Physical model-informed automated reaction optimization within a digital platform

Shuyuan Zhang[1] and Alexei A. Lapkin[1,2]

[1]*Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge, CB3 0AS, UK*

[2]*Cambridge Centre for Advanced Research and Education in Singapore Ltd., 1 Create Way, CREATE Tower #05-05, Singapore 138602, Singapore*

Correspondence to: Alexei A. Lapkin aal35@cam.ac.uk

1. Introduction

Reaction optimization remains a major bottleneck in modern chemical process development, driven by the growing emphasis on sustainable, resource-efficient production [1]. The empirical trial-and-error experimentation approach is costly, time-consuming, and environmentally burdensome, highlighting the urgent need for highly effective digital methods.

Automated robotics and optimization algorithms, particularly those with physical model basis, offer a means to bridge the traditionally disconnected stages of reaction discovery and process development.

Conventional combinatorial reaction libraries, typically conducted on well plates, explore limited design spaces under mild conditions, often requiring substantial redesign for manufacturing. Recent advances in flow chemistry have enabled high-throughput experimentation with extended operating windows and inline analytical instrumentation, establishing a solid foundation for generating reproducible quantitative data and constructing predictive models [2].

Incorporating mechanism-aware physical models can further enhance data interpretability and optimization efficiency. Moreover, such digital platforms also establish the knowledge base that facilitates collaboration across distributed automated laboratories operating on a shared reaction-modelling scaffold.

2. Methodology

Here we propose a digital platform that couples physical models and automated chemical robotics to advance chemical reaction development. The platform allows for the application of mechanistic understanding of chemical processes within self-driving laboratories to maximise yield, selectivity, and process robustness while minimising resource usage.

Fig. 1 gives an overview of the proposed digital platform which operates across both cyber and physical spaces, forming a closed-loop system that integrates first-principles modelling, data-driven optimization, and automated experimentation. Physical model-based Bayesian optimization (BO) is developed to analyse reaction data, quantify uncertainty, and propose new experimental conditions. These suggested conditions are executed autonomously by robotic systems in self-driving laboratories, and the resulting data are continuously fed back to refine and update the underlying models.

To support physical modelling, the platform incorporates dedicated modules for proposing reaction mechanisms, defining experimental setups, and querying physicochemical properties. Reaction mechanisms are generated through connections to reaction ontologies, literature databases, and AI models. Experimental configurations are defined through superstructural representations and designed chemical spaces, allowing the system to specify reaction configuration and parameters in a structured, machine-readable format. The platform interfaces with established chemical databases – such as ChemSpider, PubChem, and ChEMBL – and deployed AI models to obtain essential physicochemical information [3–5].

3. Results

The overall proposed workflow has been tested on a nucleophilic substitution aromatic reaction (S_NAr) benchmark targeting at ortho-substituted fluoronitrobenzene production [6]. Fig. 2 shows the reaction network inferred by the reaction mechanism generation module using simplified molecular-input line-entry system (SMILES) and SMILES Arbitrary Target Specification (SMARTS), with side reactions successfully identified. Parameters for describing the flow reactor is automatically extracted from the model knowledge graph based on the reactor phenomenon description [7,8]. Basic physicochemical properties, such as density and viscosity, are automatically retrieved by autonomous agents from public databases for physical modelling (Fig. 3). The physical model based reaction optimisation method vastly outperforms the conventional BO baseline over 100 runs (Fig. 4) [9].

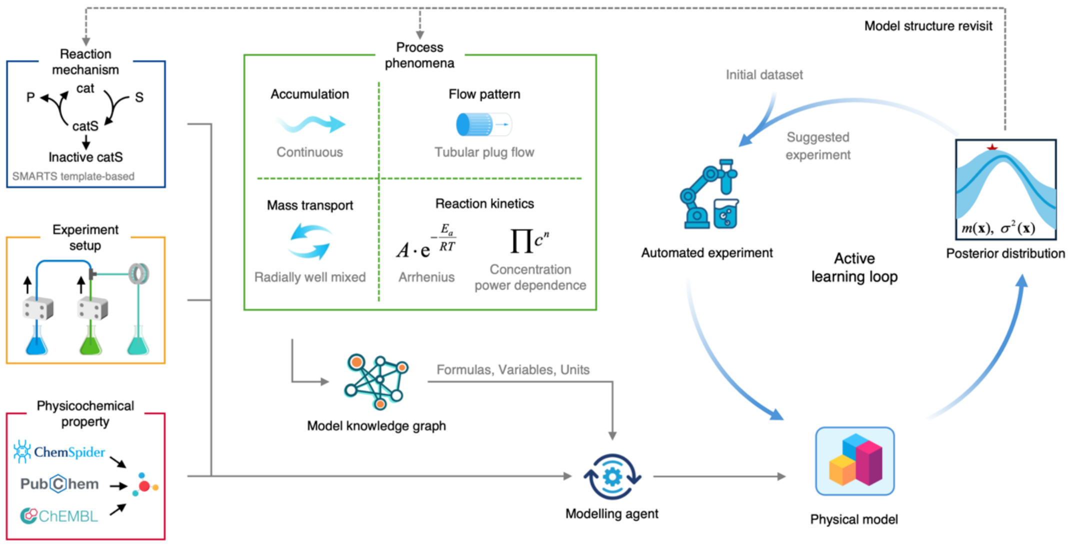

Fig. 1: An overview of the proposed digital platform coupling physical models and chemical robotics.

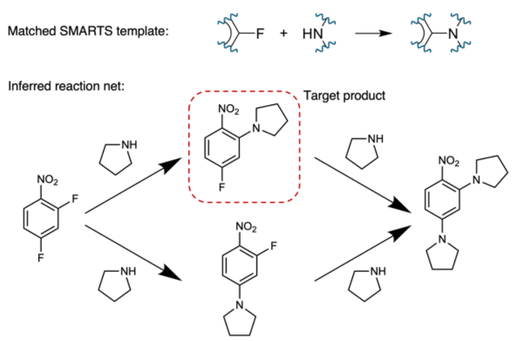

Fig. 2: Matched SMARTS template and inferred reaction network.

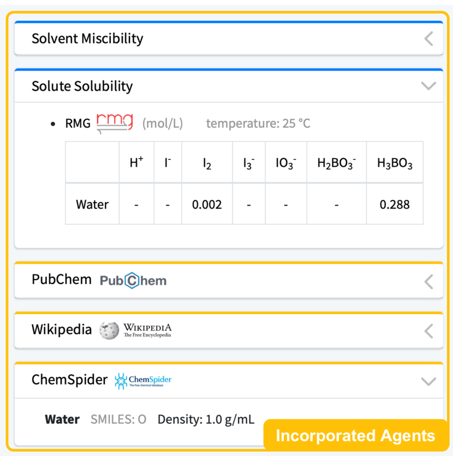

Fig. 3: Incorporated agents for physicochemical information retrieval.

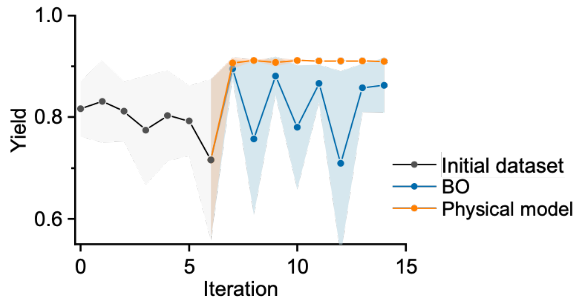

Fig. 4: Benchmark result against the BO baseline over 100 runs.

4. Future work

Looking ahead, we foresee the proposed digital platform can be expanded to multi-objective and multi-site reaction optimisation to advance sustainable molecule manufacturing. This tool also holds the promise to accelerate the development of new chemistry by further integrating an additional module for reaction proposal and experiment design.

Acknowledgments

This work was funded by UKRI via the project EP/Z531339/1 “Accelerated Development of Pharmaceutical Processes Through Digitally Coupled Reaction Screening and Process Optimisation”.

References

[1] P. Fantke, C. Cinquemani, P. Yaseneva, J. De Mello, H. Schwabe, B. Ebeling, A.A. Lapkin, Transition to sustainable chemistry through digitalization, Chem 7 (2021) 2866–2882. https://doi.org/10.1016/j.chempr.2021.09.012.

[2] F.M. Akwi, P. Watts, Continuous flow chemistry: where are we now? Recent applications, challenges and limitations, Chem. Commun. 54 (2018) 13894–13928. https://doi.org/10.1039/C8CC07427E.

[3] H.E. Pence, A. Williams, ChemSpider: An Online Chemical Information Resource, J. Chem. Educ. 87 (2010) 1123–1124. https://doi.org/10.1021/ed100697w.

[4] Q. Li, T. Cheng, Y. Wang, S.H. Bryant, PubChem as a public resource for drug discovery, Drug Discov. Today 15 (2010) 1052–1057. https://doi.org/10.1016/j.drudis.2010.10.003.

[5] F.H. Vermeire, Y. Chung, W.H. Green, Predicting Solubility Limits of Organic Solutes for a Wide Range of Solvents and Temperatures, J. Am. Chem. Soc. 144 (2022) 10785–10797. https://doi.org/10.1021/jacs.2c01768.

[6] C.A. Hone, N. Holmes, G.R. Akien, R.A. Bourne, F.L. Muller, Rapid multistep kinetic model generation from transient flow data, React. Chem. Eng. 2 (2017) 103–108. https://doi.org/10.1039/C6RE00109B.

[7] A. Hogan, E. Blomqvist, M. Cochez, C. D’amato, G.D. Melo, C. Gutierrez, S. Kirrane, J.E.L. Gayo, R. Navigli, S. Neumaier, A.-C.N. Ngomo, A. Polleres, S.M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, A. Zimmermann, Knowledge Graphs, ACM Computing Surveys 54 (2021) 71:1-71:37. https://doi.org/10.1145/3447772.

[8] J. Morbach, A. Wiesner, W. Marquardt, OntoCAPE—A (re)usable ontology for computer-aided process engineering, Comput. Chem. Eng. 33 (2009) 1546–1556. https://doi.org/10.1016/j.compcheng.2009.01.019.

[9] C.J. Taylor, A. Pomberger, K.C. Felton, R. Grainger, M. Barecka, T.W. Chamberlain, R.A. Bourne, C.N. Johnson, A.A. Lapkin, A Brief Introduction to Chemical Reaction Optimization, Chem. Rev. 123 (2023) 3089–3126. https://doi.org/10.1021/acs.chemrev.2c00798.