A Mobile Robotic Chemist

Benjamin Burger, Phillip M. Maffettone, Vladimir V. Gusev, Catherine M. Aitchison, Yang Bai, Xiaoyan Wang, Xiaobo Li, Ben M. Alston, Buyi Li, Rob Clowes, Nicola Rankin, Brandon Harris, Reiner Sebastian Sprick and Andrew I. Cooper*
*aicooper@liverpool.ac.uk

Leverhulme Centre for Functional Materials Design, Materials Innovation Factory and Department of Chemistry, University of Liverpool, 51 Oxford Street, Liverpool, L7 3NY, United Kingdom

Technologies such as batteries, biomaterials, and heterogeneous catalysts have functions that are defined by mixtures of molecular and mesoscale components. As yet, this multi-length scale complexity cannot be fully captured by atomistic simulations, and the design of such materials from first principles is still rare[1-5]. Likewise, experimental complexity scales exponentially with the number of variables, restricting most searches to narrow areas of materials space. Robots can assist in experimental searches[6-14] but their widespread adoption in materials research is challenging because of the diversity of sample types, operations, instruments and measurements that is required. Here we use a mobile robot to search for improved photocatalysts for hydrogen production from water[15]. The robot operated autonomously over 8 days, performing 688 experiments within a 10-variable experimental space, driven by a batched Bayesian search algorithm[16-18]. This autonomous search identified photocatalyst mixtures that were six times more active than the initial formulations, selecting beneficial components and deselecting negative ones. Our strategy uses a dexterous[19,20] free-roaming robot[21-24], automating the researcher rather than the instruments. This modular approach could be deployed in conventional laboratories for a range of research problems beyond photocatalysis.

The mobile robot platform is shown in Figure 1a and Extended Data Fig. 1. It can move freely in the laboratory and locates its position using a combination of laser scanning coupled with touch feedback for fine positioning (Methods & Supplementary Movie S1). This gave an (x, y) positioning precision of ± 0.12 mm and an orientation precision of θ ± 0.005 degrees within a standard laboratory environment with dimensions 7.3 m x 11 m (Figure 1b; Extended Data Fig. 2; Supplementary Figs. S1–10). This precision allows the robot to carry out dexterous manipulations at the various stations in the laboratory (Figure 1; Extended Data Fig. 3) that are comparable to those performed by human researchers, such as handling sample vials and operating instruments. The robot has human-like dimensions and reach (Fig. 1a,d) and it can therefore operate in a conventional, unmodified laboratory. Unlike many automated systems that can dispense only liquids, this robot dispenses both insoluble solids and liquid solutions with high accuracy and repeatability (Supplementary Figs. S12, S13, S16–S20), broadening its utility in materials research. Factoring in the time needed to recharge the battery, this robot can operate for up to 21.6 h per day with optimal scheduling. The robot uses laser scanning and touch feedback, rather than a vision system. It can therefore operate in complete darkness, if needed (Supplementary Movie S2), which is an advantage when carrying out light-sensitive photochemical reactions, as here. The robot arm and the mobile base comply with safety standards for collaborative robots, allowing human researchers to work within the same physical space (section 1.5, Supplementary Information). A movie of the robot operating an autonomous experiment over a 48-hour period is shown in Supplementary Movie S1.

The benefits of combining automated experimentation with a layer of artificial intelligence (AI) have been demonstrated for flow reactors[25], photovoltaic films[13], organic synthesis[8-10,14], perovskites[26], and for formulation problems[18]. However, to date, no approaches have integrated mobile robotics with AI for chemical experiments. Here, we built Bayesian optimization[16-18] into a mobile robotic workflow to conduct photocatalysis experiments within a 10-dimensional space. Semiconductor photocatalysts that promote overall water splitting to produce both hydrogen and oxygen are still quite rare[15]. For many catalysts, a sacrificial hole scavenger is needed to produce hydrogen from water, such as triethylamine (TEA)[27] or triethanolamine (TEOA)[28], and these amines are irreversibly decomposed in the reaction. It has proved difficult to find alternative hole scavengers that compete with these organic amines[29].

Our objective was to identify bioderived hole scavengers with efficiencies that match petrochemical amines and that are not irreversibly decomposed, with the long-term aim of developing reversible redox shuttles. The photocatalyst that we chose was P10, a conjugated polymer that shows good hydrogen evolution rates in the presence of TEOA[28]. We first used the robot to screen 30 candidate hole scavengers (Extended Data Fig. 4). This was done using a screening approach, without any AI. Initially, the robot loads a solid-dispensing station that weighs any solid components into sample vials (Figure 1c), in this case the catalyst, P10. Next, the vials are transported 16 at a time in a rack to a dual liquid dispensing station (Extended Data Fig. 3c), where the liquid components are added; here, 50 g L^{-1} aqueous solutions of the candidate hole scavengers (Supplementary Movies S3 & S4). The robot then places the vials into a capping station, which caps the vials under nitrogen (Supplementary Fig. S21; Supplementary Movie S5). Optionally, the capped vials are then placed into a sonication station (Supplementary Fig. S23; Supplementary Movie S3) to disperse the solid catalyst in the aqueous phase. The vials are then transported to a photolysis station, where they are illuminated with a mixture of ultraviolet and visible light (Fig. 1a; Extended Data Fig. 3b; Supplementary Fig. S24; Supplementary Movie S6). After photolysis, the robot transfers the vials to a head space gas chromatograph (GC) station where the gas phase is analyzed for hydrogen (Figure 1d) prior to storage of completed samples (Figure 1e). Except for the capping station and the photolysis station, which were built specifically for this workflow, the other stations used commercial instruments with no physical hardware modifications: the robot operates them in essentially the same way that a human researcher would.

Conditional automation was used in this hole scavenger screen to repeat any hits; that is, samples that showed a hydrogen evolution rate (HER) of >200 µmol g^{-1} h^{-1} were automatically re-analyzed five times. Most of the 30 scavengers produced little or no hydrogen (Extended Data Fig. 4), except for L-ascorbic acid (256 ± 24 µmol g^{-1} h^{-1}) and L-cysteine (1201± 88 µmol g^{-1} h^{-1}). Analysis by ^{1}H NMR spectroscopy showed that L-cysteine was cleanly converted to L-cystine (Supplementary Fig. S32), indicating that it may have potential as a reversible redox shuttle in an overall water splitting scheme[30].

While it showed promise as a hole scavenger, L-cysteine produced much less hydrogen than an aqueous solution of TEOA at the same gravimetric concentration (2985 ± 103 µmol g^{-1} h^{-1} at 50 g L^{-1}). We therefore sought to increase the hydrogen evolution rate of the P10 / L-cysteine system by using an autonomous robotic search based on five hypotheses (Figure 2a). The first hypothesis was that dye sensitization might improve light absorption and hence the HER, as found for the structurally related covalent organic framework, FS-COF[31]. Here, three dyes were investigated (Rhodamine B, Acid Red 87[31], and methylene blue). Second, we hypothesized that pH might influence the catalytic activity (NaOH addition). The third hypothesis was that ionic strength could also be important[32] (NaCl addition). Catalyst wettability is known to be factor in photocatalytic hydrogen evolution using conjugated polymers[33], so the addition of surfactants (sodium dodecyl sulfate, SDS, and

polyvinylpyrrolidone, PVP) formed our fourth hypothesis. Fifth, we speculated that sodium disilicate might act as a hydrogen-bonding anchor for the scavenger, L-cysteine, or for the dyes, based on the observation that it aids in the absorption of dyes onto the surface of carbon nitride[34].

These five hypotheses had the potential to be synergistic or anti-synergistic; for example, ionic strength could either enhance or decrease dye absorption onto the surface of the photocatalyst. We therefore chose to explore all five hypotheses at once. This involved the simultaneous variation of the concentration of P10, L-cysteine, the three dyes, NaOH, NaCl, the two surfactants, and sodium disilicate, which equates to a 10-variable search space (Figure 2a). The space was constrained by the need to keep a constant liquid volume (5 mL) and therefore head space for GC analysis and by the minimal resolution for liquid dispensing module (0.25 mL) and solid dispensing module (1 mg).

Problems of this type are defined by a simplex that scales exponentially with size (Figure 2b). For this specific search space, there were more than 98 million points. Full exploration of such a space is unfeasible, so we developed an algorithm that performs Bayesian optimization based on Gaussian process regression and parallel search strategy[35] (see Methods). To generate a new batch, we build a surrogate model predicting hydrogen evolution rate of potential formulations based on the measurements performed so far and quantify the uncertainty of prediction. Subsequent sampling points are chosen using a capitalist acquisition strategy, where a portfolio of upper confidence bound functions is generated on an exponential distribution of greed to create markets of varying risk aversity, which are searched for global maxima. Each market is given an agent that searches to return a global maximum, or batch of k-best maxima. The uneven distribution of greed allows for some suggested points to be highly exploitative, some to be highly explorative, and most to be balanced, thus making the strongest use of the parallel batch experiments.

The output from this autonomous robotic search is shown in Figure 3a. The baseline hydrogen evolution rate for P10 and L-cysteine only (5 mg P10 in 5 mL of 20 g L^{-1} L-cysteine) was 3.36 ± 0.30 µmol h^{-1}. Given that the robot would operate autonomously over multiple days, this two-component mixture was repeated throughout the search (two samples per batch) to check for long-term experimental stability (black squares in Figure 3a). Initially, the robot started with random conditions and discovered multicomponent catalyst formulations that were mostly less active than P10 and L-cysteine alone (first 22 experiments, Figure 3a). The robot then discovered that adding NaCl provides a small improvement to the HER, validating the hypothesis that ionic strength is important. In the same period, the robot found that maximizing both P10 and L-cysteine increased the HER. In further experiments (15–100), the robot discovered that none of the three dyes nor the two surfactants improves the HER; indeed, they are all detrimental, counter to our first and fourth hypotheses. These five components were therefore deselected after around 150 experiments (Figure 4); that is, after about 2 days in real experimental time (Fig. 3a). Here, P10 differs from the structurally related crystalline framework, FS-COF, where the addition of Acid Red 87 was increased the HER[31]. After 30 experiments, the robot learned that adding sodium disilicate improves the HER significantly in the absence of dyes (up to 15 µmol after 300 experiments), while deprioritizing the addition of NaOH and NaCl. After 688 experiments, which amounted to 8 days of autonomous searching, the robot found that the optimum catalyst formulation is a mixture of NaOH, L-cysteine, sodium disilicate, and P10, giving a HER of 21.05 µmol h^{-1}, which was six times higher than the starting conditions.

A number of scientific conclusions can be drawn from these data. Increased ionic strength is beneficial for hydrogen production (NaCl addition), but not as beneficial as increasing the pH (NaOH / sodium disilicate addition), which also increases the ionic strength. We had not investigated surfactant addition before, but for the two surfactants studied here, at least, the effect on catalytic activity is purely negative. Intriguingly, the dye sensitization that we

observed for a structurally similar covalent organic framework, FS-COF,[31] does not translate to this polymer, P10, possibly because the COF is porous whereas P10 is not.

To explore the dependence of the algorithmic search performance on the random starting conditions, we carried out 100 *in silico* virtual searches, each with a different random starting point, using a regression model and random noise to return virtual results (Supplementary Information, section 7). Around 160 virtual experiments were needed, on average, to find solutions with 95% of the global maximum HER (Figure 5).

We estimate that it would have taken a human researcher several months to explore these five hypotheses in the same level of detail using standard, manual approaches (Supplementary Fig. S31). Manual hydrogen evolution measurements require about 0.5 days of researcher time per experiment (1000 experiments = 500 days). The semi-automated robotic methods that we developed recently[33] can perform 100 experiments per day (half-day to set up, plus half-day for automatic dispensing and measurement; 1000 experiments = 10 days = 5 days of dedicated researcher time). The autonomous robot that we present here also requires half a day to set it up initially, but it then runs unattended over multiple days (1000 experiments = 0.5 days of researcher time). Hence, the autonomous workflow is 1000 times faster than manual methods, and at least 10 times faster than semi-automated but non-autonomous robotic workflows. It is unlikely that a human researcher would have persevered with this multivariate experiment using manual approaches since it might have taken 50 experiments / days to locate even a modest enhancement in the HER (Fig. 3a). The platform allows us to tackle search spaces of a size that would otherwise be impossible, which is an advantage for problems where our current level of understanding does not allow us to reduce the number of candidate components to a more manageable number. There were 10 components in the example given here, but search spaces with up to at least 20 components should be tractable with some modifications to the algorithm.

It took a significant investment of time to build this workflow (approximately 2 years), but once operating with a low error rate (Supplementary Fig. S38), it can be used as a routine tool. The time required to implement this approach in another laboratory would be much lower, since much of the 2-year development timescale involved core protocols and software that are transferable to other research problems. Also, this modular approach to laboratory automation uses instruments in a physically unmodified form and it will be straightforward to add further modules, such as NMR or X-ray diffraction, now that the basic principles are in place. This modularity makes our strategy applicable to a wide range of research problems beyond chemistry. The speed and efficiency of the method allow the exploration of large multivariate spaces, and the autonomous robot has no confirmation bias[36]; this raises the prospect of emergent function in complex, multi-component materials that we could not design in the conventional sense. Autonomous mobile robots could also have extra advantages for experiments with especially hazardous materials, or where traceability and auditing are important, such as in pharmaceutical processes.

This approach also has some limitations. For example, the Bayesian optimization is blind in that all components have equal initial importance. The robotic search does not capture existing chemical knowledge, nor include theory or physical models: there is no computational brain. Also, this autonomous system does not at present generate and test scientific hypotheses by itself.[37] In the future, we propose to fuse theory and physical models with autonomous searches: for example, computed structures and properties[1-5] could be used to bias searches towards components that have a higher likelihood of yielding the desired property. This will be important for search spaces with even larger numbers of components where purely combinatorial approaches may become inefficient. To give one example, energy-structure-function maps[38] could be computed for candidate crystalline components to provide Boltzmann energy weightings[39] for calculated properties, such as a charge transport or optical gap, to bias the robotic search.

Author contributions. B.B. developed the workflow, developed and implemented the robot positioning approach, wrote the control software, designed the bespoke photocatalysis station, and carried out experiments. P.M.M. and V.V.G. developed the optimiser and its interface to the control software. X.L. advised on the photocatalysis workflow. C.M.A., Y.B. and X.L. synthesized materials. Y.B. performed kinetic photocatalysis experiments. X.W. performed NMR analysis and synthesized materials. B.L. carried out initial scavenger screening. R.C. and N.R. helped to build the bespoke stations in the workflow. B.H. analysed the robustness of the system, assisted with the development of control software, and operated the workflow during some experiments. B.A. helped to supervise the automation work. R.S.S. helped to supervise the photocatalysis work. A.I.C. conceived the idea, set up the five hypotheses with BB, and coordinated the research team. Data was interpreted by all authors and the manuscript was prepared by A.I.C., B.B., P.M.M., V.V.G., and R.S.S

METHODS

Robot specifications. The robot used was a KUKA Mobile Robot (KMR) mounted on a KUKA Mobile Platform (KMP) base (Fig. 1a; Extended Data Fig. 1). The robot arm has a maximum payload of 14 kg and a reach of 820 mm. The KMP base can carry payloads of up to 200 kg. The robot arm and the mobile base combined has a mass of approximately 430 kg. The movement velocity of the robot was restricted to 0.5 m s^{-1} for safety reasons (Supplementary Information, section 1.5). A multipurpose gripper was designed to grasp 10 mL GC sample vials, solid dispensing cartridges, and a 16-position sample rack (Extended Data Fig. 5), thus allowing a single robot to carry out the whole variety of tasks required for this workflow. This robot was specified to be a flexible platform for a wide range of research tasks beyond those exemplified here; for example, the 14 kg payload capacity for the arm is not fully utilized in these experiments (one rack of filled vials has a mass of 580 g), but it could allow for manipulations such as opening and closing the doors of certain equipment. Likewise, the height and reach of the robot allows for operations such as direct loading of samples into the GC instrument (Fig. 1d). By contrast, a smaller and perhaps less expensive robot platform might require an additional, dedicated robot arm to accomplish this, or inconvenient modifications to the laboratory, such as lowering bench heights.

Robot navigation. In a process analogous to simultaneous localisation and mapping (SLAM)[40], the robot tracks a cloud of possible positions, and updates its position to the best fit between the output of its laser scanners and the map for each position in the cloud. The position of the robot is determined by x & y and θ, where x & y describe its position on the map, and θ its orientation angle. Histograms of the robot position measured over 563 movements are shown in Supplementary Figs. S2–S5, which show that the (x, y) positioning precision was better than ± 10 mm and the orientation precision was less than ± 2.5 degrees, as achieved within a real, working laboratory environment. This level of precision allows navigation to the various experimental stations in the laboratory, but it does not allow fine manipulations, such as placement of sample vials. The precision was therefore enhanced by using a touch-sensitive 6-point calibration method. Here, the robot touches six points on a cube that is associated with each experimental station to find the position and orientation of the cube relative to the robot (Supplementary Figs. S7–S11). This increased the positioning precision to ± 0.12 mm and the orientation precision to ± 0.005 degrees. This makes it possible for the robot to operate instruments and to carry out delicate manipulations

such as vial placements at a level of precision that is broadly comparable to a human operator.

Experimental stations. The workflow comprised six steps, each with its own station. Solid dispensing was carried out with a Quantos QS30 instrument (Mettler Toledo) (Fig. 1c; Supplementary Fig. S11; Extended Data Fig. 3a; Supplementary Movie S3). Liquid dispensing was carried out with bespoke system that used a 200 series Mini Peristaltic Pump (Williamson) and a PCG 2500-1 scale (Kern), to dispense liquids gravimetrically using feedback loop (section 2.2 in Supplementary Information; Supplementary Movie S5). This system showed excellent precision and accuracy for a range of aqueous and non-aqueous liquids over 20,000 dispenses (Figures S16–S17, S19, S20). A bespoke instrument was built (Labman) to allow both for sample inertization (to exclude oxygen) and cap crimping in one step. It would be straightforward to modify this platform to allow other gases to be introduced; for example, to study photocatalytic CO_2 reduction. The instrument used caps from a vibratory bowl feeder to cap crimp 10 mL headspace vials (section 2.2 in Supplementary Information; Supplementary Movie S5). If required, a sonication station was used to disperse the solid photocatalyst in the aqueous solution, prior to reaction (Supplementary Fig. S23). Photolysis was carried out in bespoke photolysis station (Fig. 1a) that uses vibration to agitate liquids and a light source that is composed of BL368 tubes and LED panels (Extended Data Fig. 5b; Supplementary Fig. S24; Supplementary Movie S6). GC measurements were performed with 7890B GC with a 7697A Headspace Sampler from Agilent GC (Supplementary Movie S3; Extended Data Fig. 3d). The experimental stations were controlled by a process management system (PMS) module, which contains all of the process logic for controlling the labware. Communication between the PMS and the stations was achieved using various communication protocols (TCP/IP over WIFI/LAN; RS-232), as detailed in section 2.7 in the Supplementary Information (Supplementary Fig. S28).

Autonomous search procedure and scheduling. The robot worked with batches of 16 samples per sample rack and ran 43 batches (688 experiments) during the search. Of these 688 experiments, 11 results were discarded because of workflow errors or because the system flagged that the oxygen level was too high (faulty vial seal). It took, on average, 183 minutes to prepare and photolyze each batch of samples and then 232 minutes per batch to complete the GC analysis. The detailed timescales for each of the step in the workflow are shown in Extended Data Fig. 6. The work was heuristically scheduled in parallel, with the robot starting the oldest available scheduled job. While the robot was working on one job, other instruments, such as the solid dispenser, the photolysis station, and the GC, worked in parallel. This system can process up to six batches at once, but given the timescales for this specific workflow, where the preparation/reaction time is approximately equal to the analysis time, the robot processed two batches simultaneously. That is, it prepared samples and ran photolysis for one batch while analyzing the hydrogen produced for the second batch using GC. The robot recharged its battery automatically in between two jobs when the battery charge reached a 25% threshold. The robot was charged but idle for approximately 32% of the time in this experiment, largely because of time spent waiting for the GC analysis, which is the slow step. In principle, this time could be utilized to run other experiments in parallel. The autonomous workflow was programmed to alert the operator automatically when the system is out of stock (*e.g.*, it ran out of sample vials, or stock solutions were low), or if a part of the workflow failed (Supplementary Information, section 8). Most errors could be reset remotely without being in the laboratory since all stations were equipped with 24/7 CCTV cameras (Supplementary Fig. S39).

Bayesian search algorithm. The AI guidance for the autonomous mobile robot was a batched, constrained, discrete Bayesian optimization algorithm. Traditionally, Bayesian optimization is a serial algorithm tasked with finding the global maximum of an unknown objective function[16]. Here, this equates to finding the optimal set of concentrations in a multicomponent mixture for photocatalytic hydrogen generation. The algorithm builds a

model that can be updated and queried for the most promising points to inform subsequent experiments. This surrogate model is constructed by first choosing a functional prior $\phi_{prior}(\theta)$, informed by existing chemical knowledge (if any). Given data $\mathcal{D}$ and a likelihood model $\phi_{likelihood}(\mathcal{D}|\theta)$, this yields a posterior distribution of models using Bayes' theorem:

$$(1) \quad \phi_{posterior}(\theta|\mathcal{D}) = \frac{\phi_{likelihood}(\mathcal{D}|\theta)\phi_{prior}(\theta)}{\phi(\mathcal{D})}$$

The Gaussian process prior used a Matern similarity kernel, constant scaling, and homoscedastic noise[41]. This composite kernel allows for variable smoothness, catalytic activity, and experimental noise. The form and respective hyperparameters were refined using cross-validation on other, historical photocatalysis datasets (350 experiments). Other alternatives for a functional prior included Bayesian neural networks[17]; but Gaussian processes were selected here for robustness and flexibility[42]. An acquisition function, α_{UCB}, was assembled from the posterior distribution by considering the posterior mean, $\mu(x)$, and uncertainty, $\sigma(x)$. The maximum of this function was then used as the next suggested experiment. To balance exploitation (prioritizing areas where the mean is expected to be largest) and exploration (prioritizing areas where the least is known), we used an upper confidence bound that is dependent on a single hyperparameter, β, to govern how 'greedy' (exploitative) the search is:

$$(2) \quad \alpha_{UCB}(x; \mathcal{D}) := \mu(x) + \beta\sigma(x)$$

The portfolio of acquisition functions for different values of β, which we call markets, was used to generate a batch. This 'capitalist' approach has the advantage of simple parallelization and is robust across variable batch sizes[35]. Our method allowed us to constrain the sum of all liquid components to 5 mL to allow a constant gas headspace volume for GC analysis. The sum total volume constraint was handled during the market searches; discretization, which was determined by instrument resolution, was handled after the market searches. The market search was completed using a large initial random sampling followed by a batch of seeded local maximizations using the SLSQP algorithm. This maximization occurs on a continuous space, and the results are placed into discrete bins following the experimental precision. The explored space is tracked as continuous and discrete variables for respective model building and acquisition function maximization. The algorithm was implemented using the scikit-learn and in scipy packages[43].

Materials and synthetic procedures. The polymeric photocatalyst P10 was synthesized and purified according to a modification on a literature procedure[44] (Supplementary Information, section 10). For solid dispensing, the polymer was ground with mortar and pestle before use. Sodium disilicate was obtained as a free sample from Silmaco. Tap water was purified with PURELAB Ultra System. All other materials were purchased from Sigma-Aldrich and used as received.

Data availability. The implementation of liquid dispense station, photolysis station, and workflow, along with 3D designs for labware developed in the project, are available at https://bitbucket.org/ben_burger/kuka_workflow, the code for the robot at and the Bayesian optimizer is available at https://github.com/Taurnist/kuka_workflow_tantalus and https://github.com/CooperComputationalCaucus/kuka_optimizer. Additional design details can be obtained from the authors upon request.

Figure 1. Autonomous mobile robot and experimental stations. a, Photograph showing robot loading samples into the photolysis station. **b**, Map of the lab generated by laser scanning showing positions of the eight stations; the orange crosshairs indicate recorded navigation locations, the robot position is indicated by the green rectangle. Inputs 1–3 are areas for the storage of empty vials or completed sample racks. **c**, Robot loading empty sample vials into the solid-dispensing station prior to dispensing the photocatalyst. **d**, Loading the gas chromatograph (GC) station with a new rack of samples for analysis. **e**, Storing racks of completed samples in Input Station 1 after GC analysis.
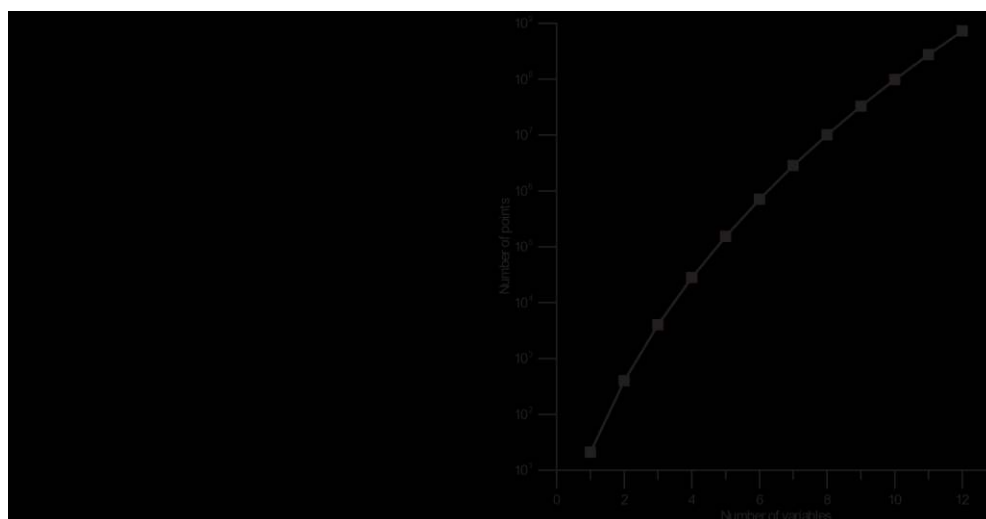
Figure 2. Hypothesis-led autonomous search strategy. a, The robot searches chemical space to optimize the activity of the photocatalyst / scavenger combination according to five separate hypotheses. It does this by simultaneously varying the concentration of the ten chemical species shown here. **b**, Plot showing the size of the simplex, or the search space, created with a discretization of 19 concentrations for each liquid and 21 concentration levels for the solid catalyst, P10, which corresponds to the solid / liquid dispensing precision over the constrained space of the experiment. For this 10-component problem, the full simplex has 98,423,325 points.
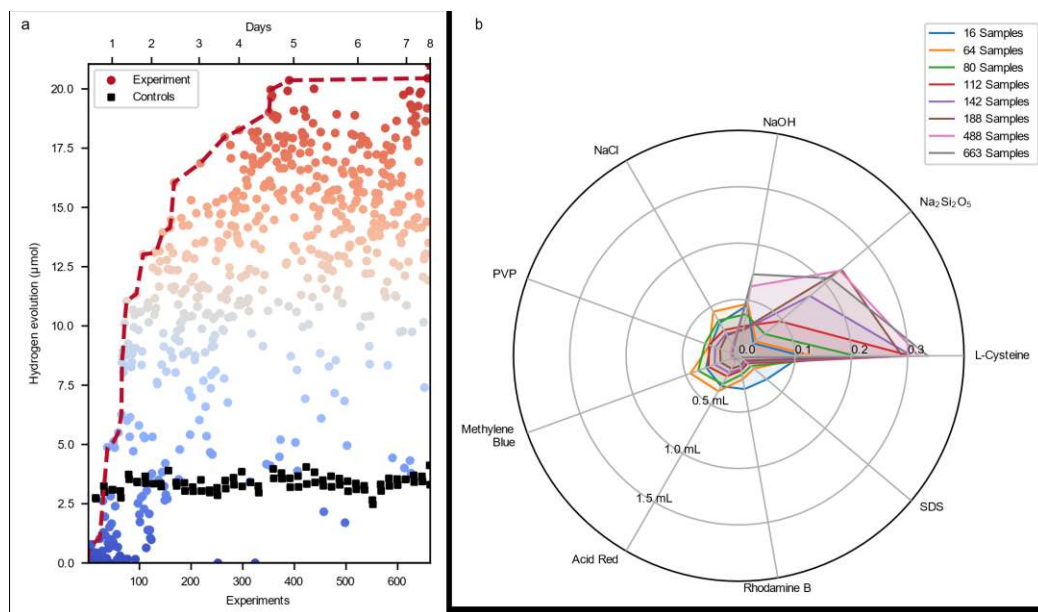
Figure 3. Output from the autonomous robotic search. a, Plot showing hydrogen evolution achieved per experiment in an autonomous search that extended over 8 days. Sixteen experiments were performed per batch, along with two baseline controls. The baseline hydrogen evolution was 3.36 ± 0.30 μmol (black squares). The maximum rate attained after 688 experiments was 21.05 μmol h^{-1}. The robot made 319 moves between stations and travelled a total distance of 2.17 km during this 8-day experiment. **b**, Radar plot showing the evolution of the average sampling of the search space in mL; scale denotes fraction of maximum solution volume dispensed. The starting conditions (Batch 1) were chosen randomly. The best catalyst formulation found after 43 batches contained P10 (5 mg), NaOH (6 mg), L-cysteine (200 mg) and Na$_2$Si$_2$O$_5$ (7.5 mg) in water (5 mL).
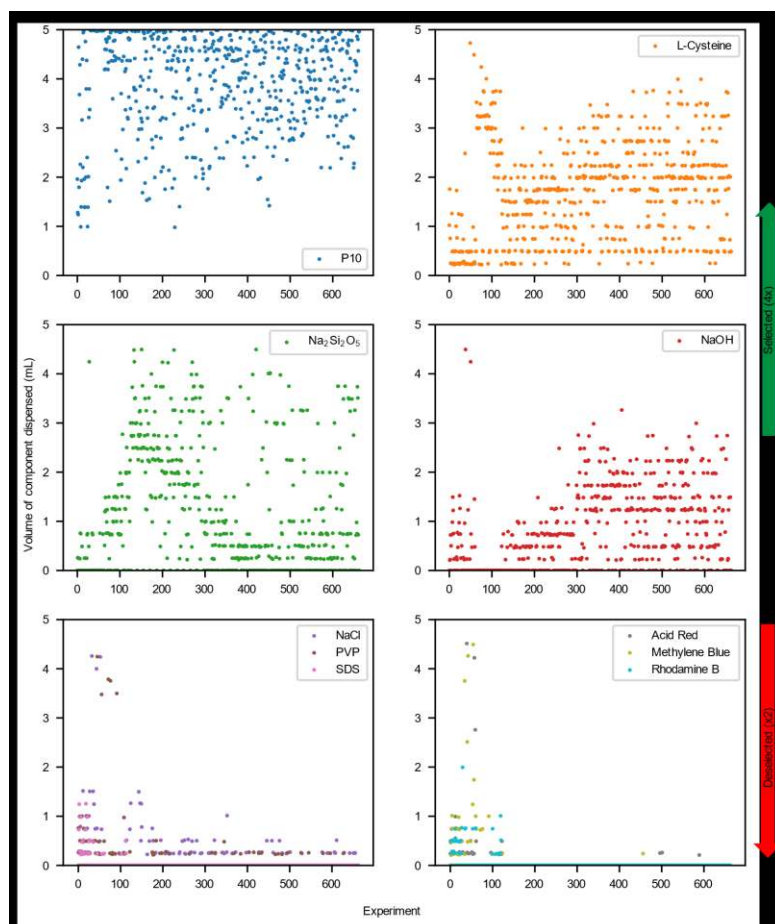
Figure 4. Selection and deselection of photocatalyst formulation components. Plots showing the mass (mg, P10) or volume (mL, all other components) dispensed for the various components in the search space as a function of experiment. The photocatalyst, P10, and the scavenger, L-cysteine, are selected, along with sodium disilicate ($Na_2Si_2O_5$) and NaOH. All other components were deselected after around 150 experiments. The three dyes and the two surfactants had a negative effect on the hydrogen evolution rate. NaCl had a small positive effect, but less so than the four selected components, and it was therefore deselected. Note that NaOH was initially deselected, and not included in experiments 15–283 (see black arrow), while $Na_2Si_2O_5$ and L-cysteine were favoured. The positive effect of NaOH was initially masked by negative components such as the dyes. Later in the search, NaOH was favoured, ultimately in preference to $Na_2Si_2O_5$, illustrating the benefit of using an uneven distribution of greed in the search.
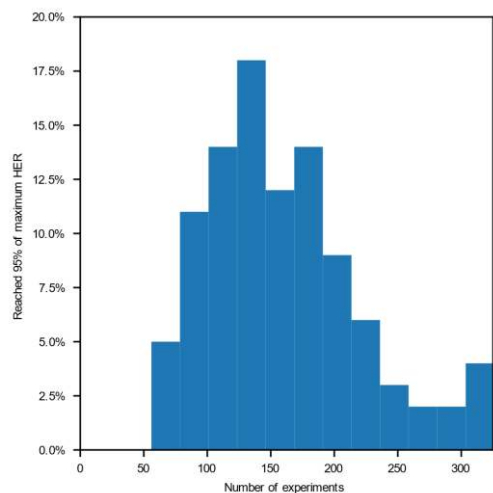
Figure 5. Virtual, *in silico* experiments. Histogram showing the number of virtual experiments needed to reach 95% of the optimal hydrogen evolution rate (HER), as determined by carrying out 100 *in silico* searches, each with a different random starting point.
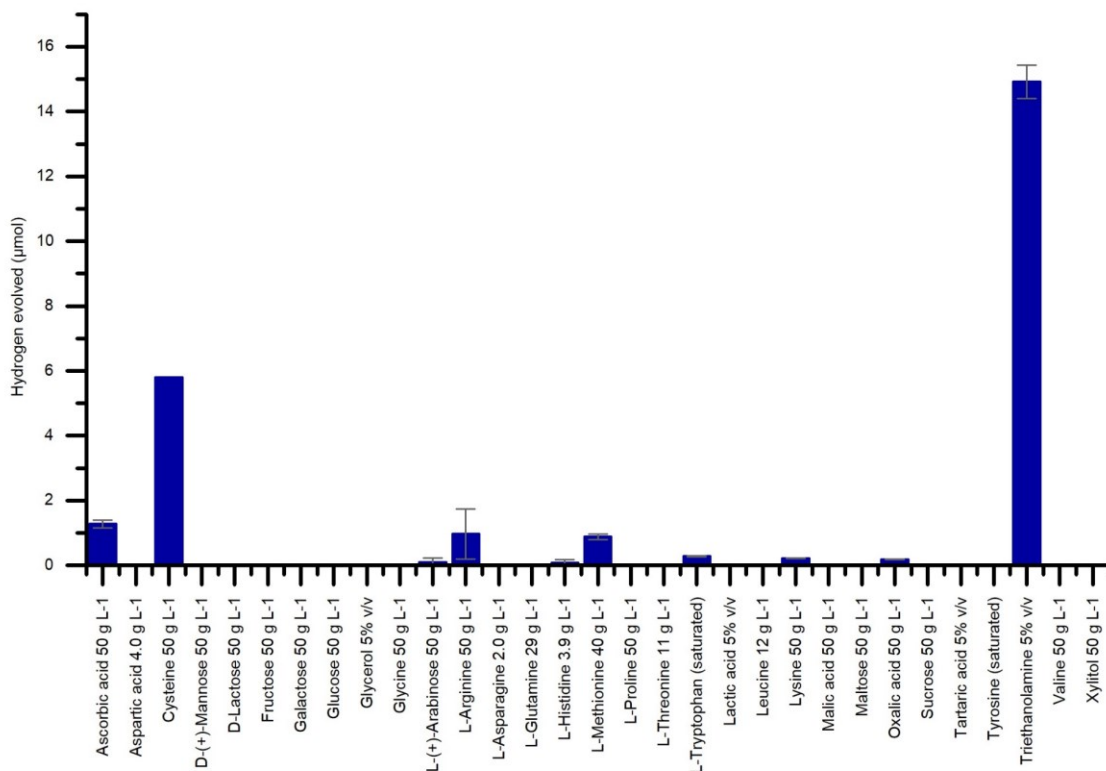
Extended Data Fig. 1. Mobile robotic chemist. The mobile robot used for this project, shown here performing a 6-point calibration with respect to the black cube that is attached to the bench, in this case associated with the solid cartridge station (see also Supplementary Fig. S11; Extended Data Fig. 3a).
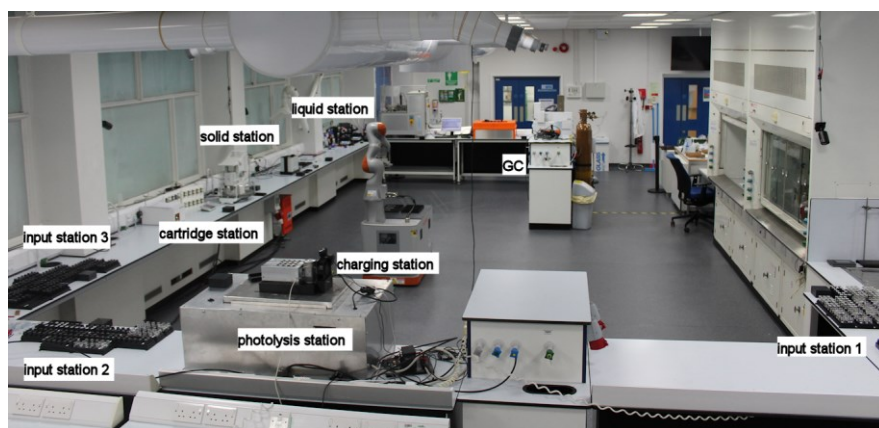
Extended Data Fig. 2. Laboratory space used for the autonomous experiments. The key locations in the workflow are labeled. Other than the black location cubes that are fixed to the benches to allow positioning (see also Extended Data Figure 1), the laboratory is otherwise unmodified.
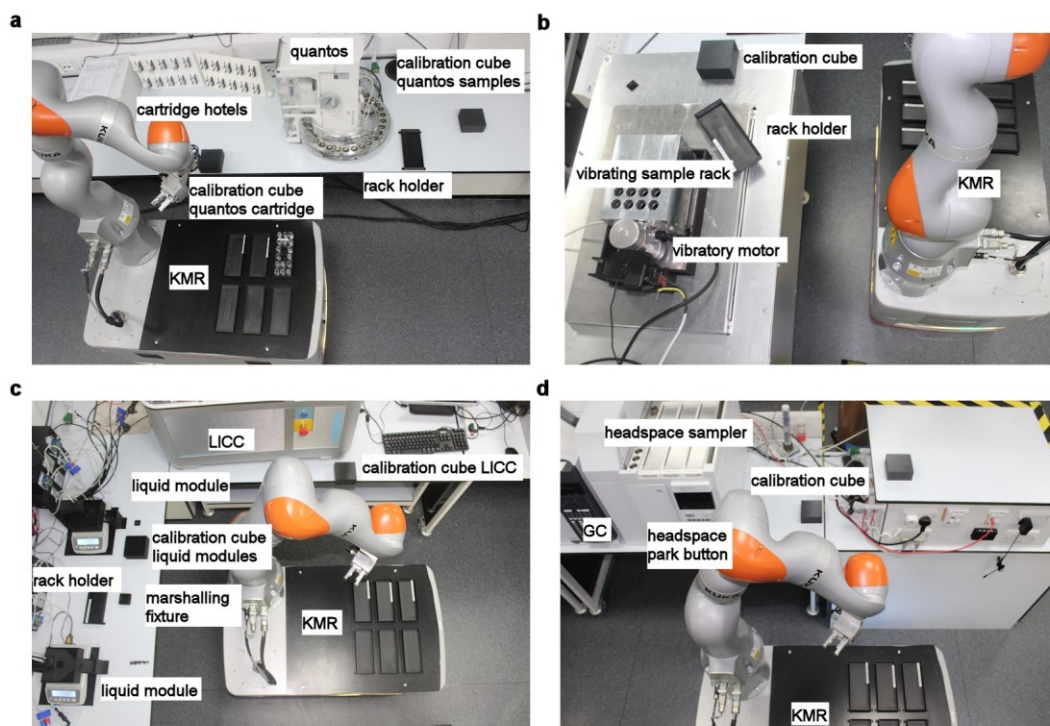
Extended Data Fig. 3. Stations in the workflow. a, Photograph showing the robot at the solid dispensing / cartridge station. The two cartridge hotels can hold up to 20 different solids; here, four cartridges are located in the hotel on the left. The door of the Quantos dispenser is opened using software prior to loading the correct solid dispensing cartridge into the instrument (Supplementary Movie S3). Since the KMR robot is free-roaming and has an 820 mm reach, it would be simple to extend this modular approach to 100's or even 1000's of different solids given sufficient laboratory space. **b**, Photograph showing the KMR robot at the photolysis station (see also Supplementary Movies S3, S6). **c**, Photograph showing the KMR robot at the combined liquid handling / capping station. The robot can reach both the liquid stations and the capper-crimper station after 6-point positioning, such that liquid addition, headspace inertization, and capping can be carried out in a single coordinated process (see Supplementary Movies S3, S5), without any position recalibration. **d**, Photograph of the KMR robot parked at the headspace gas chromatography (GC) station. The GC instrument is a standard commercial instrument and was unmodified in this workflow.

**Extended Data Fig. 4. Hydrogen evolution rates for candidate bioderived sacrificial hole scavengers.** Results of a robotic screen for sacrificial hole scavengers using the mobile robot workflow. Of the 30 bioderived molecules trialed, only cysteine was found to compete with the petrochemical amine, triethanolamine. Scavengers are labeled with the concentration of the stock solution that was used (5 mL volume; 5 mg P10).

Extended Data Fig. 5. Multipurpose gripper used in the workflow. The gripper is shown grasping various objects. **a**, The empty gripper; **b**, gripper holding a capped sample vial (top grasp); **c**, gripper holding an uncapped sample vial (side grasp); **d**, gripper holding a solid-dispensing cartridge, and; **e**, gripper holding a full sample rack using an outwards grasp that locks into recesses in the rack. The same gripper was also used to activate the gas chromatograph using a physical button press (see Supplementary Movie S3; 1 min 52 sec).
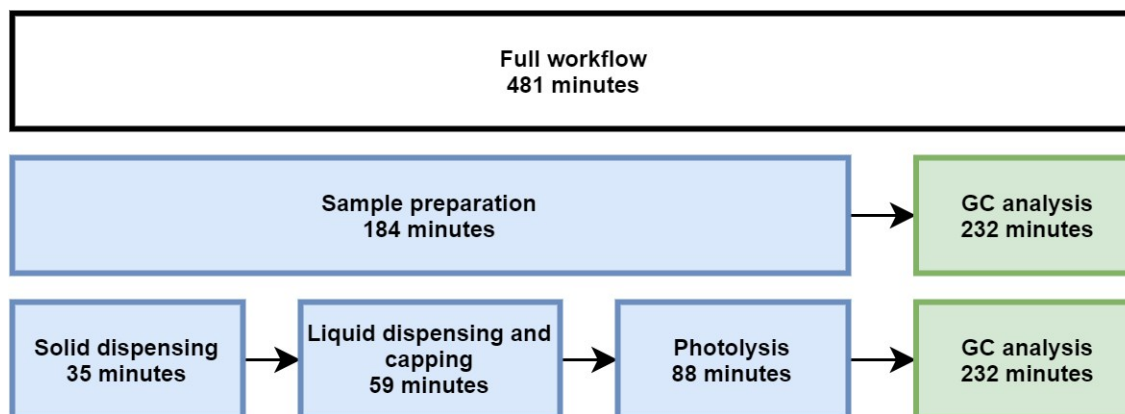
Extended Data Fig. 6. Timescales for steps in the workflow. Average timescales for the various steps in the workflow (sample preparation, photolysis and analysis) for a batch of 16 experiments. These averages were calculated over 46 separate batches. These average times include the time taken for the loading and unloading steps (*e.g.*, the photolysis time itself was 60 minutes; loading and unloading takes an average of 28 minutes per batch). The slowest step in the workflow is the GC analysis.

References

1   Kang, K., Meng, Y. S., Bréger, J., Grey, C. P. & Ceder, G. Electrodes with high power and high capacity for rechargeable lithium batteries. *Science* **311**, 977–980 (2006).
2   Woodley, S. M. & Catlow, R. Crystal structure prediction from first principles. *Nature Mater.* **7**, 937–946 (2008).
3   Gómez-Bombarelli, R. *et al.* Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **15**, 1120–1127 (2016).
4   Collins, C. *et al.* Accelerated discovery of two crystal structure types in a complex inorganic phase field. *Nature* **546**, 280-284 (2017).
5   Davies, D. W. *et al.* Computer-aided design of metal chalcohalide semiconductors: from chemical composition to crystal structure. *Chem. Sci.* **9**, 1022-1030 (2018).
6   King, R. D. Rise of the robo scientists. *Sci. Am.* **304**, 72–76 (2011).
7   Li, J. *et al.* Synthesis of many different types of organic small molecules using one automated process. *Science* **347**, 1221–1226 (2015).
8   Dragone, V., Sans, V., Henson, A. B., Granda, J. M. & Cronin, L. An autonomous organic reaction search engine for chemical reactivity. *Nat. Commun.* **8**, 15733 (2017).
9   Bédard, A.-C. *et al.* Reconfigurable system for automated optimization of diverse chemical reactions. *Science* **361**, 1220–1225 (2018).
10   Granda, J. M., Donina, L., Dragone, V., Long, D.-L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377–381 (2018).
11   Tabor, D. P. *et al.* Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**, 5–20 (2018).
12   Langner, S. *et al.* Beyond ternary OPV: High-throughput experimentation and self-driving laboratories optimize multi-component systems. *arXiv preprint arXiv:1909.03511* (2019).
13   MacLeod, B. P. *et al.* Self-driving laboratory for accelerated discovery of thin-film materials. *arXiv preprint arXiv:1906.05398* (2019).
14   Steiner, S. *et al.* Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211 (2019).
15   Wang, Z., Li, C. & Domen, K. Recent developments in heterogeneous photocatalysts for solar-driven overall water splitting. *Chem. Soc. Rev.* **48**, 2109–2125 (2019).
16   Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & Freitas, N. d. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* **104**, 148–175 (2016).
17   Häse, F., Roch, L. M., Kreisbeck, C. & Aspuru-Guzik, A. Phoenics: A Bayesian optimizer for chemistry. *ACS Central Sci.* **4**, 1134-1145 (2018).
18   Roch, L. M. *et al.* ChemOS: orchestrating autonomous experimentation. *Sci. Robot.* **3**, eaat5559 (2018).
19   Chen, C.-L., Chen, T.-R., Chiu, S.-H. & Urban, P. L. Dual robotic arm "production line" mass spectrometry assay guided by multiple Arduino-type microcontrollers. *Sensors Actuators B: Chem.* **239**, 608–616 (2017).
20   Fleischer, H. *et al.* Analytical measurements and efficient process generation using a dual–arm robot equipped with electronic pipettes. *Energies* **11**, 2567 (2018).
21   Liu, H., Stoll, N., Junginger, S. & Thurow, K. Mobile robot for life science automation. *Int. J. Adv. Robot. Syst.* **10**, 288 (2013).

22   Liu, H., Stoll, N., Junginger, S. & Thurow, K. A fast approach to arm blind grasping and placing for mobile robot transportation in laboratories. *Int. J. Adv. Robot. Syst.* **11**, 43 (2014).

23   Abdulla, A. A., Liu, H., Stoll, N. & Thurow, K. A new robust method for mobile robot multifloor navigation in distributed life science laboratories. *J. Control Sci. Eng.* **2016**, 2 (2016).

24   Dömel, A. *et al.* Toward fully autonomous mobile manipulation for industrial environments. *Int. J. Adv. Robot. Syst.* **14**, 1 (2017).

25   Schweidtmann, A. M. *et al.* Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives. *Chem. Eng. J.* **352**, 277-282 (2018).

26   Zhi, L. *et al.* Robot-accelerated perovskite investigation and discovery (RAPID): 1. Inverse temperature crystallization. doi:10.26434/chemrxiv.10013090.v1 (2019).

27   Matsuoka, S. *et al.* Photocatalysis of oligo (*p*-phenylenes): photoreductive production of hydrogen and ethanol in aqueous triethylamine. *J. Phys. Chem.* **95**, 5802–5808 (1991).

28   Shu, G., Li, Y., Wang, Z., Jiang, J.-X. & Wang, F. Poly(dibenzothiophene-*S,S*-dioxide) with visible light-induced hydrogen evolution rate up to 44.2 mmol h^{-1} g^{-1} promoted by K_2HPO_4. *Appl. Catal. B: Environ.* **261**, 118230 (2020).

29   Pellegrin, Y. & Odobel, F. Sacrificial electron donor reagents for solar fuel production. *C. R. Chim.* **20**, 283–295 (2017).

30   Sakimoto, K. K., Zhang, S. J. & Yang, P. Cysteine–cystine photoregeneration for oxygenic photosynthesis of acetic acid from CO_2 by a tandem inorganic–biological hybrid system. *Nano Lett.* **16**, 5883–5887 (2016).

31   Wang, X. *et al.* Sulfone-containing covalent organic frameworks for photocatalytic hydrogen evolution from water. *Nature Chem.* **10**, 1180-1189 (2018).

32   Schwarze, M. *et al.* Quantification of photocatalytic hydrogen evolution. *PCCP* **15**, 3466–3472 (2013).

33   Bai, Y. *et al.* Accelerated discovery of organic polymer photocatalysts for hydrogen evolution from water through the integration of experiment and theory. *J. Am. Chem. Soc.* **141**, 9063–9071 (2019).

34   Zhang, J. *et al.* H-bonding effect of oxyanions enhanced photocatalytic degradation of sulfonamides by *g*-C_3N_4 in aqueous solution. *J. Hazard. Mater.* **366**, 259-267 (2019).

35   Hutter, F., Hoos, H. H. & Leyton-Brown, K. *Parallel Algorithm Configuration*. pp. 55-70 (Springer Berlin Heidelberg, 2012).

36   Mynatt, C. R., Doherty, M. E. & Tweney, R. D. Confirmation bias in a simulated research environment: an experimental study of scientific inference. *Q. J. Exp. Psychol.* **29**, 85–95 (1977).

37   King, R. D. *et al.* Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **427**, 247–252 (2004).

38   Pulido, A. *et al.* Functional materials discovery using energy–structure–function maps, *Nature* **543**, 657–664 (2017).

39   Campbell, J. E., Yang, J. & Day, G. M. Predicted energy–structure–function maps for the evaluation of small molecule organic semiconductors. *J. Mater. Chem. C* **5**, 7574-7584 (2017).

Methods References

40   Fuentes-Pacheco, J., Ruiz-Ascencio, J. & Rendón-Mancha, J. M. Visual simultaneous localization and mapping: a survey. *Artif. Intell. Rev.* **43**, 55–81 (2015).

41   Williams, C. E. R. a. C. K. I. *Gaussian Processes for Machine Learning* (The MIT Press, 2006).

42   Matthews, A. G. d. G., Rowland, M., Hron, J., Turner, R. E. & Ghahramani, Z. Gaussian process behaviour in wide deep neural networks. *arXiv preprint* arXiv:1804.11271 (2018).

43   Millman, K. J. & Aivazis, M. Python for scientists and engineers. *Comput. Sci. Eng.* **13**, 9–12 (2011).

44   Sachs, M. *et al.* Understanding structure-activity relationships in linear polymer photocatalysts for hydrogen evolution. *Nat. Commun.* **9**, 4968 (2018).