AUTOMATIC REACTION CONDITION OPTIMIZATION FOR ORGANIC CHEMISTRY

Reported by Jingdan Chen September 23, 2025

INTRODUCTION

Automatic reaction condition optimization extends beyond the fine-tuning of experimental settings such as temperature and concentration to encompass the selection of reagents, including catalysts and solvents, to maximize reaction efficiency. Beyond its widely usage in industrial manufacturing, this data-driven approach is increasingly applied in modern academic laboratories, revolutionizing traditional research workflows. Historically, optimization relied on chemists' intuition and laborious one-factor-at-a-time (OFAT) experimentation, an inefficient method that fails to capture crucial interactions between variables. The emergence of statistical methods such as Design of Experiments (DoE) provided a more systematic framework for exploring reaction space.[1–3] The contemporary evolution, however, is driven by the integration of automated, high-throughput experimentation (HTE) platforms with sophisticated machine learning (ML) algorithms, notably Bayesian optimization (BO), which leverage vast amounts of accumulated reaction data.[4] These integrated systems facilitate "self-optimization" or "closed-loop" workflows in which algorithms iteratively design, execute, and analyze experiments, striking a balance between exploring novel conditions and exploiting known high-performing regions on systems such as the Buchwald-Hartwig reaction (Figure 1A).[5] More recently, Large Language Models (LLMs) have been incorporated to serve as accessible interfaces between chemists and complex optimization programs, creating powerful yet flexible human-AI collaboration. Despite these advances, two key challenges remain, which are co-optimization of continuous and discrete variables, and multi-objective co-optimization.[6,7]

CONTINUOUS AND DISCRETE VARIABLE CO-OPTIMIZATION

Optimizing continuous variables is a well-understood process, while the selection of discrete variables such as ligands and solvents, remains a significant challenge (Figure 1B). The ideal solution involves a substrate-adaptive model capable of leveraging prior knowledge from extensive datasets to predict conditions for novel reactant pairs. Despite the importance of seamlessly co-optimizing continuous and discrete variables within a unified framework, this continues to be a formidable task. While BO excels at continuous variables, it typically necessitates the featurization of discrete molecular choices into a continuous vector space, often employing DFT-derived descriptors for chemical interpretability.[4] Nevertheless, the effectiveness of the model is ultimately limited by descriptors selected. Conversely, LLMs are proficient at selecting discrete variables through few-shot learning, in which a handful of examples is given for domain-generalization, but struggle with numerical optimization.[6] Given the distinct

strengths and weaknesses of these approaches, many current closed-loop systems simplify the problem by keeping continuous variables constant while surveying discrete components. This limitation highlights the need for more powerful, integrated algorithms that can holistically optimize complex reaction systems such as photoredox catalysis.[8]

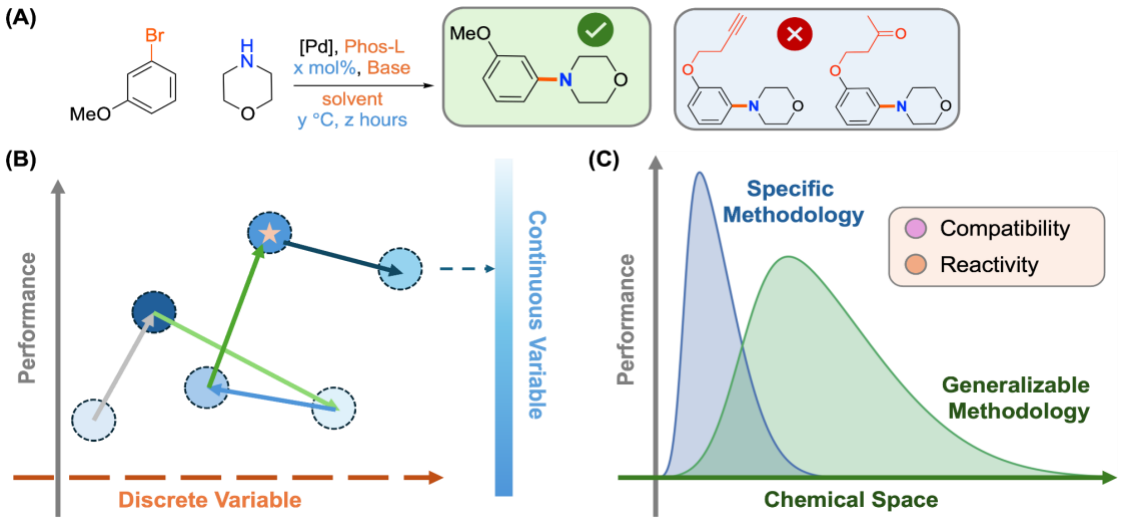

Figure 1. Automatic reaction condition optimization for the Buchwald-Hartwig (B-H) reaction. (A) B-H coupling with both continuous and discrete conditions to optimize and limited chemical space application. (B) Co-optimization of continuous and discrete variables. (C) Difference between specific reaction with high performance on limited chemical space and generalizable method giving acceptable performance on broader chemical space.

MULTI-OBJECTIVE CO-OPTIMIZATION

A major challenge hindering the industrial adoption of novel academic methods is their limited generalizability across a broad spectrum of substrates (Figure 1C). Automatic reaction condition optimization can help benchmark and improve the applicability of a discovered reaction by approaching the optimal compromise between performance (*e.g.* yield) and generalizability. However, quantifying "generalizability"[9] as an optimizable metric is a complex task. One data-driven approach involves unbiased selection of a structurally diverse set of substrates for evaluation. An alternative experimental method is the "robustness screen,"[5] which assesses the reaction's functional group tolerance by introducing a panel of additives. Despite the existing efforts, formal integration as a direct, quantifiable objective within automated, multi-objective optimization workflows remains a key challenge for the community.

REFERENCES

(1) Taylor, C. J.; et al. *Chem. Rev.* **2023**, 123 (6), 3089–3126. (2) Denmark, S. E.; et al. *J. Am. Chem. Soc.* **2008**, 130 (11), 3690–3704. (3) Wall, B. J.; et al. *ACS Catal.* **2025**, 15 (11), 8885–8893. (4) Shields, B. J.; et al. *Nature* **2021**, 590 (7844), 89–96. (5) Collins, K. D.; et al. *Nat. Chem.* **2013**, 5 (7), 597–601. (6) Zhang, Y.; et al. *arXiv* April 25, **2025**. (7) Zeng, T.; et al. *arXiv* June 26, **2025**. (8) Jorayev, P.; et al. *Org. Process Res. Dev.* **2025**, 29 (6), 1411–1422. (9) Rana, D.; et al. *ACS Cent. Sci.* **2024**, 10 (4), 899–906.