Closed-loop Optimization of Mono-functionalization via Suzuki-Miyaura Reaction

Seiji Akiyama,[1,2] Hayaka Akitsu,[2] Ryo Tamura,[3,4,5] Wataru Matsuoka,[1,2,6] Satoshi Maeda,[1,2,6] Koji Tsuda,[3,4,5] Yuuya Nagata*[1,2,7]

[1]Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Sapporo, Hokkaido 001-0021, Japan. [2]JST, ERATO Maeda Artificial Intelligence in Chemical Reaction Design and Discovery Project, Sapporo, Hokkaido 060-0810, Japan. [3]Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa, Chiba, 277-8561, Japan. [4]Center for Basic Research on Materials, National Institute for Materials Science (NIMS), 1-1 Namiki, Tsukuba, Ibaraki, 305-0044, Japan. [5]RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027, Japan. [6]Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan. [7]Research Center for Macromolecules and Biomaterials, National Institute for Materials Science (NIMS), 1-2-1 Sengen, Tsukuba, Ibaraki, 305-0047, Japan

Correspondence to: Yuuya Nagata NAGATA.Yuuya@nims.go.jp

1. Introduction

Autonomous experimentation is reshaping reaction development in synthetic chemistry. By integrating automated synthesis platforms with machine learning, closed-loop systems can execute reactions, analyze outcomes, update predictive models, and propose new conditions without direct human intervention. Early efforts focused primarily on optimizing continuous variables such as temperature and concentration. More recently, attention has expanded to discrete reaction parameters, including ligand, base, and solvent identity.[1,2]

Here we report a fully integrated autonomous optimization system designed specifically to explore discrete chemical parameters. The platform integrates a Chemspeed automated synthesis robot, supercritical fluid chromatography for rapid analysis, and Bayesian optimization based on Gaussian process regression. The system operates in a closed loop, enabling iterative experimentation and model refinement.

2. Experiments and Results

A major bottleneck in autonomous reaction optimization is analytical throughput. Conventional high performance liquid chromatography often requires long analysis times and may not sufficiently resolve closely related products. To overcome this limitation, we directly connected the synthesis robot to a supercritical fluid chromatography system. Because supercritical carbon dioxide has low viscosity and high diffusivity, separations can be performed much faster than with conventional liquid chromatography. The robot was modified to emulate manual operation of the the SFC system by generating analog contact signals that trigger sample injection. This strategy enabled seamless integration with standard analytical equipment. As a result, a complete optimization cycle including reaction setup, synthesis, analysis, and machine learning update was completed in approximately three hours.

Fig. 1: A photograph of the automatic autonomous optimization system. Green; a workstation for robot operation and machine learning analysis. Blue; the robot. Orange; supercritical fluid chromatography system.

As a model system, we selected the selective mono-functionalization of 2,7-dibromo-9,9-dihexylfluorene via the Suzuki Miyaura coupling reaction. Site selective mono coupling of polyhalogenated arenes provides access to unsymmetric building blocks from symmetric precursors. However, controlling mono versus di functionalization in this transformation remains challenging due to competing second coupling events.

To define a discrete search space, we selected 15 phosphine ligands, 10 bases, and 10 solvents, corresponding to 1500 possible combinations. Rather than treating these components as simple categorical variables, we encoded them using chemically meaningful descriptors. For ligands, we employed Tolman electronic parameters, cone angles, and calculated ^{31}P NMR chemical shifts derived from density functional theory calculations. For bases, we included the pK_a

of the conjugate acid and the ionic radius of the cation. For solvents, we adopted Hansen solubility parameters representing dispersion forces, polarity, and hydrogen bonding ability. This descriptor-based representation embeds chemical knowledge into the machine learning model and allows interpolation across related components.

The autonomous phase consisted of 4 initial randomly selected conditions followed by 47 iterative cycles, each proposing four new experimental conditions. In total, 192 reactions were attempted. After excluding 14 failed experiments due to chromatographic errors, 178 valid datasets were analyzed. The yield of the desired mono functionalized product increased from 9.5 percent in the first experiment to 49 percent by experiment number 102. Beyond this point, no higher yield was identified, suggesting convergence of the optimization process.

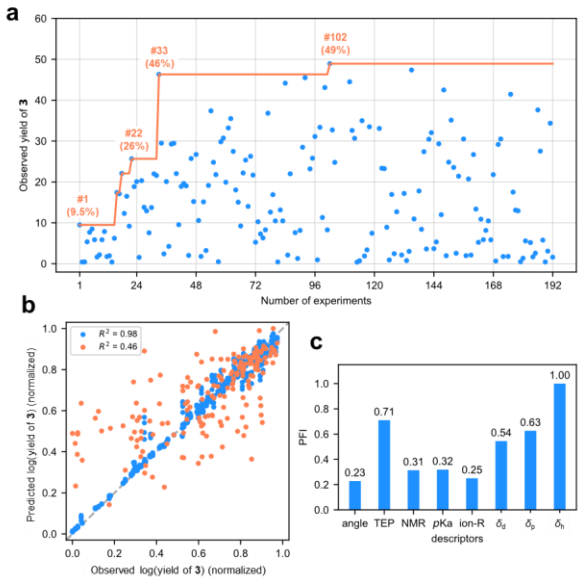

Fig. 2: Results of the autonomous optimization process (4 initial conditions, 4 suggested conditions × 47 iteration cycles). **a** The autonomous optimization process. Blue makers show the yield of a mono-functionalized product in each experiment, and orange lines show the maximum yield obtained prior to that experiment. **b** An evaluation of the data obtained from the automated experiments evaluated via a five-fold cross-validation. Blue and orange makers represent the training and test sets, respectively. **c** Contributions of each descriptor to the model as analyzed using the permutation feature importance (PFI) value. Angle and ion-R mean cone angle and Pauling ionic radius, respectively.

Cross validation of the full dataset showed an R^2 value of 0.98 for the training set and 0.46 for the test set. Although the test performance reflects the complexity of modeling discrete combinatorial chemistry, the model successfully captured trends sufficient to guide optimization. Feature importance analysis revealed that the electronic parameter of the ligand, particularly the Tolman electronic parameter, and the hydrogen bonding parameter of the solvent were the most influential descriptors. Ligands that provided higher selectivity exhibited intermediate electronic properties. Strongly electron-donating ligands tended to favor formation of the disubstituted product, whereas strongly electron withdrawing ligands reduced overall reactivity.

3. Conclusion

In summary, we developed an autonomous optimization system capable of navigating a discrete chemical space defined by ligand, base, and solvent identity. By integrating rapid supercritical fluid chromatography with Bayesian optimization and chemically informed descriptors, we achieved selective mono-functionalization of a difunctional substrate and demonstrated its application to the synthesis of advanced organic materials. We anticipate that such approaches will accelerate reaction discovery and expand the scope of machine learning driven synthetic chemistry.

Acknowledgments

This work was supported by JST-ERATO (JPMJER1903) and the Institute for Chemical Reaction Design and Discovery (ICReDD), which was established by the World Premier International Research Initiative (WPI), MEXT, Japan. Support was also provided under JSPS KAKENHI Grant Numbers JP21H01924 and JP23H03810. The computation was partly performed using the Research Center for Computational Science, Okazaki, Japan (Projects: 23-IMS-C119 and 24-IMS-C114).

References

[1] Christensen, M.; Yunker, L. P. E.; Adedeji, F.; Häse, F.; Roch, L. M.; Gensch, T.; dos Passos Gomes, G.; Zepel, T.; Sigman, M. S.; Aspuru-Guzik, A.; Hein, J. E. *Commun. Chem.* **2021**, *4*, 112.

[2] Angello, N. H.; Rathore, V.; Beker, W.; Wołos, A.; Jira, E. R.; Roszak, R.; Wu, T. C.; Schroeder, C. M.; Aspuru-Guzik, A.; Grzybowski, B. A.; Burke, M. D. *Science* **2022**, *378* (6618), 399-405.