoproject.com>
4. M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In *Advances in Neural Information Processing Systems* 33, 2020.
5. <https://github.com/langchain-ai/langgraph>
6. <https://pypi.org/project/azure-mgmt-appcontainers/>

3. Supporting Figures and Tables

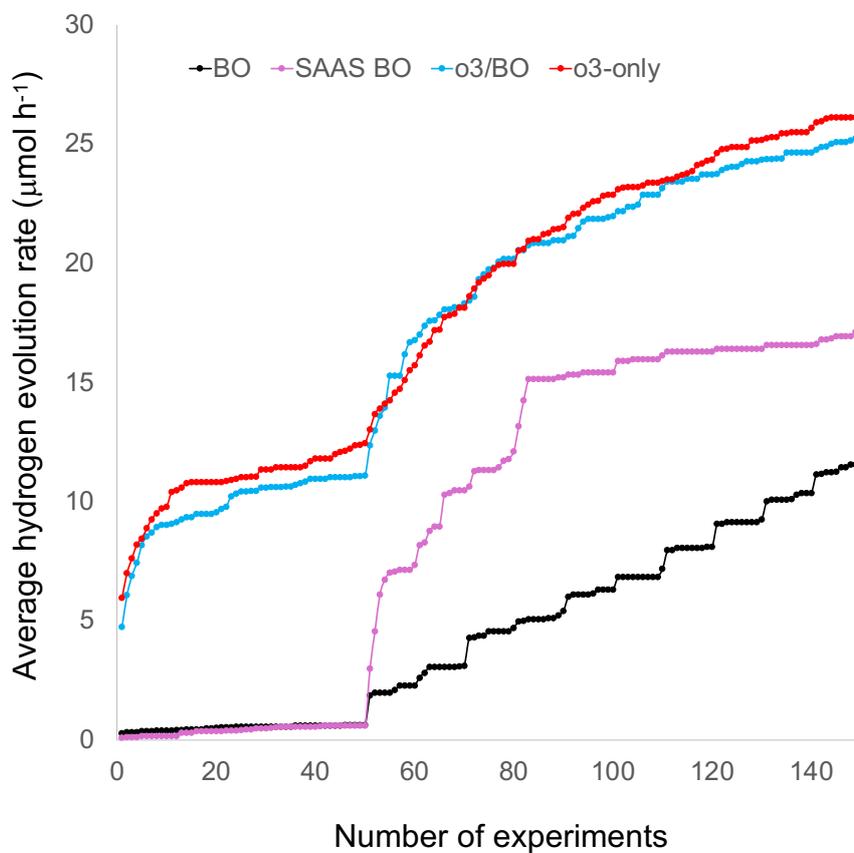


Fig. S1. Comparison of LLM-only optimisation (o3), hybrid o3 (med) / BO optimisation (BORA), and uncategoryed BO and SAAS BO for a 10-dimensional photocatalysis problem. Each run involved 15 batches of 10 experiments (150 experiments). There were 5 batches (50 experiments) prior to adaptive sampling, and a total of 20 repeat optimisations was run in each case. The plots show the average best hydrogen evolution rate (HER) value found so far. The maximum possible HER in this test is $28.37 \mu\text{mol h}^{-1}$.

Do not check any literature throughout this optimisation.

That is, do NOT call on Google or other search tools to find literature relating to the photocatalysis experiment.

This overrides any other instructions.

Fig. S2. Prompt (attached PDF) used to instruct the LLM not to carry out literature searches in hybrid gpt-5 (med) / BO optimisations (see also Fig. S3).

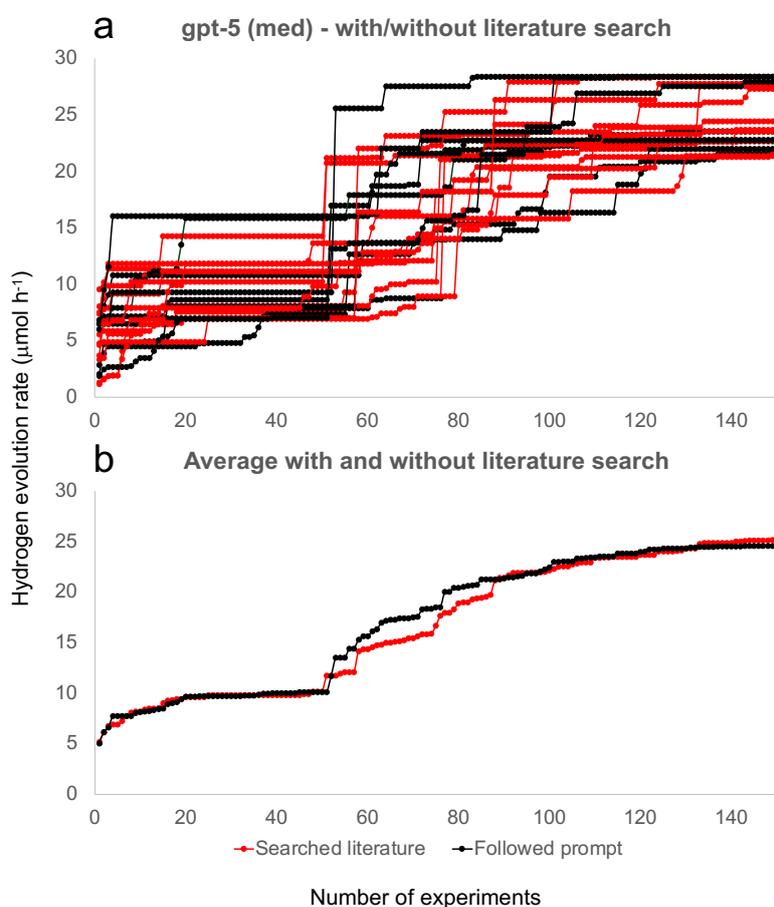


Fig. S3. Results of instructing the LLM not to carry out literature searches in hybrid gpt-5 (med) / BO optimisations (prompt shown in Fig. S2). Each run involved 15 batches of 10 experiments (150 experiments). There were 5 batches prior to adaptive sampling in each case, and a total of 20 repeat optimisations was run. (a) Plot showing all 20 runs; 10/20 of these runs followed the instruction through to the end of the optimisation (black points). (b) Plots showing the average best hydrogen evolution rate (HER) value found so far for the runs that followed the prompt (black points) and those that failed to do this (red points). Within standard deviation, there is little difference between the two cases.

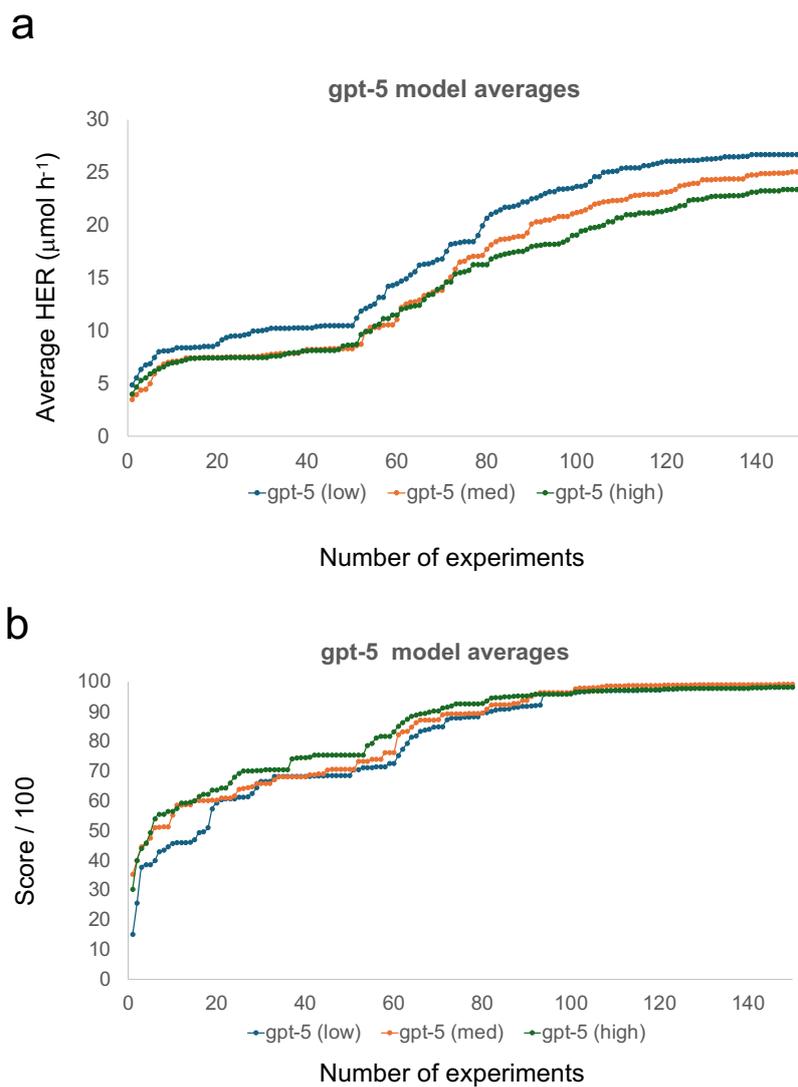


Fig. S4. Effect of reasoning level settings (low/medium/high) in hybrid gpt-5 / BO optimisations (BORA) for (a) the photocatalysis experiment and (b) the pétanque physics simulation. Each run involved 15 batches of 10 experiments (150 experiments). There were 5 batches prior to adaptive sampling in each case, and a total of 20 repeat optimisations was run.

Model	Average HER after 25 expt	STD after 25 expt	Average HER after 150 expt	Average STD after 150 expt	# times reached max
gpt-5 (low)	9.5	2.9	26.7	2.3	9
gpt-5 (med)	7.5	3.5	25.1	4.2	9
gpt-5 (high)	7.5	4.2	23.4	5.6	1

Table S1. Effect of reasoning level settings (low/medium/high) in hybrid gpt-5 / BO optimisations for the photocatalysis experiment. Each run involved 15 batches of 10 experiments (150 experiments). There were 5 batches prior to adaptive sampling in each case, and a total of 20 repeat optimisations was run.

Model	Average score after 25 expt	STD after 25 expt	Average score after 150 expt	Average STD after 150 expt
gpt-5 (low)	61.3	28.8	98.2	2.3
gpt-5 (med)	63.9	35.1	99.2	0.9
gpt-5 (high)	69.3	26.4	98.2	4.6

Table S2. Effect of reasoning level settings (low/medium/high) in hybrid gpt-5 / BO optimisations for the pétanque physics simulation. Each run involved 15 batches of 10 experiments (150 experiments). There were 5 batches prior to adaptive sampling in each case, and a total of 20 repeat optimisations was run.

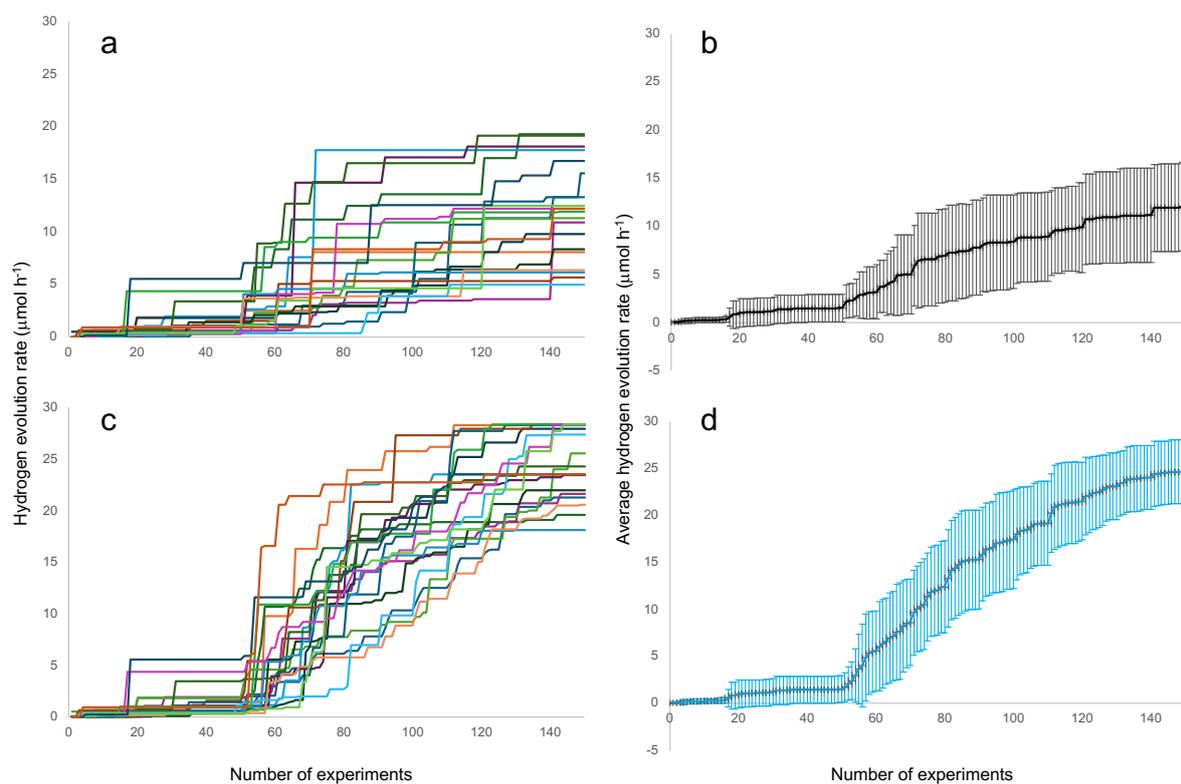


Fig. S5. Comparison of BO (a, b) and LLM-only optimisations (o3, med) (c, d) for the photocatalysis experiment. Each run involved 15 batches of 10 experiments (150 experiments). There were five batches prior to adaptive sampling, identical in all runs; these were selected using the Kennard–Stone algorithm to ensure maximum dissimilarity among the conditions. A total of 20 repeat optimisations was run. (a) and (c) show all individual runs; (b) and (d) show the average over 20 runs with standard deviations.

	Best score	pitch	yaw	v0	spin	spin axis	h0	mass
Run 1	99.773	27.076	0	24.641	0	0	2	9
Run 2	96.166	28.05	0	31.95	1500	88.56	1.5	2
Run 3	98.85	24.5	0	42.1	1790	90.6	2	0.6
Run 4	98.397	20.115	0	27.115	0	0	2	10
Run 5	99.919	24.5	0	27.1	0	0	2	0.5
Run 6	99.44	38.086	0	24.004	100	90	1.05	2.005
Run 7	98.752	32.675	0	23.405	0	0	2	10
Run 8	98.799	30.467	0	23.806	0	0	2	10
Run 9	98.977	22.105	0	26.365	0	-90	2	7.15
Run 10	99.75	31.882	0	23.922	0	0	1.001	10
Run 11	99.085	35.46	-0.1	35.46	1495	94.95	2	2.05
Run 12	99.967	39.135	0	22.747	0	0	2	10
Run 13	99.558	20.064	0	25.001	800	-90	1.481	6
Run 14	99.831	24.021	0	25.546	0	90	2	10
Run 15	99.507	33.834	0	21.787	790	-90	2	2
Run 16	99.939	27.68	0	31.32	1895.8	-90	2	0.1
Run 17	99.981	26.001	0	25.191	0	-90	1.424	10
Run 18	99.94	33.244	0	23.998	0	0	2	2
Run 19	99.764	38.013	0	22.996	120	90	2	10
Run 20	98.537	34.17	0	23.08	3000	-90	1.5	0.9

Table S3. Best solutions found for 20 hybrid LLM/BO optimisations for the 7-dimensional pétanque simulation using gpt-5, settings as per Fig. 3 in main text (15 batches of 10 experiments; the first 5 batches used the LLM to warm start the optimisation, after which the hybrid optimiser switched between LLM and BO modes). Note that no two solutions found are identical. Except for the “yaw” dimension (left/right angle in horizontal plane), which was mostly ignored, there is a wide spread of values for each of the other six input variables. This supports the interpretation that the LLM is performing active reasoning, rather than simply “memorising” a solution extrapolated from some related literature source. Note that some of the recurrent values represent maximum or minimum allowed values; for example, the maximum ball mass was 10 kg, and the maximum release height above ground level (h0) was 2 m.

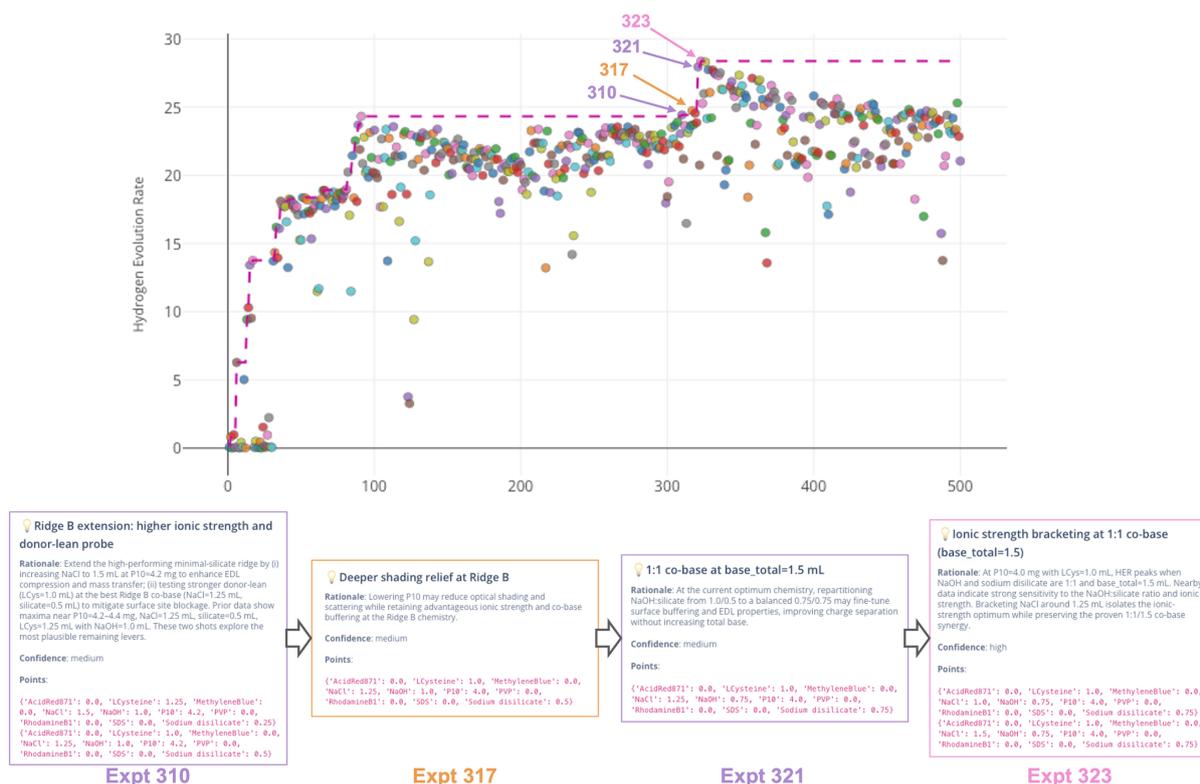
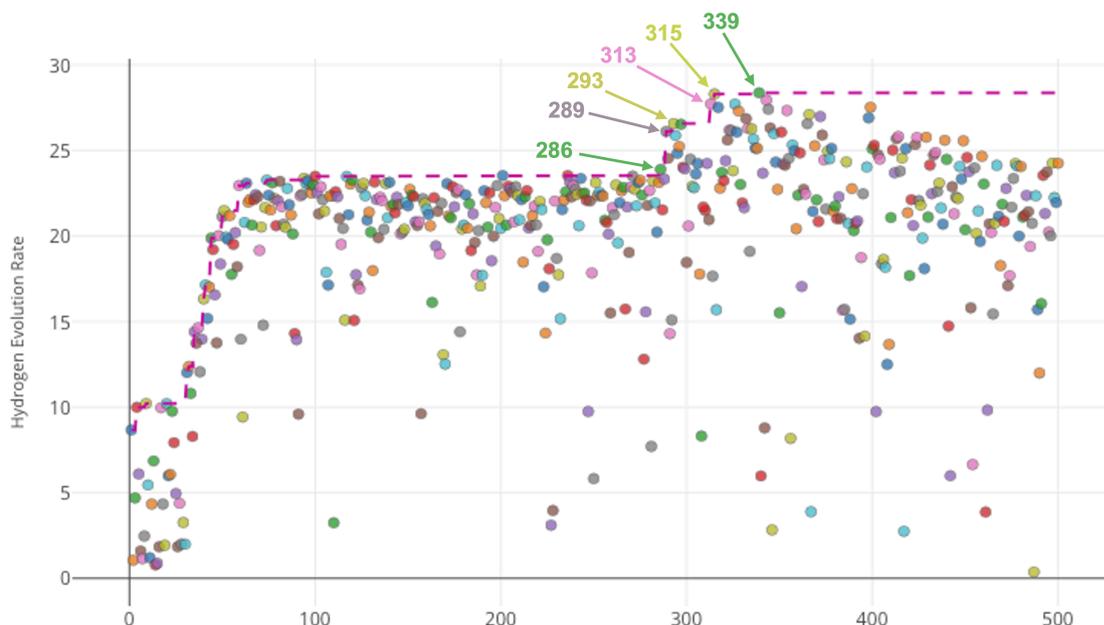


Fig. S6. LLM-only optimisation (gpt-5(med), batch size = 1; initialisation phase = 30 experiments; 500 experiments in all). The optimum HER ($28.37 \mu\text{mol h}^{-1}$) is reached at experiment 323 after a prolonged plateau. The hypotheses for the four labelled experiments that break the plateau are also shown; the full optimisation log is also included in the data repository (Fig_S6_log.pdf).



💡 **4.4 mg P10, lean donor & mild base**

Rationale: Combine current catalyst loading with reduced donor (1 mL) and slightly lower NaOH (0.75 mL) to test if over-alkalinity limits HER.

Confidence: medium

Points:

```
{'AcidRed871': 0.0, 'LCysteine': 1.0,
'MethyleneBlue': 0.0, 'NaCl': 1.25, 'NaOH': 0.75,
'P10': 4.4, 'PVP': 0.0, 'RhodamineB1': 0.0, 'SDS':
0.0, 'Sodium disilicate': 0.25}
```

Expt 286

💡 **Champion + High Silicate**

Rationale: Add 0.25 mL extra sodium disilicate to improve buffering and surface complexation while keeping donor/base optimal.

Confidence: medium

Points:

```
{'AcidRed871': 0.0, 'LCysteine': 1.0,
'MethyleneBlue': 0.0, 'NaCl': 1.25, 'NaOH': 0.75,
'P10': 4.4, 'PVP': 0.0, 'RhodamineB1': 0.0, 'SDS':
0.0, 'Sodium disilicate': 0.5}
```

Expt 289

💡 **High Silicate 0.75**

Rationale: Silicate has shown monotonic benefit up to 0.50 mL; testing 0.75 mL with optimal base may further buffer holes and disperse P10.

Confidence: medium

Points:

```
{'AcidRed871': 0.0, 'LCysteine': 1.0,
'MethyleneBlue': 0.0, 'NaCl': 1.25, 'NaOH': 0.75,
'P10': 4.4, 'PVP': 0.0, 'RhodamineB1': 0.0, 'SDS':
0.0, 'Sodium disilicate': 0.75}
```

Expt 293

💡 **Lower P10, same additives**

Rationale: Reducing catalyst from 4.4 → 4.2 mg may cut scattering and improve photon absorption without sacrificing active sites.

Confidence: medium

Points:

```
{'AcidRed871': 0.0, 'LCysteine': 1.0,
'MethyleneBlue': 0.0, 'NaCl': 1.0, 'NaOH': 0.75,
'P10': 4.2, 'PVP': 0.0, 'RhodamineB1': 0.0, 'SDS':
0.0, 'Sodium disilicate': 0.75}
```

Expt 313

💡 **Lower P10**

Rationale: Reducing P10 below current best may alleviate self-shielding and improve photon utilisation.

Confidence: medium

Points:

```
{'AcidRed871': 0.0, 'LCysteine': 1.0,
'MethyleneBlue': 0.0, 'NaCl': 1.0, 'NaOH': 0.75,
'P10': 3.8, 'PVP': 0.0, 'RhodamineB1': 0.0, 'SDS':
0.0, 'Sodium disilicate': 0.75}
```

Expt 315

💡 **Higher P10 4.0**

Rationale: Catalyst loading increase has been beneficial up to 3.8 mg; 4.0 mg may further raise HER.

Confidence: medium

Points:

```
{'AcidRed871': 0.0, 'LCysteine': 1.0,
'MethyleneBlue': 0.0, 'NaCl': 1.0, 'NaOH': 0.75,
'P10': 4.0, 'PVP': 0.0, 'RhodamineB1': 0.0, 'SDS':
0.0, 'Sodium disilicate': 0.75}
```

Expt 339

Fig. S7. LLM-only optimisation (o3(med), batch size = 1; initialisation phase = 30 experiments; 500 experiments in all). The optimum HER ($28.37 \mu\text{mol h}^{-1}$) is reached at experiment 339, again after a plateau. The hypotheses for the six experiments that break the plateau are shown; the full optimisation log is also included in the data repository (Fig_S7_log.pdf). The decline in performance after experiment 339 suggests a potential termination condition for the optimiser, although we note that such a condition might also have terminated the optimisation prematurely in the example given in Fig. S6 (see performance decline in experiments 100–200).

Team Ideas 2

Our research team has discussed this problem just now and we've got the following thoughts.

1. Addition of dyes will decrease hydrogen production
2. Addition of surfactants will increase hydrogen production
3. The amount of P10 added should be high (>4 mg)
4. The amount of cysteine added should be low to zero
5. We don't know what to think about NaCl or bases

Please note, we are not sure about all these hypotheses, but reasonably confident in Hypothesis 3.

Existing constraints such as keeping the total reaction volume equal to 5 mL must still be obeyed.

Team Ideas 3

Our research team has discussed this problem, and we have come to the following conclusions:

1. Do not add any dyes to any batches; these will decrease hydrogen production
2. Do not add any surfactants to any batches; these will decrease hydrogen production
3. The amount of P10 added should always be high (>4 mg)
3. The amount of cysteine added should never be lower than 0.75 mL
4. There should always be a base present (NaOH or sodium disilicate), and preferably both bases should be present at the same time
5. We are not sure about the role of NaCl – could be either good or bad for hydrogen production – but more likely, a small amount of NaCl will increase ionic strength and be good.

We are quite confident about these hypotheses. Existing constraints such as keeping the total reaction volume equal to 5 mL must still be obeyed.

Fig. S8. The 'mixed' hypotheses (Team Ideas 2; 1 = good; 3 = good (though not optimal), 2, 4 = bad, 5 = neutral) and 'good' hypotheses (Team Ideas 3) used to prompt the hybrid LLM (o3) / BO optimiser (section 3.3.1 in main text; see Fig. 10, Table 5). Upon analysis, we noted that both the 'mixed' and 'good' prompts specify ">4 mg" P10. In fact, the optimal amount for maximum HER in the model is 4 mg. As such, while Team Ideas 3 is broadly a helpful prompt, it is not perfect; we changed this to ">3.5 mg" (Fig. S14) for the test shown in Figs. S15 & S16.

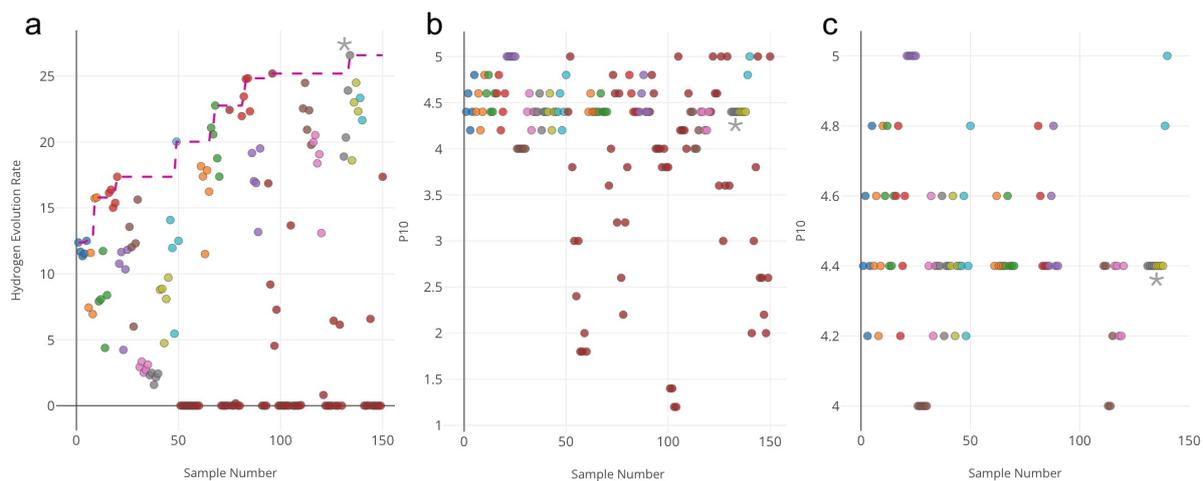


Fig. S9. Example of BO/o3(med) hybrid optimisation (batch size = 10, 15 batches, 5 batch warm start) with ‘good’ hypotheses prompt (PC_20_run_10_5_15_o3_prompt_3_Run_5.pdf). The best HER located was $26.58 \mu\text{mol h}^{-1}$ (asterisked point), below the maximum value of $28.37 \mu\text{mol h}^{-1}$. (a) HER versus experiment number; (b) Amount of P10 dispensed (mg) versus experiment number showing both LLM- and BO-derived points. (c) Amount of P10 dispensed (mg) versus experiment number showing LLM-derived points only (that is, BO points not shown). For the LLM points, (c), most experiments involved a P10 addition of >4 mg, in line with the “good” hypotheses (Fig. S8); this was the general trend throughout the 20 repeat optimisation trials. In this case, BO sampling does not manage to break the hybrid optimiser out of this regime.

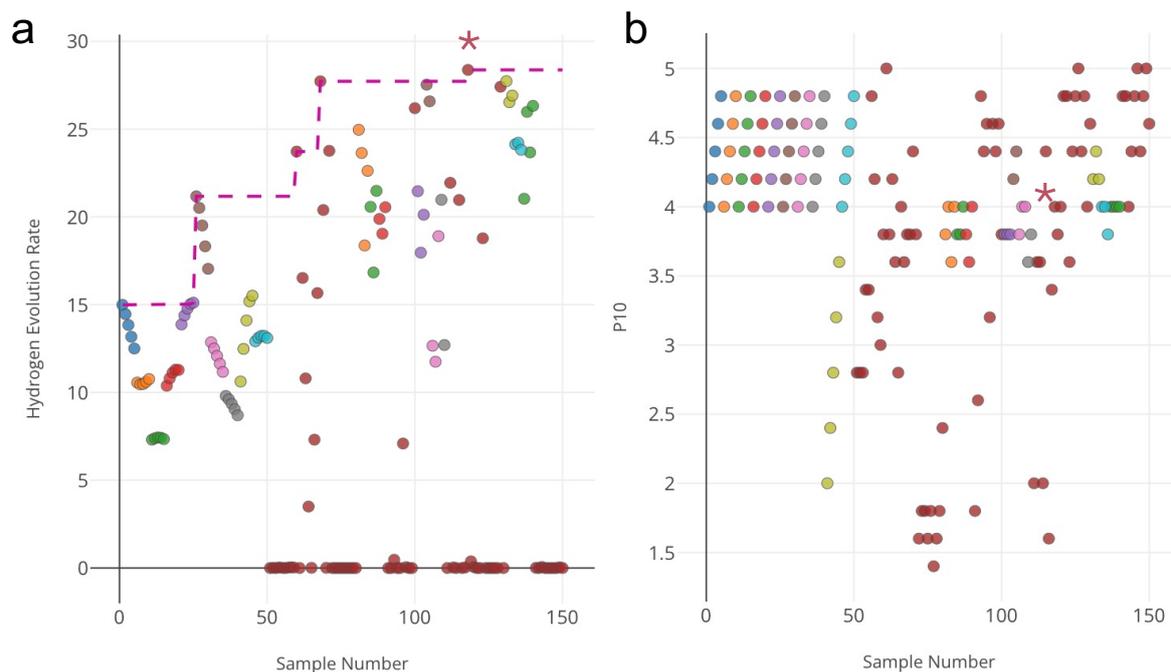


Fig. S10. Second example of BO/o3(med) hybrid optimisation (batch size = 10, 15 batches, 5 batch warm start) with “good” hypotheses prompt (PC_20_run_10_5_15_o3_prompt_3_Run_2.pdf). Here the optimiser located the maximum value of $28.37 \mu\text{mol h}^{-1}$ (asterisked point). (a) HER versus experiment number; (b) Amount of P10 dispensed (mg) versus experiment number showing both LLM- and BO-derived points. For the maximum HER (asterisked), the P10 addition was 4 mg, contravening the >4 mg suggestion in the “good” hypotheses (Fig. S8). Here, BO sampling breaks the hybrid optimiser out of the initial team hypothesis regime. Note also the lines and parabolas of points in the warm start phase (first 50 experiments)—these are effectively systematic mini-experiments suggested by the early LLM hypotheses, in this case sweeping the effect of P10 addition, mostly in the range 4.0–4.8 mg (Fig. S8b) with respect to other variables.

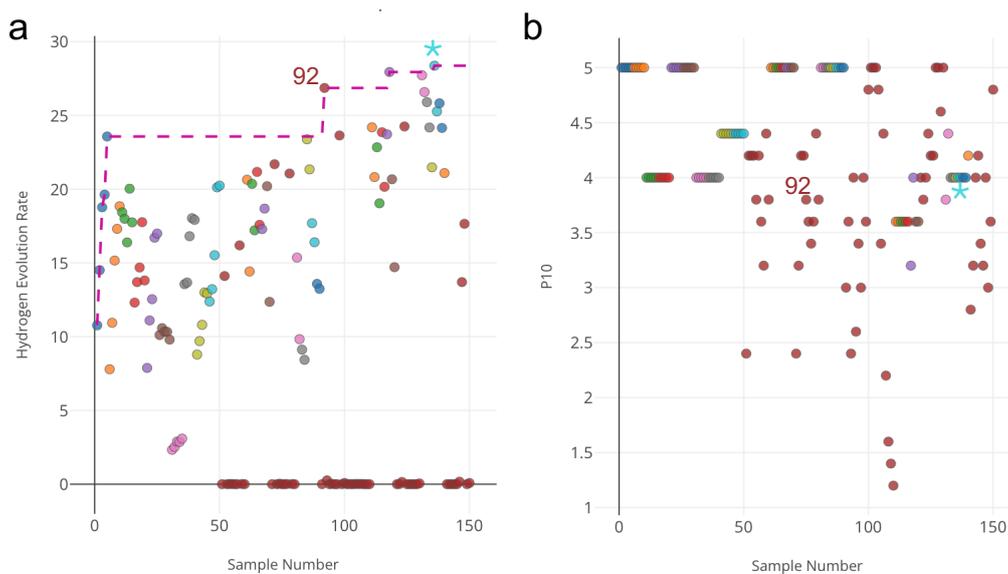


Fig. S11. Third example of BO/o3(med) hybrid optimisation (batch size = 10, 15 batches, 5 batch warm start) with “good” hypotheses prompt (PC_20_run_10_5_15_o3_prompt_3_Run_9.pdf). Again, the optimiser located the maximum value of $28.37 \mu\text{mol h}^{-1}$ (asterisked point). (a) HER versus experiment number; (b) Amount of P10 dispensed (mg) versus experiment number showing both LLM- and BO-derived points. For the maximum HER (asterisked), the P10 addition was 4 mg, contravening the >4 mg suggestion in the “good” hypotheses (Fig. S6). In this case, it is an LLM hypothesis that located this maximum. However, before this occurred, BO sampling (experiment 92, labelled) broke the optimiser out of the >4 mg regime. Excerpts from the reasoning log are shown in Fig. S12.

Hydrogen-evolution in this platform is driven by the conjugated-polymer photocatalyst P10 that absorbs UV/Vis light, while a sacrificial hole scavenger removes photogenerated holes and bases keep pH alkaline. A 5-mL aqueous suspension is irradiated and H₂ is quantified.

KEY FINDINGS

- Literature on P10 and related sulfone-containing polymers shows HER rises monotonically with catalyst loading up to at least 0.1 mg mL⁻¹. For our 5-mL vial this suggests ≥4 mg P10 (team note: ">4 mg").
- L-cysteine is an efficient sacrificial agent that can covalently bind to catalyst or co-catalyst and speed hole removal. Effective ranges reported correspond to ≈0.75–2 mL of our stock per 5 mL total. Team insists ≥0.75 mL.
- Alkaline pH is essential. Either NaOH or sodium-disilicate (silicate also buffers) should be present; the combination often outperforms either alone. Target 1–2 mL of each while respecting the 5-mL cap.
- Ionic strength (NaCl): small doses (<1 mL) can boost charge transport; larger doses reduce activity. Expect a hump-shaped dependence.
- Surfactants & dyes (SDS, PVP, Acid Red 87, Methylene Blue, Rhodamine B): literature plus team memo indicate they compete for light or block sites; routinely lower HER unless used during nanoparticle synthesis. Initial plan is to set them to 0, but allocate a small exploratory budget (≤10 % of points, ≤0.5 mL) for verification.
- Volume constraint: liquid reagents (everything except P10) must sum ≤5 mL. P10 mass does not contribute.

VARIABLE PRIORITISATION (descending importance) P10 > L-cysteine > NaOH & sodium-disilicate > NaCl > SDS/PVP/dyes.

DESIGN GUIDELINES FOR INITIAL BO BATCH • Fix P10 at 4.0 and 5.0 mg across designs. • Ensure every point contains ≥0.75 mL L-cysteine and at least one base (preferably both, 1–2 mL each). Adjust NaCl 0–1 mL. Keep dyes/surfactants at 0 except in a few exploration points. • Use Latin-hypercube within these narrowed ranges to honour volume limit; include a couple of intentionally bad controls (e.g. no scavenger) to help model learning.

10/08/2025, 15:55:12

☺ Data show HER rises with (i) maximal P10 (5 mg), (ii) abundant L-cysteine (≥2 mL) and (iii) a strongly alkaline, well-buffered medium. The current best (HER 23.6) uses NaOH 1.0 mL + sodium-disilicate 1.25 mL, NaCl 0.5 mL, L-cys 2 mL. Trends suggest three tunable levers for further gain: 1) pushing the buffer capacity of silicate beyond 1.25 mL while holding NaOH at 1 mL may raise steady-state pH and improve charge separation; 2) keeping total base at the 2.5–3 mL level but balancing NaOH ≈ disilicate (1.25 mL each) can enhance both alkalinity and buffering; 3) increasing the hole-scavenger above 2 mL could accelerate hole removal; 4) probing both higher and zero NaCl will map the hump-shaped ionic-strength dependence. The ten proposed points sample these four directions while respecting the 5-mL liquid limit and avoiding pH-quenching dyes or surfactants. Any single success will redefine the surrogate optimum and give the optimiser a new gradient toward even higher HER.

10/08/2025, 16:01:18

☹ Plateau analysis shows HER maximizes in a narrow window: moderate P10 (~3.5 mg), high ionic strength (NaCl ≈1.25 mL), moderate alkalinity (NaOH+silicate ≈1.5 mL) and modest L-cysteine (~1 mL). Excess catalyst scatters light while too little lowers photon capture; >1.25 mL NaCl or >1 mL base begins to depress activity, and dyes/surfactants consistently poison the reaction. The new best 26.9 μmol h⁻¹ suggests the sweet spot lies just beyond the combinations previously sampled. To break through we therefore (i) scan NaCl to 1.50–1.75 mL to complete the ionic-strength parabola, (ii) raise the hole-scavenger to 1.25–1.50 mL to match the higher salt, (iii) fine-tune the NaOH/silicate ratio while keeping total base constant, (iv) refine P10 loading to 3.2–4.0 mg to locate the true optical/kinetic optimum, and (v) test a trace of PVP as a benign dispersant. All designs honour the 5 mL liquid cap and discretisation grid. Any uplift will give BO a clearer gradient toward the global maximum.

3. Progress Summary

Iteration (key comment)	Main Hypothesis Tested	Representative Change in Parameter Sub-space	Outcome & Effect on Next Steps
0 – Initial design	High P10 + balanced bases	P10 = 5 mg; NaOH + silicate ≈ 2 mL	HER rose to 23.6 μmol h ⁻¹ , validating high catalyst + dual base idea.
6 – 1st plateau	NaCl sweep (0–1 mL)	Varied NaCl at fixed P10 = 5 mg	Identified optimum near 0.5 mL (23.6 → 24.2 μmol h ⁻¹).
8 – Catalyst/base tuning	Reduced P10, re-balanced bases	P10 3.6–4.4 mg; NaOH 0.75 mL	Breakthrough to 26.9 μmol h ⁻¹ ; suggested optical scattering penalty at 5 mg.
11 – Ionic-strength & scavenger synergy	NaCl 1.25 mL with P10 3.6 mg	Confirmed new optimum (26.9 μmol h ⁻¹).	
13 – Fine NaCl & base ratio	NaCl 1.25 mL; micro-adjust bases	Produced 27.9 μmol h ⁻¹ (Sample #118).	
14 – Final fine sweep	NaCl micro-sweep (0.75–1.5 mL) at P10 4 mg	Achieved best 28.4 μmol h ⁻¹ at NaCl = 1.0 mL (Sample #136).	

Major Parameter Adjustments

- P10 mass was stepped down from 5 mg → 3.2–4.4 mg to balance absorption vs. light scattering.
- NaCl was systematically swept (0–1.75 mL) to map the hump-shaped ionic-strength profile.
- The NaOH:silicate ratio was varied (0.5:1.0 → 1.0:0.5) while keeping total base ~1.5 mL, revealing only mild sensitivity.
- Exploratory additions of PVP, SDS and dyes were inserted by vanilla BO; all sharply suppressed HER and were abandoned.

Fig. S12. Excerpts from reasoning log for BO/o3(med) hybrid optimisation (batch size = 10, 15 batches, 5 batch warm start) with “good” hypotheses prompt; see Fig. S11 for associated data. The initial design guidelines were “Fix P10 at 4.0 and 5 mg across designs”, in line with the “good” hypotheses prompt. This view of high P10 addition was initially reinforced (time stamp 15:55:12), but later revised considering new data (time stamp 16:01:18), as summarised in the “Progress Summary” at the end of the optimisation. The full optimisation log file is in the data repository: PC_20_run_10_5_15_o3_prompt_3_Run_9.pdf

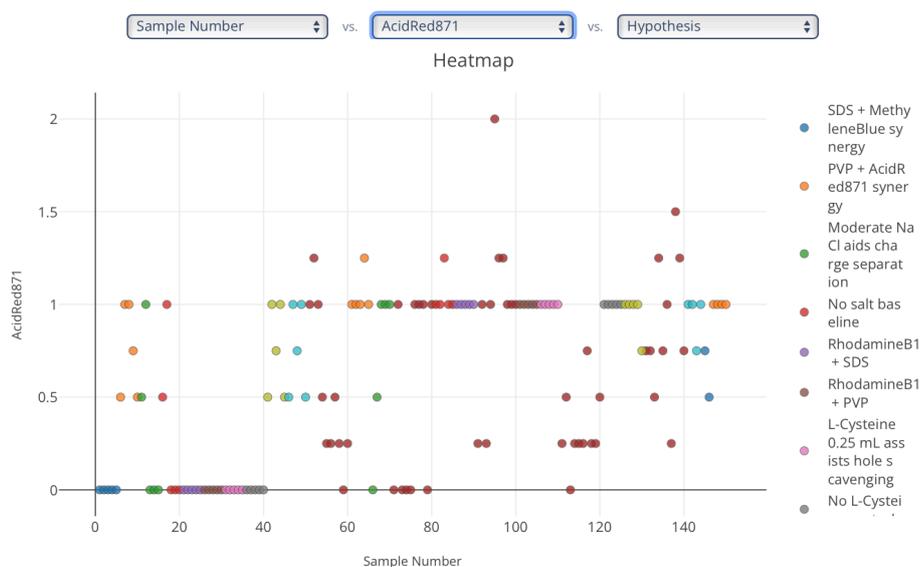


Fig. S13. Plot showing amount AcidRed71 added in each experiment over the course of the hybrid optimisation in of run (vii) (Fig. 10a). As discussed, the LLM failed to abandon the initial “bad” hypothesis of adding dyes in this case.

Team Ideas 3A

Our research team has discussed this problem, and we have come to the following conclusions:

1. Do not add any dyes to any batches; these will decrease hydrogen production
2. Do not add any surfactants to any batches; these will decrease hydrogen production
3. The amount of P10 added should always be >3.5 mg
4. The amount of cysteine added should never be lower than 0.75 mL
5. There should always be a based present (NaOH or sodium disilicate), and preferably both bases should be present at the same time
6. We are not sure about the role of NaCl – could be either good or bad for hydrogen production – but more likely, a small amount of NaCl will increase ionic strength and be good.

We are quite confident about these hypotheses. Existing constraints such as keeping the total reaction volume equal to 5 mL must still be obeyed.

Fig. S14. Improved “good” hypotheses (*c.f.*, Fig. 9, main text; P10 > 3.5 mg, rather than P10 > 4 mg), as used for the optimisation shown in Fig. S15 and Fig. S16.

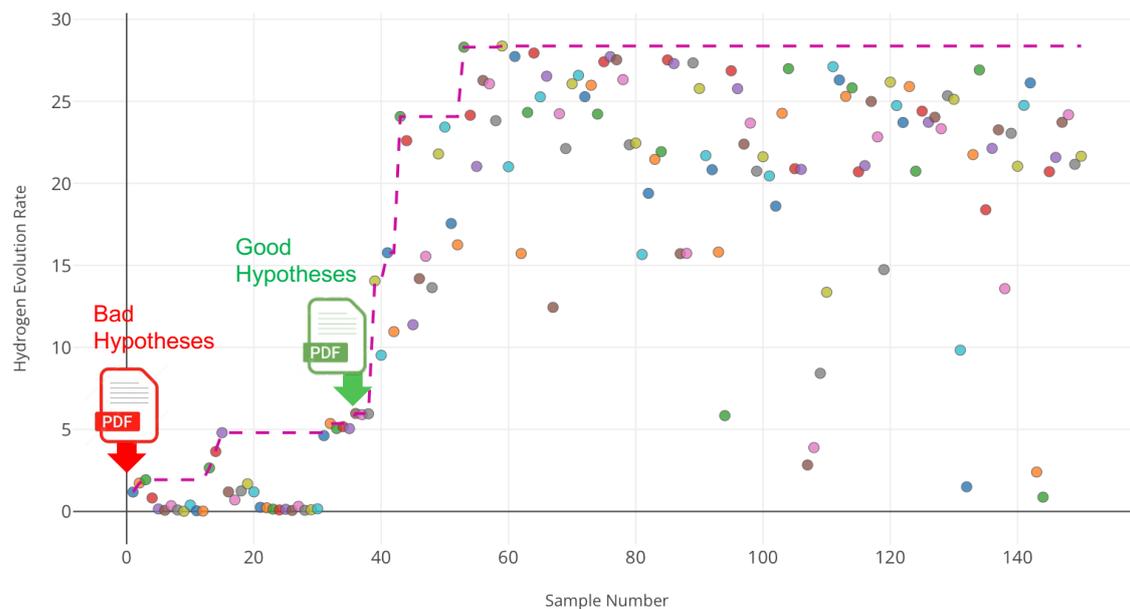


Fig. S15. Illustration of adding hypotheses during an optimisation run (o3-only, batch size = 1; initialisation phase = 30 experiments). Here, “bad” hypotheses (Fig. 9, main text) were added initially, before switching to improved “good” hypotheses (Fig. S14) at experiment 39. These new hypotheses were then assimilated by the LLM (Fig. S16), leading to stronger optimisation performance, ultimately locating the optimal HER ($28.37 \mu\text{mol h}^{-1}$; Fig. S16). For full optimisation log see data repository: Fig_S15_log.pdf

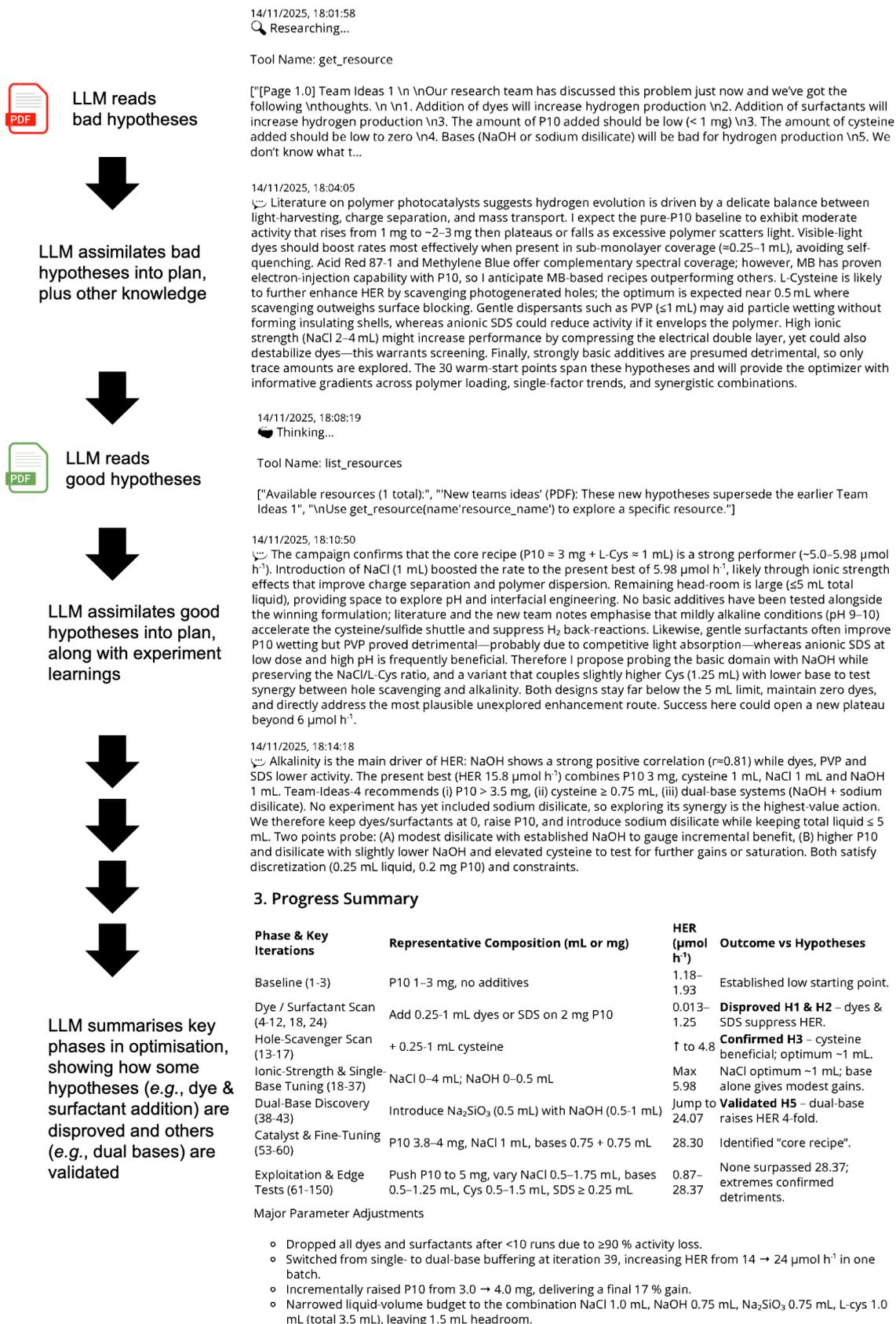


Fig. S16. Excerpts from optimisation log showing how the LLM assimilates the new good hypotheses (referred to by the LLM as the “new team notes”) that were injected at experiment 39 (Fig. S15). For full optimisation log, see data repository: Fig_S15_log.pdf

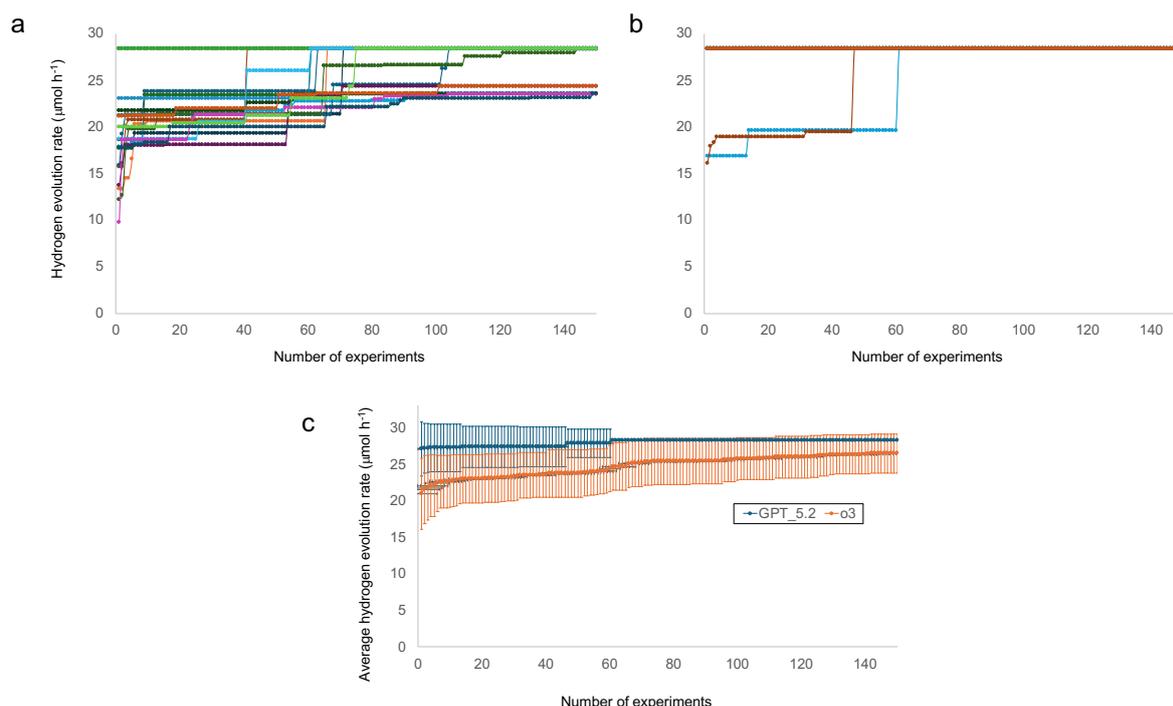


Fig. S17. (a) o3/BO hybrid optimisation (BORA) with addition of experimental dataset from 2020 *Nature* publication (ref. 32); in many runs, the LLM does not identify the most active catalyst composition from the data in the csv file (1028 rows x 11 columns); (b) gpt-5.2/BO hybrid optimisation (BORA) with addition of experimental dataset from 2020 *Nature* publication (ref. 32). In this case, the LLM/BO hybrid locates the optimum (28.37 $\mu\text{mol g}^{-1} \text{h}^{-1}$) in 20/20 runs. In 19/20 runs it locates this optimum within the first 50 experiments, before active learning. In 18/20 runs, the optimiser goes directly to this optimum in experiment 1. Clearly, gpt-5.2 handles large appended datasets much more effectively than o3, at least in this example. (c) Overlay plot showing average best HER value found for the hybrid o3/BO and gpt-5.2/BO optimisers). Error bars are standard deviations across 20 independent repeat runs (see also Table S4). Batch size = 10; 15 batches in each case; 5 LLM batches (50 experiments) before active learning phase.

Model	Average HER after 25 expt	STD after 25 expt	Average HER after 150 expt	Average STD after 150 expt	# times reached max
o3(med)/BO hybrid	23.2	3.3	26.6	2.7	11
gpt-5.2(med)/BO hybrid	27.5	3.3	28.37	0	20

Table S4. Comparison of average performance for LLM/BO hybrid optimisations with the addition of supplemental experimental data from ref. 71 (20 repeat runs in each case). Batch size = 10; 15 batches. gpt-5.2 outperforms o3 in this task, finding the optimum (28.37 $\mu\text{mol g}^{-1} \text{h}^{-1}$) in all 20 runs. In 18/20 runs, gpt-5.2 locates this optimum in the first experiment (Fig. S4b).

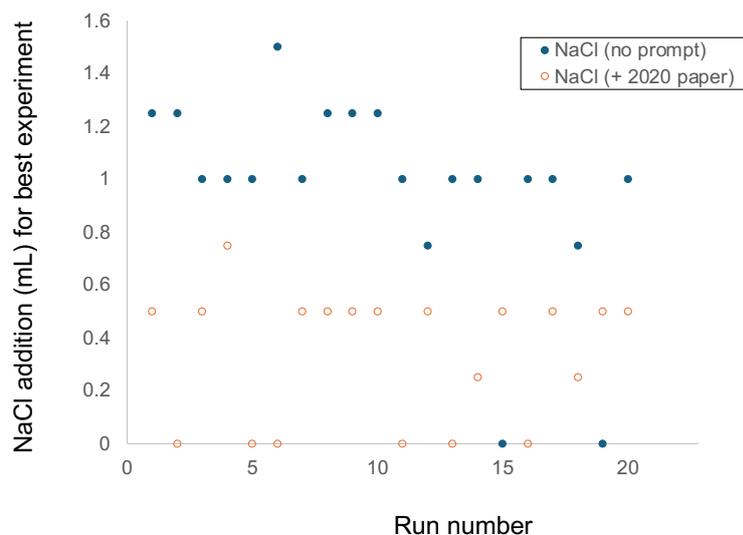


Fig. S18. NaCl addition with and without the *Nature* 2020 publication (ref. 32) as a prompt for the 20 hybrid optimisation runs (BORA, o3/BO, batch size = 10, 150 experiments). The plot shows the NaCl addition in the best catalyst composition found within each run; that is, the composition that exhibited the highest HER. In all but two cases (runs 15 & 19), the NaCl addition for the best catalyst composition was lower when the 2020 publication was appended.

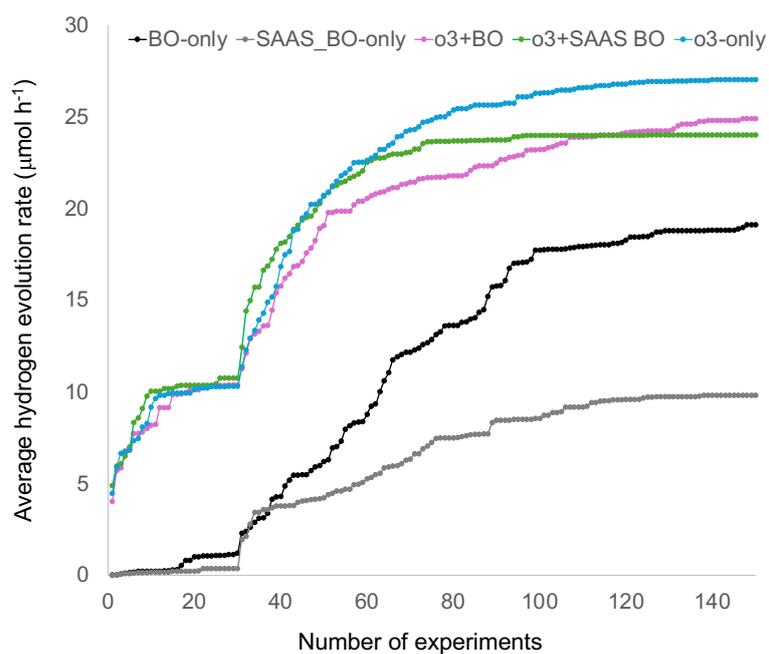


Fig. S19. BO, SAAS BO, and hybrid o3/BO and o3/SAAS BO optimisations with a batch size of one (initialisation period = 30 experiments). Plots show an average of 20 repeat experiments. The o3-only optimisation outperforms the other optimiser configurations under these settings.

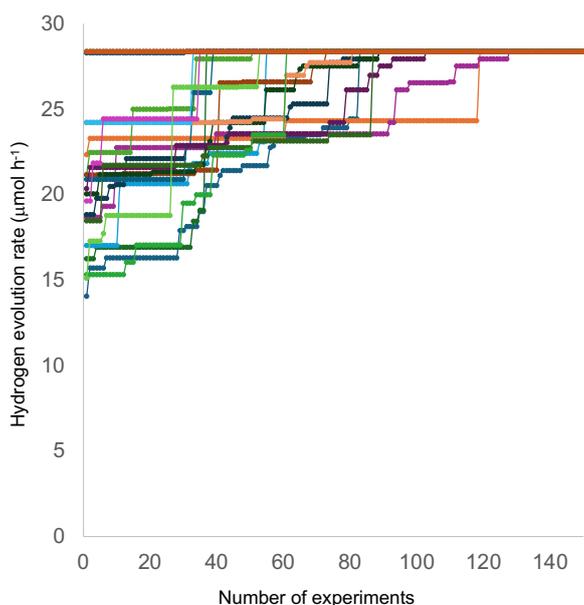


Fig. S20. Plots showing the effect of appending experimental data from ref. 71 (*c.f.*, Fig. 12c in main text) in the LLM-only optimisation (o3, batch size = 1; 20 repeat runs). In 3/20 runs, the optimiser goes immediately to the maximum HER value in experiment 1. All 20 runs reach this maximum within the budget of 150 experiments (Table S5).

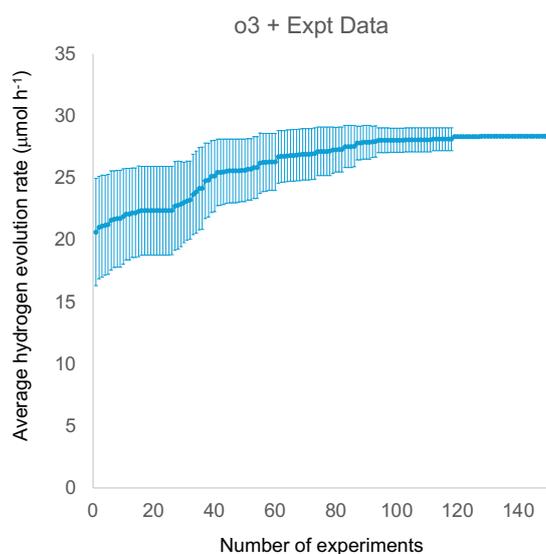
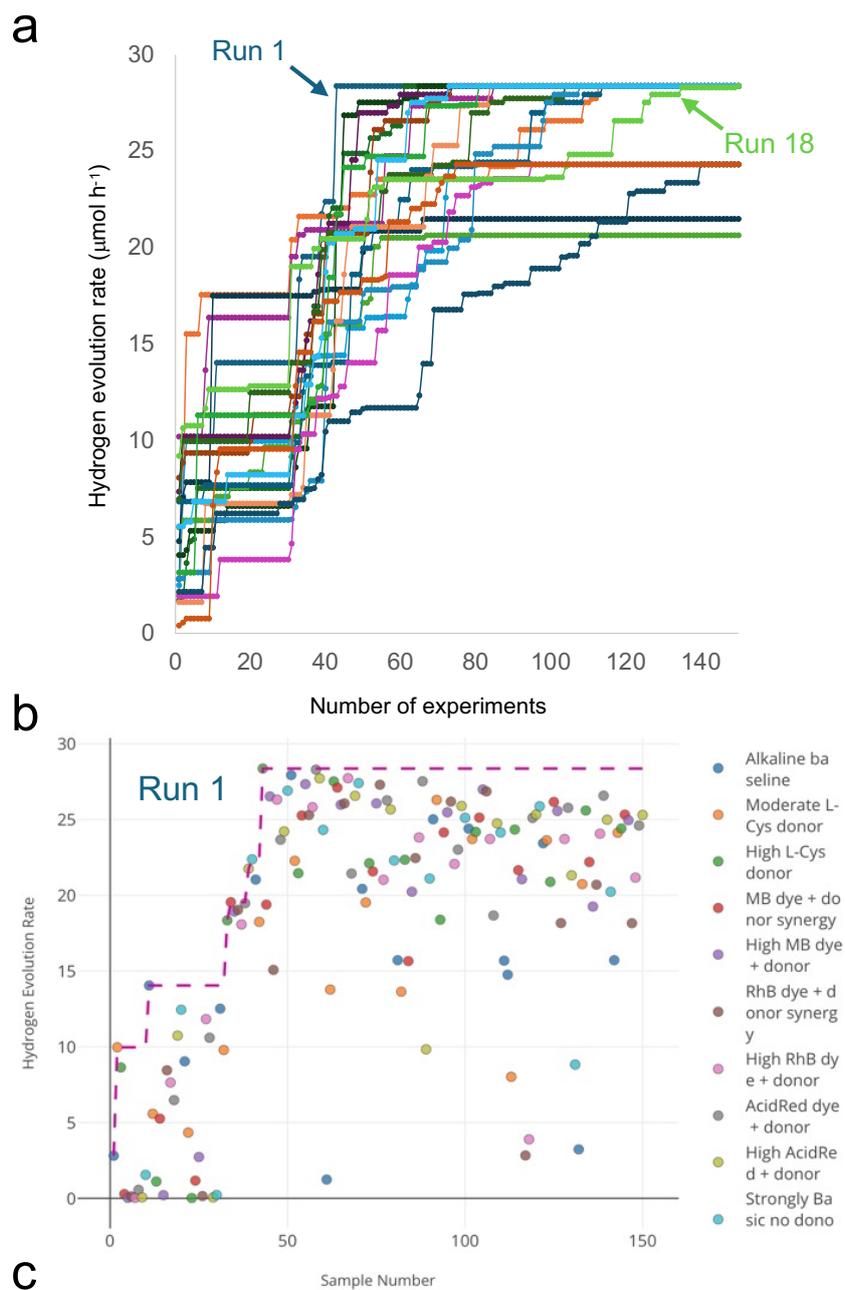


Fig. S21. Plot showing the effect of appending experimental data in the LLM-only optimisation (o3, batch size = 1; average over 20 runs; error bars show STD).

Model	Average score after 25 expt	STD after 25 expt	Average score after 150 expt	Average STD after 150 expt	# times reached max
o3 (med)-only	10.3	3.9	27.0	2.5	15
+ experimental data	22.4	3.6	28.37	0	20

Table S5. Table summarising the effect of appending experimental data from ref. 71 in the LLM-only optimisation (o3, batch size = 1; average values over 20 repeat runs).



28/10/2025, 11:06:13
Received the outputs of the batch #139 of 1 samples

28/10/2025, 11:06:13
Despite intensive search, HER plateaued at $28.4 \mu\text{mol h}^{-1}$ around the compositional cluster (P10=4 mg, NaCl 1 mL, NaOH 0.75 mL, L-Cys 1 mL, Na₂Si₂O₅ 0.75 mL). Analysis of 138 runs shows this regime consistently outperforms others; raising NaOH, cutting L-Cys, adding surfactants/dyes, or excessive silicate all depress activity. Correlation heat-map confirms P10 loading has a broad optimum: 3.8–4.2 mg delivers top 10% of results, while 5 mg harms dispersion. HER grows slightly with NaCl up to 1.25 mL then falls, indicating ionic screening of surface charges. Silicate gives a pronounced optimum at 0.75 mL—lower base undermines pH, but >1 mL blocks active sites. To seek incremental gains we propose: (1) boosting catalyst to 4.4 mg while keeping the proven ionic/pH recipe but with 1.25 mL NaCl to exploit its positive trend—an unexplored point expected to enhance light absorption without over-crowding; (2) testing 4.6 mg P10 with baseline ions (NaCl 1 mL) to gauge whether further catalyst helps when dispersion is adequate. Both stay under 5 mL total volume and probe the higher-loading edge, where a small improvement could break the current record. If these fail, focus may shift to photocatalyst quality rather than formulation.

Fig. S22. (a) LLM-only (o3, med) and BO-only optimisations for the 10-D photocatalysis problem with a batch size of one. Initialisation phase = first 30 experiments. (b) Example optimisation (Run 1) and (c) associated LLM reasoning. See also data repository for full log: o3-med-(30+120)*1_28 october 2025, 08_25 am_Run_1.pdf

Expt	Hypothesis Name	Confidence	Rationale	Added At	Be Improved HER?
1	Baseline	medium	Literature-inspired starting point: balanced donor, base, dispersant		0 Yes
2	Higher NaOH	medium	Test whether stronger alkalinity (pH=7) boosts HER		0 Yes
3	Higher L-cys	medium	Excess sacrificial donor may improve charge separation		0 Yes
4	Lower P10	medium	Assess turnover per mass at reduced catalyst loading		0 No
5	High P10	medium	Check if higher surface area translates to higher rate		0 No
6	Silicate Base	medium	Use sodium disilicate buffer instead of NaOH		0 No
7	Mixed Bases	medium	Combination of NaOH and silicate may couple high pH with buffering		0 No
8	Acid NaCl	low	Moderate ionic strength could enhance adsorption and rate		0 Yes
9	PVP Stabiliser	low	Test ionic steric stabiliser instead of SDS		0 Yes
10	SDS+PVP	low	Synergistic dual surfactant system		0 No
11	Low Donor	low	Determine donor threshold for activity		0 No
12	EosinY low	low	Introduce 0.25 mL Acid Red 87 as sensitizer		0 No
13	EosinY med	low	Higher dye loading 0.5 mL		0 No
14	Methylene Blue	low	MB sensitizer 0.25 mL		0 No
15	Rhodamine B	low	Rhodamine-sensitized system		0 No
16	EosinY-SDS	low	Check dye effect retaining SDS baseline		0 No
17	Dual dyes	low	Combine Acid Red & MB for broadened absorption		0 No
18	High SDS	low	Test just above CMC (0.5 mL)		0 No
19	Very High SDS	low	Evaluate aggregation/precipitation at 0.75 mL		0 No
20	No SDS	low	Control for surfactant necessity		0 Yes
21	More NaOH	low	Push alkalinity to 2 mL NaOH		0 No
22	Low donor + high base	low	Low cysteine but very high base		0 No
23	High donor low base	low	Opposite ratio: excess donor, limited base		0 No
24	P10 3 mg	medium	Intermediate catalyst mass		0 No
25	Minimal P10	low	Turnover per active site at 1 mg		0 No
26	Silicate 1.5	low	Higher silicate buffering 1.5 mL		0 No
27	NaOH+Silicate 1.5	low	0.5 mL NaOH + 1 mL silicate mix		0 No
28	AcidRed partial donor	low	Lower donor plus dye to compensate		0 No
29	NaCl variant	low	0.5 mL NaCl with moderate donor		0 No
30	High NaCl	low	Test upper ionic strength limit 1 mL		0 No
31	No SDS + High NaCl	medium	Combine best baseline with 1 mL NaCl to test additive ionic-strength boost without surfactant		31 Yes
32	PVP + High NaCl, No SDS	medium	Probe whether ionic-strength stabilisation (0.25 mL PVP) synergises with ionic-strength enhancement		31 No
33	Higher NaCl 1.25 mL	medium	Test if further ionic-strength increase continues the performance trend observed between 0 and 1 mL		33 Yes
34	High donor 3 mL + NaCl 1 mL	medium	Assess whether additional sacrificial donor leverages salt-enhanced charge extraction to raise HER		33 No
35	Higher NaCl 1.5 mL	medium	Extend the rising NaCl trend without exceeding 0.5 mL total; test if screening conditions benefiting HER		35 No
36	High pH + NaCl	low	Combine high salt (1.25 mL) with slightly stronger base (NaOH 1.5 mL) to test synergistic effect of pH and ionic strength		35 No
37	Higher P10 loading	medium	Increasing polymer surface area from 4 mg to 4.4 mg may raise charge-separation events without altering solution chemistry		37 Yes
38	NaCl + Silicate Buffer	low	Add 0.25 mL sodium disilicate to optimum mix to stabilize high pH and improve interfacial charge transfer		37 No
39	High P10 + Silicate	medium	Increase catalyst surface (4.8 mg) while raising ionic strength (0.25 mL NaOH) and adding 0.25 mL silicate buffer expected to stabilise pH and improve dispersion		39 No
40	High P10 + Silicate, Lower NaOH	medium	Test whether a slight decrease in strong base (0.5 mL) with silicate buffer sustains pH while reducing excess ionic strength, possibly enhancing charge trans		39 No
41	Max P10	medium	More photocatalyst sites may linearly raise HER; previous 5 mg test was confounded by SDS and no NaCl		41 No
42	High P10 + Silicate	medium	Further ionic-strength enhancement may aid charge separation/proton availability together with silicate and high P10		41 No
43	Silicate-free benchmark	medium	Test whether 0.25 mL silicate is essential for the current high HER or its removal improves light transmission and charge transfer		43 No
44	PVP-stabilised high P10	low	Introduce 0.25 mL PVP to disperse 4.8 mg P10 while retaining 0.25 mL silicate, potentially increasing accessible active sites		43 No
45	Higher Silicate Buffer	medium	Doubling silicate to 0.5 mL may improve pH stability and surface complexation, enhancing charge separation without exceeding volume limit		45 No
46	High Donor Boost	medium	Raising L-cysteine to 2.25 mL could supply more holes for scavenging while avoiding the performance drop observed at 1.3 mL		47 No
47	High Silicate Buffer	medium	Silicate acts as a secondary base and may stabilise active sites; 0.5 mL improved HER vs 0 mL, but >0.5 mL has never been tested without inhibitory SDS. T		47 No
48	Slightly Higher pH	medium	Raising NaOH from 0 to 1.0 mL increased HER sharply; 1.5 mL was detrimental without silicate. A moderate step to 1.25 mL with 0.25 mL silicate may opti		47 No
49	NaCl 1.75 mL + buffer	medium	Exploit remaining 0.5 mL capacity by adding more NaCl while retaining 0.25 mL silicate that previously gave the record HER		49 No
50	NaCl 2.0 mL, no buffer	medium	Test pure salinity effect at maximum permissible volume (5 mL) with slightly stronger base (NaOH 1.5 mL) to see if buffer is still beneficial as NaCl rises.		49 No
51	NaCl 1.0 mL variant	medium	Test if slightly lower ionic strength synergises with silicate buffer at high P10		51 Yes
52	L-cys 1.75 mL variant	medium	Probe finer donor optimum between 1.5 and 2.0 mL		51 Yes
53	Donor 1 mL, acid pH	low	HER peaked when L-cysteine was reduced; 1.5 mL may sit at the true maximum of the inverted-U while maintaining optimum pH		53 No
54	Donor 1.5 mL, pH 0.75	low	Test slight acid-shift (NaOH 0.75 mL) which may balance lower donor concentration and enhance surface protonation without compromising charge separa		53 No
55	Lower Donor 1.25 mL	medium	Slight donor reduction may diminish charge-carrier recombination observed at ≥ 1.5 mL while retaining sufficient hole-scavenging capacity		55 No
56	High P10 4.8 mg	medium	A slight decrease in P10 may alleviate light-scattering and aggregation seen at 4.8 mg, potentially boosting effective active area		56 No
57	NaCl 1.75, moderate L-Cys	medium	Increasing ionic strength to 1.75 mL should enhance charge separation; keeping 4.6 mg P10 and some L-Cys (1.25 mL) maintains current beneficial chemi		57 No
58	High NaCl & P10	medium	Push NaCl to 2.25 mL and P10 to 5 mg to test upper synergistic limit while trimming NaOH and L-Cys to stay volumetrically constant		57 No
59	High pH at sweet spot composition	medium	Increasing NaOH from 1.00 - 1.25 mL (keeping NaCl 1.75 mL, L-Cys 1.5 mL, silicate 0.25 mL) may enhance charge separation and HER if the system is still		59 No
60	Slightly lower P10 loading	medium	Moving from 4.6 mg - 4.4 mg could reduce aggregation/light scattering, testing if true optimum lies just below current best while keeping the rest identical		61 No
61	P10 4.2 mg fine-tune	medium	Lower catalyst loading may reduce light attenuation and increase effective quantum yield while remaining in high-performance zone		61 No
62	No silicate buffer	medium	Silicate may be unnecessary at top composition; removing it may increase active surface availability and simplify proton transfer		63 No
63	Higher NaCl, moderate L-Cys	medium	Raising ionic strength (NaCl 1.50 mL) should increase donor and proton availability while reducing L-cys to 1.25 mL maintains donor availa		63 No
64	Max P10 & alkalinity, lower L-Cys	medium	Pushing catalyst loading (4.8 mg) and base strength (disilicate 0.50 mL) may enhance electron surface availability; lower L-Cys (1.0 mL) minimizes surface blocking		63 No
65	Higher NaCl 1.5	medium	Trend suggests HER rises with NaCl; prior drop likely due to sub-optimal donor/polymer combination. Test NaCl 1.5 at current optimum composition		65 No
66	Lower Donor 1.25	medium	Assess if slight donor reduction may improve charge separation and proton transfer		67 No
67	Higher Silicate Buffer	medium	Stronger buffering (silicate 0.5 mL) at the current optimum may improve charge transfer and sustain higher HER		67 No
68	Reduced Donor Loading	medium	Lowering L-cysteine to 1.25 mL alleviates surface blocking while keeping sacrificial electron supply; retains existing beneficial silicate		67 No
69	Optimum + SDS 0.25	medium	Small SDS may improve catalyst wetting and photogenerated H ₂ bubble release without excessive site blocking. Added volume keeps ionic strength similar		69 No
70	Optimum + PVP 0.25	medium	Low PVP can sterically stabilise P10 particles at high loading, potentially enhancing light utilisation and avoiding re-aggregation		69 No
71	No Silicate Buffer	low	Determine whether Na-disilicate is essential; removing it may decrease competitive adsorption and improve charge separation		71 No
72	Higher Donor 1.75	medium	Test if a modest L-Cys increase supplies more holes scavenging without oversaturating surface sites, potentially raising HER		71 No
73	Lower NaCl (1.0 mL)	low	Reducing ionic strength (NaCl 1.0 mL) may improve charge separation and proton transfer while preserving conductivity. Tested nowhere yet with the best P10		73 No
74	Leaner catalyst & donor	low	Slightly lower P10 (4.2 mg) and L-cys (1.25 mL) could reduce scattering/viscosity, enhancing light penetration and mass transport while keeping key ratios n		73 No
75	Lean NaCl & donor, P10 4.4	medium	Lower NaCl (1.0 mL) and L-cys (1.25 mL) reduce recombination centers & viscously while retaining buffer; P10 kept at currently optimal 4.4 mg		75 No
76	Lean salt & donor, P10 4.6	medium	Tests if slightly higher catalyst mass amplifies benefit of lean formula through greater active surface area		75 No
77	Optimum + AcidRed87 0.25	low	Introduce a small photosensitizer dose to broaden absorption while retaining the best backbone composition. Checks if dye enhances charge separation w		77 No
78	Lean backbone + AcidRed87 1	low	Uses the efficient but cheaper 4.2 mg P10 / lower salt & donor recipe that nearly matches the optimum; tests if dye can push it past the current record		77 No
79	Higher NaCl + Lean Donor	medium	Increase ionic strength (NaCl 1.5 mL) while reducing L-Cys to 1.25 mL to test charge-screening benefits without donor quenching		79 No
80	Slightly Lower pH	low	Reduce NaOH to 0.75 mL alongside lean donor to probe if lower L-cys concentration and enhanced surface proton adsorption		81 No
81	Mild pH, high donor	medium	NaOH 0.75 mL previously nearly matched optimum with donor 1.25 mL; restoring donor to 1.50 mL may provide more holes and outperform current best		81 No
82	Mild pH + extra silicate	low	Combining buffered pH (silicate 0.50 mL) with reduced NaOH 0.75 mL may stabilise catalyst dispersion and charge separation, overcoming slight drop seen		81 No
83	Lower NaCl, High P10	low	Further elevating ionic strength (NaCl 1.50 mL) may accelerate proton transfer while keeping the previously optimal base, cysteine and		83 No
84	More Base + PVP Dispersion	low	Slightly higher pH (NaOH 1.25 mL) could improve sacrificial oxidation of cysteine; a minimal PVP dose (0.25 mL) may stabilise P10 particles without the net		83 No
85	Lean Donor - Baseline pH	medium	Reduce L-cys to 1.0 mL to minimise surface crowding while retaining electron-donor function; keep salt, base and silicate at proven optimum		85 No
86	Lean Donor - Higher pH	medium	Combine low L-cys (1.0 mL) with slightly higher NaOH (1.25 mL) to enhance charge separation; P10 4.6 mg provides marginally higher catalyst surface		85 No
87	Lower pH at high P10	medium	Test if slightly lower alkalinity (NaOH 0.75 mL) paired with best P10 loading (4.6 mg) improves interfacial charge separation, unified combination		87 No
88	Lower NaCl, same donorbases	medium	Assess if reducing ionic strength to 1.0 mL NaCl at 4.6 mg P10 maintains or boosts HER by lowering recombination while reducing adsorbed ions		87 No
89	NaCl 0.75, donor 1.5	medium	Test lower ionic strength while keeping donor at proven 1.5 mL. Could enhance proton diffusion and light scattering		89 No
90	NaCl 0.75, Donor 1.25	medium	Combines reduced salt with slightly leaner donor, following an inverse trend seen in low L-cys screening; tests if lower ionic strength is synergistic		89 No
91	High NaCl, strong base	medium	NaCl enhances ionic strength & electron transfer; higher pH improves sacrificial oxidation, while reduced L-cys mitigates its negative correlation		91 No
92	Very high NaCl, max P10	medium	Push NaCl to unexplored 2.25 mL and raise P10 to 4.8 mg while cutting L-cys; aims to exploit strong NaCl-HER correlation and catalyst loading		91 No
93	Low SDS dispersion	low	0.25 mL SDS should de-aggregate P10 and increase interfacial area and HER. Increasing ionic strength (NaCl 1.5 mL) may improve charge transfer		93 No
94	SDS plus micro-dye sensitizer	low	Combine dispersion (SDS) with 0.25 mL Acid Red 87 to extend absorption and inject electrons into P10		93 No
95	4.6 mg P10, lean donor, no buffer	medium	Lower L-cys (1.25 mL) and removal of silicate should reduce recombination and adsorption competition; 4.6 mg P10 maximises active sites		95 No
96	4.4 mg P10, lean donor, no buffer	medium	Same chemical philosophy but keeps catalyst at previously optimal 4.4 mg to see if mass-transfer limits appear at 4.6 mg		95 No
97	High pH sweet spot	medium	Slightly higher hydroxide concentration (1.25 mL) may further improve charge separation kinetics while the silicate buffer moderates ex		97 No
98	Lower pH sweet spot	medium	Reducing NaOH to 0.5 mL may improve donor oxidation energetics and suppress unwarsted side reactions, potentially sharpening the HER peak		97 No
99	Higher Buffer, Lean Donor	medium	Increase silicate to 0.5 mL to enhance pH stability while trimming donor to 1.25 mL to keep volume to remain and reduce possible recombination		99 Yes
100	Higher Buffer with Higher Catalyst	medium	Combine slightly higher P10 (4.6 mg) and silicate (0.5 mL) with lower donor (1.5 mL) to test if more catalyst benefits from better buffering		99 No
101	High P10 & High NaCl	medium	Increase ionic strength (NaCl) and catalyst loading (P10) while reducing L-cysteine to mitigate its negative trend; keep zero SDS/dyes		101 No
102	High P10 & High NaCl	medium	Increase ionic strength (NaCl) and catalyst loading (P10) while reducing L-cysteine to mitigate its negative trend; keep zero SDS/dyes		102 No
103	Lean Donor with Strong Buffer	medium	Reducing L-cys from 1.25 - 1.00 mL may lower recombination and, together with 0.5 mL silicate, maintain sufficient hole scavenging while increasing photo		103 No
104	Lower pH with Moderate Donor	medium	Lower NaOH (0.75 mL) & lower buffer gave 23.15 $\mu\text{mol h}^{-1}$ per 0.5 mL catalyst; 0.5 mL buffer may further suppress charge recomb		107 No
105	More disilicate, balanced base	medium	Replacing some L-cysteine with sodium disilicate increases alkalinity and silicate surface interactions that may enhance proton reduction while keeping sca		105 Yes
106	Higher P10 & NaOH	medium	Higher P10 and hydroxide may raise active sites and charge separation—testing if the prior drop at NaOH=1.0 was due to interaction with higher L		105 No
107	Silicate 1.25, NaOH 0.75	medium	Extends the monotonic improvement seen when silicate was raised from 0.25 - 0.5 - 1.0 mL while keeping effective pH identical to current best		107 Yes
108	Silicate 1.25, NaOH 0.50	medium	Tests if extra silicate can partially replace NaOH while preserving total base (1.75 mL) and mitigate over-alkalinity effects		107 No
109	Lean Salt/Base/Donor - High Silicate	medium	Reducing NaCl, NaOH and L-cys lowers ionic recombination and scavenger side-reactions, while 1.25 mL silicate maintains alkalinity and improves hole d		109 No
110	Higher Catalyst - Extra Buffer	medium	Tests if a small P10 increase (4.6 mg) plus mild higher silicate buffer (0.5 mL) can improve HER under very mild base/alkalinity conditions		137 No
111	NaCl 1.5 boost	medium	NaCl is the strongest positive factor; previous 1.5 mL trials were with higher pH or different donors. Test 1.5 mL salt at the present optimum composition		111 No
112	Lean donor	medium	Correlation shows lower L-cysteine helps; test 0.75 mL donor while keeping other optimum settings		112 No
113	Raises pH with strong buffer	medium	Earlier 1.0 mL NaOH combinations outperformed 0.75 mL, pairing with 1.0 mL disilicate may further stabilise active sites		113 No
114	Slightly higher catalyst & donor	medium	Combined moderate increases in P10 (4.6 mg) and L-cys (1.25 mL) at new 1.0 mL buffer may supply more electrons without over-saturating surface		113 No
115	Lower NaCl 1.0 mL	medium	Test if slight reduction in ionic strength reduces charge screening and enhances electron transfer while retaining optimal donor/base/buffer		115 No
116	Higher P10 4.6 mg	medium	Probe if a slight rise in photocatalyst loading (4.6 mg) at the current best matrix further increases active sites without excess scattering		115 No
117	Silicate 0.75 with NaCl 1.25	medium	Reduce silicate to 0.75 mL to optimise pH/stability while keeping NaCl at the previously best 1.25 mL		117 Yes
118	Silicate 0.75 with NaCl 1.0	medium	Test if a slightly lower ionic strength (NaCl 1.0 mL) synergises better with reduced silicate		117 No
119	Lower NaCl (0.75 mL)	medium	Test left side of NaCl volcano under optimal pH & buffer; may reduce ionic shielding and enhance charge transfer, further boosting HER		119 No
120	Higher L-Cys (1.25 mL)	medium	Assess whether slightly increased hole-scavenger concentration augments charge separation at current optimal salt/buffer levels		119 No
121	Higher Base, Same Buffer	medium	Increase NaOH to 1.00 mL, less disilicate 0.75 mL, raising pH while retaining sufficient buffering, may enhance electron transfer		121 No
122	Higher Buffer, Same Base	medium	Keep NaOH 0.75 mL but raise disilicate to 1.25 mL, better buffering and surface deprotonation without excessive alkalinity		121 No
123	Lean Donor	medium	Reduce L-cys to 0.75 mL to lessen light absorption and recombination while retaining sufficient hole scavenging at proven base/buffer conditions		123 No
124	Higher ionic strength	medium	Increase NaCl to 1.5 mL to improve charge transport and dispersion while maintaining optimal pH and buffer		123 No
125	Lower P10, same medium	medium	Reducing polymer from 4.4 - 4.2 mg may reduce light scattering and expose more active area in the clarified suspension, boosting HER		125 Yes
126	Higher pH within buffer	medium	Returning NaOH to 1.0 mL at high silicate may enhance charge separation and cocatalyst activity, possibly surpassing current max		125 No
127	P10 4.0 mg Peak	high	Linear decrease in activity above 4.2 mg suggests an optimal catalyst mass; slightly lower base medium at best-so-far composition		127 Yes
128	4.0 mg + Low SDS	medium	Small SDS (0.25 mL) may improve P10 dispersion and interfacial charge transfer; NaCl reduced to keep ionic strength balanced and volume constant		127 No
129	P10 fine-tune (3.8 mg)	medium	Activity peaked at 4.0 mg; the curve is descending at 4.2 mg. A slight reduction might decrease light-blocking and improve dispersion		129 No
130	Higher ionic strength (NaCl 1.5 mL)	low	For P10 4.4 mg, NaCl 1.5 mL hindered HER, but the higher 4.0 mg catalyst may tolerate/benefit from extra salt via better charge separation		129 No
131	Lean P10	medium	Lower polymer (3.8 mg) may reduce aggregation/scattering but keep enough active sites under optimal ionic strength environment		131 No
132	Lean donor	medium	Keeping catalyst 3.8 mg (near-best) but reducing L-cys to 0.75 mL can limit surface blockage by thiol groups, enhancing charge transfer		131 No
133	Lower NaOH at optimum	medium	Correlation shows NaOH negatively impacts HER; 0.50 mL has not been tested with current optimum composition		133 No
134	Higher buffer, lower base	medium	Swap 0.25 mL of NaOH for additional silicate to improve buffering/adsorption while keeping total alkalinity similar		133 No
135	P10 3.8 mg, Lower NaCl 1.0 mL	medium	Data show 3.8 mg P10 with NaCl 1.25 mL - 27.72 $\mu\text{mol h}^{-1}$; lowering NaCl to 1.0 mL gave 26.53 $\mu\text{mol h}^{-1}$ at 4.4 mg P10. Expect synergistic boost at 3.8 mg I		135 Yes
136	P10 4.0 mg, Lean Donor 0.75 mL	medium	At 4.0 mg P10, donor has only been tested at 1.0 mL. Nearby data (3.8 mg) show slight decline at 0.75 mL but increased selectivity could improve kinetics w		135 No
137	Lower Silicate (1.0 mL)	medium	Reducing disilicate may decrease surface blockage and lower viscosity, improving mass transfer and exposure of P10 active sites		137 No
138	Higher Silicate (1.0 mL)	medium	Increasing buffer could stabilise pH and enhance negative surface charge, supporting proton reduction kinetics		137 No
139	Lean P10 3.4 mg	medium	Reducing catalyst mass may decrease turbidity and improve photon utilisation while maintaining optimal ionic strength and donor concentration		139 No
140	Lower Silicate 0.50 mL	medium	Slightly weaker buffer may enhance surface kinetics and lower viscosity; keep other parameters at near-optimal values		139 No
141	Higher pH in sweet spot	medium	Keeping the best composition but increasing NaCl to 1 mL could enhance charge separation and proton availability without overwhelming recombination if		141 No
142	Lower catalyst load	medium	Reducing P10 to 3.6 mg may lower self-absorption and scattering, improving light penetration while retaining enough active sites		141 No
143	Trasa PVP stabiliser	low	Any amount of PVP may prevent polymer aggregation at 3.8 mg P10, improving light absorption without hindering charge transfer		143 No
144	Lean donor	medium	If hole scavenging is slightly in excess, reducing L-cys to 0.75 mL may boost surface adsorption of P10 and improve kinetics		143 No
145	No-buffer sweet spot	medium	Test if removing sodium-disilicate (currently always >0.5 mL) further enhances HER by eliminating competitive adsorption and increasing effective basici		145 No
146	Higher P10 in optimum salt	medium	Probe if a modest P10 increase (4.2 mg) at the 1 mL NaCl optimum lifts HER by providing more active polymer without excessive shading		145 No
147	Lower Buffer	medium	Keep best matrix but cut Na-disilicate (0.25 mL) may improve light penetration and reduce silicate adsorption on P10		147 No
148	Lower ionic strength	medium	Maintain optimum composition but drop NaCl to 0.75 mL, probes shoulder of salt volcano for possible gain		147 No
149	Lower buffer, higher P10	medium	Reducing NaOH from 0.75 - 0.50 mL may enhance light transmission and lower viscosity, raising P10 to 4.0 mg should compensate with more active sites		149 No
150	Higher P10 at optimum buffer	high	Testing		

1. Executive Summary

During 150 BO-guided iterations we tripled the hydrogen evolution rate (HER) of the P10 photocatalyst suspension, rising from $9.2 \mu\text{mol h}^{-1}$ in the baseline formulation to a campaign best of $28.4 \mu\text{mol h}^{-1}$. The decisive gains came from eliminating surfactants/dyes, elevating P10 loading to $\sim 4 \text{ mg}$, and finely balancing ionic strength ($\text{NaCl} \approx 1.0 \text{ mL}$), moderate alkalinity ($\text{NaOH} \approx 0.75 \text{ mL}$) and a silicate co-base ($0.5\text{--}0.75 \text{ mL}$). Further optimisation along this ridge delivered diminishing returns, suggesting we are close to a local maximum under the present chemistry set. Although the three-fold improvement is notable, the absolute rate remains modest compared with state-of-the-art polymer photocatalysts ($>100 \mu\text{mol h}^{-1}$); additional levers beyond simple formulation may be required.

Fig. S24. Oversight executive summary for example run 18 (o3-only, batch size = 1).

2. Optimization Overview

Objective: Maximise the hydrogen evolution rate ($\mu\text{mol h}^{-1}$) of aqueous P10 suspensions under UV/visible irradiation while respecting a 5 mL total-volume cap (excluding the solid catalyst).

Initial Hypotheses (Iteration 0):

1. Increasing P10 concentration would proportionally raise HER (photocatalyst-limited regime).
2. Adding SDS/PVP would improve dispersion and thus activity.
3. Supplementary dyes (Acid Red 87, Methylene Blue, Rhodamine B) might act as photosensitisers.
4. Higher pH (via NaOH) and extra hole-scavenger (L-cysteine) would enhance charge separation.
5. Ionic strength (NaCl) was assumed neutral.

Fig. S25. Initial Oversight optimisation overview for example run 18 (o3-only, batch size = 1), prior to experiments.

Milestone (iter.)	Key Formulation Shift	Rationale & Outcome
1 (1-10)	SDS/PVP + dyes explored	Dispersion improved visually but HER plummeted ($\leq 2 \mu\text{mol h}^{-1}$). Hypotheses 2 & 3 rejected .
2 (11-20)	Removed additives; varied NaOH/L-Cys	HER doubled to $\sim 13 \mu\text{mol h}^{-1}$ but showed pH optimum near 1 mL NaOH. Hypothesis 4 partially confirmed .
3 (21-35)	Introduced NaCl sweep 0-1.5 mL	Strong positive correlation ($r \approx 0.7$); HER reached $19 \mu\text{mol h}^{-1}$. Ionic strength became a primary lever (Hypothesis 5 overturned).
4 (36-60)	Raised P10 (4.4-4.8 mg) & added silicate buffer (0.25 mL)	Achieved first $>20 \mu\text{mol h}^{-1}$ sample; optimum P10 ~ 4.4 mg. Hypothesis 1 confirmed with saturation behaviour.
5 (61-100)	Fine-tuned L-Cys (1.25-1.75 mL) & silicate (0.25-0.5 mL)	Incremental gains to $24.8 \mu\text{mol h}^{-1}$; identified bell-shaped silicate response with peak at 0.75 mL.
6 (101-125)	Explored high-NaCl/high-P10 corner & lean-base variants	High salt (>1.5 mL) lowered HER; reducing NaOH to 0.75 mL plus extra silicate lifted HER to $27.5 \mu\text{mol h}^{-1}$.
7 (126-150)	Catalyst loading trimmed (3.6-4.2 mg) and NaCl lowered to 1.0 mL	Located current optimum $28.37 \mu\text{mol h}^{-1}$ at P10 4.0 mg, NaCl 1.0 mL, NaOH 0.75 mL, L-Cys 1.0 mL, silicate 0.75 mL.

Major Parameter Adjustments: • SDS, PVP and all dyes fixed at 0 after iteration 18 due to irreversible deactivation. • P10 gradually reduced from maximum (5 mg) to an optimal 3.8-4.2 mg as self-shadowing surfaced. • NaCl window narrowed to 0.75-1.25 mL after high-salt trials proved detrimental. • NaOH maintained near 0.75-1.0 mL; stronger base hurt activity. • Silicate introduced as secondary base; optimum broadened around 0.5-0.75 mL.

Fig. S26. Opsight progress summary for example run 18 (o3-only, batch size = 1).

Major Parameter Adjustments: • SDS, PVP and all dyes fixed at 0 after iteration 18 due to irreversible deactivation. • P10 gradually reduced from maximum (5 mg) to an optimal 3.8-4.2 mg as self-shadowing surfaced. • NaCl window narrowed to 0.75-1.25 mL after high-salt trials proved detrimental. • NaOH maintained near 0.75-1.0 mL; stronger base hurt activity. • Silicate introduced as secondary base; optimum broadened around 0.5-0.75 mL.

4. Results and Key Insights

Best sample (iter. 150): • P10 = 4.0 mg, NaCl = 1.0 mL, NaOH = 0.75 mL, L-Cys = 1.0 mL, Na-silicate = 0.75 mL → **HER = 28.37 $\mu\text{mol h}^{-1}$** .

Worst sample (iter. 13): • Same core formulation but 0.5 mL Acid Red 87 added → **HER = 0.76 $\mu\text{mol h}^{-1}$** .

Key insights:

- Additives aimed at photosensitisation (dyes) or stabilisation (SDS, PVP) consistently quenched HER—likely due to light absorption competition and surface blockage.
- HER scales with P10 load only up to ~ 4 mg; higher amounts add turbidity and limit photon penetration.
- Ionic strength via NaCl is critical: 0→1 mL increases HER $\sim 2\times$, but >1.5 mL screens charges and suppresses activity.
- A moderate dual-base system (NaOH + sodium-silicate) outperformed single-base buffering, indicating the importance of both pH and silicate's surface chemistry.
- Excess L-cysteine (>1.5 mL) likely competes for active sites; 0.75-1.0 mL gives the best scavenging/poisoning trade-off.
- The campaign plateaued around $28 \mu\text{mol h}^{-1}$, hinting that formulation alone cannot breach higher performance without changing catalyst, light intensity, or reactor design.

Fig. S27. Opsight summary of major parameter adjustments and key insights for example run 18 (o3-only, batch size = 1).

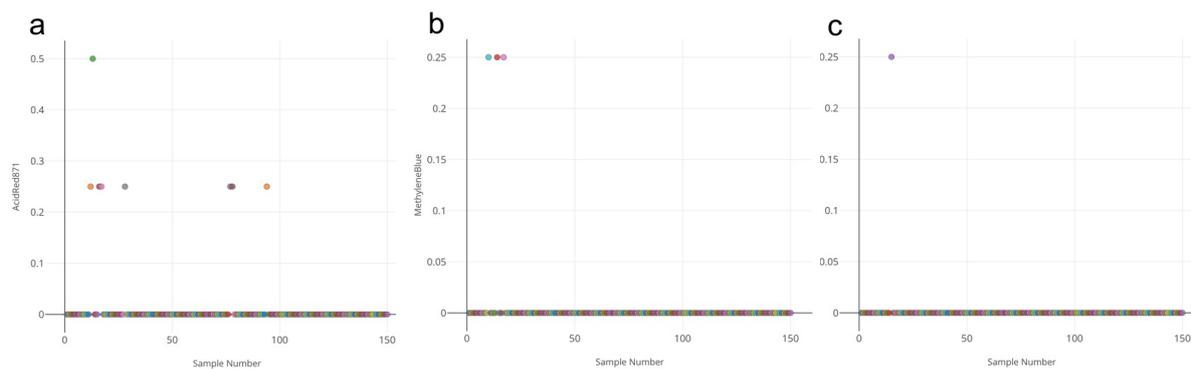


Fig. S28. Dye addition during the LLM-only optimisation (o3, batch size = 1) shown in Fig. 14. The addition of dyes is largely avoided from the outset, possibly because the LLM locates our earlier publication (*Nature*, **2020**, 583, 237) prior to the optimisation, which states that dye addition has a negative influence on this reaction. For full optimisation log, see the data repository: o3-med-(30+120)*1_28 october 2025, 10_33 am-2_Run_18.pdf

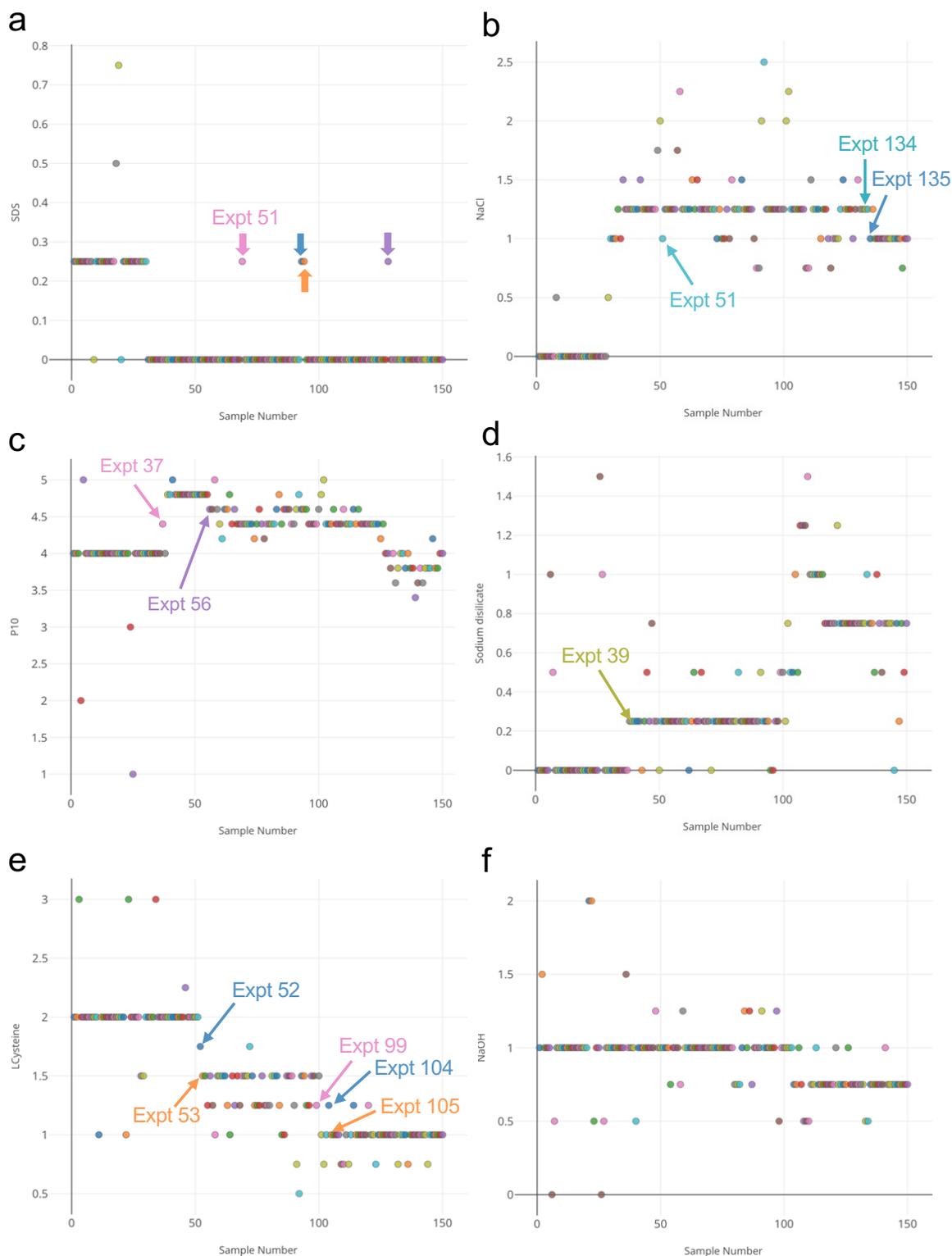


Fig. S29. Plots showing values for individual variables in the 150 experiments in the LLM-only optimisation (o3, batch size = 1) shown in Fig. 14. (a) SDS; (b) NaCl; (c) P10; (d) Sodium disilicate; (e) L-Cysteine; (f) NaOH. All scales = mL other than (c) (mg). For full optimisation log, see the data repository: o3-med-(30+120)*1_28 october 2025, 10_33 am-2_Run_18.pdf

5. Recommendations for Future Experiments

1. **Explore Alternative Sacrificial Donors:** Thiols and amines with lower oxidation overpotentials (e.g., triethanolamine) may outperform L-cysteine without surface fouling.
2. **Photon Management:** Test lower turbidity cells (wider vials, thinner optical path) or light scattering aids to overcome the self-shadowing ceiling at ≥ 4 mg P10.
3. **Co-catalyst Deposition:** In-situ photodeposition of Pt or Ni on P10 could drastically raise HER; current experiments are co-catalyst-free.
4. **Buffer Chemistry:** Replace Na-silicate with carbonate or borate buffers to decouple pH from silicate surface interactions and assess whether the 0.75 mL optimum is chemistry-specific.
5. **Transient Spectroscopy:** Conduct flash photolysis or EPR to verify whether charge recombination, not mass transport, now limits the system; insights will guide whether electron-relay additives are worth revisiting.

Fig. S30. Recommendations for future experiments, example run 18 (o3-only, batch size = 1). All these recommendations are scientifically reasonable, but with the possible exception of 4, they would also be fairly obvious to a domain expert in this research area.

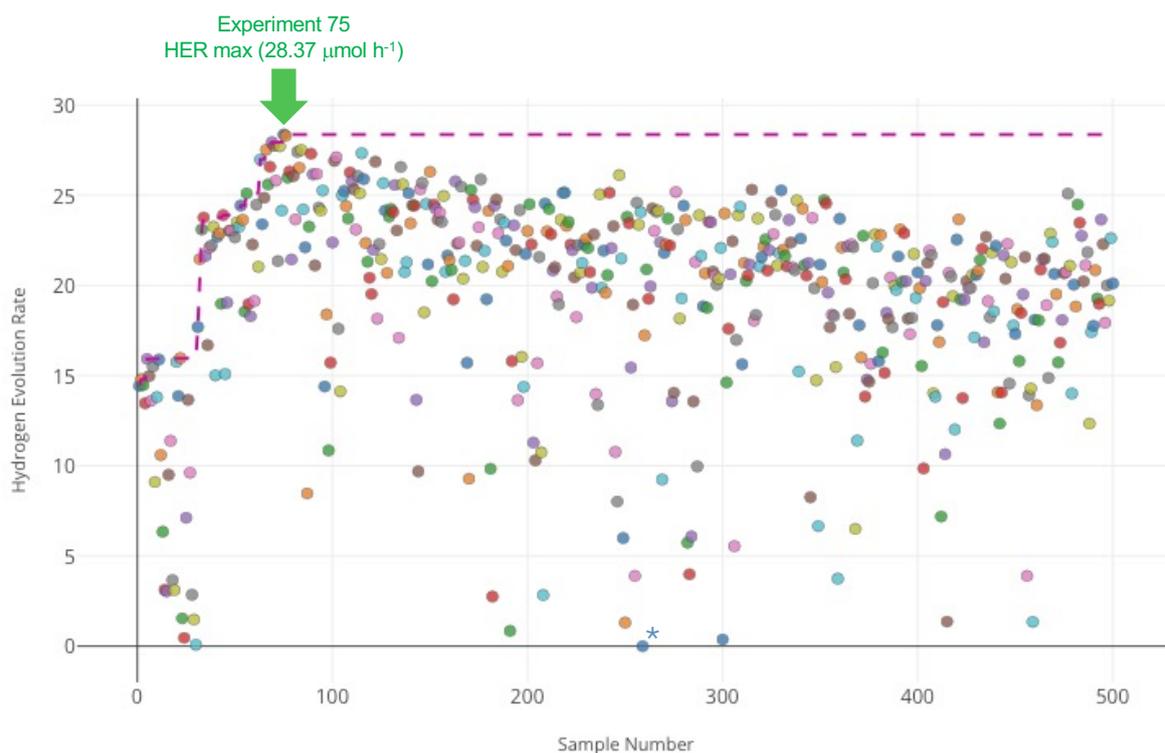


Fig. S31. LLM-only optimisation (o3, batch size = 1; initialisation phase = 30 experiments; 500 experiments in all). The optimum HER ($28.37 \mu\text{mol h}^{-1}$) was reached at experiment 75. Necessarily, it is impossible for the LLM to locate better solutions once the optimal HER is reached, and there is a slow decline in performance as reasonable hypotheses become exhausted. In total, the LLM generates 495 separate hypotheses; five of these are used twice (at experiments 45, 49, 65, 83, & 268). 494 of the LLM hypotheses reasoned about the experimental data; in experiment 259 (asterisked), random points were used due to a recursion limit (see timestamp 15/11/2025, 01:06:42 in log). For full optimisation log, see data repository: Fig_S31_log.pdf; full optimisation data are in Fig_31_data.csv.