Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Digital Discovery*, 2025, **4**, 2737

Autonomous organic synthesis for redox flow batteries *via* flexible batch Bayesian optimization

Clara Tamura, [a] Heather Job, [b] Henry Chang, [a] Wei Wang, [bc] Yangang Liang *[bc] and Shijing Sun*[a]

Traditional trial-and-error methods for materials discovery are inefficient to meet the urgent demands posed by the rapid progression of climate change. This urgency has driven the increasing interest in integrating robotics and machine learning into materials research to accelerate experimental learning. However, idealized decision-making frameworks to achieve maximum sampling efficiency are not always compatible with high-throughput experimental workflows inside a laboratory. For multi-step chemical processes, differences in hardware capacities can complicate the digital framework by introducing constraints on the maximum number of samples in each step of the experiment, hence causing varying batch sizes in variable selection within the same batch. Therefore, designing flexible sampling algorithms is necessary to accommodate the multi-step synthesis with practical constraints unique to each high-throughput workflow. In this work, we designed and employed three strategies on a high-throughput robotic platform to optimize the sulfonation reaction of redox-active molecules used in flow batteries. Our strategies adapt to the multi-step experimental workflow, where their formulation and heating steps are separate, causing varying batch size requirements. By strategically sampling using clustering and mixed-variable batch Bayesian optimization, we were able to iteratively identify optimal conditions that maximize the yields. Our work presents a flexible approach that allows tailoring the machine learning decision-making to suit the practical constraints in individual high-throughput experimental platforms, followed by performing resource-efficient yield optimization using available open-source Python libraries.

Received 15th January 2025
Accepted 20th August 2025

DOI: 10.1039/d5dd00017c

rsc.li/digitaldiscovery

Introduction

Materials exploration for energy storage systems plays a critical role in advancing toward a carbon-free power grid.[1] Although solar and wind power are pivotal for decarbonization, their status as intermittent energy sources necessitates using large and robust storage solutions to ensure grid stability.[2] Redox flow batteries (RFBs) have demonstrated great potential for grid storage due to their high energy density properties and lower costs compared to their inorganic counterparts.[2] In particular, aqueous RFBs provide a sustainable and safe solution for large-scale energy storage. However, their progress has been hindered by the scarcity of organic compounds that combine high solubility in water with reversible redox behavior within the water stability window.[3,4] Feng *et al.* achieved a notable breakthrough recently by applying molecular engineering to modify 9-fluorenone, an inexpensive redox-active molecule.[4] Through the introduction of sulfonate ($-SO_3^-$) groups, the solubility of fluorenone derivatives was significantly improved in aqueous electrolytes, enabling efficient and stable two-electron redox reactions without the need for catalysts. Moving forward, developing milder conditions for sulfonation reactions that minimize or eliminate the need for excessive fuming sulfuric acid is of great interest. Such advancements are critical to overcoming the scalability challenges of fluorenone-based aqueous RFBs, enabling their broader adoption for large-scale energy storage applications.[3–9]

Effectively screening and identifying optimal synthesis conditions across high-dimensional design spaces has long been a fundamental challenge in chemical science. Traditional human-centric approaches to exploring large chemical spaces are hindered by the limited number of samples that can be prepared manually in a single round, impeding the iterative processes of condition optimization and introducing batch-to-batch variations due to non-standardized sample handling.[10] Recent advancements in robotics and machine learning (ML) have begun to address these challenges through the development of self-driving labs (SDLs).[11–19] In SDLs, high-throughput experimentation (HTE) platforms are often employed to conduct chemical synthesis and characterization without human intervention.[14,17,20] These systems are capable of handling multiple samples in parallel with high reproducibility,

[a]*University of Washington, Department of Mechanical Engineering, 3900 E Stevens Way NE, Seattle, WA 98195, USA. E-mail: shijing@uw.edu*

[b]*Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA 99354, USA. E-mail: yangang.liang@pnnl.gov*

[c]*Energy Storage Research Alliance, Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA*

significantly accelerating the research timelines. For example, Liang *et al.* demonstrated the use of the high-throughput system to create large-scale high-quality solubility databases of redox-active materials for RFBs.[11]

It is worth noting that brute-force screening using HTEs alone often leads to excessive raw material consumption and prolonged experiment times, making it incompatible with sustainable research practices. A critical aspect of self-driving laboratories is therefore the development of decision-making methods, increasingly powered by machine learning (ML), to enable smart automation that aimed to achieve efficient learning with minimal experiments.[10] In the past decade, Bayesian Optimization (BO), an active learning method, has gained significant traction in AI-guided chemical experiments due to its efficiency and versatility for noise-heavy experiments.[21–25] BO iteratively updates a stochastic surrogate model, such as Gaussian Process (GP) regression, and employs analytical acquisition functions to determine the next set of experimental parameters. Among BO approaches, Batch BO (BBO) is particularly well-suited for HTE systems, as it can suggest multiple per round of experimentation.[26]

Despite several recent studies successfully employing BBO to guide chemical synthesis,[14,17,20] a critical yet overlooked issue is the disconnect between hardware constraints and algorithm design. Inside a chemistry laboratory, synthesis typically involves multi-step processes requiring more than one piece of equipment. For instance, a liquid handling robot can prepare a 96-well plate each round, but all at a single, fixed temperature on a conventional lab bench. Existing BBO algorithms and software packages typically operate under idealized assumptions, enforcing a fixed batch size per sampling round across all dimensions of interest, ignoring the complexity in reality, such as the number of compositions that can be explored per round is limited by the number of available wells, while temperature constraints depend on the number of heaters. Current approaches fail to account for the practical limitations imposed by hardware capabilities. As a result, these algorithms often struggle to adapt to real-world laboratory setups, leading to inadequate, if not wasteful, experimental plans where the algorithm recommendations exceed physical capabilities in the system or operate with a suboptimal allocation of hardware resources. To bridge this gap, resource-aware experimental design methods are urgently needed for autonomous sampling on high-throughput platforms.

In this study, we developed three flexible BBO frameworks to address sampling challenges in high-dimensional design spaces where some dimensions are subject to different batch size constraints. Our objective was to optimize the sulfonation reaction to enhance the solubility of fluorenone-based aqueous RFBs using an HTE platform. By iteratively varying the concentrations of the sulfonating agent and fluorene analyte, reaction time, and temperature, we identified 11 conditions achieving high reaction yields (yield > 90% under mild conditions of <170 °C), to mitigate the hazards associated with fuming sulfuric acid. To accommodate different batch size requirements between compositions and temperature sampling, each framework employed a two-stage BO approach within a four-dimensional (4D) design space, utilizing strategies of (1) post-BO clustering, (2) post-BO temperature redistribution, and (3) temperature pre-selection, respectively. The frameworks successfully identified optimal synthesis conditions and were evaluated based on their optimization efficiency and predictive accuracy. We introduce flexible decision-making frameworks to bridge the gap between idealized optimization strategies and practical sampling constraints, which we hope to shed light on sustainable autonomous chemical research.

Results and discussion

High-throughput experimentation platform

The explored chemical space of the sulfonation reaction consists of two formulation parameters and two process parameters spanning four dimensions. The variables of interest are reaction time (min), reaction temperatures (°C), sulfuric acid (%), and the concentration of fluorenone analyte (mg ml^{-1}). For simplicity, we refer to these variables as time, temperature, sulfonating agent, and analyte, respectively, throughout the paper. The sampling boundaries are as follows: time (30.0–600 min), temperature (20.0–170.0 °C), sulfonating agent (75.0–100.0%), and analyte (33.0–100 mg mL^{-1}). These boundaries of the search space were selected based on prior literature (see more details in the section "Chemical Insight" in SI). State-of-the-art synthesis involves sulfonation using fuming sulfuric acid, which often leads to a release of sulfur trioxide fumes at high synthetic temperatures, posing challenges in reactivity control and energy efficiency.[4] In this paper, we address these issues by optimizing reaction conditions under mild temperature ranges, aiming to reduce excessive fuming, enhance energy efficiency, and maintain high reaction yield and product quality.

Fig. 1 illustrates the digital and experimental workflow for the conducted experiment. The HTE synthesis system is equipped with liquid handlers for formulation, robotics arms for sample transfers, and three heating blocks for temperature control. Each heating block can accommodate up to 48 samples per plate. Assuming three replicates per condition and three controls, the total number of unique conditions we are able to generate per batch is 15 conditions with 45 specimens, hence we designed the experimental workflow for 15 conditions per batch. For initializing the optimization process, the first round of conditions was generated using four-dimensional (4D) Latin Hypercube Sampling (LHS),[27] where 15 unique sets of conditions were generated. Since the synthesis hardware only has three heating blocks, the capacity is limited to three temperature values, and therefore, we need to modify the LHS-generated conditions. The LHS-generated temperatures were clustered to determine three temperatures, where the centroids of the clusters were determined. LHS is a form of confined random sampling designed to span the parameter space with a degree of symmetry; hence, the temperatures in Round 1 were evenly spaced. The original temperatures from the initial LHS were reassigned with the newly found centroid temperatures by proximity. After all 45 specimens were synthesized, they were transported to a high-performance liquid chromatography

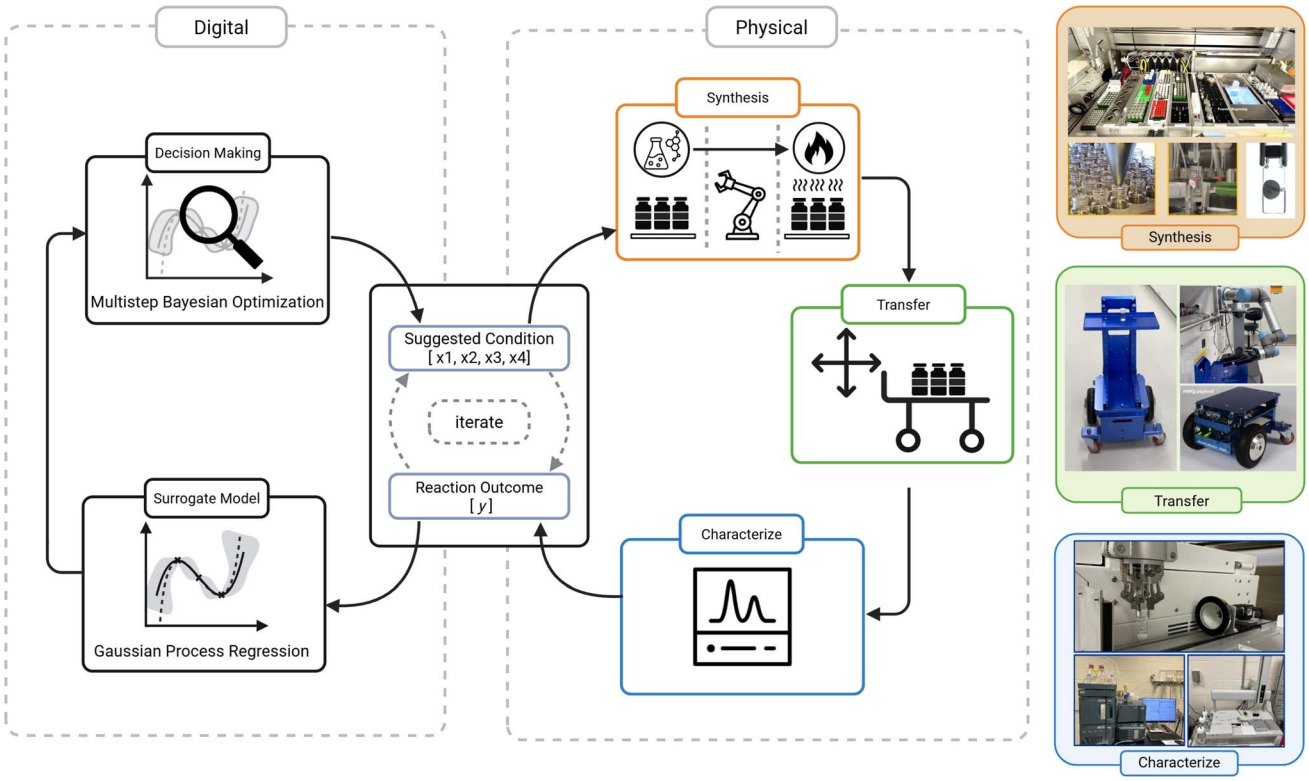

Fig. 1 Illustration of the closed-loop experimental workflow with the digital and physical framework. First, a set of conditions is sent to a multi-process synthesis robotic platform. Synthesized samples are transferred to a characterization station for data collection and analysis to extract the desired reaction outcome. The results from the characterization are used to create a Gaussian Process Regression surrogate model, and a multi-step BO is conducted as the decision-making step to generate the next set of suggested conditions for sampling. This iteration of generated conditions and collecting the reaction outcome is continued until the desired target is reached.

(HPLC) system for automatic characterization, as shown in Fig. 1.

Decision-making models

After the first round of HPLC data was collected, feature extraction for each HPLC was conducted to determine the percent yield of the product. The desired features from the areas of interest are the peaks corresponding to product, reactant, acid, and byproducts. The percent product yield was calculated using the areas determined under each peak. These yields are the outputs of interest for training the model. To prepare the data, the mean and variance of the three repeated specimens per condition were calculated and implemented to train the surrogate GP model. The output is the mean values, and the variance is the noise. The optimization goal for our algorithms is to find conditions with the highest product yield. Consequently, all our algorithms are designed for single-task BO. For benchmarking purposes, we used the same acquisition functions across all models (see Methods: digital).

A key challenge in our application is that temperature is constrained to three values per batch due to hardware limitations, while for the rest of the three dimensions of analyte, sulfonating agent, and reaction time, 15 values can be sampled per variable per batch. Automated HPLC characterization was only conducted after the completion of both formulation and heating. When designing the sampling strategy for our BBO framework, the question we are trying to answer is: How do we handle varying batch sizes within a single round of sampling?

Model A: post-BO clustering. In the first round of sampling, discrepancies in variable sizes were addressed by reassigning the LHS-generated temperatures with the proximal centroids of the clustered temperatures. This same clustering approach in Round 1 can also be applied to conditions sampled from a 4D BBO with a fixed batch size. In Model A, a GP surrogate model was trained on the first round of data collection, and a 4D BBO of batch size 15 was performed to identify the next 15 optimal conditions. The generated conditions were then clustered into three temperature groups, and the centroids of the clustered groups replaced the BBO-suggested temperatures (Fig. 2: Model A). These ML-suggested conditions were then sent to the high-throughput experimentation platform for data collection.

Model B: redistribution of chosen temperatures. For this approach, we compare the effects of post-BO and pre-BO clustering to determine which method is suitable for condition generation. While Model A resolves the hardware issue for varying batch size, reassigning temperatures post-BO sampling raises concerns about whether conditions are optimally selected. By replacing the BO-suggested temperatures with clustering centroids, we overlooked the possibility that the variables sampled from the other three dimensions of sulfonating agent, time, and analyte in the generated sets are no

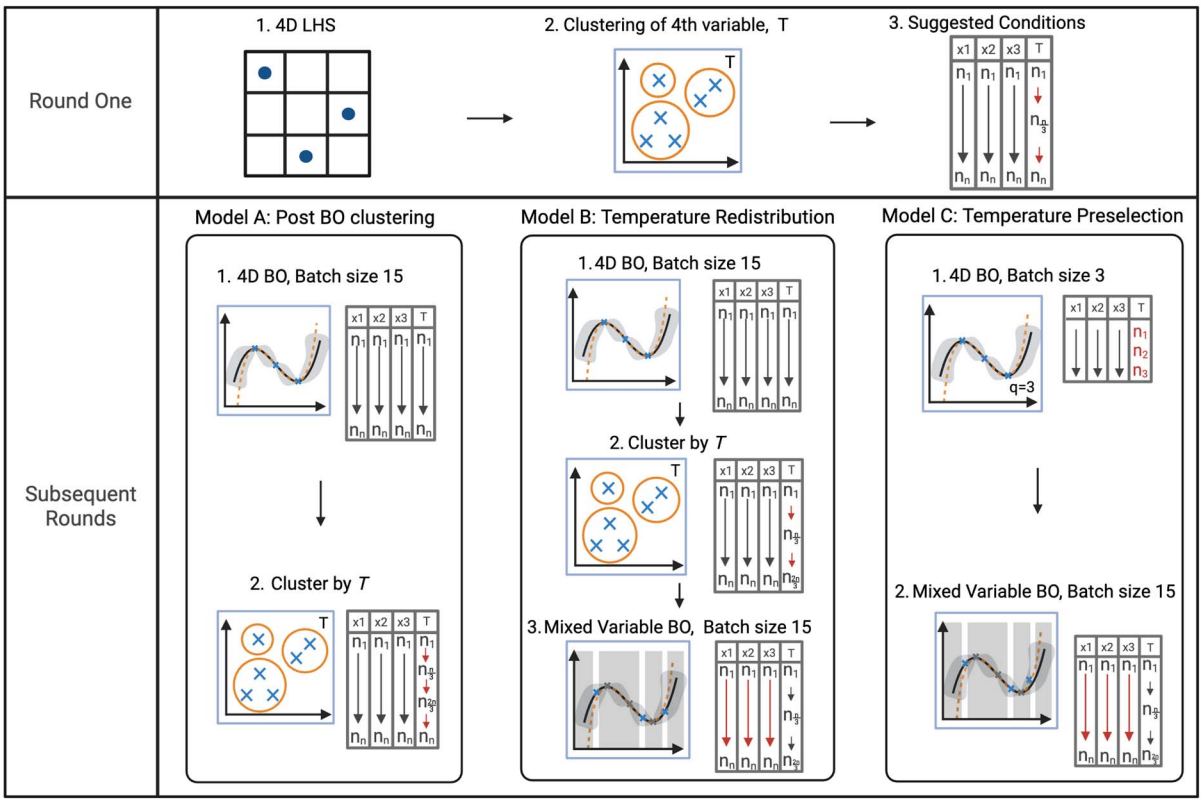

Fig. 2 Illustration of sampling strategies for round one conditions and the subsequent rounds. Three sampling strategies labeled Model A, Model B, and Model C show the multi-step process that was applied for the respective strategies in the subsequent rounds. Round one: (1) $n = 15$ conditions were generated using four-dimensional (4D) latin hypercube sampling, (2) generated conditions are clustered into three temperature groups, (3) shape of suggested conditions. Model A: (1) 4D batch Bayesian Optimization (BBO) with a Gaussian Process (GP) surrogate model with batch size = 15, (2) generated conditions are clustered by temperature (*T*). Model B: (1) 4D BBO with a GP Surrogate with batch size 15 to generate a full set of conditions, (2) generated conditions are clustered by *T* (same process as Model A), (3) conditions for ×1, ×2, ×3 are regenerated using mixed-variable BO of batch size 15 with temperatures from the cluster. Model C: (1) 4D BBO with a GP Surrogate with batch size = 3, (2) conditions for ×1, ×2, ×3 are regenerated using mixed-variable BO of batch size 15 with temperatures chosen from previous BO.

longer optimal at the newly assigned temperatures. This potentially could affect the exploration and the resulting optimization. This realization prompted us to develop a second approach that regenerates the other variables, assuming pre-selected temperatures. After reviewing various approaches to BO, we found mixed-variable BO as a possible method for generating sampling conditions given prior temperatures.[28,29] Mixed-variable BO is an algorithm that allows a mixture of continuous and categorical inputs, making it well-suited for real-world engineering problems where not all variables are continuous.[28] Hence, by discretizing the temperature variable, mixed-variable BO effectively convolves the temperature categories with the remaining continuous variables. The same temperatures from Model A were used for constructive comparison to determine whether this approach optimizes the system effectively.

Model C: temperature pre-selection. A challenge with mixed-variable BO in our application is that temperature is, in practice, a continuous variable, not discrete. Mixed-variable BO assumes that categorical variables remain fixed from their initialization, thus creating a mismatch between how temperature is selected and updated each round. Keeping the same constrained temperatures each round would limit the exploration of the chemical space, which negatively impacts our efficiency in finding the optimal conditions in the entire design space. Ideally, we wanted a BO method that could update the category at each iteration; however, current mixed-variable BO models assume pre-selected categories, so categories remain constant in subsequent rounds of sampling. To address this issue, we adopted a sequential BO approach that uses two BO steps in the framework, which would enable temperature categories to update after each iteration. The Model C BBO framework consists of two BBOs of batch sizes 3 and 15 that first choose 3 temperatures, then redistribute the temperatures using mixed variable BBO. We performed a 4D BBO with a batch size of 3 to select the temperatures. Then, the chosen temperatures were incorporated into a mixed variable BO of batch size 15, so that the other variables have been optimized based on the 3 temperature inputs. This approach preserves BO's flexibility in exploring the chemical space and utilizes the same condition regeneration strategy as Model B.

Model summary and experimental results

Two rounds of data collection following each subsequent sampling strategy were conducted, resulting in 45 iteratively

sampled conditions in each surrogate model training. Since the BBO approach was used, the Gaussian process regression models were updated after each round of data collection, instead of after each individual data point collection. All the models combined, we collected 105 unique synthesis conditions and generated 315 specimens across three rounds of data collection. For a visual representation of the entire space for each model and each round, a four-dimensional (4D) representation was constructed using 3D surface plots with slices along the *z*-axis (Fig. 3a, 4a and 5a). These plots visualize the optimization process and outcomes across different rounds and show how the model navigates the 4D input space of time, temperature, sulfonating agent, and analyte to predict the product yield. The three continuous variables, time, sulfonating agent, and analyte, are represented by the 2D surface plots where the *x*-axis and *y*-axis are time and sulfonating agent, and each slice is the analyte. The fourth dimension, temperature, is represented as a constant across the entire GP model. To visualize the effects of temperature, we spanned the temperature by plotting each 3D surface slice at min, max, mean, and interquartile ranges (SI Fig. S9–S11). All axes are from 0 to 1 as all variables have been normalized. All models show a consistent trend where temperatures at the upper quartile are optimal for high product yield. Additionally, the optimal regions are observed at the upper corners, but do not increase linearly, indicating that the optimizations are effective for this chemical space. The progression of the exploitation of the space provides insight into the differences between the sampling strategies. As observed in Model A, the subsequent rounds have light areas that are narrowed after every iteration, which suggests that the algorithm successfully exploits to locate high-yield regions. On the other hand, Model C appears to explore the space more broadly, as each successive round shows minimal change, suggesting a slower learning rate. The differences in learning rate are further supported by the model uncertainty figures, which are located in the SI (Fig. S9–S11). Fig. S9 displaying Model A's uncertainty shows that the highest uncertainty in the subsequent rounds is located at the temperature boundaries. The uncertainty for Model B is similar to that of Model A but exhibits more exploration, indicated by the lower uncertainties at the boundaries. Model C has the lowest uncertainty among the other models, and the uncertainty decreases in subsequent rounds, suggesting that Model C exhibits high exploration and a slow learning rate. Panels b of Fig. 3–5 display the 2D contour plots of the posterior mean and histograms of one of the slices in the final round (Round 3), showing all variable pairs with fixed constants set at the upper quartile values. They depict the most distinct regions of high yield and visualize the impact to the variable pairs on the product yield. The histograms show the distribution of selected parameters per variable, visualizing the differences in condition selection. We observe that for the same chemical space, the resulting final round posteriors from each model show different contours. Generally, there seems to be an agreement for locations of high yield, evidenced by the overlap in the high-yield region. However, Model A has clear, narrow regions compared to Models B and C, which indicate stronger exploitation. These observations from the contour plots and histograms reveal the differences in sampling strategies, as the shape of the contours reflects variations in the selected conditions.

Optimization methods comparison

We structured the optimization experiments to compare two key aspects of the sampling process in our flexible BBO framework development: (1) the effects of pre- and post-clustering in BBO, and (2) the impact of different temperature selection methods. To evaluate the effectiveness of the three optimization methods, bar graphs showing the number of samples greater than 90%, 70%, and 50% yield were constructed. Additionally, we displayed the raw data from all three models against each parameter with the 90% and 50% yield threshold lines to visualize the impact of individual parameters on the outcome.

Pre- and post-clustered temperature. Models A and B determine three temperature levels by clustering temperatures generated from 4D BBO. Since the methods of selecting temperatures are the same, we used the same temperatures for both models. The key difference between the models is that conditions in the other three dimensions of the sulfonating agent, analyte, and time in Model B are regenerated to optimize at the given temperatures, whereas Model A modifies only the temperatures. The result of the effects is apparent in the histograms of the inputs seen in Fig. 3b and 4b. While both models use the same temperature levels, the distribution of all the inputs is different. Notably, Model A tends to sample longer reaction times, temperature splits between the mid and high ranges, and analyte around the median value. On the other hand, Model B appears to demonstrate exploration as the distribution of the histograms for each variable is spread out, while Model A conditions tend to be skewed.

Referring to the bar graph in Fig. 6a, we observe the cumulative count of conditions yielding products greater than 50%, 70%, and 90%. These results highlight the model's ability to locate conditions of high yield. Both models identify conditions with high yield across all rounds, but Model A consistently finds more conditions with high yield. Overall, out of the 45 sampled conditions per model, Model A identified 7 conditions (15.6% of samples in A), while Model B identified 4 conditions above the 90% threshold (8.88% of samples in B). For the >50% threshold, Model A identifies 19 conditions, accounting for 42.2% of the space, while Model B identifies 14 conditions covering 31.1%.

Clustered and BO-generated temperature selection. The comparison of Model B and C focuses on the strategies for temperature selection. Both models use mixed-variable BO to regenerate conditions using predetermined temperatures. The key difference is in the temperature selection strategy, where one employs clustering and the other employs a BBO of batch size 3. As observed in the posterior contours from Fig. 5(a and b), Model C identifies a broader range of high-yield regions compared to the other two models. Additionally, the uncertainty figures for Model C are much lower than the other two models with a maximum uncertainty at round 2 being 0.04 (Fig. S11). This suggests that Model C has more exploration. When looking at the performance of identifying high-yielding conditions

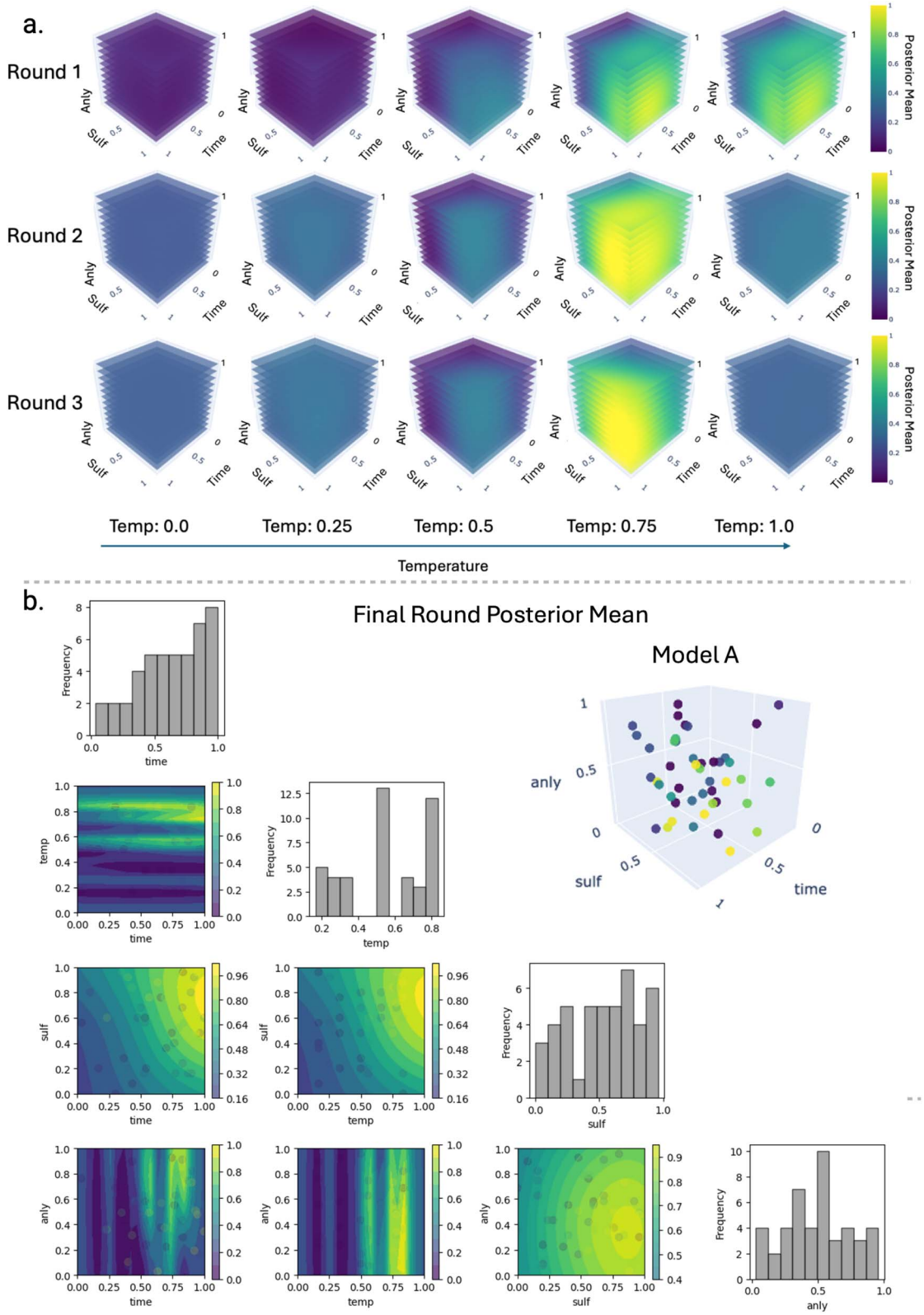

Fig. 3 Model A: T-clustering, (a) four-dimensional (4D) visualization of posterior means for each round (*x*-axis: time, *y*-axis: sulfonating agent, z-slices: analyte, constants: temperature. All axes are normalized and plotted between 0 and 1) (b) two-dimensional posteriors mean contours selected from the final rounds plotted against every input pair (hidden variables are constant at the upper quartile value (0.75). All axes are normalized and plotted between 0 and 1).

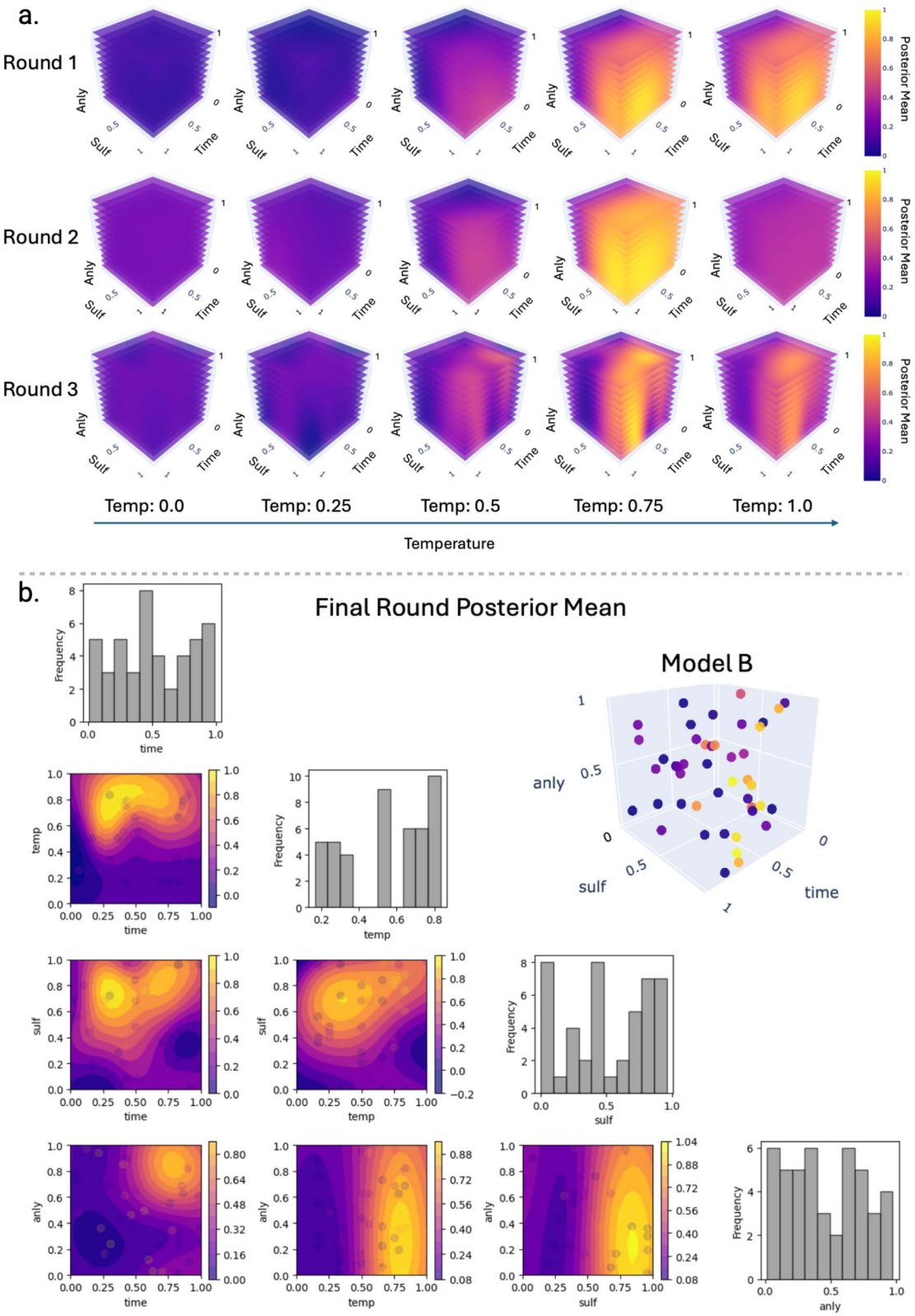

Fig. 4 Model B: T-redistribution, (a) four-dimensional (4D) visualization of posterior means for each round (*x*-axis: time, *y*-axis: sulfonating agent, z-slices: analyte, constants: temperature. All axes are normalized and plotted between 0 and 1) (b) two-dimensional posteriors mean contours selected from the final rounds plotted against every input pair (hidden variables are constant at the upper quartile value (0.75). All axes are normalized and plotted between 0 and 1).

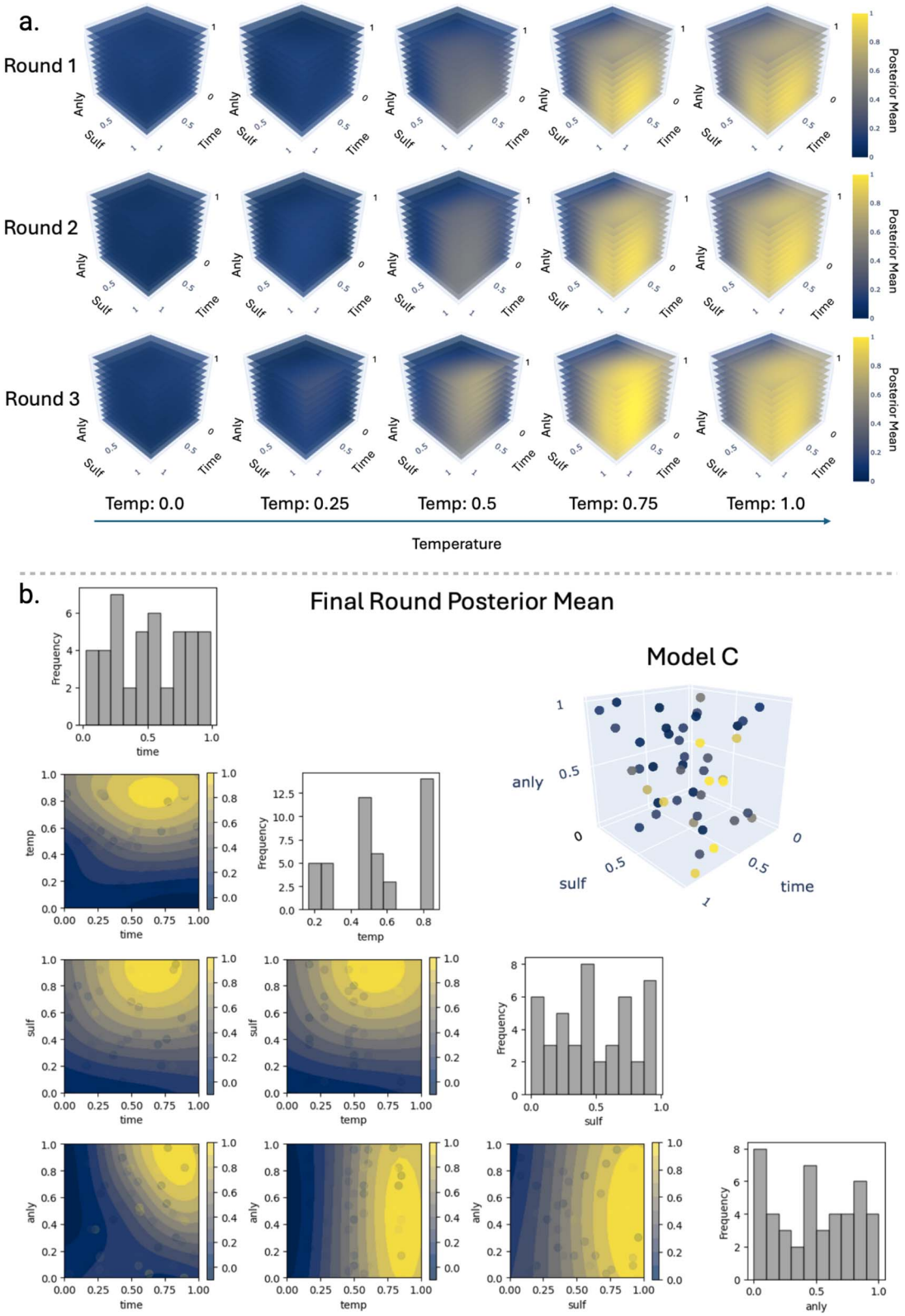

Fig. 5 Model C: T-preselection, (a) four-dimensional (4D) visualization of posterior means for each round (*x*-axis: time, *y*-axis: sulfonating agent, z-slices: analyte, constants: temperature. All axes are normalized and plotted between 0 and 1) (b) two-dimensional posteriors mean contours selected from the final rounds plotted against every input pair (hidden variables are constant at the upper quartile value (0.75). All axes are normalized and plotted between 0 and 1).

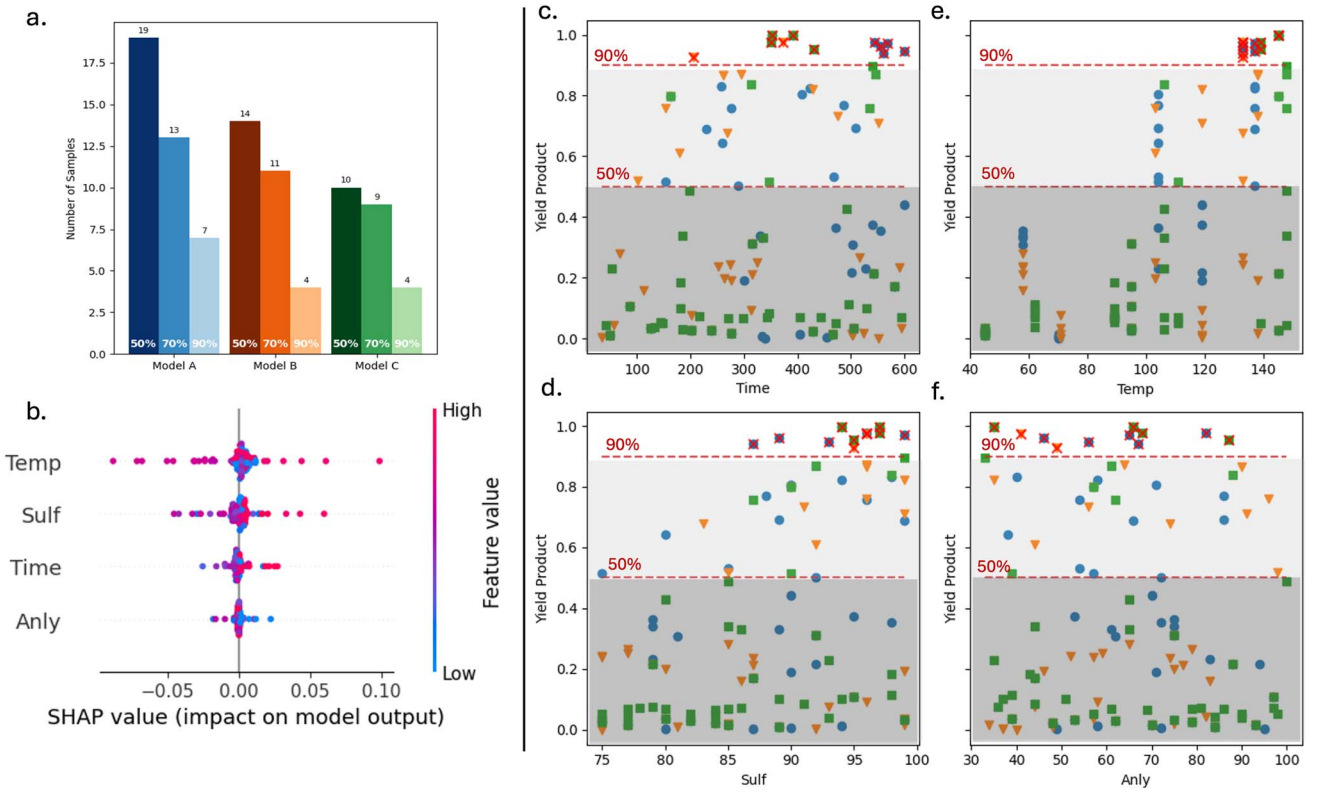

Fig. 6 Visualization of overall results for comparison. (a) Bar graph counting samples that yielded greater than 50%, 70% and 90% yield, (b) SHAP analysis for Gaussian Process Regression with full dataset, Fig. 4c–f plots the data of all three models and yields greater than 90% are marked with a red cross: (c) reaction time *vs.* product-yield of all collected samples, (d) temperature *vs.* product-yield of all collected samples, (e) sulfonating-agent *vs.* product-yield of all collected samples, (f) analyte *vs.* product-yield of all collected samples.

shown in Fig. 6a, Model C struggles to identify high-yielding conditions, evidenced by the low count of high-yielding samples compared to other models. Additionally, we can see that the second sampling round was unsuccessful in finding any conditions greater than 90% yield. Overall, the model identified 10 conditions greater than 50% yield (22.2% of samples in C) and 4 conditions greater than 90% yield (8.8% of samples in C). The lack of exploitation is further confirmed by the GP model's posterior visualizations (Fig. 5a), showing minimal refinement of high-yield regions between each round. In contrast, Models A and B display clearer narrowing onto high-yield regions. Our results suggest that clustering is more effective for temperature selection in this chemical space. On the other hand, the two-stage BBO approach, where mixed-variable BBO is used for resampling, proved less efficient, as Model A overperformed relative to Model B during our campaign of optimizing synthesis conditions of redox-flow molecules.

Selected synthesis conditions

Beyond evaluating the performance of the three flexible BBO models, the scientific goal of this study was to identify an optimal synthesis condition with milder reaction conditions yet producing a high product yield. We are interested in the conditions that generate the highest yield yet minimize the input conditions. However, since we conducted a single objective BBO for each model, we cannot conclude whether the conditions found from this BBO optimization are the optimal conditions minimized. Nevertheless, the high-yield conditions generated from this experiment can give us insight into probable conditions for future work.

A summary of all conditions yielding over 90% product yield is provided in Table 1, which highlights the high-performing outcomes. We observe that there is a slight trade-off between time and temperature, where lower temperatures require longer reaction times and *vice versa*. We also observe that the highest yields were achieved by conditions with the highest temperatures. Generally, there appear to be three distinct groups of conditions. One is characterized by high temperatures, short reaction times, high concentrations of sulfonating agents, and large amounts of analyte, and the other is characterized by longer reaction times, lower temperatures, lower concentrations of sulfonating agents, and higher amounts of analyte. The third grouping exists at lower temperatures, shorter reaction times, and high concentrations of the sulfonating agent, along with low amounts of the analyte. This indicates that there are several local maxima that result in a high yield of product. Referencing the plots of the output yield *vs.* the individual input variables, we can identify the most influential factor affecting the reaction output for chemical insight. Additionally, Fig. 6b presents a SHAP analysis performed on the Gaussian Process model trained with all 105 samples, which provides insight into

Table 1 All synthesized conditions that yielded greater than 90% product

Model	Time (min)	Temp (°C)	Sulf (%)	Anly (mg mL^{-1})	Yield (%)
ABC	353	145	94	66	97
ABC	391	145	97	35	97
A	545	133	96	82	95
A	556	133	89	46	91
A	561	133	87	67	91
A	569	137	99	65	97
A	600	137	93	56	94
B	206	133	95	49	91
B	372	133	96	41	92
C	350	139	97	68	98
C	430	139	95	87	95

the variable influence on the model and chemical space. SHAP (SHapley Additive exPlanations) is a technique used to quantify the contribution of each input variable to the model's predictions and provide insights into the relative importance and influence the input features have on the output. Using these two panels from Fig. 6 to rank the variables by influence, we can see that temperature has the greatest impact, followed by sulfonating agent, reaction time, and analyte. Fig. 6e shows a clear high-yield operating range, where temperatures above 130 °C are essential for generating conditions exceeding 90%, suggesting that lower temperatures fail to activate the substrate efficiently. Referencing Table 1, we can see that the lowest temperature producing high-yield was 133 °C. The sulfonating agent is the second influential variable among the reaction conditions, with an effective concentration range above 87%. Narrowing these search bounds for temperature and sulfonating agents could accelerate optimization. Reaction time ranks third in impact on the reaction outcome. Based on Fig. 4c, the optimal operating ranges for reaction time are between 200 and 600 minutes, providing insight into the trade-off between reactivity and throughput. The shortest reaction time identified in this experiment was 206 minutes, which achieved 91% yield. Finally, the analyte had the least impact on the reaction outcome. Its range spans the defined boundaries without significantly influencing the yield, indicating that it is not a critical variable for optimization. These findings demonstrate that temperature and sulfonating agent concentration are the most influential factors, whereas analyte concentration plays a minor role. The minor role that the analyte plays in the reaction outcome suggests that its primary function is in mass balance rather than mechanistic control, allowing for flexibility in batch sizing during scale-up. A particularly significant finding is that using 90% *vs.* 99% sulfonic acid results in no significant decrease in product yield, which is a crucial insight for scaled-up manufacturing, as it suggests that slight dilution does not impair the reaction outcomes. Achieving high yields with regular sulfuric acid at moderate temperatures and reduced reaction times is particularly promising for scalability and process efficiency. Overall, the results align with expectations while offering some unexpected simplifications and efficiencies, reinforcing the utility of the multi-step BO method.

Surrogate model evaluation

As an active learning method, BO follows an iterative learning process, where our understanding of the design space improves as more data is collected. With an increasing number of data points, the surrogate models gradually capture the ground truth with reduced uncertainty. The uncertainty from the models seen in Fig. S9–S11 provides insight into the confidence of the prediction, as the GP models give the variance of the posterior that quantifies the uncertainty associated with the projections.

As one method of evaluation, we compare the GP surrogate models trained on data collected using each sampling strategy against a GP model built using the combined dataset from all strategies (A, B, and C models). This comparison allows us to evaluate the effectiveness of each model in capturing the underlying design space. Fig. 7 shows the parity plots comparing the actual experimental outputs to the predicted posterior mean evaluated by each model. For each parity plot, the root mean square errors (RMSE) are determined. The Model ABC shown in Fig. 7a is trained with the full set of data (105 samples) and serves as a baseline for how well the total space is explored from this campaign. The resulting RMSE value for the full model is 0.10, which suggests that the predictions are very close to the actual experimental yields. The partial set of data shows that the models tend to under-predict the yields. This prediction variability is potentially due to the noise between the repeated specimens. The parity plots shown in Fig. 7b, c, and d are constructed by training the models with their corresponding partial data consisting of 45 samples. The test data used for evaluation is the combined set of data from all models (105 samples), including the training samples. This allows us to evaluate how well each model is able to predict the actual yield that was collected from other sets, indicating how well the space is defined by that model. The results of the parity plots show that Model A has the lowest RMSE value of 0.18 and Model B has the highest RMSE of 0.25. This suggests that Model A represents the chemical space closest to the true space and suggests that Model A performed the best and is fastest at learning, as evidenced by the combined results from Fig. 6a and 7a.

Due to BO's stochastic nature, the optimization outcome differs based on the initial training conditions affecting the model's acquisition. Therefore, robustness testing of the BBO frameworks is essential to ensure generalization. To address this, we conducted a pool-based BO where we randomly selected the BBO frameworks, but rather than sampling for new data, we queried the 105 collected samples. The pool-based BO was repeatedly run 50 times with different initial training sets for robustness testing. For baseline comparison, we also included a pool-based random search. The results comparing the pool-based BO with the experimental campaign are shown in Fig. S1. We observed that Model A and Model B outperform the random search, while Model C performs worse. However, this is not a fair comparison as the random search draws samples

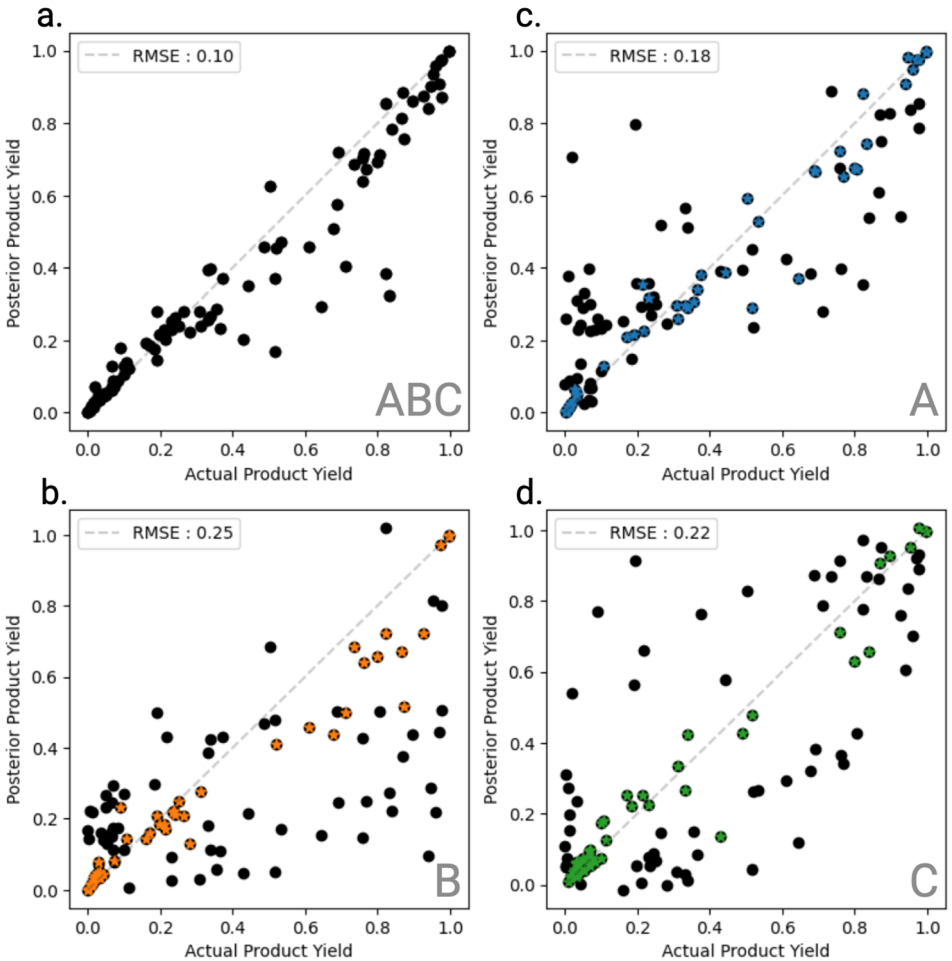

Fig. 7 Parity plots to evaluate model performance and Root Mean Square Errors (RMSE) calculated from all data (a) full data set model: training data of 105 conditions. Test data of 105 conditions, (b) partial data set using Model A data points: training data of 45 conditions colored in blue, test data of 105 conditions, (c) partial data set using Model B data points: training data of 45 conditions colored in orange, test data of 105 conditions, (d) partial data set using Model C data points: training data of 45 conditions colored in green, test data of 105 conditions.

from a pool of 105 samples, while the Models for the experimental campaign query from the total chemical space, which has 152 438 728 possible samples. For a better comparison, Fig. S1b presents the pool-based results for random search and all models querying the 105 sample pool. The results show that all model frameworks outperform random search. Between models, we observe that the number of high-yielding samples is relatively similar, but Model C performs slightly better, as suggested by having the lowest average RMSE and a higher mean count for high-yield samples. This suggests that Model C may have stronger potential for generalization across different training sets.

To further assess the generalizability as well as the scalability of the BBO frameworks, we applied them to an external, open-source dataset. The open-source dataset we used was a 3D printed cross-barrel toughness dataset reported by Gongora *et al.*, where the input variables are number of struts, n, displacement angle, theta (degrees), strut radius, r (mm), and strut thickness t (mm), and the output was mechanical toughness.[30] Since this data set has the same 4D space as our own experiments of redox-flow molecular synthesis, we structured the pool-based BOO setup for the experimental campaign using a batch size of 15 over three iterations. Robustness testing against the random search was also conducted for this pool-based analysis. The results showed that all three model frameworks, Model A, B, and C, outperformed the random search, which is consistent with the pool-based results from the experiments' dataset. Additionally, we observed that the RMSE rankings between the Gaussian process regression models remained consistent with the campaign's results, where Model A has the lowest RMSE and Model B has the highest. Our findings suggest that within the first three rounds using a batch size of 15, Model A was able to form the most accurate surrogate model. To complement these generalizability and robustness tests, we conducted additional benchmarking to investigate the influence of batch size on optimization performance. We compared traditional single-point BO with BBO using batch sizes of 3, 5, and 15 and constrained the number of samples to 50, including the training data, for fair comparison. Fig. S2 shows the results of this benchmarking, comparing BO and

BBO across iteration counts of 45, 15, 9, and 3, respectively. We observed that traditional single-point BO converges faster to the optimal compared to the larger batch sizes, highlighting the trade-off between convergence and parallel sampling. However, when considering hardware and synthesis time, larger batch sizes can be inevitable.

It is worth noting that the ideal model for our system is one that can select temperatures and other parameters non-sequentially. Current BO libraries lack models that can adapt dynamically to variable-size design spaces. The closest available approach we found was mixed-variable BO, which assumes a fixed categorical variable, which this work demonstrated. The poor performance of the mixed-variable BO model approaches tested in this work is likely attributed to this assumption. We have found prior work on variable-size BO for mixed-variable and variable-size design space problems. However, this work was designed for a highly complex single-objective system and reported no experimental validation.[31] Future work for non-sequential approaches is to design BO for variable-size design spaces or mixed-variable BO with updatable categories.

Experimental efficiency

The effectiveness of our BBO approach is further highlighted by comparing the outcomes of the models to a brute-force method, which is if the experiment were run by sampling every point in the space using a grid. The detailed calculation of the sample space and durations can be found in the SI under section "Efficiency Comparison". To collect samples for the entire space, we would need to collect 152 438 728 samples for a step size of 1, and since our system can only run 45 conditions at a time, we would have to run 3 387 528 rounds of data collection. The synthesis time for a single round of experiments takes approximately 12 to 14 hours, and the characterization is an additional 5 to 6 hours. Hence, a single round of data collection takes between 17 and 20 hours. Hence, running 3 million rounds is impossible. Increasing the step size to 5 would reduce the number of samples requiring 6655 rounds, but it is still impractical as it would take about 15 years of continuous data collection, which is highly inefficient in terms of the allocation of time and resources. Comparably, our BBO approach explored the space and identified several conditions that fit our target of milder conditions in only three rounds. The uncertainty of the Gaussian model is relatively low, which suggests that the space is adequately defined. Hence, our BBO approach is more effective in defining the space and locating high-yield and, consequently, more effectively optimizing targets in high-yield regions.

Conclusions

In this work, we developed an experimental workflow leveraging BO-guided HTE systems to optimize the synthesis conditions of fluorenone derivatives for RFBs. The experimental design involved a two-step process where samples were formulated and subsequently heated. Hence, our workflow entailed varying variable sizes in a single batch due to differing hardware capacity constraints. To overcome the challenge of constrained variables, we developed an innovative, flexible approach for our BBO to fit our HTE workflow needs. We demonstrated the implementation of Batch BO in HTE platforms employing three distinct sampling strategies and evaluated their effectiveness in exploring and exploiting the chemical space. Our results showed that the post-clustering strategy (Model A) was the most effective as it yielded the highest number of conditions with high product yield and performed the best predictions for our data set. The models that performed worse could be attributed to the improper use of mixed-variable BO, as mixed-variable BO assumes a static categorical input. Nevertheless, our BBO framework enabled us to achieve our goal of determining high-yield conditions under milder sulfonation conditions. We found that from the 45 conditions sampled in Model A, 7 conditions reached over 90% yield, with the lowest temperature at 133 °C and the lowest concentration of sulfonating agent at 87%. In Model B, we identified 4 conditions greater than 90% yield out of the 45 sampled, with the lowest temperatures at 133 °C and the lowest sulfonating agent at 95%. Finally, from the 45 conditions sampled in Model C, we identified 4 conditions with yield greater than 90%, with the lowest temperature at 139 °C and with the lowest sulfonating agent at 90%. Overall, all conditions found with greater than 90% yield had temperatures below 150 °C. As demonstrated in this study, the presence of small amount of water does not adversely affect the reaction, provided that the concentration of sulfuric acid exceeds 90% and the temperature is in the range of 130 °C to 150 °C. Additionally, this study provides valuable insight into the emerging field of SDLs, highlighting the strategies for implementing BO-based decision-making in predefined hardware. We demonstrate the overcoming of the challenges of hardware and software misalignment.

Methods: digital

Traning data

The three models used the same training set, where conditions were generated using Latin hypercube sampling (LHS). LHS was the most effective sampling method for the four variable spaces since grid search sampling showed that it minimally covered the space for 15 conditions, and random sampling effectively covered the space more than the grid search. Still, the spread between values was worse than LHS. We used the Python package pyDOE 0.3.8, which produced a 2D array of size 4×15. Each variable was given a minimum and maximum value from which the *lhs* function sampled values between bounds. The bounds for each variable were determined from preliminary experiments. Three identical vials were made for each condition and three controls were made for 48 samples. The conditions after clustering were sent to the HTE system for synthesis and characterization.

Clustering

The clustering was done using the Scikit-learn 1.6.0 Python package, which employed the *k*-means clustering algorithm.[32]

The *k*-means clustering function parameters were set to cluster 3 groups with an initial value of $n_{\text{init}} = 95$, the mid value of the temperature bounds.

Data extraction and data preparation

Feature extraction from the High-Performance Liquid Chromatography (HPLC) data was performed using the Python package hplc-py 0.2.7. The software in this Python package is built to fit the peaks detected in the HPLC, where the area under the curve (AUC) is calculated by selecting the regions of interest. Regions of interest are selected at the hypothesized retention times (RT).[33] The three peaks of interest for our experiment were the product peak at 0.29 min RT, the reactant peak at 3.92 min RT, and anything unknown, which tends to be found at around 4.3 min RT. The AUC was extracted from the regions of interest to calculate the total product yield, which is determined by (eqn (1)) product area/total area, where the total area is the sum of the product area, reactant area, and area of the unknown. The mean and standard deviation were calculated since each generated condition had three repeated samples. The mean of the three samples is the training output for the model, and the standard deviation is the noise when training the GP models.

$$\text{Product yield} = \frac{\text{product AUC}}{\text{product AUC} + \text{reactant AUC}} \quad (1)$$

GP models and batch BO functions

Our GP models and BO framework were all done in BOTorch.[34,35] SingleTask GP is the default model of the BOTorch library, pulled from GPyTorch,[34,35] that uses a Matern kernel and works best with normalized values. Hence, all inputs were normalized before being implemented into their respective GP models using BOTorch *botorch.utils.transforms.normalize*. The two GP models we used for our models were *heteroskedasticSingleTaskGP* and *MixedSingleTaskGP*. Our system used Heteroskedastic SingleTask GP, which is much like the singletask GP but treats noise independently. We used *heteroskedasticSingleTaskGP* in Model A and the temperature selection portion of Model C. *MixedSingleTaskGP* were used for Models B and C to generate optimized parameters post-temperature selection. The mixed-variable BO in BOtorch is based on Hammington's distance to convolve the categorical and continuous variables. We used *optimize_acqf* and *optimize_acqf_mixed* for the minimization function for their respective models. While all frameworks presented in this study use GP regression as the surrogate model, other surrogate models can also be applied within the presented frameworks. This is exemplified in the benchmarking analysis comparing GP regression and Random Forest Regression as detailed in the SI under section "Pool-Based Bayesian optimization" subsection "Surrogate Model Comparison". The BBO framework in this study employed the qExpected Improvement (qEI) acquisition function, which was selected after initial pool-based simulation comparing three analytical acquisition functions (qExpectedImprovement, qUpperConfidenceBound, qProbabilityOfImprovement). qEI is the batch version of the expected improvement for our acquisition function since regular expected improvement only takes a batch size of 1.[26,36] In the preliminary acquisition testing, the acquisition functions resulted in similar results; therefore, we chose qEI, as it naturally balances exploration and exploitation. This pool-based analysis can be found in SI under "Acquisition Function".

Methods: experimental

Sample collections were done at PNNL's automated robotics for energy storage.

ARES: PNNL high-throughput facilities.

Video: autonomous organic synthesis video.

Robotic platform

An automated material handling system from Unchained Labs (Big Kahuna model) was used for sample preparation and reaction execution. The automated system is equipped with an analytical balance (0.1 mg readability), solid dispensers (1 mg to 25 g), liquid handlers including the positive displacement pipetting for viscous liquids/slurries (10 μL to 10 mL), capping/uncapping station, on-deck magnetic stirrer with heating/cooling (−20 °C to 180 °C) function, and vortex mixer (Fig. S7). The robotic platform, including liquid/solid handling, thermal control, and stirring, was benchmarked in prior work, with routine calibrations ensuring <5% relative errors. Each data point in this study represents the average of triplicate samples, with variability monitored and outliers excluded. ^{1}H NMR confirmed product identity against a manual benchmark, and cross-lab HPLC validation confirmed the reproducibility of results across independent instruments.

High-throughput automated sulfonation workflow

Following a typical sulfonation synthesis protocol described in the literature,[37,38] we implemented a high-throughput automated workflow using a modular robotic platform. Each experimental batch consisted of 48 vials processed concurrently in approximately 10 hours. The workflow was orchestrated *via* automation software to ensure seamless operation and precise execution. Three pre-calibrated heating modules were maintained at the target reaction temperatures, while sample preparation was conducted on an ambient deck. In each vial, 10 mg of 9-fluorenone-2-carboxylic acid (2CLF) was dispensed *via* a powder hopper, followed by the addition of 5 μL of water using a fixed-volume syringe and 132 μL of sulfuric acid (99.99%, Sigma-Aldrich) *via* a disposable positive displacement pipette. The vials were then capped, stirred at 700 rpm, and transferred to the heated modules based on the experimental design. After the reaction time elapsed, the vials were sequentially moved to the ambient deck for cooling, and this iterative process was repeated until all reactions were complete.

Post-reaction, 1 mL of dimethyl sulfoxide (DMSO) was added to each reaction vial under stirring to dissolve the products. An equal volume of a 60 : 40 (v/v) acetonitrile/water solution was added to UPLC vials, followed by aliquoting 10 μL of the DMSO-diluted reaction mixture into these vials. The UPLC vials were

vortexed at 800 rpm and analyzed using a Waters ACQUITY UPLC system equipped with a BEH C18 column (130 Å, 1.7 μm, 2.1 mm × 50 mm). This fully automated workflow demonstrated precise control over reagent handling, reaction timing, and sample preparation, enabling reproducible and high-quality data acquisition. These capabilities underscore the effectiveness of high-throughput automation in accelerating reaction optimization and materials discovery.

Automation experimental procedure

The automated experiments were performed on a 150 μL total volume scale. Analyte, solvent, and sulfonating agent were added to 2 mL glass vials, capped and then transferred to one of three pre-heated reactor blocks. The system was then programmed to transfer the vials from heat to a room-temperature deck at designated time intervals. At the end of the run, the samples were dissolved in dimethylsulfoxide before diluting with 60/40 acetonitrile/water mixture for UPLC analysis. An example library design is provided.

Multi-robot sample transferring

Transfer of samples between the synthesis and characterization units is achieved through multi-robot collaboration under the Robot Operating System (ROS) framework. The mobile robot (Ubiquity) autonomously navigates the laboratory, efficiently performing sample transfer tasks as specified by the algorithm. A Universal Robot UR3e facilitates the loading and unloading of samples between the material handling system and the mobile robot. Similarly, a Universal Robot UR5e manages the loading and unloading of samples between the characterization system (*e.g.*, HPLC) and the mobile robot. Both UR robotic arms are equipped with wrist cameras to precisely locate the mobile robot's docking configuration for precision load transfer.

Note that a workstation is connected to and monitors all the aforementioned devices *via* the ROS network, serving as the master device for receiving experiment suggestions generated by BO and executing the plan through the multi-robot system. Additionally, the mobile robot is equipped with advanced perception units, including lidar and sonar, enabling it to detect and avoid collisions when humans or other moving obstacles are present in the environment.

Author contributions

Experimental design was done by H. J. and Y. L. and S. S. Data collection was done by H. J. HPLC extraction, model design, and model evaluation were done by C. T. Transport system execution, and code review was done by H. C. Conceptualization was done by Y. L. and S. S. and W. W., Y. L., and S. S. supervised the project.

Conflicts of interest

There are no conflicts to declare.

Data availability

Data and processing scripts for this paper, including figure generation, are available at **https://doi.org/10.5281/zenodo.16895908**, or from the GitHub repository UWSunLab/BO-RFB. Pool-based Bayesian optimization for scalability testing utilized open-source data from DOI: **https://doi.org/10.1126/sciadv.aaz1708** which is available at **https://www.kablab.org/data**.

Supplementary information: pool-based simulations, additional figures for the results, and additional chemical insights. See DOI: **https://doi.org/10.1039/d5dd00017c**.

Acknowledgements

The authors acknowledge Dr Sterling Baird, Dr Osman Mamun, Dr Cory Simon, and Dr Steven Torrisi for fruitful discussions on handling the varying batch size issues in a single round of BO sampling, as well as Dr Maher Alghalayini on the constrained BO framework. The work conducted at the University of Washington was funded startup fund from the Department of Mechanical Engineering, University of Washington, and by the Energy Storage Materials Initiative (ESMI) at the Pacific Northwest National Laboratory (PNNL) under Contract No. 711979. The work at PNNL is supported as part of the Energy Storage Research Alliance (ESRA), an Energy Innovation Hub funded by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES), under award number DE-AC02-06CH11357, PNNL FWP 82132; and the ESMI program at PNNL, a Laboratory Directed Research and Development project under Contract No. DE-AC05-76RL01830 (idea generation, initial experiment). Generative AI applications including ChatGPT and GitHub Copilot were used to aid in plotting figures. The illustrations in Fig. 1 and 2 were made using BioRender **https://www.biorender.com**.

References

1 R. Jayabal, *Results Eng.*, 2024, **24**, 103121.
2 P. Leung, A. Shah, L. Sanz, C. Flox, J. Morante, Q. Xu, M. Mohamed, C. Ponce De León and F. Walsh, *J. Power Sources*, 2017, **360**, 243–283.
3 K. Wedege, E. Dražević, D. Konya and A. Bentien, *Sci. Rep.*, 2016, **6**, 39101.
4 R. Feng, X. Zhang, M. Vijayakumar, A. M. Hollas, Y. Chen, X. Wei, Z. Nie, V. Sprenkle, J.-G. Zhang and W. Wang, *Science*, 2021, **372**, 836–840.
5 J. Rodriguez, C. Niemet and L. D. Pozzo, *ECS Trans.*, 2019, **89**, 49–59.
6 N. W. Stauffer, *Flow batteries for grid-scale energy storage*, MIT News, Massachusetts Institute of Technology, 2023, **https://news.mit.edu/2023/flow-batteries-grid-scale-energy-storage-0407**.
7 C. Ye, A. Wang, C. Breakwell, R. Tan, C. Grazia Bezzu, E. Hunter-Sellars, D. R. Williams, N. P. Brandon, P. A. A. Klusener, A. R. Kucernak, K. E. Jelfs, N. B. McKeown and Q. Song, *Nat. Commun.*, 2022, **13**, 3184.

8 G. Kwon, S. Lee, J. Hwang, H.-S. Shim, B. Lee, M. H. Lee, Y. Ko, S.-K. Jung, K. Ku, J. Hong and K. Kang, *Joule*, 2018, **2**, 1771–1782.
9 P. Arévalo-Cid, P. Dias, A. Mendes and J. Azevedo, *Sustainable Energy Fuels*, 2021, **5**, 5366–5419.
10 M. Abolhasani and E. Kumacheva, *Nat. Synth.*, 2023, **2**, 483–492.
11 Y. Liang, H. Job, R. Feng, F. Parks, A. Hollas, X. Zhang, M. Bowden, J. Noh, V. Murugesan and W. Wang, *Cell Rep. Phys. Sci.*, 2023, **4**, 101633.
12 Y. Cao and A. Aspuru-Guzik, *Nat. Comput. Sci.*, 2024, **4**, 89–91.
13 A. Jain, I. A. Shkrob, H. A. Doan, K. Adams, J. S. Moore and R. S. Assary, *ACS Appl. Mater. Interfaces*, 2023, **15**, 58309–58319.
14 K. McCullough, T. Williams, K. Mingle, P. Jamshidi and J. Lauterbach, *Phys. Chem. Chem. Phys.*, 2020, **22**, 11174–11196.
15 S. M. Mennen, C. Alhambra, C. L. Allen, M. Barberis, S. Berritt, T. A. Brandt, A. D. Campbell, J. Castañón, A. H. Cherney, M. Christensen, D. B. Damon, J. Eugenio De Diego, S. García-Cerrada, P. García-Losada, R. Haro, J. Janey, D. C. Leitch, L. Li, F. Liu, P. C. Lobben, D. W. C. MacMillan, J. Magano, E. McInturff, S. Monfette, R. J. Post, D. Schultz, B. J. Sitter, J. M. Stevens, I. I. Strambeanu, J. Twilton, K. Wang and M. A. Zajac, *Org. Process Res. Dev.*, 2019, **23**, 1213–1242.
16 S. Langner, F. Häse, J. D. Perea, T. Stubhan, J. Hauch, L. M. Roch, T. Heumueller, A. Aspuru-Guzik and C. J. Brabec, *Adv. Mater.*, 2020, **32**, 1907801.
17 A. O. Oliynyk, E. Antono, T. D. Sparks, L. Ghadbeigi, M. W. Gaultois, B. Meredig and A. Mar, *Chem. Mater.*, 2016, **28**, 7324–7331.
18 A. D. Clayton, A. M. Schweidtmann, G. Clemens, J. A. Manson, C. J. Taylor, C. G. Niño, T. W. Chamberlain, N. Kapur, A. J. Blacker, A. A. Lapkin and R. A. Bourne, *Chem. Eng. J.*, 2020, **384**, 123340.
19 X. Chen, X. Liu, X. Shen and Q. Zhang, *Angew. Chem., Int. Ed.*, 2021, **60**, 24354–24366.
20 L. Zhichao, M. Dong, L. Xiongjun and Z. Lu, *Commun. Mater.*, 2024, **5**, 76.
21 B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, J. M. Janey, R. P. Adams and A. G. Doyle, *Nature*, 2021, **590**, 89–96.
22 B. Ru, A. S. Alvi, V. Nguyen, M. A. Osborne and S. J. Roberts, *Proceedings of the 37th International Conference on Machine Learning*, 2020.
23 A. M. K. Nambiar, C. P. Breen, T. Hart, T. Kulesza, T. F. Jamison and K. F. Jensen, *ACS Cent. Sci.*, 2022, **8**, 825–836.
24 Y. Wu, A. Walsh and A. M. Ganose, *Digital Discovery*, 2024, **3**, 1086–1100.
25 M. B. Alghalayini, D. Collins-Wildman, K. Higa, A. Guevara, V. Battaglia, M. M. Noack and S. J. Harris, *Cell Rep. Phys. Sci.*, 2025, **6**, 102543.
26 N. Hunt, PhD thesis, Massachusetts Institute of Technology, 2020.
27 S. Dutta and A. H. Gandomi, *Handbook of Probabilistic Models*, Elsevier, 2020, pp. 369–381.
28 E. Daxberger, A. Makarova, M. Turchetta and A. Krause, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020, pp. 2633–2639.
29 Y. Zuo, A. Dezfouli, I. Chades, D. Alexander and B. W. Muir, Bayesian Optimisation for Mixed-Variable Inputs using Value Proposals, *arXiv*, 2022, preprint, arXiv:2202.04832 [stat], DOI: **10.48550/arXiv.2202.04832**.
30 A. E. Gongora, B. Xu, W. Perry, C. Okoye, P. Riley, K. G. Reyes, E. F. Morgan and K. A. Brown, *Sci. Adv.*, 2020, **6**, eaaz1708.
31 J. Pelamatti, L. Brevault, M. Balesdent, E.-G. Talbi and Y. Guerin, Bayesian optimization of variable-size design space problems, *arXiv*, 2020, preprint, arXiv:2003.03300 [math], DOI: **10.48550/arXiv.2003.03300**.
32 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Muller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, Scikit-learn: Machine Learning in Python, *arXiv*, 2018, preprint, arXiv:1201.0490 [cs], DOI: **10.48550/arXiv.1201.0490**.
33 G. Chure and J. Cremer, hplc-py: A Python Package For Rapid Peak Quantification in Complex Chromatograms, 2023, **https://chemrxiv.org/engage/chemrxiv/article-details/6520405345aaa5fdbb709f2f**.
34 M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson and E. Bakshy, BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization, *arXiv*, 2020, preprint, arXiv:1910.06403 [cs], DOI: **10.48550/arXiv.1910.06403**.
35 J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger and A. G. Wilson, GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration, *arXiv*, 2021, preprint, arXiv:1809.11165 [cs], DOI: **10.48550/arXiv.1809.11165**.
36 W. Lyu, F. Yang, C. Yan, D. Zhou and X. Zeng, *Proceedings of the 35th International Conference on Machine Learning*, 2018, pp. 3306–3314.
37 G. Rieveschl and F. E. Ray, *Chem. Rev.*, 1938, **23**, 287–389.
38 H. Sharghi, P. Shiri and M. Aberi, *Beilstein J. Org. Chem.*, 2018, **14**, 2745–2770.