

Perspective

A practical guide to Bayesian optimization for organic chemists

Rachel Shey,^{1,3} Sven Roediger,^{2,3} Abigail G. Doyle,^{2,*} and Julian G. West^{1,*}

¹Department of Chemistry, Rice University, Houston, TX 77005, USA

²Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA

³These authors contributed equally

*Correspondence: abigaildoyle@g.ucla.edu (A.G.D.), jgwest@rice.edu (J.G.W.)

<https://doi.org/10.1016/j.checat.2026.101889>

THE BIGGER PICTURE The development of new chemical reactions often commences with experiments that are conducted based on procedures from textbooks and manuscripts. The goal is to find conditions that provide high-performing reaction outcomes. However, relatively little attention is focused on reaction optimization as a topic, and many research chemists are unaware of state-of-the-art tools for optimization. In academic chemistry, intuition-based optimization is the norm, despite mounting evidence that reaction modeling and algorithmic techniques are more efficient. The use of data-driven techniques has the potential to not only streamline optimization campaigns but also discover improved conditions compared with conventional methods. This is especially relevant for complex systems such as the ones handled in catalytic reactions. Bayesian optimization (BO) supplies an interesting and flexible solution to this universal problem in synthetic chemistry. This field is also ripe for further investigation to understand how to best apply BO. Wise utilization of this technique is paramount, as the potential chosen parameters and specified levels are infinite. This article is intended to be a starting point for curious chemists!

SUMMARY

Bayesian optimization (BO) is a strategy for global optimization, based on iterative refinement of a machine learning model. This algorithm can aid chemists in finding the highest possible reaction outcome (e.g., yield, *er*, or turnover frequency) from a relatively small number of experiments. In recent times, BO has garnered great interest in the organic synthesis community as a computer-guided method to increase the efficiency of reaction optimization. This method complements design of experiments (DoE) and traditional “one variable at a time” (OVAT) experimentation. However, beginning a BO campaign can be an intimidating prospect to an organic chemist. To lower the barrier of entry, we wish to share some insights gleaned from experience using BO for reaction optimization.

INTRODUCTION

Academic chemists often rely on “one variable at a time” (OVAT) optimization campaigns (Figure 1), in which a single variable is changed, and a reaction target, often the yield, is measured until the best value is discovered. However, OVAT results in both inefficient sampling and a high likelihood of missing optimal experimental conditions, as only thin slices of the reaction space can be assayed and synergistic effects between variables are entirely ignored (Figure 1).¹ These are especially relevant for complex systems such as the ones handled in catalytic reactions.

There are several alternatives to OVAT, including design of experiments (DoE) and Bayesian optimization (BO) (Figure 1).²

In contrast to DoE, BO is incredibly flexible; campaigns can be augmented on the fly. Furthermore, conclusions can be drawn after exploring a relatively small region of a given search space, allowing a chemist to consider much larger search spaces. BO is especially advantageous in mixed search spaces with both categorical and continuous variables that are more challenging for DoE. BO (and DoE) is less popular than OVAT, partly because of a misconception that it requires specialized statistics and programming knowledge. This work is intended to dispel such misconceptions and provide a guide to initializing and conducting a BO campaign, targeted to organic chemists. For discussion of more technical aspects of BO for chemical reaction optimization, we refer to the literature.^{3,4}

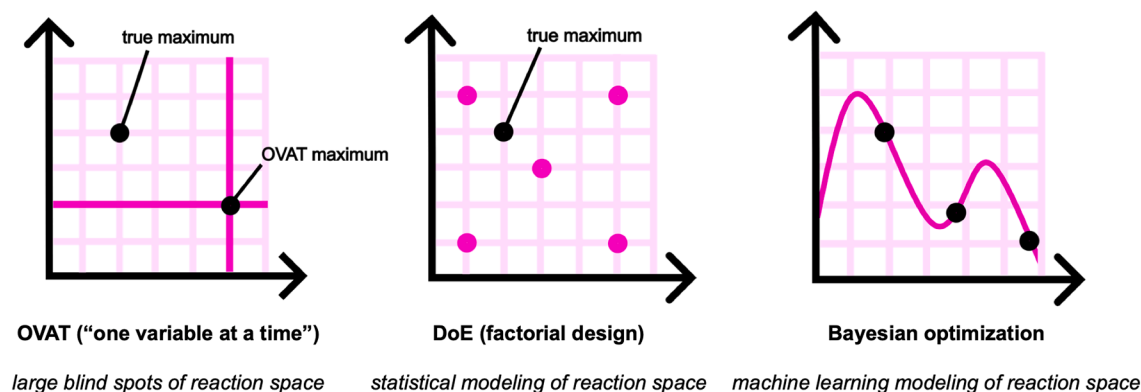


Figure 1. Depiction of OVAT optimization versus computationally aided optimization techniques

GENERAL OVERVIEW

BO includes a surrogate model, which provides predictions of the objective values across the search space (collection of considered reaction conditions). This objective is often the reaction yield, but it can also be a selectivity, catalyst turnover number, or other targets of interest. Notably, multiple objectives can also be optimized at the same time. The information from the surrogate model is then used by an acquisition function, which provides new experimental suggestions by finding a compromise between exploitation of high predictions and exploration of undersampled regions. The surrogate model and acquisition function are already implemented in software packages for BO and ready for use. The user only needs to define the search space and objectives to initialize an optimization campaign. The optimizer will provide experimental suggestions, while the chemist heads into the lab to perform the experiments and then inputs the new yields into the optimizer to obtain new suggestions (Figure 2).

Choosing an optimizer

Options abound, ranging from user-friendly apps like JMP 19,⁵ which now features BO, to Python packages, which provide ultimate flexibility, some of which have graphic user interfaces (GUIs) and interface with a webpage.

JMP 19 is an app that is free for academics and now has a BO feature. This has the benefit of automatically displaying data in a digestible format and being readily accessible for chemists who are already familiar with using JMP for DoE.

While JMP is a general-purpose data science program, Python-based BO algorithms were specifically designed for optimization in chemistry and can, therefore, be better equipped to handle the data and objectives that are used in this field. Python knowledge is optional when using established BO packages, but learning some basics can be helpful (e.g., common Python libraries like pandas). To set up a Python-based algorithm, the user should first install a package management system for Python, such as miniconda.⁶ Next, the user should download the algorithm code from GitHub. The respective GitHub pages are accompanied by installation instructions. Typically, these involve opening the terminal in the root folder of the algorithm and

executing terminal commands to create, activate, and install required packages into a virtual environment. Similar to how each room in a home is equipped with tools for certain tasks (e.g., cooking utensils in a kitchen), a virtual environment provides the tools to run a specific program, in this case a BO campaign. The environment has to be installed only once and requires simple re-activation whenever the user wants to execute the BO program. The GitHub sites of the BO algorithms also provide usage examples. Replacing the search space variables with the desired ones (*vide infra*) and executing the code allow the user to get experimental suggestions. Popular BO frameworks for chemistry include EDBO+ (featuring mono and multi-objective optimization),^{2,7} Web-BO,⁸ BayBE (boasting substantial documentation),⁹ and Gryffin.¹⁰ The first two listed examples also include a web app interface for a codeless interaction with the respective program. For reviews comparing some of these optimizers in more depth, we refer to the literature.¹¹

Creating a search space

Once an optimizer has been chosen, the next task is to define the search space, which spans the range of reaction conditions that are considered. Here is where chemical intuition has a place to shine by considering prior experiments and literature knowledge. Categorical variables (e.g., solvents, pre-catalysts, and ligands) can be implemented alongside continuous variables (e.g., reagent loadings, temperature, and light wavelength). The initial creation of the search space does not have to be a fraught decision, as the campaign can easily be paused and augmented at any point throughout the campaign. A search space for a BO campaign can be much larger than a search space for an OVAT or a DoE campaign. While there is no strict limit for the maximum size of the search space, a larger space requires more experiments to be explored. Search spaces consisting of tens of thousands or even hundreds of thousands of experiments¹² can often be explored in a reasonable number of experiments.

The basis of predictive modeling in chemistry is the description of the chemical context (such as molecules or reaction conditions) using machine-readable descriptors and their correlation with the output labels to establish a model that can learn these correlations and make predictions. In the context of BO, the

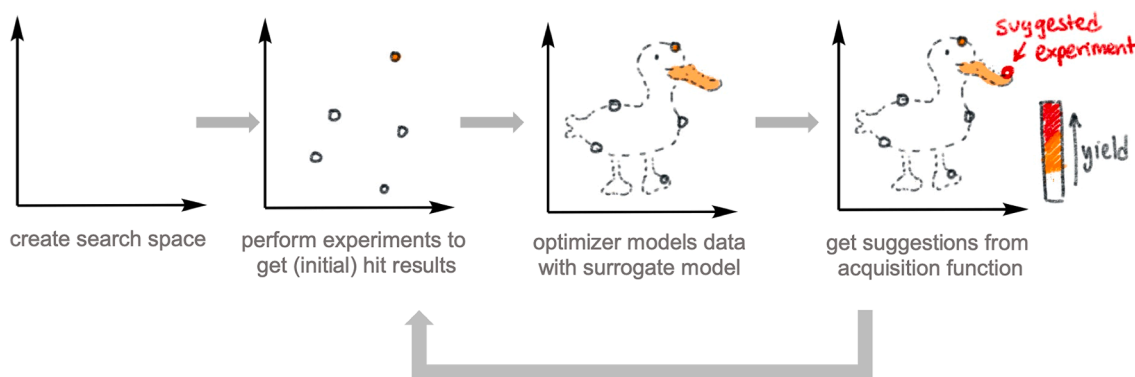


Figure 2. Bayesian optimization workflow

conditions and objectives are described using numerical values. However, there are different strategies that can be pursued. Settings for continuous variables are straightforwardly defined as their numerical values, but categorical variables can be trickier to interpret as they also need to be transformed to a numerical representation for the BO algorithm. One available avenue is to define variables with one-hot-encoding (OHE), which simply assigns each variable setting a vector that is equidistant to the other possible settings. OHE is the default handling of categorical variables in many algorithms but contains no chemical information due to the equidistant treatment. However, it is still possible to embed some chemical insights using this technique by OHE of multiple aspects of one variable. For instance, when featurizing different pre-catalysts (e.g., $\text{Pd}(\text{PPh}_3)_4$, $\text{Pd}_2(\text{dba})_3$, and $\text{Pd}(\text{OAc})_2$), the metal charge and ligand identity could be treated separately with OHE so that the algorithm would have increased information about the difference between the Pd pre-catalysts. More chemical information can be encoded by using cheminformatics approaches, such as RDKit featurization¹³ or Mordred descriptors,¹⁴ which are sometimes bundled into the software package for BO. These features describe molecules through predicted properties such as logP and counts of certain functional groups. Chemists can also use density functional theory (DFT)-based featurization to encode categorical variables representing distinct chemical compounds. DFT features consist of calculated physiochemical properties such as orbital energies, atomic charges, and steric shielding at specific atoms. It has been shown that such features can lead to improved BO performance,² but they are also more expensive to compute. There are tools available for efficient automation of DFT calculations to featurize these categorical variables, such as Auto-QChem.¹⁵ DFT descriptors can also be approximated by calculating similar descriptors at a lower level of theory¹⁶ or by predicting them using machine learning models.¹⁷

Defining the objectives

In the context of a reaction optimization, the most classical target objective is the reaction yield, defined on a linear scale with percentages from 0 to 100. Due to this nature, yield values can be directly fed into a BO model without any issue. Optimization objectives in organic chemistry can, however, also be many

different other metrics such as the absorption property of a metal-organic framework¹⁸ or the delivery performance of bioactive material by polymers.¹⁹ Additionally, multiple objectives might be optimized simultaneously by BO. This can, for instance, be helpful when considering asymmetric reactions in which both yield and enantioselectivity are important.⁷ In such scenarios, BO algorithms try to find optimal tradeoff points between the two objectives via the so-called Pareto front, instead of optimizing them separately. When modeling selectivities such as enantioselectivity, it is also worth noting that it is preferred to input the objective values as differences in Gibbs free energy instead of using enantiomeric excess (ee) directly. This is due to the non-linear nature of the ee scale that is more challenging to model.²⁰

Notably, the objectives can also be optimized for minimization, instead of maximization, by providing appropriate input information to the BO algorithm. This could, for instance, be useful when thinking about how to reduce the cost of a reaction.⁷

Initialization methods

BO is mostly agnostic to the initiation of the optimization campaign. However, an important consideration is that the algorithmic modeling depends on the available prior data. If all prior experiments performed very poorly, the algorithm will predict the same low performance across the search space. The algorithm, therefore, performs best if some hit result has already been obtained at the start of the campaign. A chemist could start with some initial experiments using OVAT or DoE methods and then use the data obtained to begin a BO campaign. Alternatively, the campaign can also be initialized using pseudo-random²¹ or geometric sampling methods⁷ that are integrated into different BO algorithms and automatically executed when there is no prior data. This flexibility in the initiation distinguishes BO from DoE, as any set of data can be used to start a campaign or augment an ongoing one.

Technical aspects

While the surrogate model and acquisition function do not have to be modified in a standard BO campaign, we provide a short summary here. For more details, we refer to resources in the literature.^{3,4}

Sigman, Doyle, Genentech (2024):

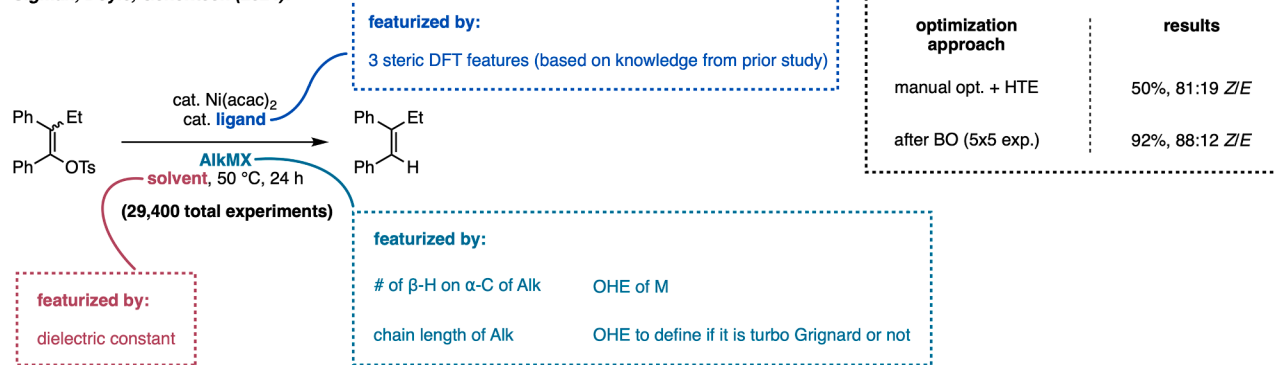


Figure 3. Example application of Bayesian optimization in catalysis

In the language of BO, the ground-truth reaction yield surface is the objective function. Because the objective function is unknown, BO algorithms use a surrogate model to approximate the objective function and predict outcomes for data points that have not yet been sampled. Thus, surrogate models are, in essence, structure-reactivity relationships, which many synthetic chemists may already know in the form of concepts like linear free energy relationships.^{22,23} Approaches vary on the basis of the featurization strategy, the prediction target, and the mathematical relationship used to connect the two. In the context of BO for chemical reactions, the most used surrogate model is the Gaussian process, but other popular methods include random forests and Bayesian linear regression. Many BO packages provide the option to customize these surrogate models. Doyle and coworkers collected evidence that the Gaussian process surrogate model is superior to a random forest surrogate model.²

The acquisition function chooses the next set of experiments, using a combination of exploration and exploitation. Exploration is forging outside the known regions and trying new conditions about which the algorithm has little information, thus reducing model uncertainty. Exploitation is the sampling of points that are expected to possess a high yield, which may lie close to points that have already been tested. This dual strategy enables global optimization, instead of staying fixed on an initially found, potentially local, maximum. It also means that some experiments are mainly explorative in nature. Poor results of some suggested experiments, therefore, do not have to indicate that the campaign is failing but that they help decrease the model uncertainty for future rounds.

Godineau and Braconi²⁴ compared different acquisition functions (“expected improvement” [EI], “Thompson sampling,” and “upper confidence bound”) for a reaction optimization campaign, and their results suggested that EI found a higher maximum for their campaign. Doyle and coworkers also benchmarked several acquisition functions and found that EI was the most successful.² EDBO+ uses EI as its default acquisition function, whereas Merck’s BayBE package uses “upper confidence bound” as its default acquisition function.

Standard acquisition functions are compatible with providing multiple suggestions per round, and software packages provide the batch size as an algorithm variable. The user can define the

batch size based on experimental constraints, as BO can tolerate batch sizes up to 96-well high-throughput experimentation (HTE) plates.²¹ Notably, the modeling quality improves with the number of the provided prior results. Therefore, the batch sizes should be kept smaller if the total experimental budget is limited so that the BO algorithm can be updated more often. For standard organic chemistry practice, we personally prefer batch sizes between 5 and 10.

Ending the campaign

The simplest endpoint to a campaign is when conditions are found that are satisfactory for the user. The BO acquisition function also provides metrics that can give the chemist insight into the value of continuing the optimization campaign. For instance, with EI, expected improvement values are produced. Decrease in the expected improvement value for unseen conditions signals that the likelihood of finding a higher yield in that search space is also decreasing. The chemist can use these data to determine when to end the optimization campaign.^{25–27} Alternatively, decreasing expected improvement values can be used as a signal to modify the search space to find more suitable conditions. For example, in a study on the flow synthesis of pyridinium salts,²⁸ the authors noticed low expected return values after several rounds of experiments. As the most promising experimental results favored high temperatures, the authors extended their search space to even higher temperatures and were able to improve their reaction outcome after more experiments.

Example applications

The power of BO in reaction optimization is illustrated by successful application in challenging catalytic settings. In a study from Sigman, Doyle, Genentech, and co-workers (Figure 3) on a stereo-convergent nickel-catalyzed enol tosylate reduction,²⁷ the authors show that the reduction of stereo-undefined enol tosylates can lead to either *E*-alkenes or *Z*-alkenes based on the nature of the ligand and further reaction parameters. After manual optimization and an extensive ligand screen by HTE, the yield for the *Z*-selective reaction was only 50%. The authors turned to BO for further optimization using a mixed featurization approach to describe the categorical reaction parameters. In particular, the ligand was described using three steric

Luo and Yang (2026)

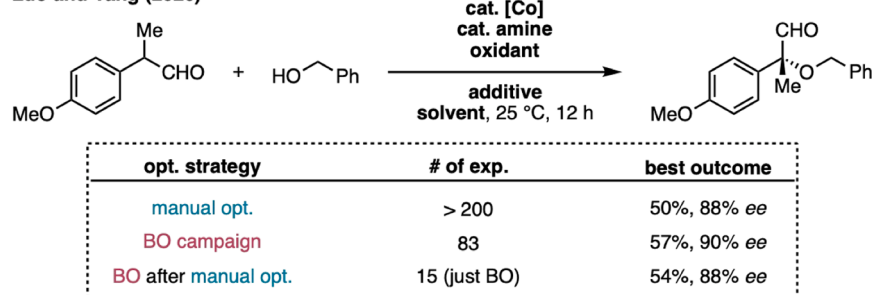


Figure 4. Comparison of Bayesian optimization and manual optimization of a chemical reaction

descriptors that were identified in a related prior study as being important for reactivity. The solvent was described by its dielectric constant. Finally, the alkylmetal species, which functions as a reductant, was described by variables that were chosen by chemical intuition such as the number of β hydrogen atoms. As the authors were not only interested in the yield of the Z-alkene product but also that it is formed in high stereoselectivity (to facilitate isolation), the percentage of Z product in the crude alkene mixture was used as a second reaction objective in addition to the Z-alkene yield. Using this setup, the authors used existing experimental data from prior optimization attempts as an initial input for the BO campaign. After five rounds of five experiments selected by BO, the yield was improved to 92%, with a contaminant improvement in selectivity from 81:19 Z/E to 88:12 Z/E.²⁷

In a second example, Luo, Yang, and co-workers reported an aldehyde α -alkoxylation under cobalt- and amine-co-catalysis, in which they aimed to maximize two objectives: reaction yield and enantioselectivity (Figure 4).¹² The reaction was optimized using different strategies. In manual optimization, optimized conditions were found after more than 200 experiments. In parallel, the authors conducted a BO campaign starting from a very large search space of more than one million experiments. The authors opted to optimize only categorical variables (cobalt and amine catalysts, additive, and solvent), all of which were featurized using a broad range of DFT descriptors. After initial experimental batches, this search space was pruned to ca. half a million samples by removing variable settings that appeared to be unpromising based on the experimental data. After further batches, the search space was again reduced, and promising new options were added based on the observations (space of 114,400 experiments remaining). This stepwise strategy resulted in optimized results after a total of 83 experiments that were also better than the conditions of the manual campaign (+7% yield and +2% ee). The authors then also tested whether the data from the manual campaign can be used as a starting point for a BO campaign to improve the results. To this end, they started a separate campaign and used the top 45 results from the manual optimization as initial samples. After 15 total samples, the algorithm found conditions that improved yield by 4%.¹²

CONCLUSION

Although the term “Bayesian optimization” was coined in 1970 by Jonas Mockus,²⁹ this statistical technique was adopted for

chemical synthesis much more recently. An early pioneer in this field is the Aspuru-Guzik group, which developed packages such as Gryffin and Anubis.^{10,30} Notably, these packages required some knowledge of Python for their operation, which posed a substan-

tial obstacle for organic chemists. Over the past few years, many alternative ways of implementing BO without prior programming experience have been developed. For instance, BO is now available in web apps or as a feature in JMP 19. The accessibility of this technique for synthetic chemists has greatly increased over time, enabling broader adoption.

Commensurately with the increase in accessibility and interest in this technique, creative approaches have evolved to applying BO for a range of situations. For instance, BO has been applied for “telescoped” continuous flow synthesis, in which material is carried through synthesis without the purification of intermediates.³¹ These types of situations in which chemical information can be sparse and difficult to interpret are ideal for the implementation of BO. Looking into the future, BO is a useful tool for researchers developing closed-loop “self-driving laboratories,” in which both experimentation and data-driven decision-making are automated.^{11,32}

ACKNOWLEDGMENTS

S.R. acknowledges the Swiss National Science Foundation (SNSF) for a Postdoc Mobility fellowship (P500PN_222270). Financial support was provided by the NSF Center for Computer Assisted Synthesis (C-540 CAS) under grant CHE-2202693 to A.G.D. and NIH NIGMS (R35GM142738) and the Welch Foundation (C-2085) to J.G.W.

AUTHOR CONTRIBUTIONS

Conceptualization, R.S.; writing – original draft, R.S., S.R., A.G.D., and J.G.W.; writing – review & editing, R.S., S.R., A.G.D., and J.G.W.; funding acquisition, A.G.D. and J.G.W.; supervision, A.G.D. and J.G.W.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Taylor, C.J., Pomberger, A., Felton, K.C., Grainger, R., Barecka, M., Chamberlain, T.W., Bourne, R.A., Johnson, C.N., and Lapkin, A.A. (2023). A Brief Introduction to Chemical Reaction Optimization. Chem. Rev. 123, 3089–3126. <https://doi.org/10.1021/acs.chemrev.2c00798>.
2. Shields, B.J., Stevens, J., Li, J., Parasram, M., Damani, F., Alvarado, J.I.M., Janey, J.M., Adams, R.P., and Doyle, A.G. (2021). Bayesian Reaction Optimization as a Tool for Chemical Synthesis. Nature 590, 89–96. <https://doi.org/10.1038/s41586-021-03213-y>.
3. Desimpel, S., Dorbec, M., Geem, K.M.V., and Stevens, C.V. (2026). Bayesian Optimization for Chemical Reactions. Chem. Soc. Rev. 55, 2731–2775. <https://doi.org/10.1039/D5CS00962F>.

4. Garnett, R. (2023). *Bayesian Optimization* (Cambridge University Press).
5. Introducing the Bayesian Optimization platform in JMP Pro.2026 JMP User Community. <https://community.jmp.com/t5/JMPPer-Cable/Introducing-the-Bayesian-Optimization-platform-in-JMP-Pro/ba-p/897807> (accessed 2026-01-12).
6. Miniconda. <https://www.anaconda.com/download/success> (accessed 2026-03-13).
7. Torres, J.A.G., Lau, S.H., Anchuri, P., Stevens, J.M., Tabora, J.E., Li, J., Borovika, A., Adams, R.P., and Doyle, A.G. (2022). A Multi-Objective Active Learning Platform and Web App for Reaction Optimization. *J. Am. Chem. Soc.* **144**, 19999–20007. <https://doi.org/10.1021/jacs.2c08592>.
8. Mroz, A.M., Toka, P.N., Del Río Chanona, E.A., and Jeffs, K.E. (2025). Web-BO: Towards Increased Accessibility of Bayesian Optimisation (BO) for Chemistry. *Faraday Discuss.* **256**, 221–234. <https://doi.org/10.1039/D4FD00109E>.
9. Fitzner, M., Šošić, A., Hopp, A.V., Müller, M., Rihana, R., Hrovatin, K., Liebig, F., Winkel, M., Halter, W., and Brandenburg, J.G. (2025). BayBE: A Bayesian Back End for Experimental Planning in the Low-to-No-Data Regime. *Digit. Discov.* **4**, 1991–2000. <https://doi.org/10.1039/D5DD00050E>.
10. Häse, F., Aldeghi, M., Hickman, R.J., Roch, L.M., and Aspuru-Guzik, A. (2021). Gryffin: An Algorithm for Bayesian Optimization of Categorical Variables Informed by Expert Knowledge. *Appl. Phys. Rev.* **8**, 031406. <https://doi.org/10.1063/5.0048164>.
11. Tom, G., Schmid, S.P., Baird, S.G., Cao, Y., Darvish, K., Hao, H., Lo, S., Pablo-García, S., Rajaonson, E.M., Skreta, M., et al. (2024). Self-Driving Laboratories for Chemistry and Materials Science. *Chem. Rev.* **124**, 9633–9732. <https://doi.org/10.1021/acs.chemrev.4c00055>.
12. Cheng, L., Tan, Z., Jia, Z., Lin, Q., Yang, Q., and Luo, S. (2025). Heuristic Data-Driven Approach for Synergistic Cobalt(IV)–Enamine Catalysis. *Nat. Synth.* **5**, 455–465. <https://doi.org/10.1038/s44160-025-00944-y>.
13. Landrum, G.; Tosco, P.; Kelley, B.; Cosgrove, D.; sriniker; gedeck; Vianello, R.; NadineSchneider; Kawashima, E.; N, D.; Jones, G., et al; jasondbiggs; strets123. Rdkit/Rdkit: 2023_03_3 (Q1 2023) Release, 2023. <https://doi.org/10.5281/zenodo.8254217>.
14. Moriwaki, H., Tian, Y.-S., Kawashita, N., and Takagi, T. (2018). Mordred: A Molecular Descriptor Calculator. *J. Cheminf.* **10**, 4. <https://doi.org/10.1186/s13321-018-0258-y>.
15. Żurański, A.M., Wang, J.Y., Shields, B.J., and Doyle, A.G. (2022). Auto-QChem: An Automated Workflow for the Generation and Storage of DFT Calculations for Organic Molecules. *React. Chem. Eng.* **7**, 1276–1284. <https://doi.org/10.1039/D2RE00030J>.
16. Jorner, K.; Turcani, L. Kjelljorner/Morfeus: V0.7.2, 2022. <https://doi.org/10.5281/zenodo.7017599>.
17. Haas, B.C., Hardy, M.A., Sowndarya, S.V.S., Adams, K., Coley, C.W., Paton, R.S., and Sigman, M.S. (2025). Rapid Prediction of Conformationally-Dependent DFT-Level Descriptors Using Graph Neural Networks for Carboxylic Acids and Alkyl Amines. *Digit. Discov.* **4**, 222–233. <https://doi.org/10.1039/D4DD00284A>.
18. Liu, T.-W., Nguyen, Q., Dieng, A.B., and Gómez-Gualdrón, D.A. (2024). Diversity-Driven, Efficient Exploration of a MOF Design Space to Optimize MOF Properties. *Chem. Sci.* **15**, 18903–18919. <https://doi.org/10.1039/D4SC03609C>.
19. Dalal, R.J., Oviedo, F., Leyden, M.C., and Reineke, T.M. (2024). Polymer Design via SHAP and Bayesian Machine Learning Optimizes pDNA and CRISPR Ribonucleoprotein Delivery. *Chem. Sci.* **15**, 7219–7228. <https://doi.org/10.1039/d3sc00692of>.
20. Ruth, M., Gensch, T., and Schreiner, P.R. (2024). Contrasting Historical and Physical Perspectives in Asymmetric Catalysis: $\Delta\Delta G^\ddagger$ versus Enantiomeric Excess. *Angew. Chem. Int. Ed.* **63**, e202410308. <https://doi.org/10.1002/anie.202410308>.
21. Sin, J.W., Chau, S.L., Burwood, R.P., Püntener, K., Bigler, R., and Schwaller, P. (2025). Highly Parallel Optimisation of Chemical Reactions through Automation and Machine Intelligence. *Nat. Commun.* **16**, 6464. <https://doi.org/10.1038/s41467-025-61803-0>.
22. Williams, W.L., Zeng, L., Gensch, T., Sigman, M.S., Doyle, A.G., and Anslyn, E.V. (2021). The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **7**, 1622–1637. <https://doi.org/10.1021/acscentsci.1c00535>.
23. Strieth-Kalthoff, F., Sandfort, F., Segler, M.H.S., and Glorius, F. (2020). Machine Learning the Ropes: Principles, Applications and Directions in Synthetic Chemistry. *Chem. Soc. Rev.* **49**, 6154–6168. <https://doi.org/10.1039/c9cs00786e>.
24. Braconi, E., and Godineau, E. (2023). Bayesian Optimization as a Sustainable Strategy for Early-Stage Process Development? A Case Study of Cu-Catalyzed C–N Coupling of Sterically Hindered Pyrazines. *ACS Sustain. Chem. Eng.* **11**, 10545–10554. <https://doi.org/10.1021/acssuschemeng.3c02455>.
25. Gallarati, S., Bucci, E.M., Doyle, A.G., and Sigman, M.S. (2026). Transferable Enantioselectivity Models from Sparse Data. *Nature* **651**, 637–646. <https://doi.org/10.1038/s41586-026-10239-7>.
26. Deolka, S., Dalmau, D., Garcia Roca, A., Vasylevskyi, S., Samha, M.H., and Sigman, M.S. (2025). Development of a Heteroleptic Palladium Catalyst for the Fluorination of Arylboronic Acids Using Data Science Tools. *J. Am. Chem. Soc.* **147**, 35705–35717. <https://doi.org/10.1021/jacs.5c11929>.
27. Romer, N.P., Min, D.S., Wang, J.Y., Walroth, R.C., Mack, K.A., Sirois, L.E., Gosselin, F., Zell, D., Doyle, A.G., and Sigman, M.S. (2024). Data Science Guided Multiobjective Optimization of a Stereoconvergent Nickel-Catalyzed Reduction of Enol Tosylates to Access Trisubstituted Alkenes. *ACS Catal.* **14**, 4699–4708. <https://doi.org/10.1021/acscatal.4c00650>.
28. Dunlap, J.H., Ethier, J.G., Putnam-Neeb, A.A., Iyer, S., Luo, S.-X.L., Feng, H., Garrido Torres, J.A., Doyle, A.G., Swager, T.M., Vaia, R.A., et al. (2023). Continuous Flow Synthesis of Pyridinium Salts Accelerated by Multi-Objective Bayesian Optimization with Active Learning. *Chem. Sci.* **14**, 8061–8069. <https://doi.org/10.1039/D3SC01303K>.
29. Mockus, J. (1989). *Bayesian Approach to Global Optimization: Theory and Applications* (Springer Netherlands). <https://doi.org/10.1007/978-94-009-0909-0>.
30. Hickman, R.J., Tom, G., Zou, Y., Aldeghi, M., and Aspuru-Guzik, A. (2025). Bayesian Optimization with Unknown Feasibility Constraints for Scientific Experimentation. *Digit. Discov.* **4**, 2104–2122. <https://doi.org/10.1039/D5DD00018A>.
31. Clayton, A.D., Pyzer-Knapp, E.O., Purdie, M., Jones, M.F., Barthelme, A., Pavey, J., Kapur, N., Chamberlain, T.W., Blacker, A.J., and Bourne, R.A. (2023). Bayesian Self-Optimization for Telescoped Continuous Flow Synthesis. *Angew. Chem. Int. Ed.* **62**, e202214511. <https://doi.org/10.1002/anie.202214511>.
32. Pilon, S., Savino, E., Bayley, O.M., Vanzella, M., Claros, M., Siasiaridis, P., Liu, J., Lukas, F., Damian, M., Tseliou, V., et al. (2026). A Flexible and Affordable Self-Driving Laboratory for Automated Reaction Optimization. *Nat. Synth.*, 1–13. <https://doi.org/10.1038/s44160-026-01053-0>.