Learning Arrow Pushing for Reaction Space Prediction and Exploration

Kye Sung Park[1]†, Shuan Chen[1]†, Taewan Kim[2], Jason E. Hein[3], K. Barry Sharpless[4], Sunkyu Han[2] and Yousung Jung[1,5]*

[*]Equal contribution [1] Department of Chemical and Biological Engineering, and Institute of Chemical Processes, Seoul National University, Seoul, South Korea [2] Department of Chemistry, Korea Advanced Institute of Science and Technology, Daejeon, South Korea, [3] Department of Chemistry, University of British Columbia, Vancouver, BC, Canada Department of Chemistry, [4] The Scripps Research Institute, La Jolla, CA, USA, [5] Institute of Engineering Research, Seoul National University, Seoul, South Korea

Correspondence to: Yousung Jung yousung.jung@snu.ac.kr

1. Introduction

Predicting and discovering chemical reactions from a mechanism-based approach remains a fundamental challenge in organic chemistry. Chemists traditionally use arrow-pushing diagrams [1] as the primary language for describing reaction mechanisms — encoding the directed flow of electrons from nucleophilic sources to electrophilic sinks through sequences of elementary steps. Despite its central role in chemical reasoning, arrow pushing has remained largely informal and qualitative in computational models. Prior approaches either relied on hand-crafted rules [2,3] or bypassed mechanism entirely, learning direct mappings from reactants to products and obscuring the underlying causal electron-flow processes [4–6]. Here, we introduce Reactron, a graph neural network that explicitly models electron flow through arrow-pushing formalism, bridging the gap between machine learning-based reaction prediction and mechanistic chemical reasoning.

2. Reactron: Model and Dataset

Reactron represents molecules as graphs where atoms and bonds serve as nodes and edges. The model learns chemical environment representations via message passing [7], identifies nucleophilic and electrophilic sites by reactive pooling [8], and predicts the most probable arrow-pushing step by comparing each source–sink combination (**Fig. 1a**). Intermediate structures are sequentially derived by applying each predicted arrow-pushing step until no further reactivity is detected (**Fig. 1b**). To train the model at scale, we employed MechFinder — a chemist-validated mechanism labeling tool [9] — to annotate 247,071 reactions from the USPTO-480k dataset [10] with 1.86 million elementary electron-movement events, covering both polar and organometallic transformations.

3. Results and Contribution

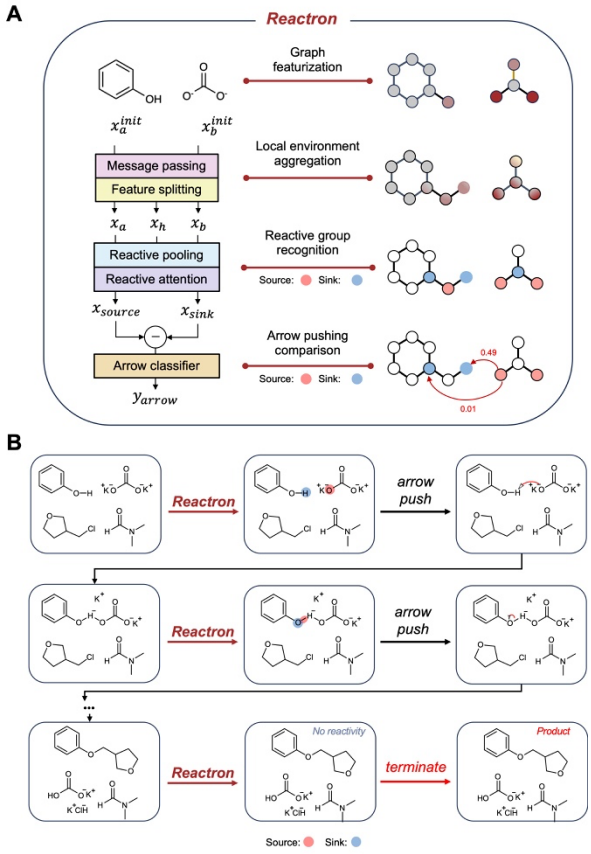


Fig. 1: Overall framework of Reactron for mechanistic prediction and reaction-space exploration. (A) Reactron model architecture. (B) Reactron inference workflow for predicting reaction outcomes via sequential arrow-pushing steps.

3.1 Superior Performance with Mechanistic Interpretability

Reactron achieves state-of-the-art performance on reaction outcome prediction while simultaneously producing interpretable mechanistic trajectories and guaranteeing chemical validity. On the full test set of 21,093 reactions, Reactron achieved 96.2% top-1 accuracy — 0.3% higher than the best product-predicting baseline (Transformer [4], Graph2SMILES [5]) and 18.8% higher than the mechanism prediction baseline FlowER [11] (**Fig. 2a**). On 582 out-of-distribution (OOD)

reactions, Reactron outperformed the best baseline by 7.3% (**Fig. 2b**). Crucially, Reactron maintains 100% atom conservation across all predictions — an intrinsic guarantee arising from its arrow-pushing formalism — whereas Transformer [4] and Graph2SMILES [5] exhibit decreasing conservation rates for higher-k predictions (**Fig. 2c**). In terms of speed, Reactron requires only 0.18 seconds per reaction, making it over 200× faster than FlowER (37.27 s/reaction) (**Fig. 2d**).

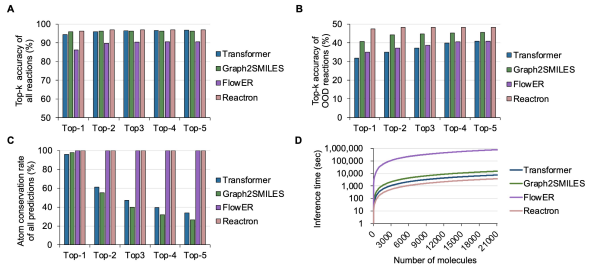


Fig. 2: The benchmark results of Reactron on the USPTO test set. (A and B) Top-k product prediction accuracies on the full test set (A) and out-of-distribution (OOD) reactions (B) compared with baseline models. (C) The top-k atom conservation (D) Total inference time required to predict the full test set; Reactron is >200-fold faster than FlowER.

3.2 Related Work

Prior machine learning approaches to reaction prediction fall into two categories: (1) template-based expert systems that encode hand-crafted mechanistic rules but fail to generalize beyond narrow chemical domains [2,3], and (2) transformer-based sequence-to-sequence models that achieve high accuracy but provide no mechanistic insight and frequently violate atom conservation [4–6]. More recent electron-redistribution models such as NERF [12] and FlowER [11] move toward mechanistic representations, but predicting net electron redistribution does not constitute a discrete causal sequence of arrow-pushing steps as used by practicing chemists. ELECTRO [13] represents the only prior explicit arrow-pushing model, yet relies on heuristically generated mechanisms that are frequently chemically incorrect. Reactron addresses all of these limitations by treating discrete arrow-pushing events as the primary object of prediction.

3.3 Few-Shot Fine-Tuning on Named Reactions and Regioselectivity

To assess transfer learning capability, Reactron was fine-tuned on just two examples per class across 16 named reaction classes [14] and tested on unseen reactions. After fine-tuning, Reactron correctly predicted 14 out of 16 named reactions — including mechanistically complex transformations such as the Cope rearrangement, Heine reaction, and Nenitzescu indole synthesis — while maintaining 100% atom balance. FlowER [11], by comparison, correctly predicted only 7. For the regioselective ring-opening of aziridines and epoxides—processes governed by subtle electronic and steric effects—Reactron correctly predicted all 4 test cases after fine-tuning on just 12 examples, whereas FlowER failed entirely.

3.4 Large-Scale Virtual High-Throughput Experimentation (vHTE)

Treating arrow pushing as a transferable reactivity representation enables Reactron to function as a reaction exploration engine. By combinatorially pairing 9,900 substrate sets and 250 reagent sets, Reactron screened 2,475,000 virtual reactions in approximately 140 GPU hours — a scale prohibitive for FlowER (~28,000 GPU hours). After excluding reactions showing no or known reactivity, 21,209 vHTE reactions exhibiting novel reactivities were identified. Literature searches confirmed that many of these predictions — while absent from USPTO training data — had been independently reported in literature spanning the 1960s to 2020s [15–17], validating Reactron's ability to extrapolate the reaction space. Laboratory experiments further confirmed Reactron's predictions for previously untrained reactivities, including a 1,2-Meisenheimer rearrangement [18], an acid-promoted lactamization, and an unexpected Lewis acid–promoted C–C bond cleavage adjacent to a benzylic substituent.

Acknowledgements

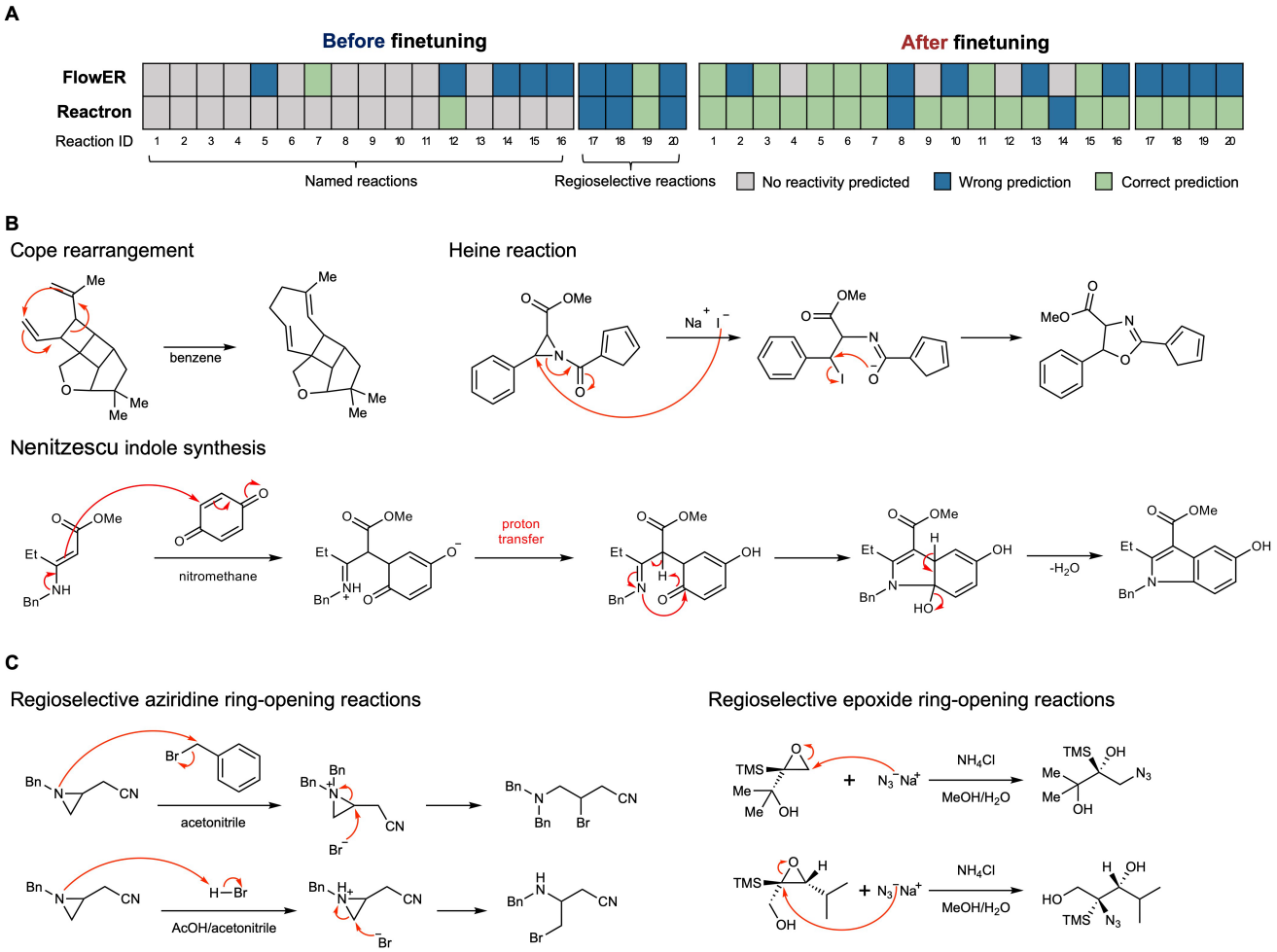
We acknowledge financial support from NRF Korea (RS-2025-00514706, RS-2024-00464386). A generous supercomputing time from KISTI is acknowledged.

References

[1] D. E. Levy. Arrow-Pushing in Organic Chemistry: An Easy Approach to Understanding Reaction Mechanisms. John Wiley & Sons, Inc., 2008.

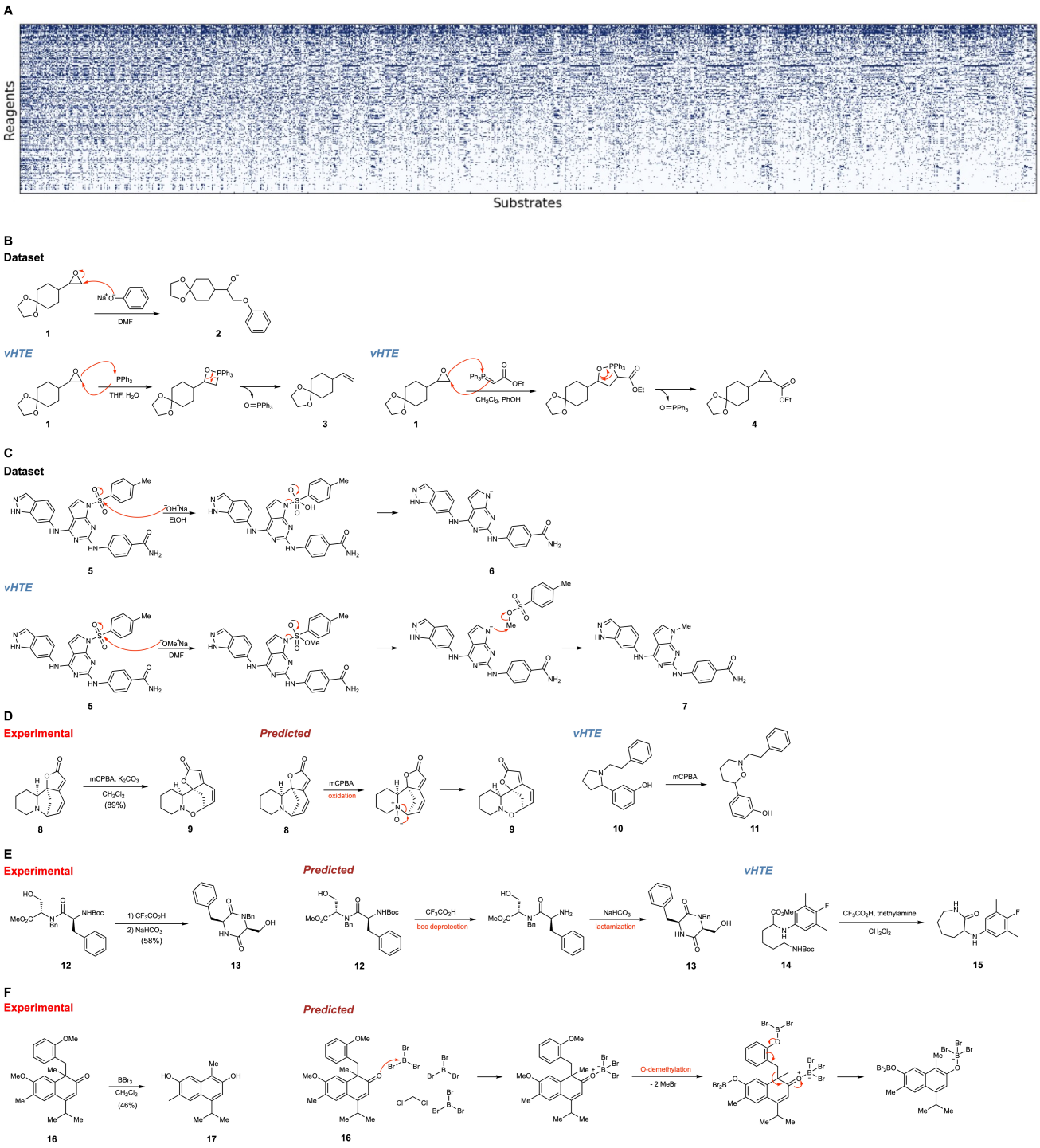
[2] J. H. Chen, P. Baldi. No Electron Left Behind: A Rule-Based Expert System To Predict Chemical Reactions and Reaction Mechanisms. J. Chem. Inf. Model., 2034–2043, 2009.
[3] T. Klucznik, L.-D. Syntrivanis, S. Baś, B. Mikulak-Klucznik, M. Moskal, S. Szymkuć, J. Mlynarski, L. Gadina, W. Beker, M. D. Burke, K. Tiefenbacher, B. A. Grzybowski. Computational prediction of complex cationic rearrangement outcomes. Nature, 508–515, 2024.
[4] P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, A. A. Lee. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. ACS Central Science, 1572–1583, 2019.
[5] Z. Tu, C. W. Coley. Permutation Invariant Graph-to-Sequence Model for Template-Free Retrosynthesis and Reaction Prediction. J. Chem. Inf. Model., 3503–3513, 2022.
[6] R. J. Miller, A. E. Dashuta, B. Rudisill, D. Van Vranken, P. Baldi. Mechanism-Aware Deep Learning for Polar Reaction Prediction. J. Am. Chem. Soc., 2025.
[7] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, G. E. Dahl. Neural Message Passing for Quantum Chemistry. arXiv:1704.01212, 2017.
[8] S. Chen, Y. Jung. A generalized-template-based graph neural network for accurate organic reactivity prediction. Nat Mach Intell, 772–780, 2022.
[9] S. Chen, R. Babazade, T. Kim, S. Han, Y. Jung. A large-scale reaction dataset of mechanistic pathways of organic reactions. Sci Data, 863, 2024.
[10] W. Jin, C. Coley, R. Barzilay, T. Jaakkola. Predicting Organic Reaction Outcomes with Weisfeiler-Lehman Network. Advances in Neural Information Processing Systems, 2017.
[11] J. F. Joung, M. H. Fong, N. Casetti, J. P. Liles, N. S. Dassanayake, C. W. Coley. Electron flow matching for generative reaction mechanism prediction. Nature, 115–123, 2025.
[12] H. Bi, H. Wang, C. Shi, C. Coley, J. Tang, H. Guo. Non-Autoregressive Electron Redistribution Modeling for Reaction Prediction. arXiv:2106.07801, 2021.
[13] J. Bradshaw, M. J. Kusner, B. Paige, M. H. S. Segler, J. M. Hernández-Lobato. A Generative Model for Electron Paths. 2019.
[14] L. Kurti, B. Czako. Strategic Applications of Named Reactions in Organic Synthesis. Elsevier, 2005.
[15] D. Cao, S. Xia, L. Li, H. Zeng, C.-J. Li. PPh3-Promoted Direct Deoxygenation of Epoxides to Alkenes. Org. Lett., 6418–6423, 2024.
[16] D. B. Denney, J. J. Vill, M. J. Boskin. A New Synthesis of Cyclopropanecarboxylic Acids. J. Am. Chem. Soc., 3944–3946, 1962.
[17] M.-H. Zhu, C.-L. Yu, Y.-L. Feng, M. Usman, D. Zhong, X. Wang, N. Nesnas, W.-B. Liu. Detosylative (Deutero)alkylation of Indoles and Phenols with (Deutero)alkoxides. Org. Lett., 7073–7077, 2019.
[18] N. Castagnoli, J. Cymerman Craig, A. P. Melikian, S. K. Roy. Amine-N-oxide rearrangements. Tetrahedron, 4319–4327, 1970.

Appendix A. Performance of Reactron on named reactions and regioselective transformations.



(A) Top-1 prediction accuracy of Reactron and baselines on 16 named reaction classes before and after few-shot fine-tuning. (B) Representative mechanistic pathways correctly predicted by Reactron, including pericyclic rearrangement (Cope) and heterocycle syntheses (Heine, Nenitzescu). (C) Accurate prediction of regioselectivity in aziridine and epoxide ring-opening reactions.

Appendix B. Large-scale reaction discovery via virtual high-throughput experimentation



(A) Subset of virtual high-throughput experimentation (vHTE) results (246 reagents × 990 substrates); reactions with novel predicted reactivity are highlighted in blue. Reagents are sorted by frequency of novel reactivity (top). (B and C) Representative vHTE predictions. Top rows ("Dataset") show established training-set transformations; subsequent rows ("vHTE") show Reactron's newly predicted reactions for alternative reagent combinations. Red arrows indicate key mechanistic steps. (D and E) Experimental validation of Reactron predictions (left) alongside analogous reactions discovered in the vHTE dataset (right). (F) Accurate prediction of unexpected benzylic C–C bond cleavage in ketone 16 by BBr_3 to yield 17, identifying a plausible mechanistic pathway.