Multi-Variable Multi-Metric Optimization of Self-Assembled Photocatalytic CO_2 Reduction Performance using Machine Learning Algorithms

Shannon A. Bonke,[a] Giovanni Trezza,[b] Luca Bergamasco,[b] Hongwei Song,[c] Santiago Rodríguez-Jiménez,[a] Leif Hammarström,[c] Eliodoro Chiavazzo[b]* and Erwin Reisner[a]*

[a] Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

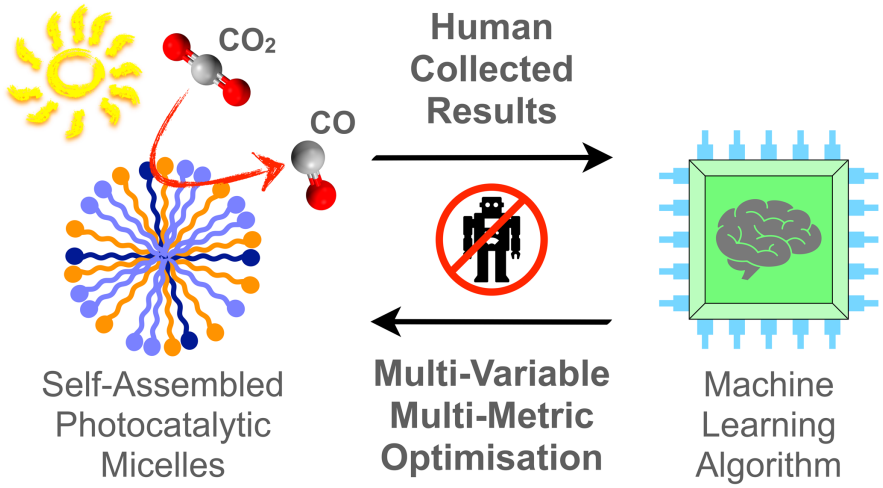
[b] Department of Energy, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy.

[c] Department of Chemistry, Ångström Laboratory, Uppsala University, Box 523, 75120 Uppsala, Sweden.

* Correspondence to: Prof E. Reisner (reisner@ch.cam.ac.uk) or Prof E. Chiavazzo (eliodoro.chiavazzo@polito.it).

Keywords: Machine Learning, Photocatalysis, CO_2 Reduction, Self-Assembly, Micelles, Shapley Additive Explanations (SHAP), Holistic Optimization

Table of Contents Figure



Abstract

The sunlight-driven reduction of CO_2 into fuels and platform chemicals is a promising approach to enable a circular economy. However, established optimization approaches are poorly suited to multi-variable multi-metric photocatalytic systems because they aim to optimize one performance metric while sacrificing the others and thereby limit overall system performance. Herein, we address this multi-metric challenge by defining a metric for holistic system performance that takes all figures of merit into account, and employ a machine learning algorithm to efficiently guide our experiments through the large parameter matrix to make holistic optimization accessible for human experimentalists. As a test platform, we employ a five-component system that self-assembles into photocatalytic micelles for CO_2-to-CO reduction, which we experimentally optimized to simultaneously improve yield, quantum yield, turnover number, and frequency while maintaining high selectivity. Leveraging the dataset with machine learning algorithms allows quantification of each parameter's effect on overall system performance. The buffer concentration is unexpectedly revealed as the dominating parameter for optimal performance, and is nearly four times more important than the catalyst concentration. The expanded use and standardization of this methodology to define and optimize holistic performance will accelerate progress in different areas of catalysis by providing unprecedented insights into performance bottlenecks, enhancing comparability, and taking results beyond comparison of subjective figures of merit.

Introduction

Catalysis relies on a complex interplay of interdependent variables that must be optimized to meet a set of performance metrics. The challenge is exemplified by multicomponent photocatalytic systems where the parameter space is increasingly difficult to navigate due to the increasing number of variables required to provide supramolecular control (*e.g.*, concentrations of reagents, additives, experimental variables). The optimization target is also unclear as there is limited standardization and a multitude of metrics to optimize (*e.g.*, yield, quantum yield, selectivity, turnover number and frequency).[1,2] Crucially, established and intuitive heuristic/human optimization approaches can only maximize 1 or 2 performance metrics simultaneously. This situation has led to selective optimizations where some metrics are prioritized in ways that have limited meaning to overall system performance because they use conditions that sacrifice other metrics; such as using a very low catalyst loading to reach a high turnover frequency (TOF) but having negligible product yield.[1] The fundamental problem is that optimization to improve all figures of merit (holistic optimization) is not possible with established protocols. The parameter space is too large to evaluate, the interactions of variables that affect multiple metrics are too complex, and no holistic figure of merit has been defined.[1,2] New strategies are required to navigate the large parameter space and extract deeper understanding into how multiple variables interact and affect each metric to control overall system performance.[3–5]

Holistic optimization first requires all figures of merit to be connected via a mathematical description (an *objective function*) that evaluates to a single scalar value representing overall system performance. Iteratively varying parameters changes the value of each figure of merit, and thereby enables the objective function to quantify the change in overall performance. However, a complex system has too many combinatorial possibilities to test them all manually, for example, testing 10 concentrations of just 5 components gives 10^5 possibilities. Finding the maximum in such a large parameter matrix could be addressed with 'brute force' high-throughput approaches to rapidly screen all possibilities until a positive result is found.[6–9] Alternatively, a learning algorithm can use Bayesian optimization to decipher how each variable affects overall system performance and vastly reduce the number of experiments required to maximize the objective function.[3,4,10–12] Because learning algorithms change multiple parameters simultaneously and capture the relationships between variables, [3,4,10–12] they can reduce the number of

experiments required to optimize a complex system into range for human experimentalists. Thereby, use of learning algorithms enables all laboratories to holistically optimize their systems without expensive robotic systems. When robotic systems become more commonplace in the future, they will be guided by learning algorithms to further enhance the rate of scientific discovery.[5,13]

Supramolecular assembly is a frontier in molecular photocatalysis and biomimicry as it can accelerate charge transfer processes and eliminate diffusion limitations by providing optimal environments for reactions, in particular photocatalytic CO_2 reduction.[14,15] While judicial design of molecular components can enable self-assembly such as through amphiphilicity as used herein,[7,16,14,15] the dependence of assembly on the components and environment results in a wide parameter space to explore and provides ideal systems to demonstrate holistic optimization of complex systems.

Herein we employ Bayesian optimization to steer an experimental campaign towards the simultaneous improvement of all key metrics for a five-component self-assembled photocatalytic molecular CO_2 reduction system. This system was selected because the interdependent variables that control supramolecular structure make it challenging to optimize and representative of complex photocatalytic systems. Defining an objective function allowed simultaneous improvement of performance metrics, namely catalyst turnover number, turnover frequency, quantum yield and moles of product, without loss of selectivity. Concurrently, the use of Bayesian optimization reduced the number of experiments required to optimize the system from a theoretical 10^5 down to ~100 and made the work accessible for a human experimentalist with standard laboratory equipment. Acquiring a systematic dataset that maps the parameter space then enabled the application of machine learning algorithms to extract relationships between the dependent variables. Furthermore, the role and importance of different parameters on the overall performance was revealed using Shapley additive explanations (SHAP).[17–19] Finally, the problem dimensionality was reduced and parameter regions likely to further enhance system performance are exposed by controlling feature grouping.[20] Thereby, the system was holistically optimized and deeper insight into the ruling parameters for each figure of merit was extracted.

Results and Discussion

Assembly of photocatalytic micelles. Systems that self-assemble into functional supramolecular structures are attractive targets, but they are also hard to optimize due to the numerous interdependent

variables. This is exemplified herein with amphiphilic self-assembly of molecular components for photocatalytic CO_2 reduction. The system exploits the tuneability of molecular components to functionalize a CO_2 reduction catalyst and photosensitizer with alkyl tails and enable supramolecular assembly into micelles. The catalyst and photosensitizer are cationic complexes, to which alkyl tails were added to the periphery to render them amphiphilic while preserving the metal coordination environment and functionality.[16] For the catalyst, cobalt tetra-methylpyridinium porphyrin ($CoPyP_{C1}$) is an active catalyst for CO_2-to-CO conversion in aqueous media,[21,22] and was made amphiphilic by replacing the methyl groups with hexadecyl groups ($CoPyP_{C16}$; Fig 1).[14] A ruthenium tris-bipyridine was selected as it is a prototypical photosensitizer previously used for photocatalytic CO_2 reduction with cobalt porphyrins and amphiphilic variants are known.[14,16,23] To minimize additional functional groups, one bipyridine ligand was functionalized with two heptadecyl groups ($Rubpy_{C17}$; Fig 1).[23] In both cases, characterization of the complexes was consistent with previous reports and the UV-Vis spectra showed minimal changes to λ_{max} to indicate that the coordination environment was preserved (details in experimental section). For photocatalytic studies, the water-insoluble $[CoPyP_{C16}](PF_6)_4$ was solubilized in micelles from a MeCN stock solution into aqueous CO_2-saturated phosphate solution (pH 6.3; MeCN removed in CO_2 purging step). The solution also contained the $Rubpy_{C17}$ photosensitizer and sodium ascorbate as a sacrificial electron donor (reductive quencher of the photosensitizer[14,21,24]); test conditions were 15 min illumination of 1 mL solution at λ = 447 nm using a 2.3 W LED at 25 °C with orbital convection.

The photosensitizer commonly serves as a single electron donor while two electrons are required for CO_2-to-CO conversion, which has led to the established practice to have excess photosensitizer in photocatalytic tests, with 20-50 times excess previously used.[14,21,24] Initial photocatalytic tests herein used 20:1 $Rubpy_{C17}$ to $CoPyP_{C16}$, which resulted in 52±2 catalyst turnovers for CO formation (TON_{CO}) to give a turnover frequency (TOF_{CO}) of 3.4±0.1 min^{-1} with selectivity for CO over H_2 (Sel_{CO}) of 84±3% (Fig 1, Table S1; see SI Note 1 for list of metrics and definitions). This is more selective than literature values for the water soluble catalyst variant ($CoPyP_{C1}$) in aqueous phosphate buffer (Sel_{CO} 11%, 89% H_2 formation),[14] and comparable to a self-assembled liposome system (Sel_{CO} 77%).[14] Ru photosensitizers are known to self-quench at high concentrations, which was mitigated herein by including the charge-neutral photocatalytically-inactive surfactant hexaethylene glycol monododecyl ether ($C_{12}E_6$; critical micelle

concentration, CMC, in H_2O = 75 μM;[25] Fig 1). The inclusion of the nominally inactive $C_{12}E_6$ surfactant increased the TON_{CO} 5-fold to 249±38 for a TOF_{CO} of 17±2 min^{-1} and Sel_{CO} of 87±2% (Fig 1), thereby demonstrating the importance of dispersing the photosensitizers in micelles to minimize self-quenching. Comparing this result to other self-assembled systems that use alkylated Ru polypyridyl photosensitizers, the TOF_{CO} exceeds that of Lehn-type Re based systems ([Re^{I}(bpy)$(CO)_3$Cl]) in liposomes (17±2 vs 0.08 or 1.1 min^{-1}),[26,27] and the $CoPyP_{C16}$ catalyst in liposomes (0.8 min^{-1} at 0.5 μM or 6.1 min^{-1} at 0.02 μM).[14] The micelle system also shows a comparable TOF_{CO} to a state-of-the-art quantum dot sensitized system that electrostatically assembles with a Co porphyrin in aqueous media (13 min^{-1}).[28] Isotopic labelling with $^{13}CO_2$ reveals that only ^{13}CO is formed to confirm that all CO is formed from CO_2 (Fig S1). Exclusion controls show that the system is inactive if any of CO_2, catalyst, photosensitizer, reductant or light are removed (Fig S2). Amphiphilic $Rubpy_{C17}$ results in a 2.4-fold higher TOF_{CO} and 30% higher CO selectivity than using its water-soluble hydrophilic analogue [Ru(bpy)$_3$]Cl_2 (Rubpy hereinafter, TOF_{CO} of 6.8±0.8 min^{-1}; Fig 1), demonstrating the activity enhancement enabled by supramolecular self-assembly of the micellar system. This self-assembly benefit is also seen when comparing to homogeneous photocatalytic systems in the literature using Co porphyrins in water, which reach comparable TOF_{CO} values (17 and 20 min^{-1}) but use a 17-fold higher Ru photosensitizer concentration (30 vs 500 μM).[21,24] This is likely due to higher local concentrations of photosensitizer in the micelle, thereby facilitating better use of the system components.

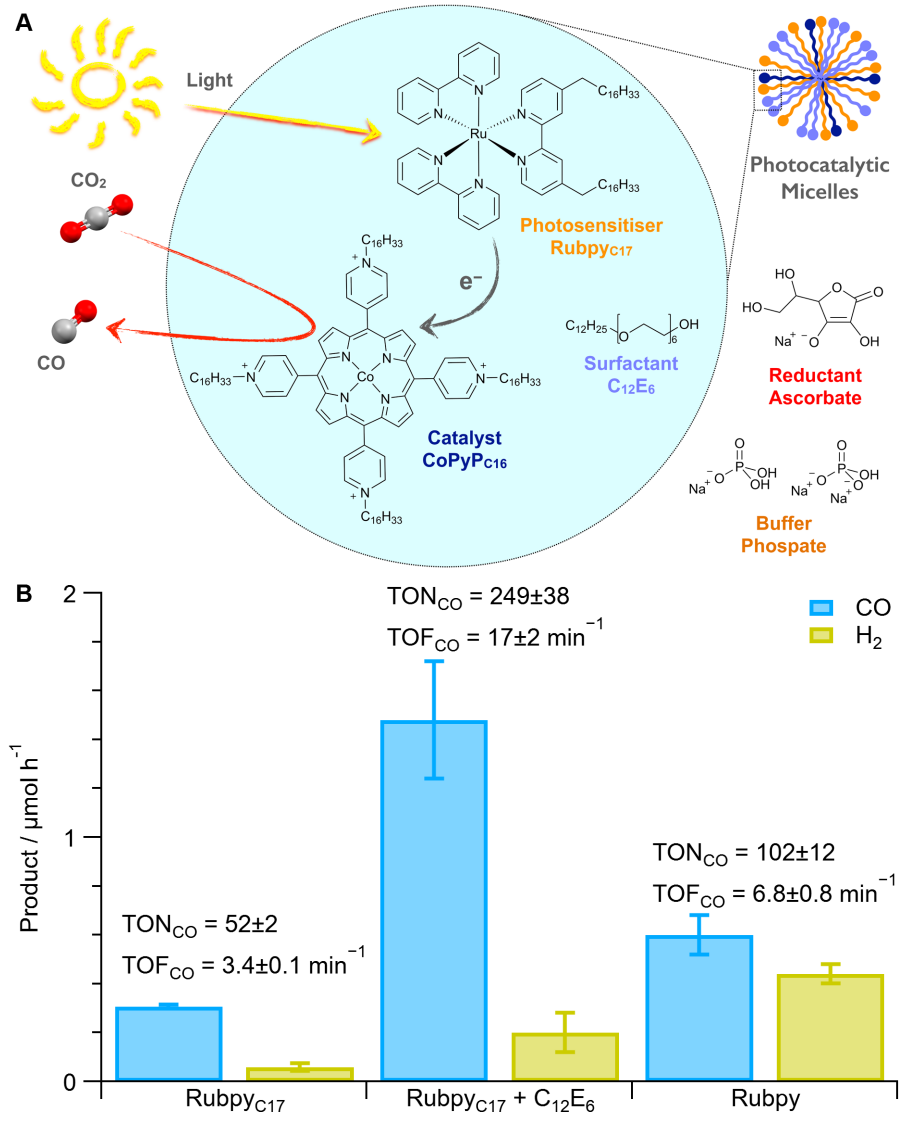


Fig 1: (A) Molecular components of photocatalytic system and reaction scheme and (B) Photocatalytic test results with 1.5 µM $CoPyP_{C16}$, 30 µM $Rubpy_{C17}$ or $[Ru(bpy)_3]Cl_2$ (Rubpy), including or excluding 225 µM $C_{12}E_6$ (~3 CMC), 100 mM NaHAsc, phosphate (0.1 M), CO_2-sat. (pH 6.3) after 15 min 447 nm illumination with 2.3 W LED at 25 °C with 250 rpm orbital shaking, 1 mL reaction volume. Product detection by GC, 3 replicates except $Rubpy_{C17}$ + $C_{12}E_6$ where n = 21 over 8 batches. Tabulated values in Table S1.

In contrast to non-ionic $C_{12}E_6$, the use of cationic cetrimonium bromide (CTAB) and anionic sodium dodecyl sulfate (SDS) surfactants inhibited catalytic turnover, with no CO detected after equivalent measurements screening from 1 to 18-fold their nominal CMCs of 1.0 and 7.8 mM for CTAB and SDS in pure water,[29] respectively (Fig S3 incl. chemical structures). UV-Vis spectra measured before and after photocatalytic tests with and without $C_{12}E_6$ show the $Rubpy_{C17}$-band intensity (λ_{max} 456 nm) decreases over time correlating with decreasing catalytic turnover (Fig S4), indicating that the intrinsic instability of Rubpy complexes through photobleaching limits long-term performance and highlights the need to move beyond Ru-polypyridyl based photosensitizers.[21,22,28] Contrastingly, no degradation of $Rubpy_{C17}$ is observed when

using CTAB or SDS surfactants within the timescale of the experiments (Fig S3), indicating that the ionic surfactants are inhibiting catalysis by preventing formation of long-lived reduced Ru states that are the most vulnerable to degradation.[21,30,31]

Photoluminescence (PL) and Transient Absorption Spectroscopy (TAS). PL and TAS were employed to understand the charge-transfer processes of the self-assembled system and lifetime of transient photosensitizer states. Visible excitation of Rubpy complexes results in a metal-to-ligand charge transfer (MLCT) to give the charge-separated excited state $[Ru^{III}(bpy^{\bullet-})(bpy)_2]^{2+}$ (Ru*).[14] Tracking the Ru* PL at 620 nm shows that $C_{12}E_6$ increases the $Rubpy_{C17}$* lifetime, but does not affect the hydrophilic Rubpy* (Fig 2A, S5-6 and Table S2-3). This longer $Rubpy_{C17}$* lifetime is attributed to $C_{12}E_6$ decreasing the local $Rubpy_{C17}$ concentration in micelles and thereby limiting self-quenching and allowing more time for reductive quenching by ascorbate, which reduces the Ru^{III} center of Ru* to form $[Ru^{II}(bpy^{\bullet-})(bpy)_2]^{+}$ (denoted Ru^{-}).[14]

Tracking Ru^{-} formation with TAS (510 nm probe) shows that adding $C_{12}E_6$ increases the $Rubpy_{C17}^{-}$ yield by 50%, resulting in a quenched photosensitizer yield that is only 25% lower than diffusionally free $Rubpy^{-}$ (17 to 30 vs 40 ΔmOD for $Rubpy_{C17}$, $Rubpy_{C17}$+$C_{12}E_6$ vs Rubpy; Fig S7 and Table S4). The $Rubpy_{C17}^{-}$ lifetime is then 18-fold longer than $Rubpy^{-}$ with $C_{12}E_6$ and phosphate buffer (282 vs 16 μs; Fig 2B, S8,9 and Table S5), allowing more time for electron transfer to the catalyst.

Adding $CoPyP_{C16}$ results in deeper absorbance bleaching because its Soret band (λ_{max} 440 nm; Fig S10) is convoluted with Rubpy (λ_{max} 456 nm). However, $CoPyP_{C16}$ increases the $Rubpy_{C17}^{-}$ lifetime by 33% rather than oxidizing it (282 to 375 μs at 510 nm and lack of $bpy^{\bullet-}$ decay at 366 nm; Fig 2C and S11). The rapid formation of $CoPyP_{C16}^{-}$ (broad response at 550-750 nm within 10 μs),[14] the longer $Rubpy_{C17}^{-}$ lifetime, and the relevant redox potentials indicates that $HAsc^{-}$ is oxidized by $Rubpy_{C17}$* to form $Asc^{\bullet}$, which then reduces $CoPyP_{C16}$ and forms dehydroascorbate.[32] This pathway prevents charge recombination between $Asc^{\bullet}$ and $Rubpy_{C17}^{-}$ by forming $CoPyP_{C16}^{-}$. It also means the system extracts two electrons from the donor per photon, with the process enabled by the high local concentrations inside the micelles. Further TAS analysis including for the ionic surfactants is provided in supplementary information (SI Note 2).

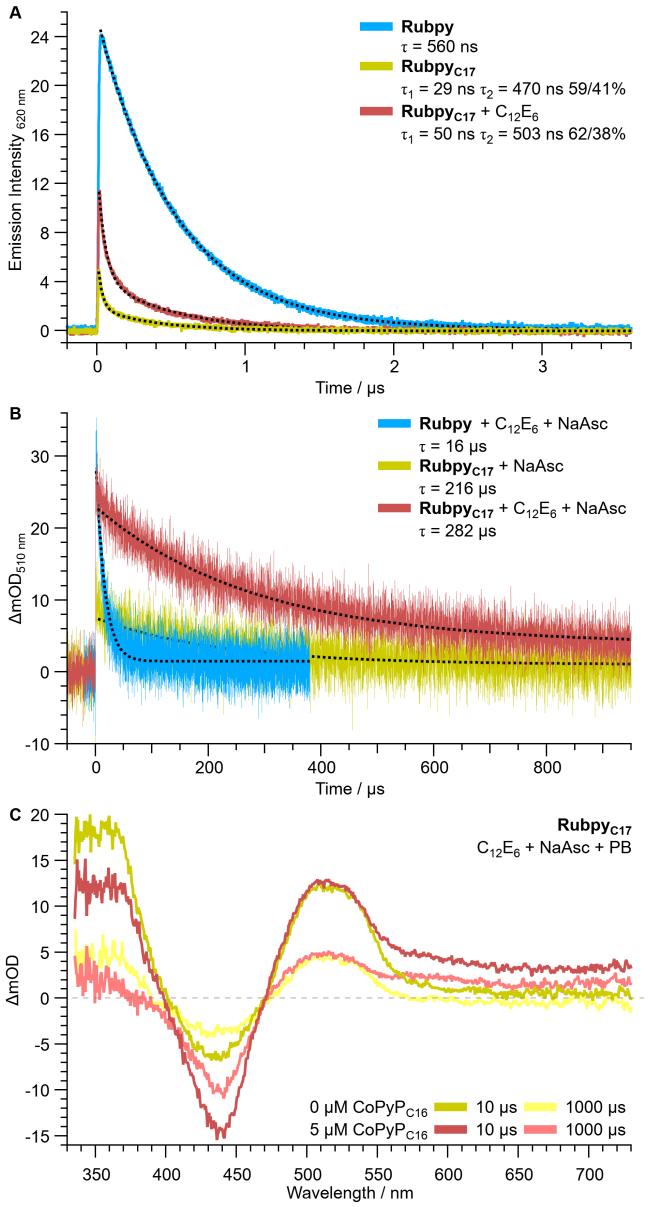


Fig 2: (A) Photoluminescence decay showing lifetime of excited photosensitizer in absence of reductant (excitation 460 nm, detection 620 nm); (B) Transient absorption spectroscopy (pump 460 nm, probe 510 nm) measuring lifetime of reductively quenched photosensitizer in presence of reductant; and (C) Transient absorption spectroscopy (pump 460 nm) measuring spectral changes after 10 and 1000 µs delay. 30 µM photosensitizer in phosphate buffer (0.1 M, pH 7.0), Ar purged. 5 µM $CoPyP_{C16}$, 225 µM $C_{12}E_6$ and/or 100 mM NaHAsc as indicated. Dashed lines show data fitting to determine lifetimes.

Heuristic human optimization. The photocatalytic system has categorical variables (variables that are divided into groups, often non-numerical such as the molecular structures) and continuous variables (a quantitative variable that can any value with a fixed range, *e.g.*, concentrations). Five continuous variables were sequentially varied in the system, namely, concentrations of catalyst, photosensitizer, surfactant, reductant and buffer, while the performance was quantified with five figures of merit, namely, $Yield_{CO}$

(moles of CO formed), QY_{CO} (percentage of incident photons forming CO), TON_{CO} (the number of catalyst turnovers forming CO), TOF_{CO} (TON_{CO} over time), and Sel_{CO} (the percentage of CO formed from the sum of CO and H_2 formed) to determine the sensitivity of each metric to the variables and start building the algorithm training set (Fig 3 and Table S6). Because the experiment time is fixed, $Yield_{CO}$ and QY_{CO} must follow the same trends, as must as TON_{CO} and TOF_{CO}. Varying the catalyst concentration shows that lower catalyst concentrations allow observation of higher TOF_{CO} (~86±8, 11±2, 2.5±0.1 min^{-1} at 0.01, 0.1, and 10 μM, respectively; Fig 3A); a trend that is consistent with general observations in the literature.[14,21,22,24] However, the high TON and TOF come at a cost of selectivity and product yield (25, 50 and 89% CO and 0.05, 0.07 and 1.5 μmol CO h^{-1}), emphasizing the importance of considering the system holistically rather than focusing on individual figures of merit.

Variation of the photosensitizer concentration shows a peak TOF_{CO} at 17±2 min^{-1} for 30 μM with clear decays at higher or lower concentration (2.0±0.7 at 3 μM and 3.4±0.2 at 200 μM; Fig 3B) with constant CO selectivity (88±1%). The peak performance occurs with concentrations able to effectively absorb light but is far below full absorption (A_{456} = 0.39), which may be due to a lack of surfactant to prevent self-quenching, lack of catalyst or reductant or a combination thereof. Keeping the photosensitizer concentration constant and increasing the $C_{12}E_6$ surfactant concentration from 0 to 225 μM (3 CMC) increases the TOF_{CO} followed by a plateau with constant 89±1% selectivity (Fig 3C). TOF_{CO} increases with the concentration of the ascorbate reductant (0.6±0.1, 17±2, 11±3 min^{-1} at 1, 100, 1000 mM, respectively; Fig 3D), with the lowest concentrations limiting photosensitizer quenching whereas the highest likely increase back-reactions.[33] With these fixed concentrations of the other components, the phosphate concentration had a minimal effect until performance decreased as surfactant solubility was visibly lowered (TOF_{CO} 12.8±0.7, 17±2, 7 min^{-1} at 1, 100, 1000 mM respectively. Fig 3E). In contrast to the catalyst concentration, the trend of the $Yield_{CO}$ follows the TOF_{CO} for the photosensitizer, surfactant, reductant and buffer concentrations. However, it is not clear how to increase both the $Yield_{CO}$ and TOF_{CO} for the system. Notably, the reductant and buffer concentrations significantly affect the solution ionic strength to indirectly influence micelle size and shape, as does the loading of catalyst and photosensitizer, yet varying the concentrations individually cannot capture these effects. Fully exploring the parameter space with 10 concentrations of each of the 5

variables would require a humanly impractical 10^5 experiments (SI Note 3). Thus, we employed Bayesian statistics to bring the number of experiments to an actionable number for a human experimentalist.

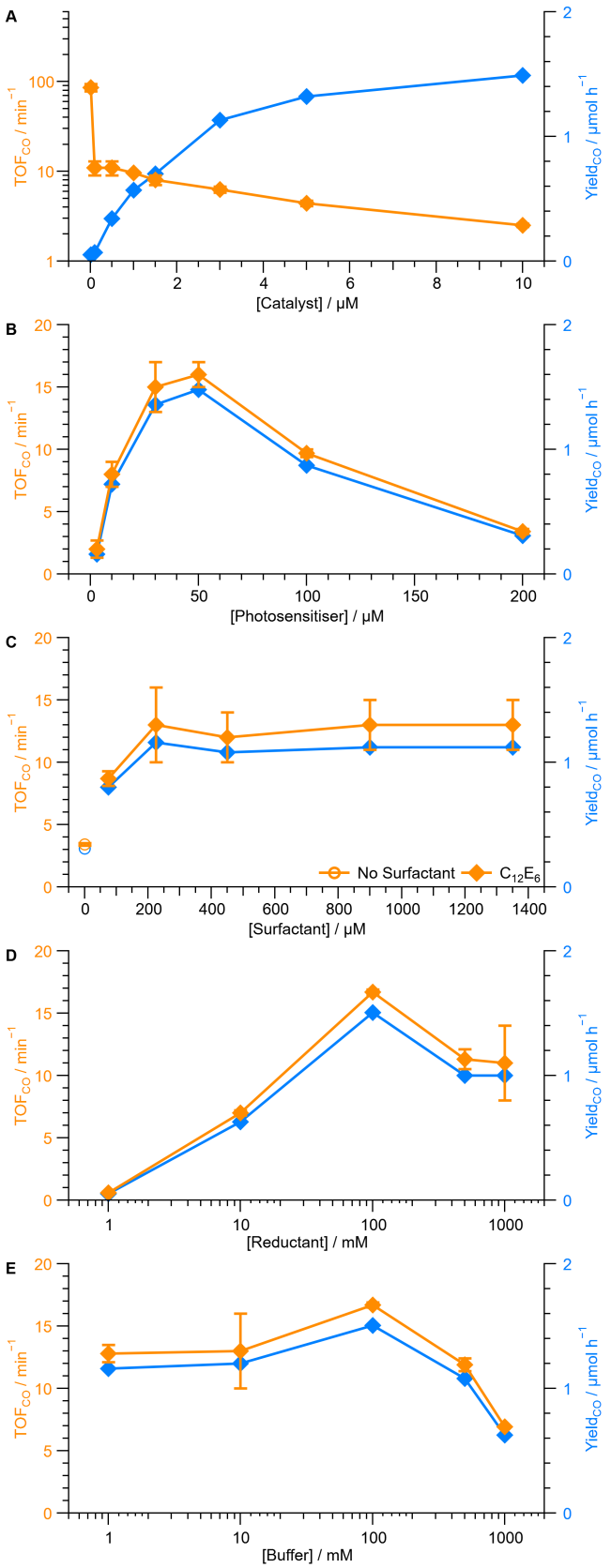


Fig 3: Heuristic optimization of photocatalytic system showing catalyst turnover frequency (TOF) and Yield of CO as a function of catalyst (A), photosensitizer (B), surfactant (C), reductant (D) and buffer (E) concentration. Unless specified, 1.5 µM $CoPyP_{C16}$, 30 µM $Rubpy_{C17}$, 225 µM $C_{12}E_6$ (~3 CMC), 100 mM NaHAsc, phosphate (0.1 M), CO_2-sat. (pH 6.3) after 15 min (60 min for catalyst series) 447 nm illumination with 2.3 W LED at 25 °C with 250 rpm orbital shaking, 1 mL volume, GC quantification, 3 replicates.

Learning algorithm (Bayesian) optimization. Our approach enables variation of >3 parameters simultaneously to increase optimization speed and find relationships between interdependent parameters, thereby allowing rapid and simultaneous optimization of multiple performance metrics. Our objective function (Eq. 1 and Fig 4) provided a single metric that incorporated key figures of merit to quantify overall system performance (holistic performance) and was defined as the sum of the weighted natural logarithms of QY_{CO}, TOF_{CO}, and TOF_{PS} to balance a high product yield against high efficiency. Sel_{CO} is indirectly considered by the QY_{CO} and TOF_{CO}. w_1, w_2 and w_3 are weighting coefficients to define the relative importance of each metric.

$$y^{\mathrm{obj}} = w_1 \ln(QY_{CO}) + w_2 \ln(TOF_{CO}) + w_3 \ln(TOF_{PS}) \qquad \text{(Eq. 1)}$$

Table 1: Original and optimized figures of merit for photocatalytic system composed of $CoPyP_{C16}$, $Rubpy_{C17}$, $C_{12}E_6$, NaHAsc, phosphate buffer, CO_2-sat. (pH 6.3) under 447 nm illumination with 2.3 W LED at 25 °C with 250 rpm orbital shaking, 1 mL reaction volume for 15 min, GC quantification.[a]

	Original	Holistic Opt.	Max $Yield_{CO}$ & QY_{CO}[b]	Max TON_{CO} & TOF_{CO}[b]
$Yield_{CO}$ / µmol h^{-1}	1.48±0.24	5.72±0.12	7.2±0.04	3.72±0.08
TON_{CO}	249±38	368±8	184±8	422±8
TOF_{CO} / min^{-1}	17±2	25±1	12±1	28±1
QY_{CO}	0.04±0.01	0.15±0.01	0.19±0.01	0.10±0.01
[Cat] / µM	1.5	3.9	9.8	2.2
[PS] / µM	30	112	98	109
[Surf.] / CMC	3	19	25	32
[Red.] / mM	100	220	348	195
[Buf.] / mM	100	462	323	509

[a] 21 replicates over 8 batches for 'original', 3 replicates for others. [b] Maximum values show performance and conditions that resulted in the highest $Yield_{CO}$ and QY_{CO} or highest TON_{CO} and TOF_{CO} during the optimization. $Yield_{CO}$ and QY_{CO} are linked, as are TON_{CO} and TOF_{CO}, because the reaction time is constant meaning they must be maximized under the same conditions.

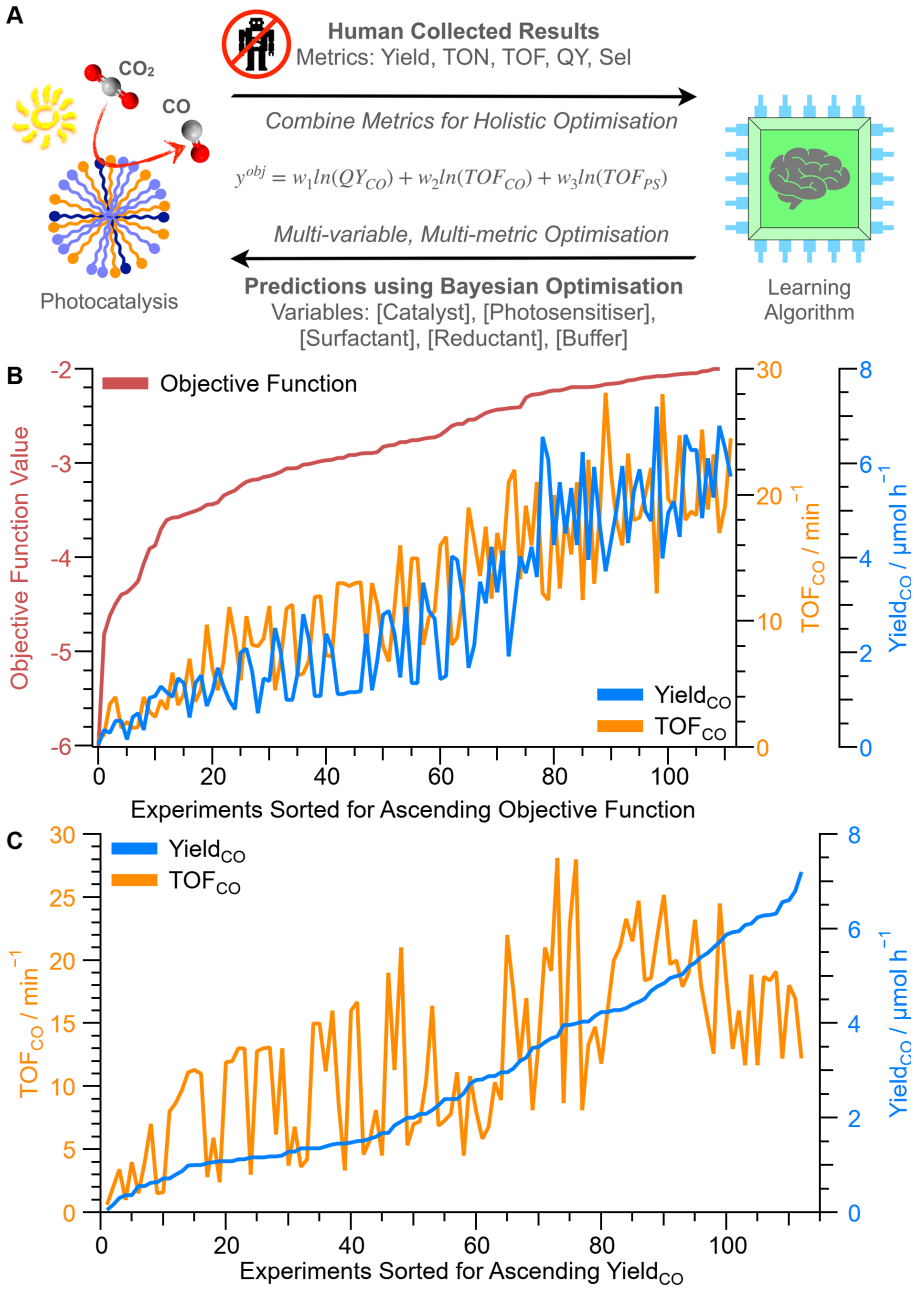


Fig 4: Holistic optimization using learning algorithms. Overview of the workflow (A); sorted data to show holistic improvement measured by objective function alongside improvement of both $Yield_{CO}$ and TOF_{CO} (B); and sorted data showing ascending $Yield_{CO}$ and corresponding TOF_{CO} (C). The data in B and C are sorted and not in chronological order. Full experimental results tabulated in Table S7. Experimental conditions: various concentrations of $CoPyP_{C16}$, $Rubpy_{C17}$, $C_{12}E_6$, NaHAsc and phosphate buffer, CO_2-sat. (pH 6.3) under 447 nm illumination with 2.3 W LED at 25 °C with 250 rpm orbital shaking, 1 mL reaction volume for 15 min, GC quantification.

The workflow comprised iterative experimental testing, from which the algorithm learnt and predicted new conditions (Fig 4A or S12 for a more technical description). The aim was to maximize the objective function (Eq. 1), which represents overall performance. The process is: (i) the algorithm learns from the initial dataset pool (subset of heuristic optimization data, Table S7); (ii) algorithm prediction of next values to test; (iii) the dataset pool is enlarged with new results; (iv) new

predictions and iteration.[34] Two Bayesian optimization methodologies (GpyOpt[35] and Gryffin[36]) were used with multiple acquisition functions to control paths through promising regions of the parameter space and exploit known peaks, while balancing the need to explore new regions (exploitation vs exploration; see experimental section). GpyOpt and Gryffin were selected as they have shown excellent performance in reaching the global optimum among Bayesian Optimization methods,[36] which are already superior to optimizations based on local gradient and Hessian approximations.[34] Holistic optimization proceeded through 10 iterations with each iteration providing five concentrations of catalyst, photosensitizer, surfactant, reductant, and buffer to test (Fig 4; each prediction is the mean of 5 replicate algorithm runs which are shown with their standard deviations in Fig S13; the mean prediction values and experimental results are shown in Table S7). Two optimizations were conducted in parallel by using different weightings of the objective function components, where y^{obj1} considered all components with weighting coefficients $w_1 = 0.4$, $w_2 = 0.4$, $w_3 = 0.2$; while y^{obj2} excluded the TOF_{CO} of photosensitizer (TOF_{PS}) with $w_1 = 0.5$, $w_2 = 0.5$, $w_3 = 0$. Equal weighting was given to the QY_{CO} and TOF_{CO} to prioritize maximization of both metrics simultaneously (the selection of weights is subjective as it depends on the prioritization of metrics meaning there is no mathematically correct choice). By optimizing the two objective functions simultaneously, this provides 10 sets of conditions (optimization paths) each iteration, which matches the capacity of the experimental equipment. The results were then all added into the dataset pool to provide all possible information to all algorithms (largest possible dataset pool at each iteration).

The optimizations including and ignoring the photosensitizer (y^{obj1} and y^{obj2}) both reached maxima with the same conditions in iteration 8/10, and displayed improvement in all figures of merit included in the objective function. Compared to the intuitively selected and literature supported starting conditions that appeared high performing based on initial tests and heuristic optimization (Table 1), the holistically optimized conditions displayed 4.2-fold higher $Yield_{CO}$ (5.72±0.12 vs 1.48±0.24 μmol h^{-1}), 57% higher TON_{CO} (368±8 vs 249±38), 47% higher TOF_{CO} (25±1 vs 17±2 min^{-1}) and 3.8-fold higher QY_{CO} (0.15±0.01 vs 0.04±0.01%) while Sel_{CO} was unchanged (88 vs 87±2%). This shows that systems can be optimized for high Yield, TON, TOF and QY rather than focusing on a single metric. These values are exceptional for photocatalytic CO_2-to-CO reduction in aqueous

media with far higher TOF_{CO} than Lehn-type Re complexes in liposomes,[26,27] higher TOF_{CO} and Sel_{CO} than high performing homogeneous Co porphyrin systems that use 17-fold higher photosensitizer concentration,[21,22,24] and higher TOF_{CO} than even a high performing quantum dot sensitized Co porphyrin system (Table S8).[28] Expectedly, the objective function (overall system performance metric) is not linearly correlated with any of the system components (r^2<0.27 in all cases; full linear correlation plots and correlation coefficient matrix in Fig S14 and S15). The highest correlation coefficient of a parameter with the objective function is 0.27 for [photosensitizer] and 0.17 for [surfactant], while correlations of the objective function with other parameters have coefficients <0.05. This confirms that increasing all parameters would not result in the observed performance improvement. TAS reveals that these conditions coincide with the $Rubpy_{C17}^-$ lifetime shortening from 375 to 190 μs in the optimized system, suggesting more efficient electron transfer to the catalyst (Fig S16).

Sorting the experiments shows that all the performance metrics improved as the objective function was maximized (Fig 4B), thereby validating the way the objective function was defined and showing a stark contrast to the response from heuristically varying the catalyst concentration (Fig 3A). However, neither the highest $Yield_{CO}$ and QY_{CO} nor highest TON_{CO} and TOF_{CO} were obtained at the holistic optimum because a compromise in conditions is required (Table 1). To reiterate, $Yield_{CO}$ and QY_{CO} are linearly correlated (Fig S14 and S15), as are TON_{CO} and TOF_{CO}, because QY is Yield over a constant and TOF is TON over a constant (with consistent reaction time and photon flux), hence they are maximized under the same conditions. At conditions for the highest $Yield_{CO}$ and QY_{CO} (7.2±0.04 μmol h^{-1} and 0.19±0.01%), the TON_{CO} and TOF_{CO} are ~50% lower than the holistic maximum (184±8 and 12±1 min^{-1}; Fig 4C). Coherently, conditions displaying the highest TON_{CO} and TOF_{CO} (422±8 and 28±1 min^{-1}) result in $Yield_{CO}$ and QY_{CO} being ~50% lower (3.72±0.08 μmol h^{-1} and 0.10±0.01%) than their ideal conditions. In other words, TOF_{CO} and $Yield_{CO}$ are not linearly correlated, especially at high values, meaning that these maxima cannot be reached by maximizing only one metric. This demonstrates how objective functions enable holistic optimization and highlights the limitations in focusing on individual figures of merit as remains prevalent in the literature.

Post-Optimization Machine Learning Analysis. The internally consistent dataset produced during the optimization was then exploited to determine the ruling parameters for each metric with machine learning

algorithms, *i.e.*, the quantitative importance of each parameter to overall system performance and to individual figures of merit was extracted. The full dataset was split into two sections, with 70% used to train/cross-validate the regression model while the model predicted results for the remaining 30%, thereby allowing the model to be tested against experimental results (Random Forest with *k*-fold cross-validation was employed to identify the most effective hyperparameters for the regression models; see SI Note 3).[37,38] The number of datapoints is low by machine learning standards with only 72 experiments for training/cross-validation and 31 experiments for testing, yet the consistency of the dataset means the regression model is accurately predictive of how the concentration of each component will affect holistic system performance (R^2 = 0.71 is suitable for a predictive model trained on experimental data;[10,39] Fig 5A).

Shapley additive explanations (SHAP) were then employed to quantify the importance of each parameter (system component) on the system performance (Fig 5A),[17–19] revealing the buffer concentration as the parameter with dominating importance (61%), far more than the catalyst or photosensitizer concentrations (16 and 12%, respectively; note the 4% difference may not be statistically significant). The non-linear effect of the buffer concentration is also visible in the correlation pair-plot while other parameters show more stochastic scatter plots (Fig S14). While the buffer concentration allows the ionic strength to be varied and thereby provides control of micelle size, its importance is seemingly contradictory to the heuristic optimization wherein the concentration of buffer did not affect the system until it lowered solubility and hindered performance (Fig 3E). However, the ubiquitous heuristic optimization did not consider dependencies between the variables, thereby exemplifying the value of the statistics driven holistic optimization demonstrated herein. Consequently, SHAP analysis can quantitatively identify the cause of improvements after Bayesian Optimization, which provides insight for allow rationalization and further system development. It is especially valuable as the causes identified may not be immediately visible to the experimentalist.

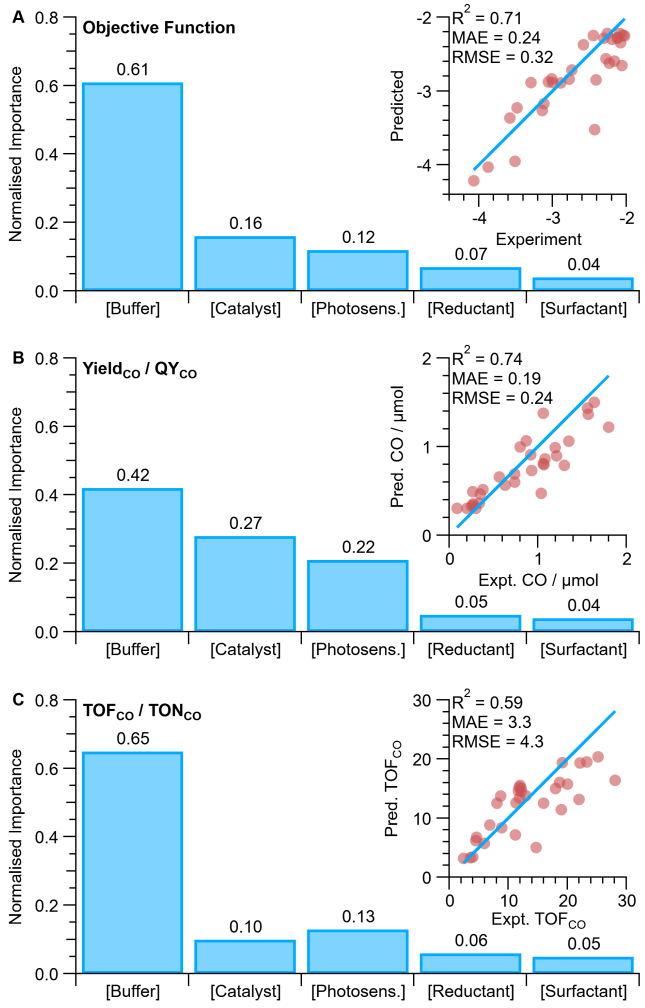


Fig 5: Machine learning analysis showing the normalized importance of each system component to: holistic optimization by objective function 1 (A), $Yield_{CO}$ and QY_{CO} (B), TON_{CO} and TOF_{CO} (C); with machine learning predictions with the Random-forest regression models (inset), where model performance is shown by coefficient of determination R^2, mean absolute error (MAE), and Root Mean Squared Error (RMSE).

Machine learning regression and the SHAP routine then extracted the importance of each parameter to individual figures of merit (Fig 5B,C), thereby providing insight into the conditions that favor high $Yield_{CO}$ and QY_{CO} vs those that favor high TON_{CO} and TOF_{CO}. In all cases the buffer concentration remains the most important factor, with comparably dominating importance to TON_{CO} and TOF_{CO} (64-65%), yet buffer concentration is far less important to $Yield_{CO}$ and QY_{CO} (42-43%) where the catalyst and photosensitizer concentrations grow in importance (27 and 22%, respectively). Thereby, the conflicting requirements to have high $Yield_{CO}$ vs high TOF_{CO} can be quantitatively linked to different parameter combinations. Consistent with these results, maximum CO yield is reached with 4.5-fold higher catalyst concentration than maximum TOF_{CO} (9.8 vs 2.2 µM), whereas higher buffer

concentration is required for maximum TOF_{CO} (509 vs 323 mM). The holistically optimized results balance these two requirements, yet a similar photosensitizer concentration is maintained for all three maxima, presumably to maximize light absorption and quantum yield.

Control Group Feature Analysis. Further leveraging the holistic optimization dataset, local performance maxima in the 5-dimensional parameter space can be identified to find conditions that significantly reduce the amount of valuable components used at a minor cost to system performance. These fruitful regions of the parameter space can be identified and visualized by reducing the five variables into a lower number of features following the recently introduced Control Group Feature analysis (herein, the two mixed features are defined as x_1 and x_2 in Fig S17; details in Methods).[20] The approach is inspired by Buckingham analysis and based on the principle that variable combinations that offer the same mixed feature value will offer comparable performance by exploiting alternate maxima in the parameter space. This analysis highlights regions of the parameter matrix containing high performance conditions as clustered high-performing values in scatter plots.

The Control Group Feature analysis shows clustered regions for values of the holistically optimized system (objective function), $Yield_{CO}$ and QY_{CO}, but no clear regions were found for TON_{CO} and TOF_{CO} (Fig S17). Focusing on the holistic performance, alternative parameter combinations (component concentrations) were identified that are predicted to provide comparable performance while using 30% less catalyst and 16% less photosensitizer (Catalyst 2.7 vs. 3.9 µM; Photosensitizer 94 vs 112 µM; Surfactant 23 vs 19 CMC; Reductant 256 vs 220 mM; Buffer 402 vs 462 mM). Thereby, the dataset is further exploited to minimize the more expensive, metal-containing components at a minor performance cost, which is a key trade-off for system commercialization (additional analysis in SI Note 3).

Wider Implications. Methodology for multi-variable multi-metric optimization is required in many fields to move beyond the focus on individual figures of merit that drives optimization away from practical conditions. Overall system development requires multiple metrics to be improved simultaneously, even though some metrics initially appear to require opposite conditions, *e.g.*, high TOF observed at low catalyst loading while high product yield requires the opposite as initially observed in this study. Even pioneering high-throughput robotic approaches have focused on single figures of merit,[6,40] including those employing machine learning.[5,13,41] New approaches should minimize expensive and space-demanding equipment

wherever possible to facilitate widespread adoption, which is a barrier that continues to hinder high-throughput robotic setups. Learning algorithms require no elaborate equipment and can accelerate multi-variable optimization so effectively that human experimentalists can complete otherwise insurmountable studies, such as multi-metric optimization of complex systems. The use of overall performance metrics can take into account a variety of factors, for example, using an expensive component will be justified if the performance improvement outweighs the cost and leads to an improvement in the overall metric. Thereby such comparisons can be objective when comparing systems and conditions. To fully exploit this approach, each field should reach a consensus on a standardized metric against which all systems should be compared. In some ways, this could be a powerful extension of the approach taken for solar cell characterization where photon to current efficiency is the standard metric of comparison.

The highly general methodology for holistic optimization shown herein can be applied in any chemistry laboratory to address the inability to complete multi-variable multi-metric optimization in many fields beyond photocatalytic CO_2 reduction (incl. materials, polymer and organic synthesis). While the approach is demonstrated with CO_2 reduction, a clear example application is optimizing organic synthesis reactions so they are high yielding, fast and general (wide substrate scope), *i.e.*, multi-variable multi-metric reaction optimization is required.[6] For an organic reaction, post-optimization machine learning would then quantify the importance of each variable to each metric and guide mechanistic understanding; *e.g.*, the importance of the buffer herein is analogous to additives that unexpectedly enhance organic reactions, which has been the focus of a prominent high-throughput study.[6]

Conclusions

We demonstrate how to optimize multi-variable multi-metric systems using five-component self-assembled photocatalytic micelle system for CO_2-to-CO reduction as a test platform. Self-assembly improved the system's photocatalytic performance 2.4-fold compared to freely diffusing components, with transient absorption spectroscopy showing this was due to an 18-fold longer reduced photosensitizer lifetime. However, standard optimization showed that the system could not be optimized for both high TOF_{CO} and high CO yield as they appeared to require opposite conditions. Therefore, we established a novel approach and defined a holist metric set that encompassed overall system performance and which was optimized using machine learning algorithms to simultaneously improve CO yield (4.2-fold), TON_{CO} (57%), TOF_{CO}

(47%) and quantum yield (3.8-fold), while maintaining 87% CO selectivity over H_2. Machine learning enabled this optimization in 103 tests compared to 10^5 combinatorial possibilities, thereby bringing multi-variable multi-metric optimization into reach for all chemistry laboratories without expensive robotic or high-throughput setups.

A predictive machine learning model was then trained on the dataset and quantified the effect of each component. This revealed surprisingly that the buffer has 61% importance to the holistic performance despite heuristic optimization indicating the buffer concentration had a marginal effect, thereby emphasizing the complexity of systems with inter-dependent variables. The buffer is 20% more important to TON_{CO}/TOF_{CO} than it is to yield of CO and quantum yield, which demonstrates the trade-offs required for holistic optimization and why maximizing a single metric has limited meaning to overall system performance.

Future standardization of holistic performance metrics will enhance comparability and focus research on meaningful system performance rather than a subjective individual metric that often relies on impractical conditions. Exploiting developments in machine learning and data science will concurrently reveal system limitations and guide the understanding that underpins progress across research in catalysis.

Experimental Section

Reagents. Commercial chemicals were used as supplied: hexaethylene glycol monododecyl ether ($C_{12}E_6$; CAS 3055-96-7, Sigma, ≥98%); sodium phosphate, dibasic (CAS 7782-85-6, Thermo Scientific, ≥99%); sodium phosphate, monobasic (CAS 10049-21-5, ACROS, >99%); sodium dodecyl sulfate (CAS 151-21-3, ThermoScientific, 99%); hexadecyltrimethylammonium bromide (CAS 57-09-0, ACROS, ≥99%); sodium L-ascorbate (CAS 134-03-2, Sigma, ≥99%); tris(2,2′-bipyridyl)dichlororuthenium(II) hexahydrate (CAS 50525-27-4, Sigma, 99.95%); hexadecyltrimethylammonium chloride (CAS 112-02-7, Sigma, 25 wt.% aqueous solution). Deionized water was used throughout (18 MΩ cm, Milli-Q), MeCN was freshly distilled. All gases were obtained from BOC.

Synthesis. **$CoPyP_{C16}$** was synthesized following reported procedures;[14] briefly, tetrapyridyl-porphyrin (PyP), was reacted with excess $C_{16}H_{33}Br$ at 130 °C for 16 h in anhydrous DMF under a N_2 atmosphere to form the tetraalkylated porphyrin (PyP_{C16}). Purification was by precipitation from $CHCl_3$:MeOH (85:15) with

acetone and reprecipitation from hot EtOH. EA: C, H, N values from all batches within 0.5% of calculated for $C_{104}H_{158}N_8Br_4.3H_2O$ (1894.1 g mol^{-1}): C 65.95, H 8.73, N 5.92%. ^{1}H NMR (DMSO-d_6) and ^{13}C NMR (DMSO-d_6) as reported.[14] Metalation of PyP_{C16} used excess $Co(OAc)_2$ at 120 °C for 3 h in anhydrous DMF under a N_2 atmosphere followed by precipitation with Et_2O. $CoPyP_{C16}$ was then precipitated from acetone:MeOH (9:1) with aqueous $NaPF_{6(sat'd)}$,and precipitated from acetone with water. EA: C, H, N values from all batches within 0.5% of calculated for $CoC_{104}H_{156}N_8P_4F_{24}.3H_2O$ (2211.3 g mol^{-1}): C 56.59, H 7.38, N 5.07%. **Rubpy$_{C17}$** was synthesized following reported procedures from $[Ru(Cl)_2(bpy)_2]$ and 4,4'-diheptadecyl-2,2'-bipyridine (bpy_{C17}).[27,42] ^{1}H and ^{13}C NMR (CH_3OD) and UV-Vis (CH_3OH) as reported. EA: C, H, N values from all batches within 0.5% of calculated for $RuC_{64}H_{92}N_6Cl_2{\cdot}7H_2O$ (1243.6 g mol^{-1}): C 61.81, H 8.59, N 6.76%.

Solution preparation. The water soluble chloride salts of photosensitizers $[Ru(bpy)_2(bpy_{C17})](Cl)_2$ or $[Ru(bpy)_3](Cl)_2$ were prepared as aqueous stock solutions and used within two days. The PF_6^- salt of $CoPyP_{C16}$, $[CoPyP_{C16}](PF_6)_4$, was prepared as an acetonitrile solution that was confirmed by UV-Vis to be stable for months. Concentrations were determined by UV-Vis (Agilent Cary 60) and known or determined molar extinction coefficients. Surfactant and ascorbate solutions were prepared fresh daily with concentrations based on weighed mass (excl. CTAC), while phosphate buffer was prepared as a larger stock from the monobasic and dibasic sodium salts ($H_2PO_4^-$:HPO_4^{2-} as 0.615:0.385 molar ratio to reach pH 7.0 in final solution).[43] Reaction solutions were prepared by diluting these stock solutions into water to reach the desired concentrations, with the minimal volume of organic solvent (MeCN) excluded from final volume as it is removed during CO_2 purging.

Photocatalysis. 1 mL aliquots of solution, (3 mL for $^{13}CO_2$ labelling experiment), were delivered into crimp-cap vials sealed with septa and purged through inlet and outlet needles for 15 min with CO_2 or N_2 including 2 vol.% CH_4 as an internal standard. CO_2 saturation was reached within 10 min and defined by the solution stabilizing at pH 6.3. The profile of the foam changes during purging as the MeCN is removed. The vials were then placed in a LED photoreactor (Treellum Technologies, Patent EP17382313) with a thermostatically controlled heating block at 25 °C with 250 rpm orbital convection; wherein the vials were illuminated with 447 nm LED light (20 nm spectral width) from 2.3 W LEDs (Luxeon Rebel ES LED 'Royal Blue').[44,45] The reaction headspace was sampled via gastight syringe (100 µL aliquot; Hamilton) and

analyzed by gas chromatography (Shimadzu Tracera GC-2010 Plus with a discharge ionization detector) using a molecular sieve column (Restek ShinCarbon micro ST) at 85 °C. H_2 and CO peak areas were converted to moles using calibrated response factors and the 2 vol.% CH_4 internal standard. Quantum yields were calculated based on photon flux previously determined by actinometry for the photoreactor as 2.1 μmol photons s^{-1}.[9] Isotopic labelling used $^{13}CO_2$ for photocatalysis followed by transferring the vial headspace into an evacuated gas infrared cell (SpecAc, 10 cm path length, KBr windows) and collecting a FT-IR transmission spectrum (ThermoScientific Nicolet iS50). Data processing used MatLab (Mathworks) while plotting used Igor Pro (Wavemetrics).

Transient Absorption Spectroscopy. Nanosecond optical excitation (pump) used a Nd:YAG laser/OPO (EKSPLA NT 342) to deliver 8 mJ $pulse^{-1}$ at 460 nm. The measurement (probe) spectrometer (Edinburgh Instruments LP 920-K) was equipped with a 450 W Xe arc lamp (Osram) and CCD camera (Oxford Instruments Andor iStar) connected to an oscilloscope (Tektronix TDS3052B). Measurements were through a 3 mL 10×10 mm quartz cuvette at ambient temperature after purging with Ar or CO_2 where specified, with 20 scans averaged for each spectrum. Data fitting used Igor Pro (Wavemetrics).

Bayesian Optimization. GpyOpt 1.2.6[35] and Gryffin[36] (dynamic implementation) methodologies were employed. Three acquisition functions were employed with GpyOpt: exploitation/exploration trade-off functions Expected Improvement (EI) and Lower Confidence Bound (LCB), and exploitation focused Maximum Probability of Improvement (MPI). Two acquisition functions were employed with Gryffin: exploitation and exploration (hyperparameter λ to +1 or −1, respectively). Predictions were made by taking the mean of 5 repeated queries for the next step, thereby avoiding limitations in deterministic prediction by GpyOpt and Gryffin.[35,36] When the CO yield = 0, the objective function is equal to $-\infty$ (Eq. 1); which was replaced with −1000 for compatibility with the algorithms. Bounds were set for each of the five variables: Catalyst 1.5:10 μM, Photosensitizer 10:180 μM, Surfactant 1:40 CMC, Reductant 1:500 mM, Buffer 10:1000 mM. Where multiple suggested points were deemed comparable, they were skipped for the sake of maximum capitalization of the experimental setup.

Machine Learning Regression and Feature Importance. Experimental data harvested during the experiments was used to train/validate five Random Forest regression-based pipelines with hyperparameter tuning in 5-fold cross-validation. Conditions with no CO, corresponding to $y^{obj\ 1} = -1000$,

were deemed outliers and the same training/testing sets were used for all the models. Specifically, the space of hyperparameters is composed by all the possible combinations of (i) the number of estimators, to be chosen among [100, 200, 500, 1000, 2000] and (ii) the number of features to keep when looking for the best split, to be chosen among ['auto', 'sqrt', 'log2']. For the database handling and for the models training/validation Pandas[46] and the Scikit-Learn[47] Python packages were respectively employed. The TreeSHAP algorithm[17,18] was used to identify feature rankings over the corresponding models' outputs as it is suitable for tree-based models like Random Forest.

Features grouping. The feature grouping separates values into classes based on a threshold for the corresponding properties, i.e., above (class 1) and below (class 0) the threshold following the methodology recently introduced by Trezza & Chiavazzo.[20] Such thresholds are −2.2 for $y^{\text{obj }1}$, 1.3 μmol for CO, 300 for TON_{CO}, 18.5 min^{-1} for TOF_{CO}, 0.14 for QY_{CO}. In particular, given the original set of 5 features ($x_1, \dots, x_5$), the corresponding dimensionless quantities are denoted $(\widetilde{x_1}, \dots, \widetilde{x_5})$, where $\widetilde{x_i} = \frac{x_i - x_{i,min}}{x_{i,max} - x_{i,min}} + 1$ representing the minimum and the maximum value of the i−th feature over the training set, respectively. A new set of two mixed features $(\overline{x_1}, \overline{x_2})$ is thus defined, where $\overline{x_j} = \prod_{i=1}^{5} \widetilde{x_i}^{\alpha_{ij}}$, with $\{\alpha_{ij}\} \in \mathbb{R}^{5\times 2}$ being a matrix computed by means of a multi-objective optimization criterion in two steps. First, the matrix $\{\alpha_{ij}\}$ is the utopia point of the Pareto front simultaneously attempting (i) the maximization of the distance between the two classes according to the Bhattacharyya distance,[48,49] (ii) the minimization of a norm of the covariance matrix of the first class distribution, (iii) the minimization of a norm of the covariance matrix of the second class distribution. Second, the solution of the first optimization step is used as the input of a nearest neighbor optimization, to minimize the number of neighbors of class 0 to each sample of class 1 within a fixed cutoff radius of 10^{-2}. Finally, the new variables normalized in the interval [0, 1] are computed as: $\frac{\bar{x}_i - \overline{x_{i,min}}}{\overline{x_{i,max}} - \overline{x_{i,min}}}$.

Associated Content

Supporting Information. Tabulated values from photocatalytic testing and learning algorithm optimization (in order of algorithm input/output) and transient absorption spectroscopy; supplementary notes elaborating on photoluminescence, regression model design and controlling feature grouping; transient absorption spectra; plots of photocatalytic results including ^{13}C labelling, exclusion controls,

measurements over time, tests with cationic and anionic surfactants; a more technical depiction of the learning algorithm optimization, correlation plot matrices and controlling feature grouping plots.

Data availability. The primary data supporting the findings of this study are available from the University of Cambridge open-access data repository (link to be added upon acceptance) with additional information available from the corresponding authors upon reasonable request.

Acknowledgements

This work was supported by an EU Horizon 2020 Future Emerging Technologies (FET) project 'Soap Film based Artificial Photosynthesis' (SoFiA, GAN 828838, DOI: 10.3030/828838) and an EU Marie Skłodowska-Curie individual fellowship to S.R.J. (GAN 891338). We sincerely thank Dongseok Kim, Miruna Constantinescu, Dr Tessel Bouwens, Dr Samuel J. Cobb, Prof Stuart M. Clarke (all University of Cambridge, UK), Dr Gabriele Falciani (Politecnico di Torino, Italy), Dr Carla Casadevall (ICIQ, Spain), Dr Indraneel Sen (Uppsala University, Sweden) and other SoFiA Consortium members for valued discussions. Dr Andrea Pannwitz and Prof Sylvestre Bonnet (University of Leiden, The Netherlands) have provided the alkylated Ru complex.

References

(1) Bonchio, M.; Bonin, J.; Ishitani, O.; Lu, T.-B.; Morikawa, T.; Morris, A. J.; Reisner, E.; Sarkar, D.; Toma, F. M.; Robert, M. Best Practices for Experiments and Reporting in Photocatalytic CO2 Reduction. *Nat. Catal.* **2023**, *6* (8), 657–665. https://doi.org/10.1038/s41929-023-00992-7.

(2) Fang, S.; Rahaman, M.; Bharti, J.; Reisner, E.; Robert, M.; Ozin, G. A.; Hu, Y. H. Photocatalytic CO2 Reduction. *Nat. Rev. Methods Primer* **2023**, *3* (1), 1–21. https://doi.org/10.1038/s43586-023-00243-w.

(3) Kitchin, J. R. Machine Learning in Catalysis. *Nat. Catal.* **2018**, *1* (4), 230–232. https://doi.org/10.1038/s41929-018-0056-y.

(4) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature* **2021**, *590* (7844), 89–96. https://doi.org/10.1038/s41586-021-03213-y.

(5) Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. A Mobile Robotic Chemist. *Nature* **2020**, *583* (7815), 237–241. https://doi.org/10.1038/s41586-020-2442-2.

(6) Prieto Kullmer, C. N.; Kautzky, J. A.; Krska, S. W.; Nowak, T.; Dreher, S. D.; MacMillan, D. W. C. Accelerating Reaction Generality and Mechanistic Insight through Additive Mapping. *Science* **2022**, *376* (6592), 532–539. https://doi.org/10.1126/science.abn1885.

(7) Sahm, C. D.; Ucoski, G. M.; Roy, S.; Reisner, E. Automated and Continuous-Flow Platform to Analyze Semiconductor–Metal Complex Hybrid Systems for Photocatalytic CO2 Reduction. *ACS Catal.* **2021**, *11* (17), 11266–11277. https://doi.org/10.1021/acscatal.1c02921.

(8) Lawson, T.; Gentleman, A. S.; Lage, A.; Casadevall, C.; Xiao, J.; Petit, T.; Frosz, M. H.; Reisner, E.; Euser, T. G. Low-Volume Reaction Monitoring of Carbon Dot Light Absorbers in Optofluidic Microreactors. *ACS Catal.* **2023**, 9090–9101. https://doi.org/10.1021/acscatal.3c02212.

(9) Casadevall, C.; Pascual, D.; Aragón, J.; Call, A.; Casitas, A.; Casademont-Reig, I.; Lloret-Fillol, J. Light-Driven Reduction of Aromatic Olefins in Aqueous Media Catalysed by Aminopyridine Cobalt Complexes. *Chem. Sci.* **2022**, *13* (15), 4270–4282. https://doi.org/10.1039/D1SC06608K.

(10) Trezza, G.; Bergamasco, L.; Fasano, M.; Chiavazzo, E. Minimal Crystallographic Descriptors of Sorption Properties in Hypothetical MOFs and Role in Sequential Learning Optimization. *Npj Comput. Mater.* **2022**, *8* (1), 1–14. https://doi.org/10.1038/s41524-022-00806-7.

(11) Schweidtmann, A. M.; Clayton, A. D.; Holmes, N.; Bradford, E.; Bourne, R. A.; Lapkin, A. A. Machine Learning Meets Continuous Flow Chemistry: Automated Optimization towards the Pareto Front of Multiple Objectives. *Chem. Eng. J.* **2018**, *352*, 277–282. https://doi.org/10.1016/j.cej.2018.07.031.
(12) Reizman, B. J.; Wang, Y.-M.; Buchwald, S. L.; Jensen, K. F. Suzuki–Miyaura Cross-Coupling Optimization Enabled by Automated Feedback. *React. Chem. Eng.* **2016**, *1* (6), 658–666. https://doi.org/10.1039/C6RE00153J.
(13) MacLeod, B. P.; Parlane, F. G. L.; Morrissey, T. D.; Häse, F.; Roch, L. M.; Dettelbach, K. E.; Moreira, R.; Yunker, L. P. E.; Rooney, M. B.; Deeth, J. R.; Lai, V.; Ng, G. J.; Situ, H.; Zhang, R. H.; Elliott, M. S.; Haley, T. H.; Dvorak, D. J.; Aspuru-Guzik, A.; Hein, J. E.; Berlinguette, C. P. Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials. *Sci. Adv.* **2020**, *6* (20), eaaz8867. https://doi.org/10.1126/sciadv.aaz8867.
(14) Rodríguez-Jiménez, S.; Song, H.; Lam, E.; Wright, D.; Pannwitz, A.; Bonke, S. A.; Baumberg, J. J.; Bonnet, S.; Hammarström, L.; Reisner, E. Self-Assembled Liposomes Enhance Electron Transfer for Efficient Photocatalytic CO2 Reduction. *J. Am. Chem. Soc.* **2022**, *144* (21), 9399–9412. https://doi.org/10.1021/jacs.2c01725.
(15) Yu, J.; Huang, L.; Tang, Q.; Yu, S.-B.; Qi, Q.-Y.; Zhang, J.; Ma, D.; Lei, Y.; Su, J.; Song, Y.; Eloi, J.-C.; Harniman, R. L.; Borucu, U.; Zhang, L.; Zhu, M.; Tian, F.; Du, L.; Phillips, D. L.; Manners, I.; Ye, R.; Tian, J. Artificial Spherical Chromatophore Nanomicelles for Selective CO2 Reduction in Water. *Nat. Catal.* **2023**, *6* (6), 464–475. https://doi.org/10.1038/s41929-023-00962-z.
(16) Graetzel, M. Artificial Photosynthesis: Water Cleavage into Hydrogen and Oxygen by Visible Light. *Acc. Chem. Res.* **1981**, *14* (12), 376–384. https://doi.org/10.1021/ar00072a003.
(17) Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2017; Vol. 30.
(18) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2* (1), 56–67. https://doi.org/10.1038/s42256-019-0138-9.
(19) Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Interpretable Machine Learning for Knowledge Generation in Heterogeneous Catalysis. *Nat. Catal.* **2022**, *5* (3), 175–184. https://doi.org/10.1038/s41929-022-00744-z.
(20) Trezza, G.; Chiavazzo, E. Leveraging Composition-Based Energy Material Descriptors for Machine Learning Models. *Mater. Today Commun.* **2023**, 106579. https://doi.org/10.1016/j.mtcomm.2023.106579.
(21) Call, A.; Cibian, M.; Yamamoto, K.; Nakazono, T.; Yamauchi, K.; Sakai, K. Highly Efficient and Selective Photocatalytic CO2 Reduction to CO in Water by a Cobalt Porphyrin Molecular Catalyst. *ACS Catal.* **2019**, *9* (6), 4867–4874. https://doi.org/10.1021/acscatal.8b04975.
(22) Zhang, X.; Cibian, M.; Call, A.; Yamauchi, K.; Sakai, K. Photochemical CO2 Reduction Driven by Water-Soluble Copper(I) Photosensitizer with the Catalysis Accelerated by Multi-Electron Chargeable Cobalt Porphyrin. *ACS Catal.* **2019**, *9* (12), 11263–11273. https://doi.org/10.1021/acscatal.9b04023.
(23) Sprintschnik, G.; Sprintschnik, H. W.; Kirsch, P. P.; Whitten, D. G. Photochemical Reactions in Organized Monolayer Assemblies. III. Photochemical Cleavage of Water: A System for Solar Energy Conversion Using Monolayer-Bound Transition Metal Complexes. *J. Am. Chem. Soc.* **1976**, *98* (8), 2337–2338. https://doi.org/10.1021/ja00424a057.
(24) Call, A.; Cibian, M.; Yamauchi, K.; Sakai, K. Visible-Light-Driven Reduction of CO2 to CO in Fully Aqueous Media Using a Water-Soluble Cobalt Porphyrin. *Sustain. Energy Fuels* **2022**, *6* (9), 2160–2164. https://doi.org/10.1039/D2SE00291D.
(25) Falciani, G.; Franklin, R.; Cagna, A.; Sen, I.; Hassanali, A.; Chiavazzo, E. A Multi-Scale Perspective of Gas Transport through Soap-Film Membranes. *Mol. Syst. Des. Eng.* **2020**, *5* (5), 911–921. https://doi.org/10.1039/C9ME00186G.
(26) Ikuta, N.; Takizawa, S.; Murata, S. Photochemical Reduction of CO2 with Ascorbate in Aqueous Solution Using Vesicles Acting as Photocatalysts. *Photochem. Photobiol. Sci.* **2014**, *13* (4), 691–702. https://doi.org/10.1039/C3PP50429H.
(27) Klein, D. M.; Rodríguez-Jiménez, S.; Hoefnagel, M. E.; Pannwitz, A.; Prabhakaran, A.; Siegler, M. A.; Keyes, T. E.; Reisner, E.; Brouwer, A. M.; Bonnet, S. Shorter Alkyl Chains Enhance Molecular Diffusion and Electron Transfer Kinetics between Photosensitisers and Catalysts in CO2-Reducing Photocatalytic Liposomes. *Chem. – Eur. J.* **2021**, *27* (68), 17203–17212. https://doi.org/10.1002/chem.202102989.
(28) Arcudi, F.; Đorđević, L.; Nagasing, B.; Stupp, S. I.; Weiss, E. A. Quantum Dot-Sensitized Photoreduction of CO2 in Water with Turnover Number > 80,000. *J. Am. Chem. Soc.* **2021**, *143* (43), 18131–18138. https://doi.org/10.1021/jacs.1c06961.
(29) Deodhar, S.; Rohilla, P.; Manivannan, M.; Thampi, S. P.; Basavaraj, M. G. Robust Method to Determine Critical Micelle Concentration via Spreading Oil Drops on Surfactant Solutions. *Langmuir* **2020**, *36* (28), 8100–8110. https://doi.org/10.1021/acs.langmuir.0c00908.
(30) Hawecker, J.; Lehn, J.-M.; Ziessel, R. Photochemical Reduction of Carbon Dioxide to Formate Mediated by Ruthenium Bipyridine Complexes as Homogeneous Catalysts. *J. Chem. Soc. Chem. Commun.* **1985**, No. 2, 56–58. https://doi.org/10.1039/C39850000056.
(31) Grant, J. L.; Goswami, K.; Spreer, L. O.; Otvos, J. W.; Calvin, M. Photochemical Reduction of Carbon Dioxide to Carbon Monoxide in Water Using a Nickel(II) Tetra-Azamacrocycle Complex as Catalyst. *J. Chem. Soc. Dalton Trans.* **1987**, No. 9, 2105–2109. https://doi.org/10.1039/DT9870002105.
(32) Creutz, C. Complexities of Ascorbate as a Reducing Agent. *Inorg. Chem.* **1981**, *20* (12), 4449–4452. https://doi.org/10.1021/ic50226a088.
(33) Bachmann, C.; Probst, B.; Guttentag, M.; Alberto, R. Ascorbate as an Electron Relay between an Irreversible Electron Donor and Ru(II) or Re(I) Photosensitizers. *Chem. Commun.* **2014**, *50* (51), 6737–6739. https://doi.org/10.1039/C4CC01500B.
(34) Snoek, J.; Larochelle, H.; Adams, R. P. Practical Bayesian Optimization of Machine Learning Algorithms. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2012; Vol. 25.
(35) *GPyOpt*. https://github.com/SheffieldML/GPyOpt (accessed 2023-07-07).
(36) Häse, F.; Aldeghi, M.; Hickman, R. J.; Roch, L. M.; Aspuru-Guzik, A. Gryffin: An Algorithm for Bayesian Optimization of Categorical Variables Informed by Expert Knowledge. *Appl. Phys. Rev.* **2021**, *8* (3), 031406. https://doi.org/10.1063/5.0048164.
(37) Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B Methodol.* **1974**, *36* (2), 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x.

(38) Seni, G.; Elder, J. *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions*; Synthesis Lectures on Data Mining & Knowledge Discovery; Morgan & Claypool Publishers, 2010.
(39) Choudhary, K.; Garrity, K. Designing High-TC Superconductors with BCS-Inspired Screening, Density Functional Theory, and Deep-Learning. *Npj Comput. Mater.* **2022**, *8* (1), 1–12. https://doi.org/10.1038/s41524-022-00933-1.
(40) Haber, J. A.; Cai, Y.; Jung, S.; Xiang, C.; Mitrovic, S.; Jin, J.; Bell, A. T.; Gregoire, J. M. Discovering Ce-Rich Oxygen Evolution Catalysts, from High Throughput Screening to Water Electrolysis. *Energy Environ. Sci.* **2014**, *7* (2), 682–688. https://doi.org/10.1039/C3EE43683G.
(41) Bai, Y.; Wilbraham, L.; Slater, B. J.; Zwijnenburg, M. A.; Sprick, R. S.; Cooper, A. I. Accelerated Discovery of Organic Polymer Photocatalysts for Hydrogen Evolution from Water through the Integration of Experiment and Theory. *J. Am. Chem. Soc.* **2019**, *141* (22), 9063–9071. https://doi.org/10.1021/jacs.9b03591.
(42) Ellison, D. K.; Iwamoto, R. T. Synthesis of Unsymmetrical 4,4′-Dialkyl-2,2′-Bipyridines. *Tetrahedron Lett.* **1983**, *24* (1), 31–32. https://doi.org/10.1016/S0040-4039(00)81318-0.
(43) Green, A. A. The Preparation of Acetate and Phosphate Buffer Solutions of Known PH and Ionic Strength. *J. Am. Chem. Soc.* **1933**, *55* (6), 2331–2336. https://doi.org/10.1021/ja01333a018.
(44) Call, A.; Casadevall, C.; Acuña-Parés, F.; Casitas, A.; Lloret-Fillol, J. Dual Cobalt–Copper Light-Driven Catalytic Reduction of Aldehydes and Aromatic Ketones in Aqueous Media. *Chem. Sci.* **2017**, *8* (7), 4739–4749. https://doi.org/10.1039/C7SC01276D.
(45) Casadevall, C.; Aragón, J.; Cañellas, S.; Pericàs, M. A.; Lloret-Fillol, J.; Caldentey, X. Development of Advanced High Throughput Experimentation Platforms for Photocatalytic Reactions. In *The Power of High-Throughput Experimentation: General Topics and Enabling Technologies for Synthesis and Catalysis (Volume 1)*; ACS Symposium Series; American Chemical Society, 2022; Vol. 1419, pp 145–165. https://doi.org/10.1021/bk-2022-1419.ch009.
(46) McKinney, W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*; O'Reilly Media, Inc., 2012.
(47) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
(48) Bhattacharyya, A. On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distribution. *Bull. Calcutta Math. Soc.* **1943**, *35*, 99–110.
(49) Bhattacharyya, A. On a Measure of Divergence between Two Multinomial Populations. *Sankhyā Indian J. Stat. 1933-1960* **1946**, *7* (4), 401–406.