Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Molecular traps reveal limitations of synthesis planning algorithms

**Yang Tian**

tiany21@mails.tsinghua.edu.cn

Tsinghua University

**Yating Deng**

Jinan University

**Peng Jia**

InsightChemy Co., Ltd

**Hongyu Zhang**

InsightChemy Co., Ltd

**Jingxin Zhai**

InsightChemy Co., Ltd

**Research Article**

**Keywords:** Computer-aided synthesis planning, Retrosynthesis, Molecular traps, Adversarial testing, Reaction selectivity

**Posted Date:** November 9th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11639503/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Molecular traps reveal limitations of synthesis planning algorithms

Yang Tian*[1,3], Yating Deng[2,3], Peng Jia[3], Hongyu Zhang[3], Jingxin Zhai[3,4]

[1]*MOE Key Laboratory of Bioinformatics, State Key Laboratory of Molecular Oncology, Beijing Frontier Research Center for Biological Structure, School of Pharmaceutical Sciences, Tsinghua University, Beijing, 100084, P. R. China.*

[2]*Institute of Nanotechnology And Intelligence (inAI), Jinan University, Guangzhou 511443, P. R. China.*

[3]*InsightChemy Co., Ltd., Jinan 250000, P. R. China.*

[4]*School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing 211198, P. R. China.*

*Corresponding author: Yang Tian (tiany21@mails.tsinghua.edu.cn)

## Abstract

Computer-aided synthesis planning (CASP) has advanced substantially over the past decade. However, gains on conventional metrics do not mean that algorithms have truly mastered comprehensive chemical knowledge or can deliver practical routes with high yield and high selectivity. In real synthetic settings, chemists must contend with key chemical constraints such as chemoselectivity, regioselectivity and steric hindrance. Current algorithms are trained on reaction datasets dominated by successful examples, and the scarcity of failed reactions makes reaction boundaries and limiting rules difficult for algorithms to learn. Meanwhile, the main evaluation metrics in use rarely examine whether all of an algorithm's predictions are feasible. Inspired by adversarial testing in machine learning, we introduce the concept of "trap molecules": molecules deliberately designed to lure synthesis-planning algorithms into characteristic errors. Trap molecules enable a systematic assessment of how well algorithms have mastered basic chemical principles. They are generated by an AI agent guided by common reaction SMARTS templates and reaction SMILES. Evaluation combines two modes: a pre-set matching test and an agent-based test. We applied the framework to three retrosynthesis algorithms built on different principles (UAlign, SimpRetro and RetroChimera) and performed tests and case studies on chemoselectivity, regioselectivity and steric hindrance. The results reveal marked weaknesses of these retrosynthesis tools on all three problem types, with error rates in top-20 predictions ranging from 66% to 100% across different problems and algorithms. Overall, trap molecules reveal limitations of current synthesis-planning algorithms, and this testing approach offers a strategy and tool for developing more practical synthesis-planning systems.


## Scientific Contribution

Conceptual contribution: We introduce molecular traps as an adversarial stress-testing framework for identifying potential selectivity- and feasibility-related failure modes in retrosynthesis predictions.

Methodological contribution: The framework combines preset reaction matching, agent-assisted screening, and manual verification to evaluate three retrosynthesis algorithms across chemoselectivity, regioselectivity, and steric-hindrance challenges.

Practical contribution: It complements conventional top-k accuracy and route-solve metrics with target-level stress-test outcomes, providing a structured approach for diagnosing model limitations and improving synthesis-planning systems.


## Keywords

Computer-aided synthesis planning; Retrosynthesis; Molecular traps; Adversarial testing; Reaction selectivity

## Introduction

Over the past decade, computer-aided synthesis planning (CASP) has developed rapidly [1, 2], and data-driven approaches to chemical synthesis planning have continued to emerge [3, 4]. Existing retrosynthesis methods include expert-rule [5–7], template-based [1, 8–12], semi-template [13–18], template-free [19–23] and hybrid [2, 24] approaches. Modern retrosynthesis algorithms can quickly propose many possible synthetic routes for organic molecules, and benchmark metrics such as top-k accuracy and route solve rate have improved steadily [1, 12, 25] (Fig. 1a). These advances provide a solid foundation for AI-assisted synthesis.

However, improvements on these metrics do not imply that algorithms have mastered all relevant chemical knowledge, nor do they guarantee that algorithms can generate practical routes [12, 15, 24]. Synthesis planning aims not only to predict reaction pathways that lead to the target product but also to find practical routes with high yield and high selectivity (Fig. 1b). In real-world synthetic analysis, chemists must handle constraints such as chemoselectivity, regioselectivity and steric hindrance. For example, catalytic hydrogenation of a nitro group may also reduce an aldehyde in the same molecule [26], so selective conditions or a protecting group may be needed [27]. When a molecule contains two or more non-equivalent reactive sites, reacting only one of them is usually difficult and readily gives a mixture of regioisomers. Because regioisomers are often difficult to separate [28], synthesis planners generally avoid such routes. Steric effects are comparatively limited in scope and are therefore easily overlooked; however, with too much bulk at the reaction center, high yields may be difficult to achieve [29].

Current synthesis-planning algorithms are trained mainly on reaction datasets [30, 31], among which USPTO is one of the most widely used open data sources [32–37]. This dataset holds almost exclusively successful reactions and lacks failed cases [37, 38]. When rules are extracted in a purely data-driven manner, algorithms therefore struggle to learn reaction boundaries and limiting rules (Fig. 1c). Moreover, current algorithm evaluation relies mainly on top-k accuracy. This metric checks whether the recorded answer is among the model predictions but does not check whether the other predictions are correct or feasible. Metrics such as solve rate check whether an algorithm can break a target molecule down to purchasable materials within a limited number of steps or time [1, 12, 39]. They likewise say little about route quality, so the solved routes may include infeasible or impractical ones.

Inspired by adversarial testing in machine learning (Fig. 1c) [40], we propose the concept of trap molecules, that is, molecules deliberately designed to induce characteristic errors in synthesis-planning algorithms. Testing with trap molecules can assess whether an algorithm has mastered basic chemical principles such as chemoselectivity, regioselectivity and steric hindrance. It thereby rapidly identifies the boundaries of an algorithm's capabilities and provides an empirical basis for subsequent improvement.

The evaluation comprises two test modes. The pre-set matching test checks whether a retrosynthesis algorithm produces the predefined erroneous results, that is, the preset errors. The agent-based test judges every prediction for chemoselectivity, regioselectivity or steric problems, thereby capturing errors beyond the preset scope. In this study, we applied the trap-molecule framework to three retrosynthesis algorithms with different design principles: UAlign [22], SimpRetro [12] and RetroChimera [24]. We performed case analyses and comparative tests for chemoselectivity, regioselectivity and steric hindrance (Figs. 3–5). The results show that all three tools have clear weaknesses on these problems, with error rates varying by trap type and test mode. The agent-based test also uncovered failure modes not covered by the pre-set matching test, including absurd predictions, indicating that agent-based review can complement rule-based preset testing.

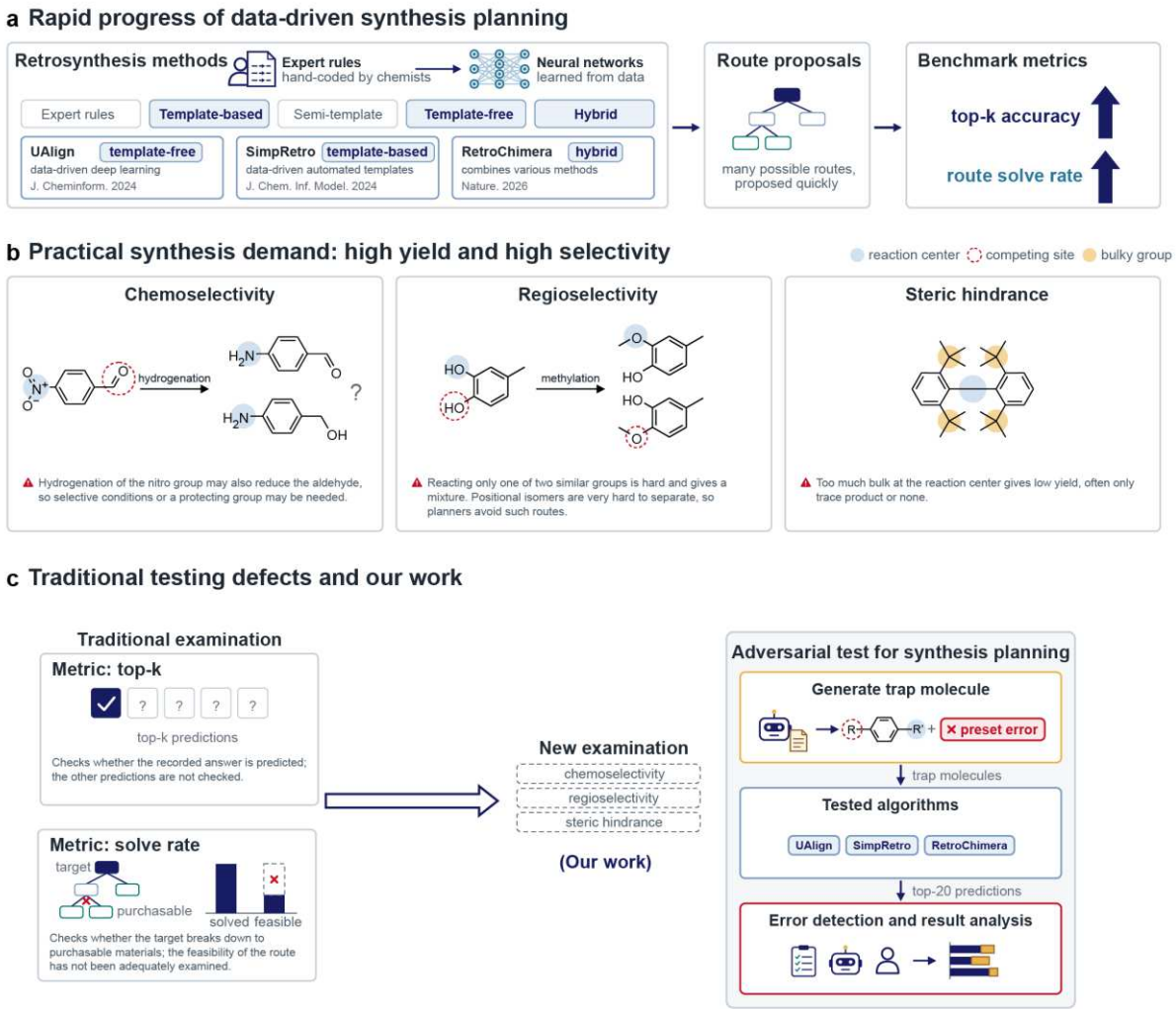



*Fig. 1 | Overview of molecular traps for evaluating synthesis-planning algorithms. a. Rapid progress of data-driven synthesis planning. b. Practical synthesis demand: high yield and high selectivity. c. Traditional testing defects and our work.*

## Methods

### Generation of trap molecules

We constructed three classes of trap molecules, targeting chemoselectivity, regioselectivity and steric hindrance. Each trap molecule was recorded in SMILES format. One or more predefined potential erroneous reactions, the preset errors, were defined at generation time for subsequent pre-set matching. The final test set contains 100 chemoselectivity, 100 regioselectivity and 27 steric trap molecules.

### Chemoselectivity and steric trap molecules

Chemoselectivity and steric trap molecules were obtained by SMARTS-guided design, using 100 common reaction SMARTS templates (Fig. 2a). These templates encode chemical reactions and are commonly used in template-based or template-assisted retrosynthesis algorithms. For each SMARTS template, an agent (Codex in this work) modified the substrates under the constraints of prebuilt generation skills. Chemoselectivity trap molecules were designed to retain the target reaction center while adding a sensitive group. Under the usual reaction conditions, such a group may also react, be consumed concurrently or interfere with the intended transformation [29, 41, 42]. Steric trap molecules were built from classic two-reactant, single-product transformations selected from the same templates. The agent introduced bulky groups (e.g., tert-butyl) around the reaction center. As a result, the attacking reagent, coupling partner or catalytic system encounters local steric crowding when approaching the reaction center [43]. For each trap molecule, preset errors were defined at generation time. These are the mistakes the tested algorithms are expected to make, derived from the design rationale and commonly used reactions.

### Regioselectivity trap molecules

Regioselectivity trap molecules were obtained by USPTO-case modification, that is, by modifying specific reaction SMILES records from USPTO (Fig. 2a). A preset SMARTS first selected a concrete reaction case matching the target transformation. The agent, constrained by the generation skills, then copied the reaction center so that the substrate contains two or more copies of it. The candidate sites had to be in similar but non-equivalent chemical environments. Modifications preferentially added a competing group of the same type to the original scaffold or moved an existing group. Overall symmetry was usually broken by small differences in peripheral short alkyl groups (e.g., Me vs Et on the ring), so that the copies differ only slightly. As with the other classes, preset errors were defined for each trap molecule at generation time for subsequent testing.

### Testing protocol

We tested three algorithms built on different design logics. UAlign is a template-free, data-driven deep-learning algorithm [22]. SimpRetro is a template-based method built on data-driven automated

templates [12], and RetroChimera is a hybrid that combines both approaches [24]. For each trap molecule, UAlign, SimpRetro and RetroChimera were each used to generate top-20 single-step retrosynthesis predictions as test data (Fig. 2b). The pre-set matching test was applied first, asking whether any of the 20 predictions exactly reproduces a preset error. Reaction SMILES were parsed with RDKit, and only the reactants>>products form was used. For each trap molecule, if any of a model's top-20 predictions matched a preset error, that molecule was counted as a hit; even if the same molecule matched multiple preset errors, it was counted only once. Clearly, this approach can only verify whether a retrosynthesis algorithm makes common, anticipated errors, so it is inherently limited and may underestimate the problems of retrosynthesis algorithms. We therefore also introduced an agent-based test, in which the agent judges the predictions autonomously and issues warnings, using skills that include some erroneous examples for reference. For molecules with no hit, the reactions that received agent warnings were checked manually; a molecule was counted as a new hit if at least one error was confirmed. Several caveats apply. Because the predicted reactions do not specify complete reaction conditions, chemoselectivity was judged mainly against commonly used conditions, which may introduce some error. For regioselectivity, we assessed whether a regioselectivity problem would arise, assuming that the predicted reaction can proceed. Some molecules in the steric test may have no efficient synthetic method at all. In addition, the retrosynthesis algorithms produced absurd predictions that are difficult to rationalize with conventional chemical knowledge and are mostly algorithmic hallucinations [24]. These cases do not involve chemoselectivity, regioselectivity or steric problems, but this does not make them correct; on the contrary, such problems should be addressed first.

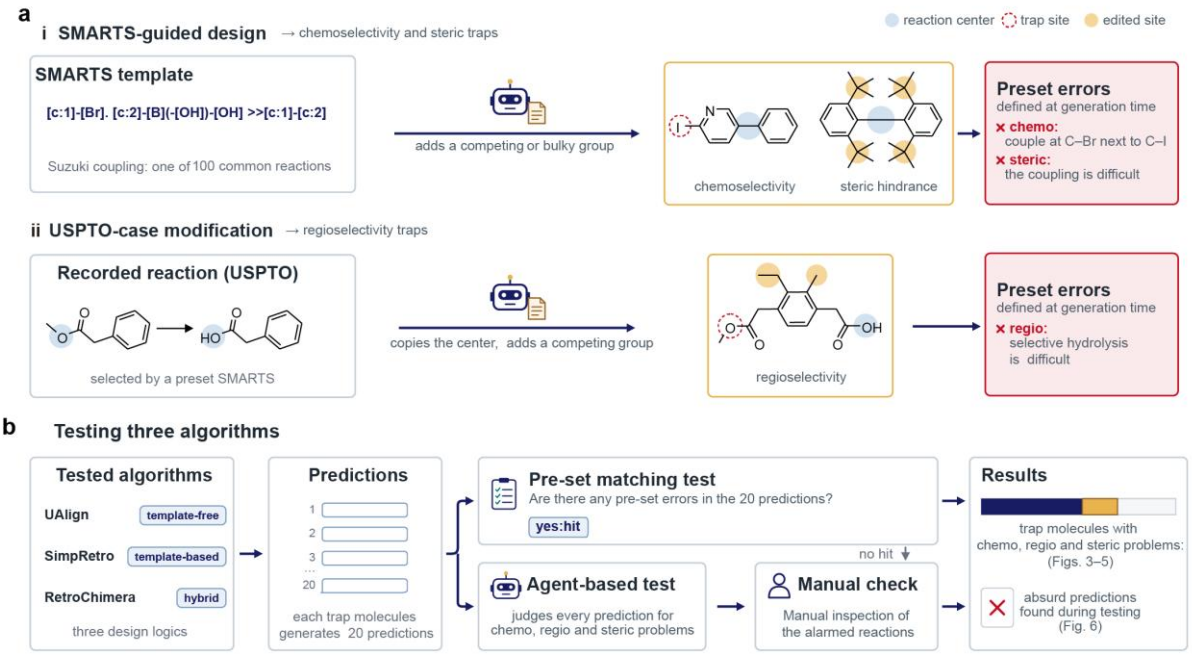



*Fig. 2 | Generation of molecular traps and the model-testing workflow. a. Generation strategies for trap molecules: SMARTS-guided design and USPTO-case modification. b. Workflow for testing the three algorithms.*

# Results

## Exploring chemoselectivity

Chemoselectivity problems arise mainly when several functional groups in a molecule may react competitively under the same reaction conditions. In retrosynthesis, chemoselectivity problems are more condition-dependent than regioselectivity problems. Because reaction conditions are not fully fixed, selectivity can be improved by adjusting them. Some apparently risky routes may still proceed selectively under specific conditions. The trap molecules constructed in this section mainly address chemoselectivity risks under the usual reaction conditions.

Trap-molecule generation was guided by 100 common reaction SMARTS templates (Fig. 3a), which are widely used in retrosynthesis algorithms. We then used prebuilt generation skills to direct the agent to generate qualifying trap molecules. The main approach was to retain the target reaction center while adding a sensitive group that also reacts under the usual conditions. For example, the SMARTS for the classic Suzuki coupling of an aryl bromide with an arylboronic acid is [c:1]-[Br].[c:2]-[B](-[OH])-[OH]>>[c:1]-[c:2]. The agent built a trap molecule by adding an iodo substituent (C–I) to the product as a competing group. The rationale is that an algorithm that knows the basic coupling rule but has not adequately mastered chemoselectivity may propose coupling the aryl bromide with the boronic acid. In such a prediction, the C–I site is left untouched even though C–

I couples faster than C–Br, so the route would give the wrong product. This yields the preset error together with the trap molecule: “Brc1ccc(I)nc1.OB(O)c1ccccc1>>Ic1ccc(-c2ccccc2)cn1”.

Preset errors cannot cover every way in which a retrosynthesis algorithm may produce a problematic prediction, so the agent-based test is necessary (Fig. 3b). For example, in an amide-synthesis case, the preset error may use an acyl chloride. The actual predictions, however, may show the same problem with another acyl donor, such as an acyl bromide, an acyl fluoride or an active ester. With no exact match with the preset error, such predictions are missed by the pre-set matching test.

We evaluated the top-20 predictions of all three retrosynthesis methods. Of the 100 chemoselectivity trap molecules, 51, 58 and 56 were hit in the pre-set matching test for UAlign, SimpRetro and RetroChimera, respectively (Fig. 3c). For each of these targets, at least one preset error appeared among the top-20 predictions. In the subsequent agent-based test, the agent judged each algorithm’s 2000 predictions one by one. For molecules with no hit, the agent warnings were then confirmed by hand to determine whether a chemoselectivity problem was present. The agent-based test plus manual check added 15 hits for UAlign, 19 for SimpRetro and 15 for RetroChimera, giving totals of 66, 77 and 71.

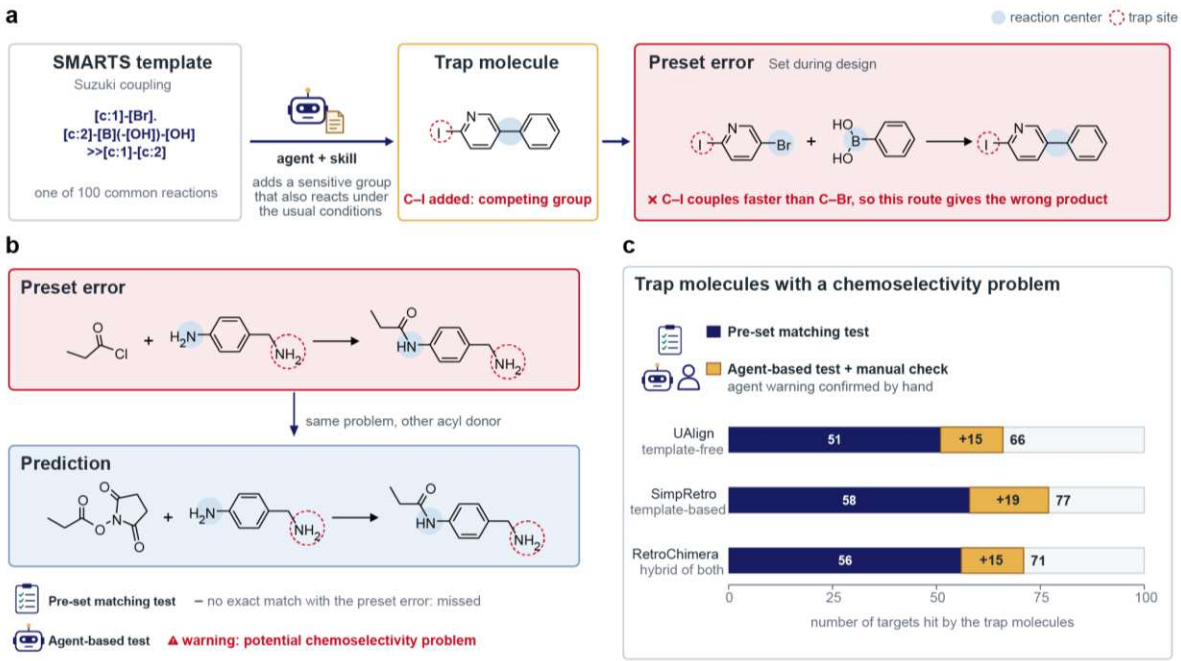



*Fig. 3 | Design and test results of chemoselectivity traps. a. Molecule-generation strategy and example. b. Typical preset error and an unexpected erroneous prediction. c. Test results: trap molecules with a chemoselectivity problem.*

## Exploring regioselectivity

Regioselectivity is an important topic in organic chemistry and a fundamental issue in synthesis planning. In practical synthesis, the main consequence of poor regioselectivity is the formation of

regioisomers, which are difficult to separate. By-products of chemoselectivity problems are usually separable by chromatography. Regioisomers, by contrast, have similar polarities, usually co-elute and in most cases are hard to separate even by chromatography (Fig. 4a). If route design requires only that the target product be obtained, it is somewhat justifiable for an algorithm to ignore regioselectivity. In real-world synthesis, however, such seemingly simple routes often lack practical value.

Regioselectivity trap molecules were generated by modifying USPTO reaction data. We used a preset SMARTS to select specific reaction cases and then modified these reactions (Fig. 4b). The core logic of the modification is to generate multiple (two or more) copies of the reaction center and ensure that they reside in as similar chemical environments as possible. For example, the recorded reaction is the hydrolysis of a phenylacetate ester to the corresponding carboxylic acid. After modification, a methyl and an ethyl group on the benzene ring tell the two methyl esters apart only slightly, leaving them in nearly the same environment. Hydrolyzing only one of the two esters is therefore difficult. Cases in which the reacting functional groups are identical but their surroundings differ substantially must be avoided. A typical failure is a coupling substrate bearing bromine at both the 2- and 3-positions of a pyridine ring. Because the 2-position may be more reactive, the two sites are not truly equivalent, and the reaction may still be selective. Trap construction must therefore consider the chemical environment within a certain radius of the reaction center rather than compare only the reacting groups themselves.

Regioselectivity traps also elicited unexpected erroneous reactions. For example, in an acid–amine condensation case, the retrosynthesis algorithm predicted anhydride aminolysis instead, a ring-opening route that likewise poses a regioselectivity problem. In another interesting case, the preset error was the reduction of a reactant containing two alkenes, but the algorithm instead predicted an alcohol dehydration. This prediction also carries a regiochemical preference, because the elimination would generally follow Zaitsev's rule (Fig. 4c).

We evaluated the top-20 predictions of all three retrosynthesis methods. Of the 100 regioselectivity trap molecules, 49, 48 and 69 were hit in the pre-set matching test for UAlign, SimpRetro and RetroChimera, respectively (Fig. 4d). In the subsequent agent-based test, the agent judged each algorithm's 2000 predictions one by one. For molecules with no hit, the agent warnings were then confirmed by hand to determine whether a regioselectivity problem was present. The agent-based test plus manual check added 36 hits for UAlign, 40 for SimpRetro and 24 for RetroChimera, giving totals of 85, 88 and 93.

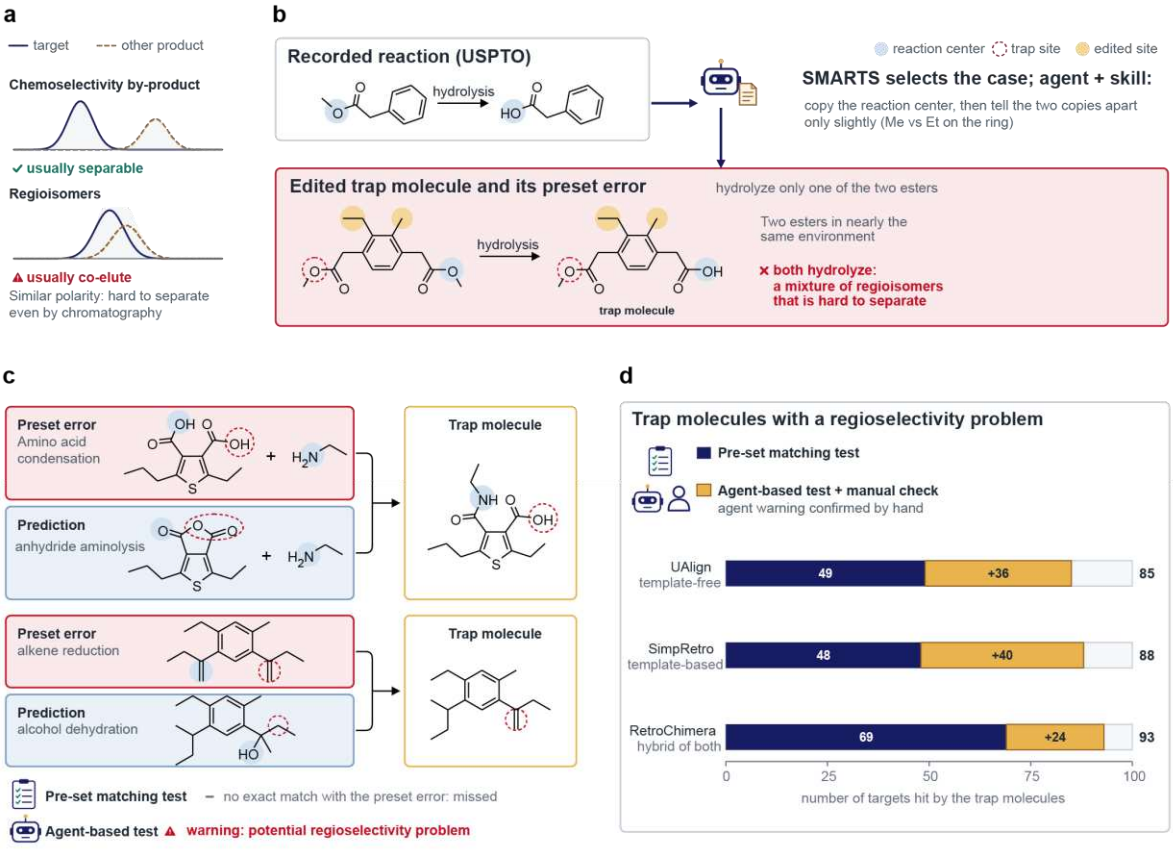



*Fig. 4 | Design and test results of regioselectivity traps. a. Regioisomers often make separation difficult. b. Molecule-generation strategy and example. c. Typical preset errors and unexpected erroneous predictions. d. Test results: trap molecules with a regioselectivity problem.*

## Exploring steric hindrance

Steric hindrance around the reaction center raises the activation barrier of the reaction [43]. When the reaction center is too crowded, the reaction may give only trace product or none at all (Fig. 5a).

For steric traps, we selected classic two-reactant, single-product reactions and designed 27 trap molecules guided by SMARTS templates. The design logic was to add bulky groups, such as *tert*-butyl, around the reaction center. For example, a classic Suzuki coupling can be converted into the highly hindered reaction shown in Fig. 5b, in which four ortho *t*-Bu groups block the coupling. Note that efficient methods for constructing such molecules may not currently exist.

This class of problem likewise includes unexpected predictions. For the reaction of a sulfonyl chloride with an amine to form a sulfonamide, the actual test returned a sulfonyl fluoride instead of the chloride (Fig. 5c). The fluoride is a less reactive sulfonyl donor and faces the same crowding.

We evaluated the top-20 predictions of all three retrosynthesis methods. Of the 27 steric trap molecules, 21, 26 and 25 were hit in the pre-set matching test for UAlign, SimpRetro and

RetroChimera, respectively (Fig. 5d). The subsequent agent-based test plus manual check added 4 hits for UAlign, none for SimpRetro and 2 for RetroChimera, giving totals of 25, 26 and 27.

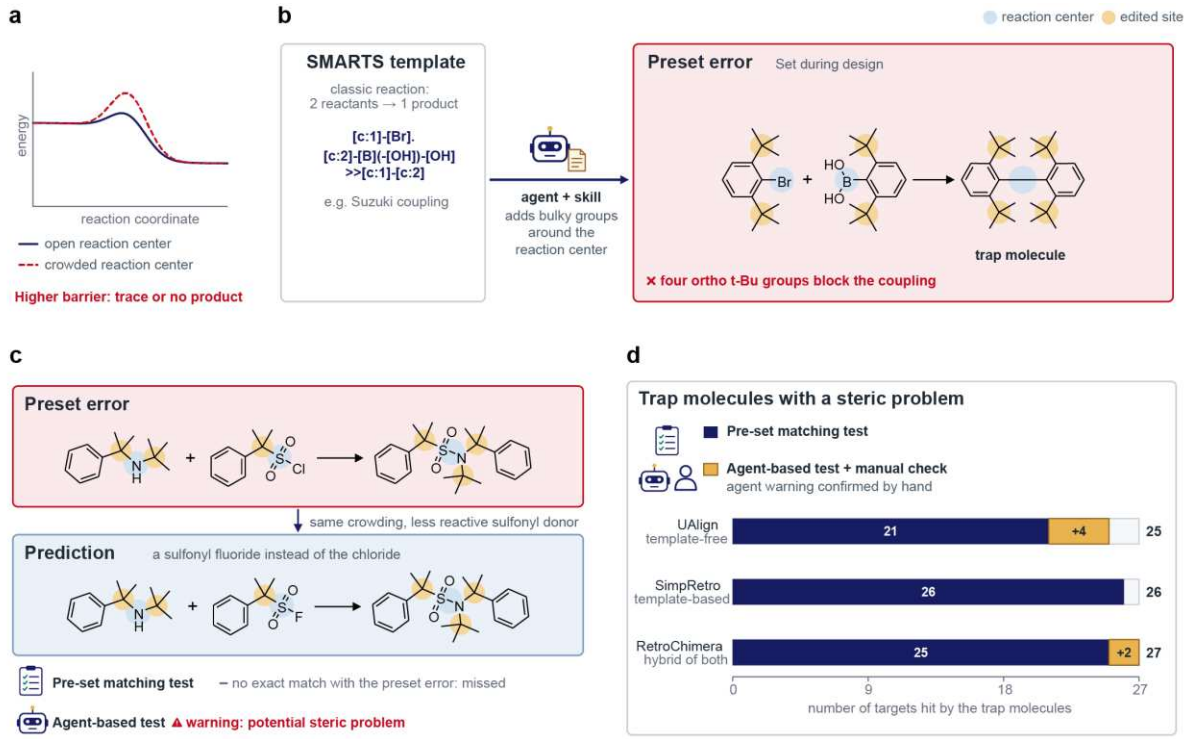



*Fig. 5 | Design and test results of steric traps. a. Steric hindrance raises the reaction barrier and often makes reactions difficult. b. Molecule-generation strategy and example. c. Typical preset error and an unexpected erroneous prediction. d. Test results: trap molecules with a steric problem.*

## Hallucinations and other issues

The agent-based test and manual check also detected several other classes of problems, which may stem from hallucinations of deep-learning algorithms [15, 19, 24]. These problems indicate that the algorithms have not learned stable reaction rules. In fact, addressing such problems should take priority over refining reaction-selectivity rules. The simpler cases include “no reaction”, in which the product appears directly among the reactants, and “magic reactions”, in which a complex compound is made from methane alone (Fig. 6a). Such problems are self-evident. Other simple cases are unreasonable changes in position: after Boc deprotection of an amino group, for example, the nitrogen atom moves to a different ring carbon. Complex hallucinations take several forms (Fig. 6b). The first is “complex precursor, simple product”, in which extremely complex reactants are used to make a simple product. Even if such a reaction is feasible, it usually violates basic principles of synthesis design. Second, reactants predicted for two different routes may be mixed into one route, giving

strange results. Third, complex reactions that are difficult to achieve may be generated, such as pyridine hydrogenation combined with ring expansion. Finally, even template-based methods can apply an incorrect template, producing strange predictions in which, for example, the coupling partner loses its dichlorophenyl ring. In summary, reducing reaction hallucinations may be an important engineering task before selectivity rules are further optimized.

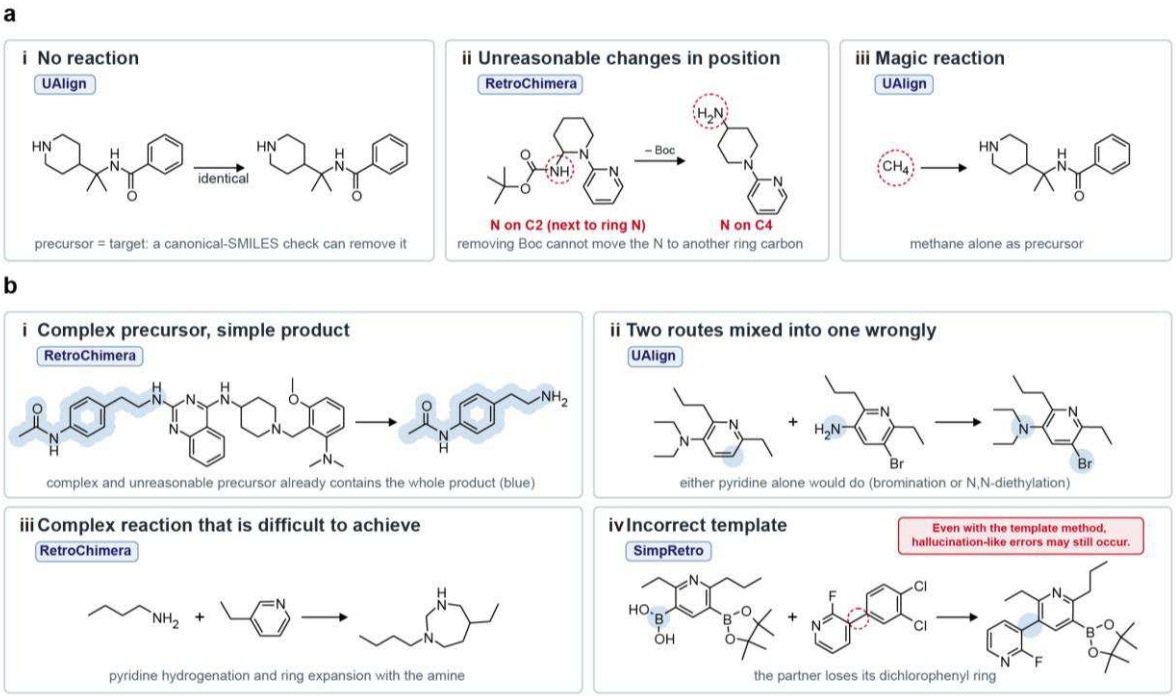



*Fig. 6 | Other absurd predictions. a. Simple cases. b. Complex cases.*

# Discussion

## Main findings and comparison of algorithms

This study introduces trap molecules as adversarial evaluation cases to probe the limitations of retrosynthesis algorithms in chemoselectivity, regioselectivity and steric hindrance. Pairing an agent with skills for generation and testing improved the efficiency of both design and testing. The combination of pre-set matching with the agent-based test and manual check makes detection both efficient and sound.

Tests of UAlign, SimpRetro and RetroChimera show that all three algorithms have weaknesses along these dimensions. The overall problem rate was 66%–77% in the chemoselectivity test, 85%–93% in the regioselectivity test and 92.6%–100% in the steric test. These results indicate that current algorithms have an incomplete command of such knowledge. They also partly support the hypothesis that reaction data lacking failures and boundary cases may give rise to such problems.

## Limitations of the molecular-trap framework

Not all organic molecules can currently be synthesized by high-yielding, highly selective methods. Some trap molecules may therefore lack an ideal route altogether. When a model is required to output predictions (i.e., it cannot choose to abstain), the probability of producing problematic cases increases accordingly. This issue should be considered when applying the framework.

## Implications for CASP evaluation

Molecular traps extend evaluation from conventional top-k accuracy and solve rate to active stress testing and can provide more specific feedback for algorithm improvement. Future ranking and evaluation metrics for retrosynthesis algorithms should be more diverse and should fully account for practical synthetic needs. The agent-based test shows that agents can assist in reviewing algorithm predictions, although they are currently limited by cost and computational speed. Developing cheaper, faster chemistry-specific judgment models based on agent technology for real-time analysis of retrosynthesis predictions may open a new direction for improving retrosynthesis algorithms.

# Conclusion

We have proposed a generate-and-test framework for trap molecules and designed trap-generation strategies based on SMARTS templates and reaction SMILES. Using the generated molecules, we performed pre-set matching and agent-based tests on retrosynthesis algorithms and systematically evaluated three algorithms with different design logics. The results show that all three tools have problems with chemoselectivity, regioselectivity and steric hindrance, leaving considerable

room for improvement. The generate-and-test framework can complement conventional evaluation metrics by assessing how well synthesis-planning algorithms have mastered basic chemical constraints. The agent-based testing and judgment scheme can also serve as a reference for developing more refined retrosynthesis algorithms.

## List of abbreviations

- AI: Artificial intelligence
- Boc: tert-Butoxycarbonyl
- CASP: Computer-aided synthesis planning
- Et: Ethyl
- Me: Methyl
- RDKit: Open-source cheminformatics software toolkit
- SMARTS: SMILES arbitrary target specification
- SMILES: Simplified molecular-input line-entry system
- t-Bu: tert-Butyl
- USPTO: United States Patent and Trademark Office

## Declarations


### Acknowledgements

We thank Mr. Zhong Tang for his investment in InsightChemy Co., Ltd.


### Authors' contributions

All authors contributed to the conception and development of the project. Y.T.: conceptualisation (lead), methodology, writing (original draft, lead), funding acquisition, project administration, resources, supervision. Y.D.: data curation, figure preparation and writing (review and editing). P.J. and H.Z.: methodology and validation. J.Z.: figure preparation and writing. All authors reviewed and approved the submitted manuscript.


### Funding

Y. T. was supported by the 2024 'Haiyou Project' Industry Leading Talent Program.


### Availability of data and materials

The molecular trap datasets and model evaluation results generated and analysed during this study are publicly available in the repository: https://github.com/Tianlab-web/Trap-Molecule.

### Competing interests

The authors declare no competing interests.

# References


1. Segler MHS, Preuss M, Waller MP (2018) Planning chemical syntheses with deep neural networks and symbolic AI. Nature 555(7698):604-610. doi:10.1038/nature25978.
2. Tu Z, Choure SJ, Fong MH, Roh J, Levin I, Yu K, Joung JF, Morgan N, Li S-C, Sun X, Lin H, Murnin M, Liles JP, Struble TJ, Fortunato ME, Liu M, Green WH, Jensen KF, Coley CW (2025) ASKCOS: open-source, data-driven synthesis planning. Acc Chem Res 58(11):1764-1775. doi:10.1021/acs.accounts.5c00155.
3. Thakkar A, Kogej T, Reymond J-L, Engkvist O, Bjerrum EJ (2020) Datasets and their influence on the development of computer assisted synthesis planning tools in the pharmaceutical domain. Chem Sci 11(1):154-168. doi:10.1039/C9SC04944D.
4. Wieczorek E, Sin JW, Tanovic S, Holland MTO, Wilbraham L, Sebastián-Pérez V, Bradley A, Miketa D, Brennan PE, Duarte F (2025) Transfer learning for heterocycle retrosynthesis. J Chem Inf Model 65(15):7851-7861. doi:10.1021/acs.jcim.4c02041.
5. Klucznik T, Mikulak-Klucznik B, McCormack MP, Lima H, Szymkuć S, Bhowmick M, Molga K, Zhou Y, Rickershauser L, Gajewska EP, Toutchkine A, Dittwald P, Startek MP, Kirkovits GJ, Roszak R, Adamski A, Sieredzińska B, Mrksich M, Trice SLJ, Grzybowski BA (2018) Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory. Chem 4(3):522-532. doi:10.1016/j.chempr.2018.02.002.
6. Mikulak-Klucznik B, Gołębiowska P, Bayly AA, Popik O, Klucznik T, Szymkuć S, Gajewska EP, Dittwald P, Staszewska-Krajewska O, Beker W, Badowski T, Scheidt KA, Molga K, Mlynarski J, Mrksich M, Grzybowski BA (2020) Computational planning of the synthesis of complex natural products. Nature 588(7836):83-88. doi:10.1038/s41586-020-2855-y.
7. Grzybowski BA, Badowski T, Molga K, Szymkuć S (2023) Network search algorithms and scoring functions for advanced-level computerized synthesis planning. WIREs Comput Mol Sci 13(1):e1630. doi:10.1002/wcms.1630.
8. Segler MHS, Waller MP (2017) Neural-symbolic machine learning for retrosynthesis and reaction prediction. Chem Eur J 23(25):5966-5971. doi:10.1002/chem.201605499.
9. Dai H, Li C, Coley CW, Dai B, Song L (2019) Retrosynthesis prediction with conditional graph logic network. In: Advances in neural information processing systems, vol 32. Curran Associates, Red Hook.
10. Chen S, Jung Y (2021) Deep retrosynthetic reaction prediction using local reactivity and global attention. JACS Au 1(10):1612-1620. doi:10.1021/jacsau.1c00246.
11. Xie S, Yan R, Guo J, Xia Y, Wu L, Qin T (2023) Retrosynthesis prediction with local template retrieval. Proc AAAI Conf Artif Intell 37(4):5330-5338. doi:10.1609/aaai.v37i4.25664.

12. Li J, Lin K, Pei J, Lai L (2024) Challenging complexity with simplicity: rethinking the role of single-step models in computer-aided synthesis planning. J Chem Inf Model 64(14):5470-5479. doi:10.1021/acs.jcim.4c00432.
13. Yan C, Ding Q, Zhao P, Zheng S, Yang J, Yu Y, Huang J (2020) RetroXpert: decompose retrosynthesis prediction like a chemist. In: Advances in neural information processing systems, vol 33. Curran Associates, Red Hook.
14. Somnath VR, Bunne C, Coley CW, Krause A, Barzilay R (2021) Learning graph models for retrosynthesis prediction. In: Advances in neural information processing systems, vol 34. Curran Associates, Red Hook.
15. Wang X, Li Y, Qiu J, Chen G, Liu H, Liao B, Hsieh C-Y, Yao X (2021) RetroPrime: a diverse, plausible and transformer-based method for single-step retrosynthesis predictions. Chem Eng J 420:129845. doi:10.1016/j.cej.2021.129845.
16. Chen Z, Ayinde OR, Fuchs JR, Sun H, Ning X (2023) G2Retro as a two-step graph generative models for retrosynthesis prediction. Commun Chem 6:102. doi:10.1038/s42004-023-00897-3.
17. Zhong W, Yang Z, Chen CY-C (2023) Retrosynthesis prediction using an end-to-end graph generative architecture for molecular graph editing. Nat Commun 14:3009. doi:10.1038/s41467-023-38851-5.
18. Zhao P-C, Wei X-X, Wang Q, Wang Q-H, Li J-N, Shang J, Lu C, Shi J-Y (2025) Single-step retrosynthesis prediction via multitask graph representation learning. Nat Commun 16:814. doi:10.1038/s41467-025-56062-y.
19. Liu B, Ramsundar B, Kawthekar P, Shi J, Gomes J, Luu Nguyen Q, Ho S, Sloane J, Wender P, Pande V (2017) Retrosynthetic reaction prediction using neural sequence-to-sequence models. ACS Cent Sci 3(10):1103-1113. doi:10.1021/acscentsci.7b00303.
20. Tetko IV, Karpov P, Van Deursen R, Godin G (2020) State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. Nat Commun 11:5575. doi:10.1038/s41467-020-19266-y.
21. Zhong Z, Song J, Feng Z, Liu T, Jia L, Yao S, Wu M, Hou T, Song M (2022) Root-aligned SMILES: a tight representation for chemical reaction prediction. Chem Sci 13(31):9023-9034. doi:10.1039/D2SC02763A.
22. Zeng K, Yang B, Zhao X, Zhang Y, Nie F, Yang X, Jin Y, Xu Y (2024) UAlign: pushing the limit of template-free retrosynthesis prediction with unsupervised SMILES alignment. J Cheminform 16:80. doi:10.1186/s13321-024-00877-2.
23. Xu L-C, Tang M-J, An J, Cao F, Qi Y (2025) A unified pre-trained deep learning framework for cross-task reaction performance prediction and synthesis planning. Nat Mach Intell 7(9):1561-1571. doi:10.1038/s42256-025-01098-4.

24. Maziarz K, Liu G, Pultar F, Gardner J, Gensch T, Helie J, Misztela H, Tripp A, Li J, Kornev A, Gaiński P, Hoefling H, Fortunato M, Gupta R, Baxter A, Poole DL, Elward JM, Krzyzanowski A, Pogány P, Pickett SD, Wall ID, Bishop CM, Humphreys PG, Lumley JA, Wiesenfeldt MP, Segler MHS (2026) Chemist-aligned retrosynthesis by ensembling diverse inductive bias models. Nature. doi:10.1038/s41586-026-11160-9.

25. Wang Y, Pang C, Wang Y, Jin J, Zhang J, Zeng X, Su R, Zou Q, Wei L (2023) Retrosynthesis prediction with an interpretable deep-learning framework based on molecular assembly tasks. Nat Commun 14:6155. doi:10.1038/s41467-023-41698-5.

26. Pandarus V, Ciriminna R, Béland F, Pagliaro M (2011) Selective hydrogenation of functionalized nitroarenes under mild conditions. Catal Sci Technol 1(9):1616-1623. doi:10.1039/C1CY00097G.

27. Jin JH, An M, Jeon M, Kim J, Kang SM, Choi I (2025) Sustainable approaches for the protection and deprotection of functional groups. Chem Eur J 31(40):e202501387. doi:10.1002/chem.202501387.

28. Sojdak CA, Polefrone DA, Shah HM, Vu CD, Orzolek BJ, Jimenez Antenucci PM, Bush MV, Kozlowski MC (2025) Direct (LC-)MS identification of regioisomers from C-H functionalization by partial isotopic labeling. ACS Cent Sci 11(2):272-278. doi:10.1021/acscentsci.4c01765.

29. Collins KD, Glorius F (2013) A robustness screen for the rapid assessment of chemical reactions. Nat Chem 5(7):597-601. doi:10.1038/nchem.1669.

30. Goodman J (2009) Computer software review: Reaxys. J Chem Inf Model 49(12):2897-2898. doi:10.1021/ci900437n.

31. Bai F, Ritter A, Madrid P, Freitag D, Niekrasz J (2022) SynKB: semantic search for synthetic procedures. arXiv:2208.07400. doi:10.48550/arXiv.2208.07400.

32. Lowe DM (2012) Extraction of chemical structures and reactions from the literature. Dissertation, University of Cambridge. doi:10.17863/CAM.16293.

33. Schneider N, Lowe DM, Sayle RA, Tarselli MA, Landrum GA (2016) Big data from pharmaceutical patents: a computational analysis of medicinal chemists' bread and butter. J Med Chem 59(9):4385-4402. doi:10.1021/acs.jmedchem.6b00153.

34. Schneider N, Stiefl N, Landrum GA (2016) What's what: the (nearly) definitive guide to reaction role assignment. J Chem Inf Model 56(12):2336-2346. doi:10.1021/acs.jcim.6b00564.

35. Jin W, Coley CW, Barzilay R, Jaakkola T (2017) Predicting organic reaction outcomes with Weisfeiler-Lehman network. In: Advances in neural information processing systems, vol 30. Curran Associates, Red Hook.

36. Schwaller P, Gaudin T, Lányi D, Bekas C, Laino T (2018) 'Found in translation': predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. Chem Sci 9(28):6091-6098. doi:10.1039/C8SC02339E.

37. Kearnes SM, Maser MR, Wleklinski M, Kast A, Doyle AG, Dreher SD, Hawkins JM, Jensen KF, Coley CW (2021) The Open Reaction Database. J Am Chem Soc 143(45):18820-18826. doi:10.1021/jacs.1c09820.

38. Rana D, Pflüger PM, Hölter NP, Tan G, Glorius F (2024) Standardizing substrate selection: a strategy toward unbiased evaluation of reaction generality. ACS Cent Sci 10(4):899-906. doi:10.1021/acscentsci.3c01638.

39. Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, Moroz YS, Mayfield J, Sayle RA (2020) ZINC20—a free ultralarge-scale chemical database for ligand discovery. J Chem Inf Model 60(12):6065-6073. doi:10.1021/acs.jcim.0c00675.

40. Goodfellow IJ, Shlens J, Szegedy C (2014) Explaining and harnessing adversarial examples. arXiv:1412.6572. doi:10.48550/arXiv.1412.6572.

41. Patel H, Ihlenfeldt W-D, Judson PN, Moroz YS, Pevzner Y, Peach ML, Delannée V, Tarasova NI, Nicklaus MC (2020) SAVI, in silico generation of billions of easily synthesizable compounds through expert-system type rules. Sci Data 7:384. doi:10.1038/s41597-020-00727-4.

42. Wołos A, Koszelewski D, Roszak R, Szymkuć S, Moskal M, Ostaszewski R, Herrera BT, Maier JM, Brezicki G, Samuel J, Lummiss JAM, McQuade DT, Rogers L, Grzybowski BA (2022) Computer-designed repurposing of chemical wastes into drugs. Nature 604(7907):668-676. doi:10.1038/s41586-022-04503-9.

43. Hamlin TA, Swart M, Bickelhaupt FM (2018) Nucleophilic substitution (SN2): dependence on nucleophile, leaving group, central atom, substituents, and solvent. ChemPhysChem 19(11):1315-1330. doi:10.1002/cphc.201701363.